This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

Compressive Sampling using Annihilating Filter-based Low-Rank Interpolation

Jong Chul Ye *Senior Member, IEEE*, Jong Min Kim, Kyong Hwan Jin, and Kiryung Lee

***Abstract*— While the recent theory of compressed sensing provides an opportunity to overcome the Nyquist limit in recovering sparse signals, a solution approach usually takes the form of an inverse problem of an unknown signal, which is crucially dependent on specific signal representation. In this paper, we propose a drastically different two-step Fourier compressive sampling framework in a continuous domain that can be implemented via measurement domain interpolation, after which signal reconstruction can be done using classical analytic reconstruction methods. The main idea originates from the fundamental duality between the sparsity in the primary space and the low-rankness of a structured matrix in the spectral domain, showing that a low-rank interpolator in the spectral domain can enjoy all of the benefits of sparse recovery with performance guarantees. Most notably, the proposed low-rank interpolation approach can be regarded as a generalization of recent spectral compressed sensing to recover large classes of finite rate of innovations (FRI) signals at a near-optimal sampling rate. Moreover, for the case of cardinal representation, we can show that the proposed low-rank interpolation scheme will benefit from inherent regularization and an optimal incoherence parameter. Using a powerful dual certificate and the golfing scheme, we show that the new framework still achieves a near-optimal sampling rate for a general class of FRI signal recovery, while the sampling rate can be further reduced for a class of cardinal splines. Numerical results using various types of FRI signals confirm that the proposed low-rank interpolation approach offers significantly better phase transitions than conventional CS approaches.**

***Index Terms*—Compressed sensing, signals of finite rate of innovations, spectral compressed sensing, low rank matrix completion, dual certificates, golfing scheme**

I. Introduction

The compressed sensing or compressive sampling (CS) theory [1]–[3] refers to the accurate recovery of unknown sparse signals from underdetermined linear measurements. In particular, the Fourier CS problem, which recovers unknown signals from sub-sampled Fourier measurements, has many important applications in imaging applications, such as magnetic resonance imaging (MRI), X-ray computed tomography (CT), optics, and others. Moreover, this problem is closely related to the classical harmonic retrieval problem that computes the amplitudes and frequencies at the off-the-grid locations of a superposition of complex sinusoids from consecutive or bunched corresponding Fourier samples. Harmonic retrieval can be solved by various methods, including Prony's method [4] and a matrix pencil algorithm [5]. These methods were proven to succeed at a minimal sample rate in a noiseless case, as they satisfy an algebraic condition called the full spark (or full Kruskal rank) condition [6], which guarantees the unique identification of an unknown signal. Typically, when operating at the critical sample rate, these methods are not robust to perturbations in the measurements due to the large condition numbers involved.

Accordingly, to facilitate the robust reconstruction of off-the-grid spectral components, CS algorithms from non-consecutively sub-sampled Fourier measurements are required. This scheme is known as *spectral compressed sensing*, which is also known as *compressed sensing off the grid* when the underlying signal is composed of Diracs. Indeed, this has been developed with a close link to recent super-resolution researches [7]–[10]. For example, Candes and Fernandez-Granda [8], [11] showed that if the minimum distance of the Diracs exceeds $2/f_c$, where f_c denotes the cut-off frequency of the measured spectrum, then a simple convex optimization can solve the locations of the Diracs. Under the same minimum distance condition, Tang in two studies [9], [12] proposed an atomic norm minimization approach for the recovery of Diracs correspondingly from random spatial and Fourier samples. Unlike these direct signal recovery methods, Chen and Chi [10] proposed a two-step approach consisting of interpolation followed by a matrix pencil algorithm. In addition, they provided performance guarantees at near-optimal levels of sample complexity (up to a logarithmic factor). One of the main limitations of these spectral compressed sensing approaches is, however, that the unknown signal is restricted to a stream of Diracs. The approach by Chen

Jong Chul Ye is with Dept. of Bio and Brain Engineering, KAIST, Daejeon 305-701, Republic of Korea. (Email:jong.ye@kaist.ac.kr.) Jong Min Kim is with Korea Science Academy of KAIST. Kyong Hwan Jin is with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Kiryung Lee is with the School of Electrical and Computer Engineering, Georgia Tech, USA.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

and Chi [10] is indeed a special case of the proposed approach, but they did not realize its potential when used to recover a much wider class of signals.

Note that the stream of Diracs is a special instance of a signal model of *signals with a finite rate of innovation (FRI)* [13]–[15]. Originally proposed by Vetterli et al. [13], the class of FRI signals includes a stream of Diracs, a stream of differentiated Diracs, non-uniform splines, and piecewise smooth polynomials. Vetterli et al. [13]–[15] proposed *time-domain* sampling schemes with these FRI signals that operate at the rate of innovation with a provable algebraic guarantee in a noise-free scenario. Their reconstruction scheme estimates an annihilating filter that cancels the Fourier series coefficients of a FRI signal at consecutive low frequencies. However, because the scheme relies on data acquisition in the time domain, the equivalent Fourier domain measurements are restricted to a bunched sampling pattern, similar to classical harmonic retrieval problems.

Therefore, one of the main aims of this paper is to generalize the scheme by Verterli et al. [13]–[15] to address Fourier CS problems that recover a general class of FRI signals from irregularly subsampled Fourier measurements. Notably, we prove that the only required change is an additional Fourier domain interpolation step that estimates missing Fourier measurements. More specifically, for the general class FRI signals introduced in earlier works [13]–[15], we show that there always exists a low-rank Hankel structured matrix associated with the corresponding annihilating filter. Accordingly, their missing spectral elements can be interpolated using a low-rank Hankel matrix completion algorithm. Once a set of Fourier measurements at consecutive frequencies is interpolated, a FRI signal can be reconstructed using conventional methods, including Prony's method and matrix pencil algorithms, as in earlier studies [13]–[15]. Most notably, we show that the proposed Fourier CS of FRI signals operates at a near-optimal rate (up to a logarithmic factor) with provable performance guarantees. Additionally, owing to the inherent redundancies introduced by the CS sampling scheme, the subsequent step of retrieving a FRI signal becomes much more stable.

While a similar low-rank Hankel matrix completion approach was used by Chen and Chi [10], there are several important differences. First, low-rankness of a Hankel matrix in earlier work [10] was found based on standard Vandermonde decomposition, which is true only when the underlying FRI signal is a stream of Diracs. Accordingly, for differentiated Diracs, the theoretical tools in that study cannot be used. Second, when the underlying signal can be converted to a stream of Diracs or differentiated Diracs by applying a linear transform which acts as a diagonal operator (i.e., element-wise multiplication) in the Fourier domain, we can still construct a low-rank Hankel matrix from the *weighted* Fourier measurements, whose weights are determined by the spectrum of the linear operator. For example, a total variation (TV) sparse signal is a stream of Diracs after the differentiation, and piecewise smooth polynomials become a stream of differentiated Diracs with the application of a differential operator. Finally, the advantage of the proposed approach becomes more evident when we model the unknown signal using cardinal L-splines [16]. In cardinal L-splines, discontinuities occur only on an integer grid, which is a reasonable model to acquire signals with a high but finite resolution. In such a case, we can show that discretization using cardinal splines makes the reconstruction significantly more stable in spite of the existence of noise in the measurements, and the logarithmic factor as well as the incoherence parameter for the performance guarantees can be improved further.

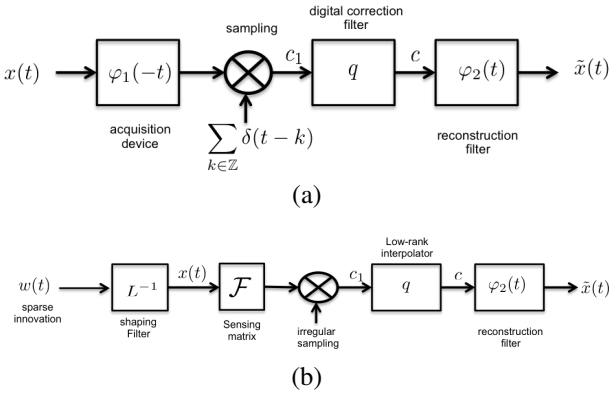

Fig. 1. Comparison with various sampling schemes: (a) generalized sampling [17], [18]; here, a continuous input signal is filtered through an acquisition device, after which uniform sampling is performed. The goal of sampling is to impose a consistency condition such that if the reconstructed signal is used as an input to the acquisition device, it can generate the same discrete sequence $\{c_1\}$. This can be accomplished using a digital correction filter q. (b) Proposed sampling scheme: here, the CS step is replaced by a discrete low-rank interpolator, and the final reconstruction is done using the reconstruction filter from fully sampled data.

It is important to note that the proposed low-rank interpolation approach is different from classical compressed sensing approaches which regard the sampling problem as an inverse problem and whose focus is directly to recover the unknown signal. Rather, the proposed approach is more closely related to the classical sampling theory, where the signal sampling step is decoupled from a signal recovery algorithm. For example, in the sampling theory for signals in shift-invariant spaces [17], [18], the nature of the signal sampling can be fully accounted for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

using a *digital correction filter*, after which signal recovery is performed by convolution with a reconstruction filter (see Fig. 1(a)). Similarly, by introducing a *low-rank interpolator*, the proposed scheme in Fig. 1(b) fully decouples signal recovery from the sampling step as a separate layer that can be optimized independently. This occurs because the same low-rank interpolator will successfully complete missing measurements regardless of whether the unknown signal is either a stream of Diracs or a stream of differentiated Diracs. In the subsequent step, analytic reconstruction methods such as Prony's method and matrix pencil algorithms can identify the signal model, as in earlier studies [13]–[15].

The proposed two-layer approach composed of Fourier domain interpolation and analytic reconstruction is very useful in real-world applications, as the low-rank interpolator can be added as a form of digital correction filter for existing systems, where the second step is already implemented. Moreover, in many biomedical imaging problems such as magnetic resonance imaging (MRI) or X-ray computed tomography (CT), accurate interpolation to fully sampled Fourier data provides the important advantage of utilizing a fully established mathematical theory of analytic reconstruction. In addition, classical preprocessing techniques for artifact removal were developed with the assumption of fully sampled measurements; hence, these steps can be readily combined with the proposed low-rank interpolation approaches. The superior advantages of the proposed scheme have been demonstrated in various biomedical imaging and image processing applications, such as compressed sensing MRI [19], [20], MR artifact correction [21], [22], image inpainting [23], super-resolution microscopy [24], image denoising [25], and others, clearly confirming the practicality of the new theory.

Nonetheless, it is remarkable that the proposed two-layer approach using low-rank interpolation achieves a near-optimal sample rate while universally applying to different signal models of the same order (e.g., streams of Diracs and streams of differentiated Diracs). Moreover, it may appear mysterious that an explicit form of the minimum separation distance as required in Fernandez-Granda [8], [11] and Tang [9], [12] is not needed for the performance guarantee. However, the proposed method is not free of limitations. Specifically, we show that the incoherence parameter in our performance guarantees is dependent upon the type of unknown signals and on the minimum separation between successive spikes. The similarity and differences of our results from the existing theory [8], [9], [11], [12] and the origin of the differences will also be discussed.

This paper is structured as follows. Section II initially discusses the main results that relate to an annihilating filter and a low-rank Hankel structured matrix, providing the performance guarantee of low-rank structured matrix completion, which will be used throughout the paper. Section III then discusses the proposed low-rank interpolation theory for the recovery of FRI signals, which is followed by low-rank interpolation for cardinal L-splines in Section IV. Section V explains the algorithmic implementation. Numerical results are then provided in Section VI, which is followed by the conclusion in Section VII.

II. Main Results

A. Notations

A Hankel structured matrix generated from an n-dimensional vector $\mathbf{x} = [x[0], \cdots, x[n-1]]^T \in \mathbb{C}^n$ has the following structure:

$$\mathscr{H}(\mathbf{x}) = \begin{bmatrix} x[0] & x[1] & \cdots & x[d-1] \\ x[1] & x[2] & \cdots & x[d] \\ \vdots & \vdots & \ddots & \vdots \\ x[n-d] & [n-d+1] & \cdots & x[n-1] \end{bmatrix} \in \mathbb{C}^{(n-d+1)\times d}.$$

where d is called a matrix pencil parameter. We denote the space of this type of Hankel structure matrices as $\mathcal{H}(n,d)$.

An $n \times d$ *wrap-around* Hankel matrix generated from an n-dimensional vector $\mathbf{u} = [u[0], \cdots, u[n-1]]^T \in \mathbb{C}^n$ is defined as:

$$\mathscr{H}_c(\mathbf{u}) = \left[\begin{array}{cccc} u[0] & u[1] & \cdots & u[d-1] \\ u[1] & u[2] & \cdots & u[d] \\ \vdots & \vdots & \ddots & \vdots \\ u[n-d] & u[n-d+1] & \cdots & u[n-1] \\ \hline u[n-d+1] & u[n-d+2] & \cdots & u[0] \\ \vdots & \vdots & \ddots & \vdots \\ u[n-1] & u[0] & \cdots & u[d-2] \end{array}\right] \in \mathbb{C}^{n\times d}. \quad (1)$$

Note that $n \times d$ wrap-around Hankel matrix can be considered as a Hankel matrix of $(d-1)$-element augumented vector from $\mathbf{u} \in \mathbb{C}^n$ with the periodic boundary expansion:

$$\tilde{\mathbf{u}} = \left[\mathbf{u}^T \quad \underbrace{u[0]\ u[1] \cdots u[d-2]}_{(d-1)}\right]^T \in \mathbb{C}^{n+d-1}.$$

We denote the space of this type of wrap-around Hankel structure matrices as $\mathcal{H}_c(n,d)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

B. Annihilating Filter-based Low-Rank Hankel Matrix

The Fourier CS problem of our interest is to recover the unknown signal $x(t)$ from the Fourier measurement:

$$\hat{x}(f) = \mathcal{F}\{x(t)\} = \int x(t)e^{-i2\pi ft}dt \ .$$

Without loss of generality, we assume that the support of $x(t)$ is $[0,1]$. Then, the sampled Fourier data at the Nyquist rate is defined by

$$\hat{x}[k] = \hat{x}(f)|_{f=k} \, .$$

We also define a length $(r+1)$-annihilating filter $\hat{h}[k]$ for $\hat{x}[k]$ that satisfies

$$(\hat{h} * \hat{x})[k] = \sum_{l=0}^{r} \hat{h}[l]\hat{x}[k-l] = 0, \quad \forall k. \tag{2}$$

The existence of the finite length annihilating filter has been extensively studied for FRI signals [13]–[15]. This will be discussed in more detail later.

Suppose that the filter $\hat{h}[k]$ is the minimum length annihilating filter. Then, for any $k_1 \geq 1$ tap filter $\hat{a}[k]$, it is easy to see that the following filter with $d = r + k_1$ taps is also an annihilating filter for $\hat{x}[k]$:

$$\hat{h}_a[k] = (\hat{a} * \hat{h})[k] \Longrightarrow \sum_{l=0}^{r} \hat{h}_a[l]\hat{x}[k-l] = 0, \quad \forall \ k, \tag{3}$$

because $\hat{h}_a * \hat{x} = \hat{a} * \hat{h} * \hat{x} = 0$. The matrix representation of (3) is given by

$$\mathscr{C}(\hat{\mathbf{x}})\bar{\hat{\mathbf{h}}}_a = \mathbf{0}$$

where $\bar{\hat{\mathbf{h}}}_a$ denotes a vector that reverses the order of the elements in

$$\hat{\mathbf{h}}_a = \left[\hat{h}_a[0], \cdots, \hat{h}_a[d-1]\right]^T, \tag{4}$$

and

$$\mathscr{C}(\hat{\mathbf{x}}) = \begin{bmatrix} \vdots & \vdots & \ddots & \vdots \\ \hat{x}[-1] & \hat{x}[0] & \cdots & \hat{x}[d-2] \\ \hline \hat{x}[0] & \hat{x}[1] & \cdots & \hat{x}[d-1] \\ \hat{x}[1] & \hat{x}[2] & \cdots & \hat{x}[d] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}[n-d] & \hat{x}[n-d+1] & \cdots & \hat{x}[n-1] \\ \hline \hat{x}[n-d+1] & \hat{x}[n-d+2] & \cdots & \hat{x}[n] \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

Accordingly, by choosing n such that $n-d+1 > r$ and defining an n-dimensional vector composed of sampled Fourier data at the Nyquist rate as:

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{x}[0] & \cdots & \hat{x}[n-1] \end{bmatrix}^T \in \mathbb{C}^n, \tag{5}$$

we can construct the following matrix equation:

$$\mathscr{H}(\hat{\mathbf{x}})\bar{\hat{\mathbf{h}}}_a = \mathbf{0},$$

where the Hankel structure matrix $\mathscr{H}(\hat{\mathbf{x}}) \in \mathcal{H}(n,d)$ is constructed as

$$\mathscr{H}(\hat{\mathbf{x}}) = \begin{bmatrix} \hat{x}[0] & \hat{x}[1] & \cdots & \hat{x}[d-1] \\ \hat{x}[1] & \hat{x}[2] & \cdots & \hat{x}[d] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}[n-d] & \hat{x}[n-d+1] & \cdots & \hat{x}[n-1] \end{bmatrix} \tag{6}$$

Then, we can show the following key result:

Theorem II.1. *Let $r+1$ denote the minimum size of annihilating filters that annihilates sampled Fourier data $\hat{x}[k]$. Assume that $\min\{n-d+1, d\} > r$. Then, for a given Hankel structured matrix $\mathscr{H}(\hat{\mathbf{x}}) \in \mathcal{H}(n,d)$ constructed in* (6)*, we have*

$$\text{RANK}\mathscr{H}(\hat{\mathbf{x}}) = r, \tag{7}$$

where $\text{RANK}(\cdot)$ *denotes a matrix rank.*

Proof. See Appendix B. □

C. Performance Guarantees for Structured Matrix Completion

Let Ω be a multi-set consisting of random indices from $\{0, \ldots, n-1\}$ such that $|\Omega| = m < n$. While the standard CS approaches directly estimate $x(t)$ from $\hat{x}[k], k \in \Omega$, here we propose a two-step approach by exploiting Theorem II.1. More specifically, we first interpolate $\hat{x}[k]$ for all $k \in \{0, \ldots, n-1\}$ from the sparse Fourier samples, and the second step then applies the existing spectral estimation methods to estimate $x(t)$ as done in [13]–[15]. Thanks to the low-rankness of the associated Hankel matrix, the first step can be implemented using the following low-rank matrix completion:

$$\begin{aligned} \underset{\mathbf{g}\in\mathbb{C}^n}{\text{minimize}} \quad & \text{RANK } \mathscr{H}(\hat{\mathbf{g}}) \\ \text{subject to} \quad & P_\Omega(\hat{\mathbf{g}}) = P_\Omega(\hat{\mathbf{x}}), \end{aligned} \tag{8}$$

where P_Ω is the projection operator on the sampling location Ω. Therefore, the remaining question is to verify whether the low-rank matrix completion approach (8) does not compromise any optimality compared to the standard Fourier CS, which is the main topic in this section.

The low-rank matrix completion problem in (8) is non-convex, which is difficult to analyze. Therefore, to provide a performance guarantee, we resort to its convex relaxation using the nuclear norm. Chen and Chi [10] provided the first performance guarantee for robust spectral compressed sensing via structured matrix completion by nuclear norm minimization and extended the result to general low-rank Hankel/Toeplitz matrix completion [10, Theorem 4]. However, parts of their proof (e.g., [10, Appendix H]) critically depend on the special structure given in the standard Vandermonde

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

decomposition. Furthermore, they also used the standard incoherence condition, which is neither assumed nor applied by their incoherence condition [10, eq. (27)]. Therefore, unlike their claim, the main results in [10], in its current forms, apply only to the spectral compressed sensing. Here, we elaborate on their results so that the performance guarantees apply to the general structured low-rank matrix completion problems which will be described in subsequent sections.

Recall that the notion of the incoherence plays a crucial role in matrix completion and structured matrix completion. We recall the definitions using our notations. Suppose that $M \in \mathbb{C}^{n_1 \times n_2}$ is a rank-r matrix whose SVD is $U\Sigma V^*$ with $U \in \mathbb{C}^{n_1 \times r}, \Sigma \in \mathbb{C}^{r \times r}$ and $V \in \mathbb{C}^{n_2 \times r}$, respectively. M is said to satisfy the *standard incoherence* condition with parameter μ if

$$\begin{aligned} \max_{1 \le i \le n_1} \|U^* \mathbf{e}_i\|_2 &\le \sqrt{\frac{\mu r}{n_1}}, \\ \max_{1 \le j \le n_2} \|V^* \mathbf{e}_j\|_2 &\le \sqrt{\frac{\mu r}{n_2}}, \end{aligned} \tag{9}$$

where $\mathbf{e}_i$ denotes the appropriate size standard coordinate vector with 1 on the i-th elements and zeros elsewhere.

To deal with two types of Hankel matrices simultaneously, we define a *linear lifting* operator $\mathcal{L} : \mathbb{C}^n \to \mathbb{C}^{n_1 \times n_2}$ that lifts a vector $\mathbf{x} \in \mathbb{C}^n$ to a structured matrix $\mathcal{L}(\mathbf{x}) \in \mathbb{C}^{n_1 \times n_2}$ in a higher dimensional space. For example, the dimension of $\mathcal{L}(\mathbf{x}) \in \mathbb{C}^{n_1 \times n_2}$ is given as $n_1 = n - d + 1$ and $n_2 = d$ for a lifting to a Hankel matrix in $\mathcal{H}(n, d)$, whereas $n_1 = n$ and $n_2 = d$ for a lifting to a wrap-around Hankel matrix $\mathcal{H}_c(n, d)$. A linear lifting operator is regarded as a synthesis operator with respect to basis $\{A_k\}_{k=1}^n, A_k \in \mathbb{C}^{n_1 \times n_2}$:

$$\mathcal{L}(\mathbf{x}) = \sum_{k=1}^{n} A_k \langle \mathbf{e}_k, \mathbf{x} \rangle,$$

where the specific form of the basis for the case of $\{A_k\}$ for $\mathcal{H}(n, d)$ and $\mathcal{H}_c(n, d)$ will be explained in Appendix C.

Then, the completion of a low rank structured matrix $\mathcal{L}(\mathbf{x})$ from the observation of its partial entries can be done by minimizing the nuclear norm under the measurement fidelity constraint as follows:

$$\begin{aligned} \underset{\mathbf{g} \in \mathbb{C}^n}{\text{minimize}} \quad & \|\mathcal{L}(\mathbf{g})\|_* \\ \text{subject to} \quad & P_\Omega(\mathbf{g}) = P_\Omega(\mathbf{x}). \end{aligned} \tag{10}$$

where $\|\cdot\|_*$ denotes the matrix nuclear norm. Then, we have the following main result.

Theorem II.2. *Let $\Omega = \{j_1, \ldots, j_m\}$ be a multi-set consisting of random indices where j_k's are i.i.d. following the uniform distribution on $\{0, \ldots, n-1\}$. Suppose $\mathcal{L}$ correspond to one of the structured matrices in $\mathcal{H}(n, d)$ and $\mathcal{H}_c(n, d)$. Suppose, furthermore, that $\mathcal{L}(\mathbf{x})$ is of rank-r and satisfies the standard incoherence condition in* (9) *with parameter μ. Then there exists an absolute constant c_1 such that $\mathbf{x}$ is the unique minimizer to* (10) *with probability $1 - 1/n^2$, provided*

$$m \ge c_1 \mu c_s r \log^\alpha n, \tag{11}$$

where $\alpha = 2$ if each images of $\mathcal{L}$ has the wrap-around property; $\alpha = 4$, otherwise, and $c_s := \max\{n/n_1, n/n_2\}$.

Proof. See Appendix E. □

Note that Theorem II.2 provides a generalized version of performance guarantee compared to the previous work [10]. Specifically, Theorem II.2 holds for other structured matrices, if the associated basis matrix $\{A_k\}$ satisfies the specific condition described in detail in Eq. (A.15). In addition, for the case of signals with wrap-around Hankel matrix (which will be explained later), the log exponentional factor α in (11) becomes 2, which reduces the sampling rate. This sampling rate reduction is novel and was not observed in the previous work [10].

Next, we consider the recovery of $\mathbf{x}$ from its partial entries with noise. Let $\mathbf{y}$ denote a corrupted version of $\mathbf{x}$. The unknown structured matrix $\mathcal{L}(\mathbf{x})$ can be estimated via

$$\begin{aligned} \underset{\mathbf{g} \in \mathbb{C}^n}{\text{minimize}} \quad & \|\mathcal{L}(\mathbf{g})\|_* \\ \text{subject to} \quad & \|P_\Omega(\mathbf{g} - \mathbf{y})\|_2 \le \delta. \end{aligned} \tag{12}$$

Then, we have the following stability guarantee:

Theorem II.3. *Suppose the noisy data $\mathbf{y} \in \mathbb{C}^n$ satisfies $\|P_\Omega(\mathbf{y} - \mathbf{x})\|_2 \le \delta$ and $\mathbf{x} \in \mathbb{C}^n$ is the noiseless data. Under the hypotheses of Theorem II.2, there exists an absolute constant c_1, c_2 such that with probability $1 - 1/n^2$, the solution $\mathbf{g}$ to* (12) *satisfies*

$$\|\mathcal{L}(\mathbf{x}) - \mathcal{L}(\mathbf{g})\|_{\mathrm{F}} \le c_2 n^2 \delta,$$

provided that (11) *is satisfied with c_1.*

Proof of Theorem II.3. Theorem II.2 extends to the noisy case similarly to the previous work [10]. We only need to replace [10, Lemma 1] by our Lemma A.6. □

Note that Theorem II.3 provides an improved performance guarantee with significantly smaller noise amplification factor, compared to n^3 dependent noisy amplification factor in the previous work [10].

III. Guaranteed Reconstruction of FRI Signals

The explicit derivation of the minimum length finite length annihilating filter was one of the most important contributions of the sampling theory of FRI signals [13]–[15]. Therefore, by combing the results in the previous

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

section, we can provide performance guarantees for the recovery of FRI signals from partial Fourier measurements.

A. Spectral Compressed Sensing: Recovery of Stream of Diracs

Consider the periodic stream of Diracs described by the superposition of r impulses

$$x(t) = \sum_{j=0}^{r-1} a_j \delta(t - t_j) \qquad t_j \in [0,1]. \tag{13}$$

Then, the discrete Fourier data are given by

$$\hat{x}[k] = \sum_{j=0}^{r-1} a_j e^{-i2\pi k t_j} \ . \tag{14}$$

As mentioned before, the spectral compressed sensing by Chen and Chi [10] or Tang [9] correspond to this case, in which they are interested in recovering (13) from a subsampled spectral measurements.

One of the important contributions of this section is to show that the spectral compressed sensing can be equivalently explained using annihilating filter-based low-rank Hankel matrix. Specifically, for the stream of Diracs, the minimum length annihilating filter $\hat{h}[k]$ has the following z-transform representation [13]:

$$\hat{h}(z) \quad = \quad \sum_{l=0}^{r} \hat{h}[l] z^{-l} = \prod_{j=0}^{r-1} (1 - e^{-i2\pi t_j} z^{-1}) \ , \tag{15}$$

because

$$\begin{aligned} (\hat{h} * \hat{x})[k] &= \sum_{l=0}^{k} \hat{h}[l]\hat{x}[k-l] \\ &= \sum_{l=0}^{r} \sum_{j=0}^{r-1} a_j \hat{h}[l] u_j^{k-l} \\ &= \sum_{j=0}^{r-1} a_j \underbrace{\left(\sum_{l=0}^{r} \hat{h}[p] u_j^{-l} \right)}_{\hat{h}(u_j)} u_j^k = 0 \end{aligned} \tag{16}$$

where $u_j = e^{-i2\pi t_j}$ [13]–[15]. Accordingly, the filter length is $r+1$, which is low ranked if $\min\{n-d+1, d\} > r$.

Therefore, by utilizing Theorem II.1 and Theorem II.2, we can provide the performance guarantee of the following nuclear norm minimization to estimate the Fourier samples:

$$\begin{aligned} \min_{\mathbf{g} \in \mathbb{C}^n} \quad & \|\mathscr{H}(\mathbf{g})\|_* \\ \text{subject to} \quad & P_\Omega(\mathbf{g}) = P_\Omega(\hat{\mathbf{x}}) \end{aligned} \tag{17}$$

where $\mathscr{H}(\mathbf{g}) \in \mathcal{H}(n,d)$.

Theorem III.4. *For a given stream of Diracs in Eq.* (13)*, $\hat{\mathbf{x}}$ denotes the noiseless discrete Fourier data in* (5)*. Suppose, furthermore, d is given by $\min\{n-d+1, d\} > r$. Let $\Omega = \{j_1, \ldots, j_m\}$ is a multi-set consisting of random indices where j_k's are i.i.d. following the uniform distribution on $\{0, \ldots, n-1\}$. Then, there exists an absolute constant c_1 such that $\hat{\mathbf{x}}$ is the unique minimizer to* (17) *with probability $1 - 1/n^2$, provided*

$$m \geq c_1 \mu c_s r \log^4 n, \tag{18}$$

where $c_s := \max\{n/(n-d+1), n/d\}$.

Proof. This is a simple consequence of Theorem II.1 and Theorem II.2, because the minimum annihilating filter size from (15) is $r+1$. □

This result appears identical to that of Chen and Chi [10]. However, they explicitly utilized the standard Vandermonde decomposition. On the contrary, the annihilating filter-based construction of low-rank Hankel matrix is more general that can cover all FRI signal models as will be shown later.

For the noisy measurement, we interpolate the missing Fourier data using the following low-rank matrix completion:

$$\begin{aligned} \min \ & \|\mathscr{H}(\mathbf{g})\|_* \\ \text{subject to} \quad & \|P_\Omega(\mathbf{g}) - P_\Omega(\hat{\mathbf{y}})\| \leq \delta \end{aligned} \tag{19}$$

where $\hat{\mathbf{y}}$ is the noisy Fourier data. Then, Theorem II.3 informs us that we can improve upon the results by Chen and Chi [10] (from n^3 to n^2):

Theorem III.5. *Suppose that the noisy Fourier data $\hat{\mathbf{y}}$ satisfies $\|P_\Omega(\hat{\mathbf{y}} - \hat{\mathbf{x}})\|_2 \leq \delta$, where $\hat{\mathbf{x}}$ denotes the noiseless discrete Fourier data in* (5)*. Under the hypotheses of Theorem III.4, there exists an absolute constant c_1, c_2 such that with probability $1 - 1/n^2$, the solution $\mathbf{g}$ to* (19) *satisfies*

$$\|\mathscr{H}(\hat{\mathbf{x}}) - \mathscr{H}(\mathbf{g})\|_{\mathrm{F}} \leq c_2 n^2 \delta,$$

provided that (18) *is satisfied with c_1.*

B. Stream of Differentiated Diracs

Another important class of FRI signal is a stream of differentiated Diracs:

$$x(t) = \sum_{j=0}^{r-1} \sum_{l=0}^{l_j - 1} a_{j,l} \delta^{(l)}(t - t_j) \ , \tag{20}$$

where $\delta^{(l)}$ denotes the l-th derivative of Diracs in the distributions sense. Thus, its Fourier transform is given by

$$\hat{x}(f) = \sum_{j=0}^{r-1} \sum_{l=0}^{l_j - 1} a_{j,l} (i2\pi f)^l e^{-i2\pi f t_j} \tag{21}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

whose discrete samples are given by

$$\hat{x}[k] := \hat{x}(k) = \sum_{j=0}^{r-1} \sum_{l=0}^{l_j-1} a_{j,l} \left(i2\pi k\right)^l e^{-i2\pi k t_j}. \qquad (22)$$

Then, there exists an associated minimum length annihilating filter whose z-transform is given by:

$$\hat{h}(z) = \prod_{j=0}^{r-1} (1 - u_j z^{-1})^{l_j} \qquad (23)$$

where $u_j = e^{-i2\pi t_j}$ [13]. Therefore, we can provide the following performance guarantees:

Theorem III.6. *For a given stream of differentiated Diracs in Eq.* (20)*, $\hat{\mathbf{x}}$ denotes the noiseless discrete Fourier data in* (5)*. Suppose, furthermore, d is given by $\min\{n-d+1, d\} > \sum_{j=0}^{r-1} l_j$. Let $\Omega = \{j_1, \ldots, j_m\}$ be a multi-set consisting of random indices where j_k's are i.i.d. following the uniform distribution on $\{0, \ldots, n-1\}$. Then, there exists an absolute constant c_1 such that $\hat{\mathbf{x}}$ is the unique minimizer to* (17) *with probability $1 - 1/n^2$, provided*

$$m \geq c_1 \mu c_s \left(\sum_{j=0}^{r-1} l_j \right) \log^4 n, \qquad (24)$$

where $c_s := \max\{n/(n-d+1), n/d\}$.

Proof. This is a simple consequence of Theorem II.1 and Theorem II.2, because the minimum annihilating filter size from (23) is $\left(\sum_{j=0}^{r-1} l_j\right) + 1$. □

Theorem III.7. *Suppose the noisy Fourier data $\hat{\mathbf{y}}$ satisfies $\|P_\Omega(\hat{\mathbf{y}} - \hat{\mathbf{x}})\|_2 \leq \delta$, where $\hat{\mathbf{x}}$ denotes the noiseless discrete Fourier data in* (5)*. Under the hypotheses of Theorem III.6, there exists an absolute constant c_1, c_2 such that with probability $1 - 1/n^2$, the solution $\mathbf{g}$ to* (19) *satisfies*

$$\|\mathscr{H}(\hat{\mathbf{x}}) - \mathscr{H}(\mathbf{g})\|_{\mathrm{F}} \leq c_2 n^2 \delta,$$

provided that (24) *is satisfied with c_1.*

C. Non-uniform Splines

Note that signals may not be sparse in the image domain, but can be sparsified in a transform domain. Our goal is to find a generalized framework, whose sampling rate can be reduced down to the transform domain sparsity level. Specifically, the signal x of our interest is a non-uniform spline that can be represented by :

$$\mathrm{L}x = w \qquad (25)$$

where L denotes a constant coefficient linear differential equation that is often called the continuous domain whitening operator in [26], [27]:

$$\mathrm{L} := b_K \partial^K + b_{K-1} \partial^{K-1} + \ldots + b_1 \partial + b_0 \qquad (26)$$

and w is a continuous sparse innovation:

$$w(t) = \sum_{j=0}^{r-1} a_j \delta(t - t_j) \quad . \qquad (27)$$

For example, if the underlying signal is piecewise constant, we can set L as the first differentiation. In this case, x corresponds to the total variation signal model. Then, by taking the Fourier transform of (25), we have

$$\hat{z}(f) := \hat{l}(f)\hat{x}(f) = \sum_{j=0}^{r-1} a_j e^{-i2\pi f t_j} \qquad (28)$$

where

$$\hat{l}(f) = b_K (i2\pi f)^K + b_{K-1}(i2\pi f)^{K-1} + \ldots + b_1 (i2\pi f) + b_0$$

Accordingly, the same filter $\hat{h}[n]$ whose z-transform is given by (15) can annihilate the discrete samples of the weighted spectrum $\hat{z}(f) = \hat{l}(f)\hat{x}(f)$, and the Hankel matrix $\mathscr{H}(\hat{\mathbf{z}}) \in \mathcal{H}(n, d)$ from the weighted spectrum $\hat{z}(f)$ satisfies the following rank condition:

$$\mathrm{RANK}\, \mathscr{H}(\hat{\mathbf{z}}) = r.$$

Thanks to the low-rankness, the missing Fourier data can be interpolated using the following matrix completion problem:

$$(P_w) \quad \min_{\mathbf{g} \in \mathbb{C}^n} \quad \|\mathscr{H}(\mathbf{g})\|_* \quad \text{subject to} \quad P_\Omega(\mathbf{g}) = P_\Omega(\hat{\mathbf{l}} \odot \hat{\mathbf{x}}) \ , \qquad (29)$$

or, for noisy Fourier measurements $\hat{\mathbf{y}}$,

$$(P'_w) \quad \min_{\mathbf{g} \in \mathbb{C}^n} \quad \|\mathscr{H}(\mathbf{g})\|_* \quad \text{subject to} \quad \|P_\Omega(\mathbf{g}) - P_\Omega(\hat{\mathbf{l}} \odot \hat{\mathbf{y}})\| \leq \delta \ , \qquad (30)$$

where $\odot$ denotes the Hadamard product, and $\hat{\mathbf{l}}$ and $\hat{\mathbf{x}}$ denotes the vectors composed of full samples of $\hat{l}[k]$ and $\hat{x}[k]$, respectively. After solving (P_w), the missing spectral data $\hat{x}[k]$ can be obtained by dividing by the weight, i.e. $\hat{x}[k] = g[k]/\hat{l}[k]$ assuming that $\hat{l}[k] \neq 0$, where $g[k]$ is the estimated Fourier data using (P_w). As for the sample $\hat{x}[k]$ at the spectral null of the filter $\hat{l}[k]$, the corresponding elements should be included as measurements.

Now, we can provide the following performance guarantee:

Theorem III.8. *For a given non-uniform splines in Eq.* (25)*, $\hat{\mathbf{x}}$ denotes the noiseless discrete Fourier data in* (5)*. Suppose, furthermore, d is given by $\min\{n-d+1, d\} > r$ and $\Omega = \{j_1, \ldots, j_m\}$ be a multi-set consisting of random indices where j_k's are i.i.d. following the uniform distribution on $\{0, \ldots, n-1\}$. Then, there exists an absolute constant c_1 such that $\hat{\mathbf{x}}$ is the unique minimizer to* (29) *with probability $1 - 1/n^2$, provided*

$$m \geq c_1 \mu c_s r \log^4 n, \qquad (31)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

where $c_s := \max\{n/(n-d+1), n/d\}$.

Theorem III.9. *Suppose that the noisy Fourier data* $\hat{\mathbf{y}}$ *satisfies* $\left\|P_\Omega(\hat{\mathbf{l}} \odot \hat{\mathbf{y}} - \hat{\mathbf{l}} \odot \hat{\mathbf{x}})\right\|_2 \le \delta$, *where* $\hat{\mathbf{x}}$ *denotes the noiseless discrete Fourier data in* (5). *Under the hypotheses of Theorem III.8, there exists an absolute constant* c_1, c_2 *such that with probability* $1 - 1/n^2$, *the solution* $\mathbf{g}$ *to* (30) *satisfies*

$$\left\|\mathscr{H}(\hat{\mathbf{l}} \odot \hat{\mathbf{x}}) - \mathscr{H}(\mathbf{g})\right\|_{\mathrm{F}} \le c_2 n^2 \delta,$$

provided that (31) *is satisfied with* c_1.

D. Piecewise Polynomials

A signal is a periodic piecewise polynomial with r pieces each of maximum degree q if and only if its $(q+1)$ derivative is a stream of differentiated Diracs given by

$$x^{(q+1)}(t) = \sum_{j=0}^{r-1} \sum_{l=0}^{q} a_{j,l} \delta^{(l)}(t - t_j) \ . \tag{32}$$

In this case, the corresponding Fourier transform relationship is given by

$$\hat{z}(f) := (i2\pi f)^{(q+1)} \hat{x}(f) = \sum_{j=0}^{r-1} \sum_{l=0}^{q} a_{j,l} (i2\pi f)^l e^{-i2\pi t_j}. \tag{33}$$

Since the righthand side of (33) is a special case of (21), the associated minimum length annihilating filter has the following z-transform representation:

$$\hat{h}(z) = \prod_{j=0}^{r-1} (1 - u_j z^{-1})^q \ . \tag{34}$$

whose filter length is given by $(q+1)r + 1$. Therefore, we can provide the following performance guarantee:

Theorem III.10. *For a given piecewise smooth polynomial in Eq.* (32), *let* $\hat{\mathbf{z}}$ *denotes the discrete spectral samples of* $\hat{z}(f) = \hat{l}(f)\hat{x}(f)$ *with* $\hat{l}(f) = (i2\pi f)^{(q+1)}$. *Suppose, furthermore,* d *is given by* $\min\{n-d+1, d\} > (q+1)r$ *and* $\Omega = \{j_1, \ldots, j_m\}$ *be a multi-set consisting of random indices where* j_k*'s are i.i.d. following the uniform distribution on* $\{0, \ldots, n-1\}$. *Then, there exists an absolute constant* c_1 *such that* $\hat{\mathbf{x}}$ *is the unique minimizer to* (29) *with probability* $1 - 1/n^2$, *provided*

$$m \ge c_1 \mu c_s (q+1) r \log^4 n, \tag{35}$$

where $c_s := \max\{n/(n-d+1), n/d\}$.

Proof. This is a simple consequence of Theorem II.1 and Theorem II.2, because the minimum annihilating filter size from (34) is $(q+1)r + 1$. □

Theorem III.11. *Suppose that noisy Fourier data* $\hat{\mathbf{y}}$ *satisfies* $\left\|P_\Omega(\hat{\mathbf{l}} \odot \hat{\mathbf{y}} - \hat{\mathbf{l}} \odot \hat{\mathbf{x}})\right\|_2 \le \delta$, *where* $\hat{\mathbf{x}}$ *denotes the noiseless discrete Fourier data in* (5). *Under the hypotheses of Theorem III.10, there exists an absolute constant* c_1, c_2 *such that with probability* $1 - 1/n^2$, *the solution* $\mathbf{g}$ *to* (30) *satisfies*

$$\left\|\mathscr{H}(\hat{\mathbf{l}} \odot \hat{\mathbf{x}}) - \mathscr{H}(\mathbf{g})\right\|_{\mathrm{F}} \le c_2 n^2 \delta,$$

provided that (35) *is satisfied with* c_1.

E. Incoherence and the Minimum Separation

Note that the proposed low-rank interpolation achieves near optimal sample rate while universally applying to different FRI signal models of the same order (e.g., stream of Diracs and stream of differentiated Diracs). Moreover, even though the concept of the minimum separation between the successive spikes was essential for the performance guarantee of super-resolution in Candes and Fernandez-Granda [8], [11], Tang [9], [12], etc, similar expression is not observable in Theorem III.4-Theorem III.11. This looks mysterious. Therefore, the main goal of this section is to show that these information are still required but hidden in the incoherence parameter μ.

Note that the proof in Theorem II.1 implies that the explicit form $\hat{x}[k]$ given by

$$\hat{x}[k] := \sum_{j=0}^{p-1} \sum_{l=0}^{l_j - 1} a_{j,l} k^l \lambda_j{}^k, \ \text{where} \quad r = \sum_{j=0}^{p-1} l_j, \tag{36}$$

is a necessary and sufficient condition to have the low-rank Hankel matrix. Here, $\lambda_j = e^{-i2\pi t_j}$ for FRI signals. Furthermore, for a Hankel matrix constructed using the signal model in (36), there exist an exponentional decomposition of Hankel matrix using *confluent* Vandermonde matrix [28]. Specifically, define a confluent Vandermonde matrix $\mathcal{V}_{n-d+1} \in \mathbb{C}^{(n-d+1)\times r}$ (resp. $\mathcal{V}_d \in \mathbb{C}^{d\times r}$):

$$\mathcal{V}_{n-d+1} = \left[C^{l_0}_{n-d+1}(\lambda_0) \quad C^{l_1}_{n-d+1}(\lambda_1) \quad \cdots \quad C^{l_{p-1}}_{n-d+1}(\lambda_{m_{p-1}})\right],$$

where the (m,l) element of the sub-matrix $C^{l_i}_{n-d+1}(\lambda) \in \mathbb{C}^{(n-d+1)\times l_i}$ is given by

$$\left[C^{l}_{n-d+1}(\lambda)\right]_{m,l} = \begin{cases} 0, & m < l \\ \frac{(m-1)!}{(m-l)!}\lambda^{i-j}, & \text{otherwise} \end{cases}. \tag{37}$$

Then, the associated Hankel matrix $\mathscr{H}(\hat{\mathbf{x}}) \in \mathcal{H}(n,d)$ with $\min\{n-d+1, d\} > r$ has the following *generalized Vandermonde decomposition* [28]–[31]:

$$\mathscr{H}(\hat{\mathbf{x}}) = \mathcal{V}_{n-d+1} \ \mathcal{B} \ \mathcal{V}_d^T, \tag{38}$$

where $\mathcal{V}_{n-d+1} \in \mathbb{C}^{(n-d+1)\times r}$ and $\mathcal{V}_d \in \mathbb{C}^{d\times r}$ are the confluent Vandermonde matrices and $\mathcal{B}$ is a $r \times r$ block diagonal matrix given by

$$\mathcal{B} = \begin{bmatrix} H_0 & 0 & \cdots & 0 \\ 0 & H_1 & \ddots & \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & H_{p-1} \end{bmatrix},$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

where H_i is the $m_i \times m_i$ upper anti-triangular Hankel matrix [29]. Because $\mathcal{B}$ in (38) is a full rank block diagonal matrix, for a given SVD of $\mathscr{H}(\hat{\mathbf{x}}) = U\Sigma V^H$ with $U \in \mathbb{C}^{(n-d+1)\times r}, V \in \mathbb{C}^{d\times r}$ and $\Sigma \in \mathbb{C}^{r\times r}$, we have

$$\text{RAN } U = \text{RAN } \mathcal{V}_{n-d+1}, \quad \text{RAN } V = \text{RAN } \mathcal{V}_d.$$

Accordingly, we can derive the following upper bound of the standard coherence:

Lemma III.12. *For a Hankel matrix $\mathscr{H}(\hat{\mathbf{x}}) \in \mathcal{H}(n,d)$ with the decomposition in* (38)*, the standard coherence μ in* (9) *satisfies:*

$$\mu \le \max\left\{ \frac{\zeta_{n-d+1}}{\sigma_{\min}\left(\mathcal{V}_{n-d+1}^*\mathcal{V}_{n-d+1}\right)}, \frac{\zeta_d}{\sigma_{\min}\left(\mathcal{V}_d^*\mathcal{V}_d\right)} \right\} \quad (39)$$

where $\sigma_{\min}(\cdot)$ denotes the least singular value and the constant $\zeta_N, N \in \mathbb{N}$ is defined by

$$\zeta_N \;=\; N\left[\frac{(N-1)!}{(N-l_{\max})!}\right]^2 \quad , \quad (40)$$

and $l_{\max} := \max\limits_{0\le j\le p-1} l_j$.

Proof. Since U (resp. $\mathcal{V}_{n-d+1}$) and V (resp. $\mathcal{V}_d$) determine the same column (resp. row) space, we can write

$$\begin{aligned} UU^* &= \mathcal{V}_{n-d+1}\left(\mathcal{V}_{n-d+1}^*\mathcal{V}_{n-d+1}\right)^{-1}\mathcal{V}_{n-d+1}^* \\ VV^* &= \mathcal{V}_d\left(\mathcal{V}_d^*\mathcal{V}_d\right)^{-1}\mathcal{V}_d^* \end{aligned}$$

Thus, we have

$$\begin{aligned} \max_{1\le i\le d}\|V^*\mathbf{e}_i\|_2^2 &= \max_{1\le i\le d}\mathbf{e}_i^*\mathcal{V}_d\left(\mathcal{V}_d^*\mathcal{V}_d\right)^{-1}\mathcal{V}_d^*\,\mathbf{e}_i \\ &\le \frac{1}{\sigma_{\min}\left(\mathcal{V}_d^*\mathcal{V}_d\right)}\max_{1\le i\le d}\|\mathcal{V}_d^*\mathbf{e}_i\|^2 \end{aligned}$$

Moreover, we have

$$\begin{aligned} \max_{1\le i\le d}\|\mathcal{V}_d^*\mathbf{e}_i\|^2 &= \|\mathcal{V}_d^*\mathbf{e}_d\|^2 \\ &= \sum_{j=0}^{p-1}\sum_{l=1}^{l_j}\left[\frac{(d-1)!}{(d-l)!}\right]^2 \\ &\le r\left[\frac{(d-1)!}{(d-l_{\max})!}\right]^2 \end{aligned}$$

where we use (37) and $r = \sum_{j=0}^{p-1} l_j$. Similarly,

$$\max_{1\le i\le n-d+1}\|U^*\mathbf{e}_i\|_2^2$$

$$\le \frac{r}{\sigma_{\min}\left(\mathcal{V}_{n-d+1}^*\mathcal{V}_{n-d+1}\right)}\left[\frac{(n-d)!}{(n-d+1-l_{\max})!}\right]^2.$$

Therefore, using the definition of μ in (9), we can obtain (39). This concludes the proof. □

Note that this is an extension of the approach in [10, Appendix C, Section III.A] which tried to bound the mutual coherence for the standard Vandermonde decomposition, (i.e. $l_0 = \cdots = l_{p-1} = 1$) by a small number. Specifically, for the cases of random frequency locations or small perturbation off the grid, they showed that the incoherence parameters become small [10]. However, the dependency of μ on the minimum separation was not explicit in their discussion.

Recently, Moitra [32] discovered a very intuitive relationship between the least/largest singular values of Vandermonde matrix and the minimum separation distance $\Delta = \min_{i\neq j}|t_i - t_j|$. Specifically, by making novel connections between extremal functions and the spectral properties of Vandermonde matrices $\mathcal{V}_N$, Moitra showed that if $N > 1/\Delta + 1$, then the least singular value is bounded as

$$\sigma_{\min}(\mathcal{V}_N^*\mathcal{V}_N) \ge N - 1/\Delta - 1.$$

If applied in our problem involving $\mathcal{V}_{n-d+1}$ and $\mathcal{V}_d$, the resulting upper bound of the coherence parameter for standard Vandermonde matrix is given by

$$\mu \le \frac{n/2}{n/2 - 1/\Delta - 1}, \quad (41)$$

which approaches to one with sufficiently large n. Eq. (41) is obtained because the matrix pencil size $d = n/2$ gives the optimal trade-off between $\sigma_{\min}\left(\mathcal{V}_{n-d+1}^*\mathcal{V}_{n-d+1}\right)$ and $\sigma_{\min}\left(\mathcal{V}_d^*\mathcal{V}_d\right)$, and $\zeta_{n/2} = n/2$ owing to $l_{\max} = 1$. Note that this coincides with the minimum separation in Fernandez-Granda [8], [11] and Tang [9], [12]. However, compared to these approaches [8], [9], [11], [12] that require the minimum separation as a *hard* constraint, our approach requires it as a *soft* oversampling factor in terms of the incoherence parameter μ.

Then, where is the difference originated ? We argue that this comes from different uses of interpolation functions. Specifically, in Candes, Fernandez-Granda [8], [11] and Tang [9], [12], dual polynomial function that interpolates the sign at the singularity locations should be found to construct a dual certificate. On other hand, in the proposed annihilating filter based approach, the interpolating function is a smooth function that has zero-crossings at the singularity locations. To see this, let $\hat{h}[k]$ is an annihilating filter that annihilates $\hat{x}[k]$. Then there exists an *annihilating function* $h(t)$ such that

$$\hat{x}[k] * \hat{h}[k] = 0, \forall k \iff x(t)h(t) = 0, \quad \forall t,$$

so $h(t) = 0$ whenever $x(t) \neq 0$. The construction of the annihilating function $h(t)$ is extremely easy and can be readily obtained by the multiplications of sinusoids (for example, to null out r-periodic stream of Diracs within $[0,1]$, we set $f(t) = \prod_{j=0}^{r-1}(e^{i2\pi t} - e^{i2\pi t_j})$). Moreover, this approach can be easily extended to have multiple

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

roots, which is required for differentiated Diracs. We believe that the "soft constraint" originated from annihilating function is one of the key ingredients that enables recovery of general FRI signals which was not possible by the existing super-resolution methods [8], [9], [11], [12].

The derivation of the least singular value for the confluence Vandermonde matrix have been also an important topic of researches [30], [31], [33]–[37]. In general, it will also depend on the minimum separation distance [33]. However, the explicit tight bound of the least singular value is not available in general, so we leave this for future work.

F. Recovery of Continuous Domain FRI Signals After Interpolation

Regardless of the unknown signal type (the stream of Diracs or a stream of differentiated Diracs), note that an identical low-rank interpolator can be used. Once the spectrum $\hat{x}[k]$ is fully interpolated, in the subsequent step, Prony's method and matrix pencil algorithm can identify the signal model from the roots of the estimated annihilator filter as done in [13]–[15]. Accordingly, our robustness guarantees on the low-rank matrix entries can be translated in terms of the actual signal that is recovered (for example, on the support or amplitudes of the spike in the case of recovery of spike superpositions). In fact, this has been also an active area of researches [29]–[33], [38], and we again exploit these findings for our second step of signal recovery. For example, see Moitra [32] for more details on the error bound for the case of modified matrix pencil approach for recovery of Diracs. In addition, Batenkov has recently generalized this for the recovery of general signals in (36) [33]. The common findings are that the estimation error for the location parameter $\{t_j\}_{j=0}^{p-1}$ and the magnitude $a_{j,l}$ are bounded by the condition number of the confluent Vandermonde matrix as well as the minimum separation distance Δ. Moreover, matrix pencil approaches such as Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) method [39] is shown to stably recovery the locations [33], [38].

Here, we briefly review the matrix pencil approach for the signal recovery [29], [40]. Specifically, for a given confluent Vandermonde matrix $\mathcal{V}_{n-d+1}$, let $\mathcal{V}^{\downarrow}_{n-d+1}$ be the matrix extracted from $\mathcal{V}_{n-d+1}$ by deleting the last row. Similarly, let $\mathcal{V}^{\uparrow}_{n-d+1}$ be the matrix extracted from $\mathcal{V}_{n-d+1}$ by deleting the first row. Then, $\mathcal{V}^{\downarrow}_{n-d+1}$ and $\mathcal{V}^{\uparrow}$ span the same signal subspace and

$$\mathcal{V}^{\uparrow}_{n-d+1} = \mathcal{V}^{\downarrow}_{n-d+1} J$$

where J is the $r \times r$ block diagonal matrix

$$J = \begin{bmatrix} J_{m_0}(\lambda_0) & 0 & \cdots & 0 \\ 0 & J_{m_1}(\lambda_1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & J_{m_{p-1}}(\lambda_{p-1}) \end{bmatrix},$$

where $J_{m_i}(\lambda_i)$ denotes the $m_i \times m_i$ Jordan block [28], [29]:

$$J_{m_i} = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ 0 & 0 & \lambda_i & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 0 & \lambda_i \end{bmatrix}.$$

In practice, the confluence Vandermonde matrix $\mathcal{V}_{n-d+1}$ is unknown, but a $(n-d+1) \times r$ matrix W that spans the signal subspace can be estimated using singular value decomposition (SVD). Then, we can easily see that

$$W^{\uparrow} = W^{\downarrow}\Phi$$

where $r \times r$ spectral matrix Φ is given by

$$\Phi = GJG^{-1}$$

for some matrix G. Finally, the matrix pencil algorithm computes the eigenvalues of Φ matrix from which the estimated poles and their multiplicities are estimated.

IV. Guaranteed Reconstruction of Cardinal L-Splines

A. Cardinal L-Spline Model

A cardinal spline is a special case of a non-uniform spline where the knots are located on the integer grid [16], [26], [27]. More specifically, a function $x(t)$ is called a *cardinal* L*-spline* if and only if

$$\mathrm{L}x(t) = w(t), \tag{42}$$

where the operator L is continuous domain whitening operator and and the continuous domain *innovation* signal $w(t)$ is given by

$$w(t) := \sum_{p\in\mathbb{Z}} a[p]\delta(t-p)\ , \tag{43}$$

whose singularities are located on integer grid.

Even though the recovery of cardinal L-splines can be considered as special instance of that of non-uniform splines, the cardinal setting allows high but finite resolution, so it is closely related to standard compressed sensing framework in discrete framework. Therefore, this section provides more detailed discussion of recovery of cardinal L-splines from partial Fourier measurements. The analysis in this section is significantly influenced by the theory of sparse stochastic processes [16], so we follow the original authors's notation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

B. Construction of Low-Rank Wrap-around Hankel Matrix

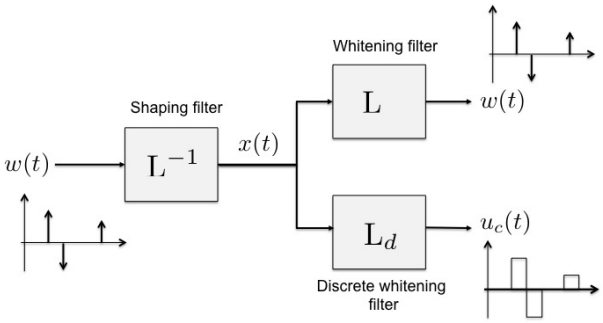

Fig. 2. The discrete innovation and continuous domain innovations generated by L_d and L, respectively.

The main advantage of using cardinal setup is that we can recover signals by exploiting the sparseness of *discrete innovation* rather than exploiting off the grid singularity. So, we are now interested in deriving the discrete counterpart of the whitening operator L, which is denoted by L_d:

$$\mathrm{L}_d\delta(t) = \sum_{p\in\mathbb{Z}} l_d[p]\delta(t-p). \tag{44}$$

Now, by applying the discrete version of whitening operator L_d to $x(t)$, we have

$$\begin{aligned} u_c(t) &:= \mathrm{L}_d x(t) = \mathrm{L}_d\mathrm{L}^{-1}w(t) = (\beta_L * w)(t) \\ &= \sum_{p\in\mathbb{Z}} a[p]\beta_L(t-p)\ . \end{aligned} \tag{45}$$

where $\beta_L(t)$ denotes a generalized B-spline associated with the operator L [16], which is defined by

$$\beta_L(t) = \mathrm{L}_d\mathrm{L}^{-1}\delta(t) = \mathcal{F}^{-1}\left\{\frac{\sum_{p\in\mathbb{Z}} l_d[j]e^{-i\omega p}}{\hat{l}(\omega)}\right\}(t)\ , \tag{46}$$

where we now use $\omega = 2\pi f$ for Fourier transform to follow the notation in [16]. As shown in Fig. 2, $u_c(t)$ is indeed a *smoothed* version of continuous domain innovation $w(t)$ in (43), because all the sparsity information of the innovation $w(t)$ is encoded in its coefficients $\{a[p]\}$, and aside from the interpolant $\beta_L(t)$, $u_c(t)$ in (45) still retains the same coefficients. Moreover, the sparseness of sampled discrete innovation on the integer grid can be identified from the discrete samples of $u_c(t)$:

$$\begin{aligned} u(t) &:= u_c(t)\sum_{p\in\mathbb{Z}}\delta(t-p) \\ &= \sum_{p\in\mathbb{Z}} u_d[p]\delta(t-p) \qquad (47) \\ &= \sum_{p\in\mathbb{Z}} (a * b_L)[p]\delta(t-p) \qquad (48) \end{aligned}$$

where

$$b_L[p] := \beta_L(t)|_{t=p}\,. \tag{49}$$

To make the discrete sample $u_d[p]$ sparse, the discrete filter $b_L[p]$ should be designed to have the minimum non-zero support. Due to the relationship (49), this can be achieved if $\beta_L(t)$ is maximally localized. The associated DFT spectrum of the discrete innovation is given by

$$\hat{u}_d[k] = \hat{u}(\omega)|_{\omega=\frac{2\pi k}{n}} = \sum_{j=0}^{r-1} u_j e^{-i\frac{2\pi k i_j}{n}} \tag{50}$$

where $\{u_j\}$ denotes the non-zero coefficient of $u[p]$ and i_j refers the corresponding index.

To exploit the sparseness of discrete innovation using the low-rank Hankel matrix, we should relate the discrete innovation to the discrete samples of the unknown cardinal L-spline $x(t)$. This can be done using an equivalent B-spline representation of $x(t)$ [16]:

$$x(t) \quad = \quad \sum_{p\in\mathbb{Z}} c[p]\beta_L(t-p)\ , \tag{51}$$

where $c[p]$ satisfies

$$a[p] = (c * l_d)[p]$$

for $a[p]$ and $l_d[p]$ in (42) and (44), respectively. Here, the equivalent B-spline representation in (51) can be shown by:

$$\mathrm{L}x(t) = \sum_{p\in\mathbb{Z}} c[p]L\beta_L(t-p) = \sum_{l\in\mathbb{Z}} \underbrace{(c * l_d)[p]}_{a[p]}\delta(t-p)\ ,$$

because $\mathrm{L}\mathrm{L}_d\mathrm{L}^{-1}\delta(t) = \mathrm{L}_d\delta(t)$. So, we have

$$\begin{aligned} u(t) &= \sum_{p\in\mathbb{Z}} (a * b_L)[p]\delta(t-p) \\ &= \sum_{p\in\mathbb{Z}} (l_d * c * b_L)[p]\delta(t-p) \\ &= \sum_{p\in\mathbb{Z}} (l_d * x_d)[p]\delta(t-p) \end{aligned}$$

where

$$x_d[p] := x(t)|_{t=p} = \sum_{l\in\mathbb{Z}} c[l]\beta(p-l) = (c * b_L)[p]\ . \tag{52}$$

Therefore, $u_d[p] = (l_d * x_d)[p]$ and the corresponding DFT spectrum is given by

$$\hat{u}_d[k] = \hat{l}_d[k]\hat{x}_d[k], \quad k = 0,\cdots,n-1. \tag{53}$$

Because the DFT data $\hat{x}_d[k]$ can be computed and $\hat{l}_d[k]$ is known, we can construct a Hankel matrix $\mathscr{H}(\hat{\mathbf{u}}_d) = \mathscr{H}(\hat{\mathbf{l}}_d \odot \hat{\mathbf{x}}_d) \in \mathcal{H}(n,d)$. Thanks to (50), the associated minimum size annihilating filter $\hat{h}[k]$ that cancels $\hat{x}_d[k]$ can be obtained from the following z-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

transform expression

$$\hat{h}(z) = \prod_{j=0}^{r-1}(1 - e^{-i\frac{2\pi k l_j}{n}} z^{-1})$$

whose length is $r+1$. Therefore, we have

$$\text{RANK}\mathscr{H}(\hat{\mathbf{u}}_d) = \text{RANK}\mathscr{H}(\hat{\mathbf{l}}_d \odot \hat{\mathbf{x}}_d) = r.$$

Moreover, due to the periodicity of DFT spectrum, we can use the following wrap-around Hankel matrix:

$$\mathscr{H}_c(\hat{\mathbf{u}}_d) =$$

$$\left[\begin{array}{cccc} \hat{u}_d[0] & \hat{u}_d[1] & \cdots & \hat{u}_d[d-1] \\ \hat{u}_d[1] & \hat{u}_d[2] & \cdots & \hat{u}_d[d] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{u}_d[n-d] & \hat{u}_d[n-d+1] & \cdots & \hat{u}_d[n-1] \\ \hline \hat{u}_d[n-d+1] & \hat{u}_d[n-d+2] & \cdots & \hat{u}_d[0] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{u}_d[n-1] & \hat{u}_d[0] & \cdots & \hat{u}_d[d-2] \end{array}\right] \in \mathbb{C}^{n\times d}$$

where the bottom block is an augmented block. Since the bottom block can be also annihilated using the same annihilating filter, we can see the rank of the wrap-around Hankel expansion is the same as the original Hankel structured matrix:

$$\text{RANK}\mathscr{H}_c(\hat{\mathbf{u}}_d) = \text{RANK}\mathscr{H}(\hat{\mathbf{u}}_d) = r.$$

Then, the missing DFT coefficients can be interpolated using the following low-rank matrix completion:

$$\begin{aligned} \min_{\mathbf{g}\in\mathbb{C}^n} \quad & \|\mathscr{H}_c(\mathbf{g})\|_* \\ \text{subject to} \quad & P_\Omega(\mathbf{g}) = P_\Omega(\hat{\mathbf{l}}_d \odot \hat{\mathbf{x}}_d)\ , \end{aligned} \tag{54}$$

or

$$\begin{aligned} \min_{\mathbf{g}\in\mathbb{C}^n} \quad & \|\mathscr{H}_c(\mathbf{g})\|_* \\ \text{subject to} \quad & \|P_\Omega(\mathbf{g}) - P_\Omega(\hat{\mathbf{l}}_d \odot \hat{\mathbf{y}}_d)\| \le \delta\ , \end{aligned} \tag{55}$$

for noisy DFT data $\hat{\mathbf{y}}_d$. Then, we have the following performance guarantee:

Theorem IV.13. *For a given cardinal L-spline $x(t)$ in Eq.* (42)*, let $\hat{l}_d[k]$ denotes the DFT of discrete whitening operator and $\hat{x}_d[k]$ is the DFT of the discrete sample $x_d[p]$ in* (52)*. Suppose, furthermore, d is given by $\min\{n-d+1, d\} > r$ and $\Omega = \{j_1, \ldots, j_m\}$ be a multi-set consisting of random indices where j_k's are i.i.d. following the uniform distribution on $\{0, \ldots, n-1\}$. Then, there exists an absolute constant c_1 such that $\hat{\mathbf{x}}$ is the unique minimizer to* (54) *with probability $1 - 1/n^2$, provided*

$$m \ge c_1 c_s \mu r \log^2 n. \tag{56}$$

where $c_s = n/d$ and μ is the incoherence parameter.

Proof. The associated Hankel matrix has wrap-around property, so the log power factor is reduced to 2, and $c_s = \max\{n/n, n/d\} = n/d$. Q.E.D. □

Theorem IV.14. *Suppose that noisy DFT data $\hat{\mathbf{y}}_d$ satisfies $\left\|P_\Omega(\hat{\mathbf{l}}_d \odot \hat{\mathbf{y}}_d - \hat{\mathbf{l}}_d \odot \hat{\mathbf{x}}_d)\right\|_2 \le \delta$, where $\hat{\mathbf{x}}_d$ is noiseless DFT data $\hat{x}_d[k]$ of $x_d[p]$ in* (52)*. Under the hypotheses of Theorem IV.13, there exists an absolute constant c_1, c_2 such that with probability $1-1/n^2$, the solution $\mathbf{g}$ to* (55) *satisfies*

$$\left\|\mathscr{H}(\hat{\mathbf{l}}_d \odot \hat{\mathbf{x}}_d) - \mathscr{H}(\mathbf{g})\right\|_{\text{F}} \le c_2 n^2 \delta,$$

provided that (56) *is satisfied with c_1.*

C. Incoherence Condition

Another advantage of using a cardinal set-up is that the coherence condition can be optimal even in the finite sampling regime. Specifically, due to the the wrap-around property, when $d = n$, the singular vectors U (resp. V) of $\mathcal{H}_c(\mathbf{x})$ are composed of r columns of a normalized DFT matrix. Thus, the standard incoherence condition is

$$\mu = \max\left\{\frac{n}{r}\max_{1\le i\le n}\|U^*\mathbf{e}_i\|_2^2, \frac{n}{r}\max_{1\le i\le n}\|V^*\mathbf{e}_i\|_2^2\right\} \quad = \quad 1, \tag{57}$$

which is optimal. Note that compared to the off the grid cases in Section III-E, the optimal mutual coherence can be obtained even with finite n.

It is also interesting to see that the corresponding separation is equal to the Nyquist sampling distance $\Delta = 1/n$, which appears smaller than the minimum separation condition $2/n$ in Section III-E. Recall that in off the grid signal reconstruction, there always exists a limitation in choosing the matrix pencil size d due to trade-off between the condition number of $\mathcal{V}_{n-d+1}$ and $\mathcal{V}_d$. However, for the cardinal set-up, thanks to the periodic boundary condition, the limitation does not exist anymore, and the net effect is doubling the effective aperture size from n to $2n$. This results in the reduction of the minimum separation in the cardinal setup.

D. Regularization Effect in Cardinal Setup

Note that in the proposed low-rank interpolation approach for the recovery of general FRI signals, the weighting factor $\hat{l}(\omega)$ used in (P_w) or (P'_w) is basically a high pass filter that can boost up the noise contribution. This may limit the performance of the overall low-rank matrix completion algorithm. In fact, another important advantage of the cardinal setup is to provide a natural regularization. More specifically, in constructing the weighting matrix for the low-rank matrix completion problem, instead of using the spectrum of the continuous domain whitening operator L, we should use $\hat{l}_d(\omega)$ of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

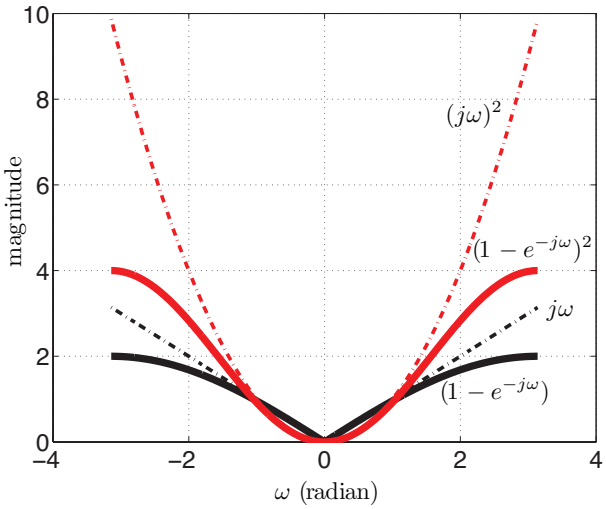

Fig. 3. Comparison of first- and second-order weights from whitening operator L and discrete counter-part L_d.

discrete counterpart L_d. As will be shown in the following examples, this helps to limit the noise amplification in the associated low-rank matrix completion problem.

1) Signals with Total Variation: A signal with total variation can be considered as a special case of (42) with $\mathrm{L} = \frac{d}{dt}$. Then, the discrete whitening operator L_d is the finite difference operator D_d given by

$$\mathrm{D}_d x(t) = x(t) - x(t-1).$$

In this case, the associated L-spline is given by

$$\beta_L(t) = \beta_+^0(t) = \mathcal{F}^{-1}\left\{\frac{1-e^{-i\omega}}{i\omega}\right\}(t) = \begin{cases} 1, & \text{for } 0 \le t < 1 \\ 0, & \text{otherwise} \end{cases} \tag{58}$$

Note that this is maximally localized spline because $b_L[p] = \beta_L(t)|_{t=p} = \delta[p]$ is a single-tap filter. Therefore, the sparsity level of the discrete innovation is equal to the number of underlying Diracs. Moreover, the weighting function for the low-rank matrix completion problem is given by

$$\hat{l}_d(\omega) = 1 - e^{-i\omega}.$$

Figure 3 compared the weighting functions that corresponds to the original whitening operator $\hat{l}(\omega) = i\omega$ and the discrete counterpart $\hat{l}_d(\omega) = 1 - e^{-i\omega}$. We can clearly see that high frequency boosting is reduced by the discrete whitening operator, which makes the low-rank matrix completion much more robust.

2) Signals with Higher order Total Variation: Consider a signal $x(t)$ that is represented by (42) with $\mathrm{L} = \frac{d^{m+1}}{dt^{m+1}}$. Then, the corresponding discrete counterpart L_d should be constructed by

$$\mathrm{L}_d\delta(t) = \underbrace{\mathrm{D}_d\mathrm{D}_d\cdots\mathrm{D}_d}_{m}\delta(t) \ .$$

In this case, the associated L-spline is given by [16]

$$\begin{aligned} \beta_+^m(t) &= \underbrace{\left(\beta_+^0 * \beta_+^0 * \cdots * \beta_+^0\right)}_{m+1}(t) \\ &= \mathcal{F}^{-1}\left\{\left(\frac{1-e^{-i\omega}}{i\omega}\right)^{m+1}\right\}(t) \\ &= \sum_{k=0}^{m+1}(-1)^k\binom{m+1}{k}\frac{(t-k)_+^m}{m!} \end{aligned}$$

with $(t)_+ = \max(0,t)$. We can see that the length of the corresponding filter $b_L[n]$ is now given by $m+1$. Hence, when the underlying signal is r-Diracs, then the sparsity level of the discrete innovation is upper bounded by

$$(m+1)r \tag{59}$$

and the corresponding weighting function is given by

$$\hat{l}_d(\omega) = (1-e^{-i\omega})^{m+1}. \tag{60}$$

Again, Figure 3 clearly showed that this weighting function is much more robust against noises compared to the original weighting $(i\omega)^{m+1}$.

Note that the relationship between the sparsity in (59) and the noise reduction by (60) clearly demonstrate the trade-off between regularization and the resolution in signal recovery. Specifically, to recover high order splines, rather than imposing the higher order weighting that is prone to noise boosting, we can use regularised weighting (60) that comes from discrete whitening operator. The catch, though, is the necessity for additional spectral samples originated from the sparsity increase.

E. Recovery of Continuous Domain Signals After Interpolation

In contrast to the recovery of general FRI signals from its spectral measurements, the reconstruction of cardinal L-spline can be done using standard B-spline signal processing tools [41], [42]. Specifically, after recovering the DFT spectrum $\hat{x}[k]$ using the Hankel structured matrix completion, a trivial application of an inverse DFT can obtain $x_d[n]$. Then, to recover $x(t)$, we use the equivalent representation Eq. (51). More specifically, the coefficient $c[n]$ in (51) can be computed by (52):

$$x_d[n] = (c * b_L)[n].$$

Because $x_d[n]$ are already computed and $b_L[n]$ is known, the unknown coefficient $c[n]$ can be obtained using the standard method in [41], [42] using recursive

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

filtering without computationally expensive matrix inversion. In case the operator L is the first differentiation, $b_L[n] = \delta[n]$, so $c[n]$ can be readily obtained as $x_d[n]$.

V. Algorithm Implementation

A. Noiseless structured matrix completion algorithm

In order to solve structured matrix completion problem from noise free measurements, we employ an SVD-free structured rank minimization algorithm [43] with an initialization using the low-rank factorization model (LMaFit) algorithm [44]. This algorithm does not use the singular value decomposition (SVD), so the computational complexity can be significantly reduced. Specifically, the algorithm is based on the following observation [43]:

$$\|A\|_* = \min_{U,V:A=UV^H} \|U\|_F^2 + \|V\|_F^2 \quad . \tag{61}$$

Hence, it can be reformulated as the nuclear norm minimization problem under the matrix factorization constraint:

$$\begin{aligned} \min_{U,V:\mathscr{H}(\mathbf{g})=UV^H} \quad & \|U\|_F^2 + \|V\|_F^2 \\ \text{subject to} \quad & P_\Omega(\mathbf{g}) = P_\Omega(\hat{\mathbf{x}}). \end{aligned} \tag{62}$$

By combining the two constraints, we have the following cost function for an alternating direction method of multiplier (ADMM) step [45]:

$$\begin{aligned} L(U,V,\mathbf{g},\Lambda) \quad := \quad & \iota(\mathbf{g}) + \frac{1}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) \\ & + \frac{\mu}{2}\|\mathscr{H}(\mathbf{g}) - UV^H + \Lambda\|_F^2 \end{aligned} \tag{63}$$

where $\iota(\mathbf{g})$ denotes an indicator function:

$$\iota(\mathbf{g}) = \begin{cases} 0, & \text{if } P_\Omega(\mathbf{g}) = P_\Omega(\hat{\mathbf{x}}) \\ \infty, & \text{otherwise} \end{cases} \quad .$$

One of the advantages of the ADMM formulation is that each subproblem is simply obtained from (63). More specifically, $\mathbf{g}^{(n+1)}, U^{(n+1)}$ and $V^{(n+1)}$ can be obtained, respectively, by applying the following optimization problems sequentially:

$$\begin{aligned} &\mathbf{g}^{(n+1)} = \\ &\quad \arg\min_{\mathbf{g}} \iota(\mathbf{g}) + \tfrac{\mu}{2}\|\mathscr{H}(\mathbf{g}) - U^{(n)}V^{(n)H} + \Lambda^{(n)}\|_F^2 \\ &U^{(n+1)} = \\ &\quad \arg\min_{U} \tfrac{1}{2}\|U\|_F^2 + \tfrac{\mu}{2}\|\mathscr{H}(\mathbf{g}^{(n+1)}) - UV^{(n)H} + \Lambda^{(n)}\|_F^2 \\ &V^{(n+1)} = \\ &\quad \arg\min_{V} \tfrac{1}{2}\|V\|_F^2 + \tfrac{\mu}{2}\|\mathscr{H}(\mathbf{g}^{(n+1)}) - U^{(n+1)}V^H + \Lambda^{(n)}\|_F^2 \end{aligned} \tag{64}$$

and the Lagrangian update is given by

$$\Lambda^{(n+1)} = \quad \mathcal{Y}^{(n+1)} - U^{(n+1)}V^{(n+1)H} + \Lambda^{(n)} \quad , \tag{65}$$

where $\mathcal{Y}^{(n+1)} = \mathscr{H}(\mathbf{g}^{(n+1)})$. It is easy to show that the first step in (64) can be reduced to

$$\mathbf{g}^{(n+1)} = P_{\Omega^c}\mathscr{H}^\dagger\left\{U^{(n)}V^{(n)H} - \Lambda^{(n)}\right\} + P_\Omega(\hat{\mathbf{x}}), \tag{66}$$

where P_{Ω^c} is a projection mapping on the set Ω^c (the complement set of Ω) and $\mathscr{H}^\dagger$ corresponds to the Penrose-Moore pseudo-inverse mapping from our structured matrix to a vector. Hence, the role of the pseudo-inverse is taking the average value and putting it back to the original coordinate. Next, the subproblem for U and V can be easily calculated by taking the derivative with respect to each matrix, and we have

$$\begin{aligned} &U^{(n+1)} = \\ &\quad \mu\left(\mathcal{Y}^{(n+1)} + \Lambda^{(n)}\right)V^{(n)}\left(I + \mu V^{(n)H}V^{(n)}\right)^{-1} \\ &V^{(n+1)} = \\ &\quad \mu\left(\mathcal{Y}^{(n+1)} + \Lambda^{(n)}\right)^H U^{(n+1)}\left(I + \mu U^{(n+1)H}U^{(n+1)}\right)^{-1} \end{aligned} \tag{67}$$

Now, for faster convergence, the remaining issue is how to initialize U and V. For this, we employ an algorithm called the low-rank factorization model (LMaFit) [44]. More specifically, for a low-rank matrix Z, LMaFit solves the following optimization problem:

$$\min_{U,V,Z} \frac{1}{2}\|UV^H - Z\|_F^2 \text{ subject to } P_I(Z) = P_I(\mathscr{H}(\hat{\mathbf{x}})) \tag{68}$$

and Z is initialized with $\mathscr{H}(\hat{\mathbf{x}})$ and the index set I denotes the positions where the elements of $\mathscr{H}(\hat{\mathbf{x}})$ are known. LMaFit solves a linear equation with respect to U and V to find their updates and relaxes the updates by taking the average between the previous iteration and the current iteration. Moreover, the rank estimation can be done automatically. LMaFit uses QR factorization instead of SVD, so it is also computationally efficient.

B. Noisy structured matrix completion algorithm

Similarly, the noisy matrix completion problem can be solved by mininimizing the following Lagrangian function:

$$\begin{aligned} L(U,V,\mathbf{g},\Lambda) \quad := \quad & \frac{\lambda}{2}\|P_\Omega(\hat{\mathbf{y}}) - P_\Omega(\mathbf{g})\|_2^2 + \frac{1}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) \\ & + \frac{\mu}{2}\|\mathscr{H}(\mathbf{g}) - UV^H + \Lambda\|_F^2 \end{aligned}$$

where λ denotes an appropriate regularization parameter. Compared to the noiseless cases, the only difference is the update step of $\mathbf{g}$. More specifically, we have

$$\begin{aligned} &\mathbf{g}^{(n+1)} = \\ &\quad \arg\min_{\mathbf{g}} \tfrac{\lambda}{2}\|P_\Omega(\hat{\mathbf{y}}) - P_\Omega(\mathbf{g})\|_2^2 + \tfrac{\mu}{2}\|\mathscr{H}(\mathbf{g}) - U^{(n)}V^{(n)H} + \Lambda^{(n)}\|_F^2 \end{aligned}$$

which can be reduced to

$$\mathbf{g}^{(n+1)} = P_{\Omega^c}\mathscr{H}^\dagger\left\{U^{(n)}V^{(n)H} - \Lambda^{(n)}\right\} + P_\Omega(\mathbf{z})$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

where $\mathbf{z} = [z[0], \cdots, z[n-1]]^T$ such that

$$z[i] = \frac{\lambda y[i] + \mu P_i\left(\mathscr{H}^*\left(U^{(n)}V^{(n)H} - \Lambda^{(n)}\right)\right)}{\lambda + \mu P_i(\mathscr{H}^*\mathscr{H}(\mathbf{e}_i))},$$

where $\mathbf{e}_i$ denotes the unit coordinate vector where the i-th element is 1, and P_i is the projection operator to the i-th coordinate.

C. Implemetation of ADMM

The alternating direction method of multipliers (ADMM) described above is widely used to solve large-scale linearly constrained optimization problems, convex or nonconvex, in many engineering fields. The convergence of ADMM algorithm for minimizing the sum of two or more nonsmooth convex separable functions have been well-studied, and Hong and Luo [46] proved the linear convergence of a general ADMM algorithm with any number of blocks under linear constraints. Even for the nonconvex problems, Hong et al [47] further showed that the classical ADMM converges to the set of stationary solutions, provided that the penalty parameter in ADMM (μ in our case) is chosen to be sufficiently large. Accordingly, to ensure the convergence of ADMM, it is usually recommended to use a sufficiently large μ; so, we chose $\mu = 10^3$ in our implementation.

Note that the computational complexity of our ADMM algorithm is crucially determined by the matrix inversion in (67). More specifically, the computational complexity in (67) in terms of multiplication is $\mathcal{O}((n-d+1)rd+r^3)$, whereas the number of multiplication required for (66) and (65) is $\mathcal{O}((n-d+1)rd)$. Thus, if the underlying signal is sparse, then we can choose sufficiently small rank estimate r and the matrix pencil size d to reduce the overall computational complexity. Another important issue in practice is the memory usage. Note that the $U, V, \mathcal{Y}$ as well as the Lagrangian parameter Λ should be stored throughout the iterations in our ADMM implementation. The associated memory requirement is at least $(n-d+1)r+rd+2(n-d+1)d$. This is not an issue in our 1-D problems, but for large size problems (especially originated for three dimensional recovery problems in medical imaging applications), the memory requirement quickly grows, which can become a dominating computational bottleneck in parallel implementation using memory limited graphic processor unit (GPU).

VI. Numerical Results

In this section, we undertake a comparative numerical study for the recovery of FRI signals on an integer grid. We then provide numerical experiments involving the reconstruction of piecewise polynomials in which the discontinuities are located in arbitrary positions.

A. Recovery of On-grid Signals

First, we perform numerical simulations using noiseless measurements. Specifically, we consider three scenarios: 1) streams of Diracs, 2) piecewise constant signals, and 3) a super-position of Diracs and piecewise constant signals. As a reference for the comparsion, the SPGL1 implementation of the basis pursuit (BP) algorithm [48] was used to recover a stream of Diracs, whereas the split Bregman method of l_1 total variation reconstruction [49] was used for recovering signals in the second and third scenarios. We assume that all of the singularities are located on an integer grid. To quantifythe recovery performances, phase-transition plots were calculated using 300 Monte Carlo runs.

1) Diracs streams: To simulate Diracs stream signals, we generated one-dimensional vectors with a length of 100, where the location of the Diracs are constrained on an integer grid. The spectral measurements were randomly sampled with a uniform random distribution, where the zero-frequency component was always included. This made the Fourier sensing matrix become a DFT matrix, allowing the use of basis pursuit using a partial DFT sensing matrix. We used the SPGL1 basis pursuit algorithm, as obtained from the original author's homepage [48]. Only thing we need to set for SPGL1 was the number of iteration to 500. For the proposed method, d was set to be $\lfloor n/2 \rfloor + 1 = 51$. The other hyperparameters for the proposed method were as following: $\mu = 10^3$, 500 iterations, $tol = 10^{-4}$ for LMaFit. For a fair comparison, we used the same iteration numbers and sampling patterns for both basis pursuit and the proposed algorithm. The phase transitions in Fig. 4 show the success ratio calculated from 300 Monte Carlo trials. Each trial from the Monte Carlo simulations was considered as a success when the normalized mean square error (NMSE) was below 10^{-3}. In Fig. 4, the proposed approach provided a sharper transition curve between success and failure than that of the basis pursuit. Furthermore, the transition curve of the proposed method (red dotted line) is higher than that by the basis pursuit (blue dotted line).

2) Piecewise-constant signals: To generate the piecewise constant signals, we initially generated Diracs signal at random locations on an integer grid and added steps between the Diracs. The length of the unknown one-dimensional vector was again set to 100. To avoid a boundary effect, the values at the end of both boundaries were set to zeros. As a conventional compressed sensing approach, we employed the 1-D version of l_1-total variation reconstruction (l_1-TV) using the split Bregman method [49], which was modified from the original 2-D version of l_1-TV from the homepage. We found that the optimal parameters for l_1-TV were $\mu = 10^2, \lambda = 1$, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

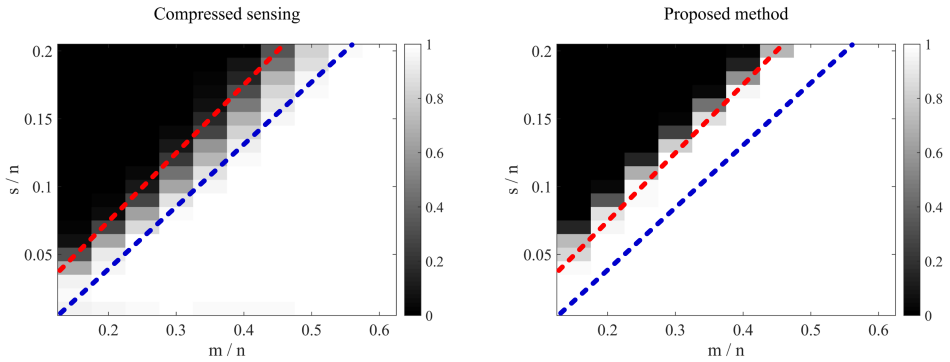

Fig. 4. Phase-transition diagrams for recovering a stream of Diracs from m random sampled Fourier samples. The size of the target signal (n) is 100 and the annihilating filter size d was set to 51. s denotes the number of Diracs. The left and right graphs correspond to the phase-transition diagrams of the basis pursuit [48] compressed-sensing approach and the proposed low-rank interpolation approach, respectively. The success ratio is obtained from the success ratio of 300 Monte Carlo runs. Two transition lines from compressed sensing (blue) and the low-rank interpolator (red) are overlaid.

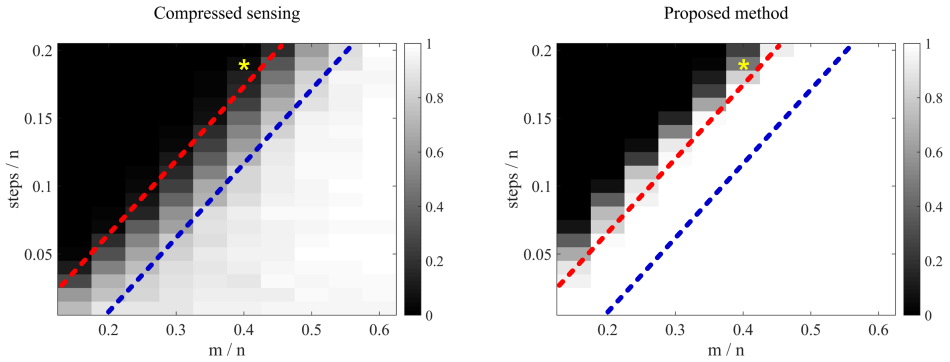

Fig. 5. Phase-transition diagrams for recovering a piecewise constant signals from m random sampled Fourier samples. The size of the target signal (n) is 100 and the annihilating filter size d was set to 51. The left and right graphs correspond to the phase-transition diagram of the l_1-TV compressed-sensing approach and the proposed low-rank interpolation approach, respectively. The success ratio is obtained from the success ratio of 300 Monte Carlo runs. Two transition lines from compressed sensing (blue) and the low-rank interpolator (red) are overlaid.

the outer and inner loop iterations of 5 and 40, respectively. The hyper-parameters for the proposed method are as follows; $\mu = 10^3$, 200 iterations, $tol = 10^{-3}$ for LMaFit. Note that we need $1 - e^{-i\omega}$ weighting for low-rank Hankel matrix completion as a discrete whitening operator for TV signals. The phase transition plots were calculated using averaged success ratio from 300 Monte Carlo trials. Each trial from the Monte Carlo simulations was considered as a success when the NMSE was below 10^{-2}. Because the actual non-zero support of the piecewise constant signals was basically the entire domain, the threshold was set higher than in previous Dirac experiments.

As shown in Fig. 5, the transition curve from the proposed method (red dotted line) provided a sharper and improved transition than the l_1 total variation approach (blue dotted line). Furthermore, even in successful cases, there were a number of unsuccessful recoveries with the conventional method, whereas the proposed method succeeded nearly every time. In Fig. 6, we also illustrate sample recovery results from the same locations in the phase-transition diagram, which are at the position marked with the yellow star in Fig. 5. We observed near-perfect reconstruction from the proposed method, whereas severe blurring was observed when using l_1-TV reconstruction.

3) Piecewise-constant signal + Diracs: We performed additional experiments on the reconstruction of the super-position of the piecewise constant signal and Dirac spikes. Note that this corresponds to the first derivative of the piecewise polynomial with a maximum order of 1 (i.e. piecewise constant and linear signals). The goal of this experiment was to verify the capability of recovering piecewise polynomials, but there was no widely used compressed-sensing solution for this type of signals. Therefore, for a fair comparison, we were inter-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

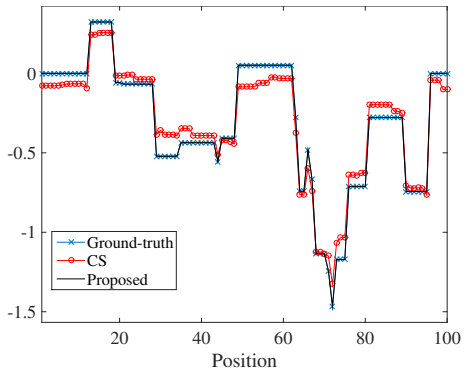

Fig. 6. Sample reconstruction results at the position marked with the yellow star in Fig. 5. Ground-truth signal (original), l_1-TV (compressed sensing) and the proposed method (low-rank interpolator) are illustrated. The parameters for the experiments are as follows: $n = 100, d = 51$, and $m = 40$ with 19 steps.

ested in recovering their derivatives, as the conventional l_1-TV approach can be still used for recovering Diracs and piecewise constant signals. The optimal parameters for l_1-TV were $\mu = 10^3, \lambda = 1$, and outer and inner loop iterations of 5 and 40, respectively. The hyper-parameters for the proposed method are as follows: $\mu = 10^3$, 200 iterations, and $tol = 10^{-3}$ for LMaFit. In this case, the sparsity level doubles at the Dirac locations when we use the l_1-TV method for this type of signal. Similar sparsity doubling was observed in our approach. More specifically, our method required the derivative operator as a whitening operator, which resulted in the first derivative of the Diracs. According to Eq.(59), this doubles the effective sparsity level. Accordingly, the comparison of l_1-TV and our low-rank interpolation approach was fair, and the overall phase transitions were expected to be inferior compared to those of piecewise constant signals. The simulation environment was set such that it is identical to those in the previous piecewise constant setups, except for the signal generation process. For signals, we generated an equal number of steps and Diracs. When the sparisty is an odd number, the numbers of Diracs was set to the number of steps minus 1.

As shown in Fig. 7, there were much more significant differences between the two approaches. Our algorithm still provided very clear and improved phase transitions, whereas the conventional l_1-TV approach resulted in very fuzzy and inferior phase transitions. In Fig. 8, we also illustrated the sample recovery results from identical locations in the phase-transition diagram, which are at the position marked by the yellow star in Fig. 7. The proposed approach provided a nearly perfect reconstruction, whereas l_1-TV reconstruction exhibits blurring. This again confirms the effectiveness of our approach.

4) Recovery from Noisy Measurements: To verify the robustness to noise of the proposed method, we performed experiments using piecewise constant signals by adding additive complex Gaussian noise to partial Fourier measurements. Fig. 9(a) shows the recovery performance of the proposed low-rank interpolation method at several signal-to-noise ratios (SNR). All setting parameters for the proposed method are identical to the parameters in previous experiments except for the addition of $\lambda = 10^5$. As expected from the theoretical results, the recovery performance was proportional to the noise level. Fig. 9(b) presents an example of reconstructions from measurements with 30dB of noise. Here, the optimal parameters for l_1-TV (CS) were $\mu = 10^3, \lambda = 1$, with outer-inner loop iterations of 5 and 40, respectively. The proposed approaches still provide accurate reconstruction results.

B. Recovery of Off-the-Grid Signals

To verify the performance of off-the-grid signal recovery, we also performed additional experiments using piecewise constant signals whose edges are located on a continuous domain in the following range $[0, 1]$. Specifically, we consider a signal composed of several rectangles whose edges are located in off-grid positions. Due to the existence of closed-form expressions of the Fourier transform of shifted rectangle functions, the measurement data could be generated accurately without using the discrete Fourier transform. Because the signal is composed of rectangles, the singularities are located at the edge positions after differentiation. Accordingly, the weighting factor for low-rank interpolation was the spectrum of the continuous domain derivative, i.e. $\hat{l}(\omega) = i\omega$. Owing to the Nyquist criterion, the sampling grid in the Fourier domain corresponds to an integer grid, and the ambient signal dimension n (which corresponds to the aperture size) was set to 100. Then, $m = 36$ Fourier samples were randomly obtained from the range of $[0, \cdots, 99]$. The parameters for the proposed low-rank interpolation were as follows: the tolerance for LMaFit = 10^{-1}, number of iterations 300, $\mu = 10^3$. Once the Fourier data were interpolated, we used the matrix pencil method [5], [40], as described in Section III-F for the second step of signal recovery.

Fig. 10(b) illustrates the interpolated measurement from 36 instances of irregularly sampled Fourier data using the proposed low-rank interpolation. Because the underlying signal is a piecewise constant signal, the optimal weighting $\hat{l}(\omega) = i\omega$ was applied for the Fourier data before the low-rank matrix completion was applied. As shown in Fig. 10(b), near-perfect interpolation was achieved. On the other hand, if the original Fourier data was used for the low-rank interpolation without

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

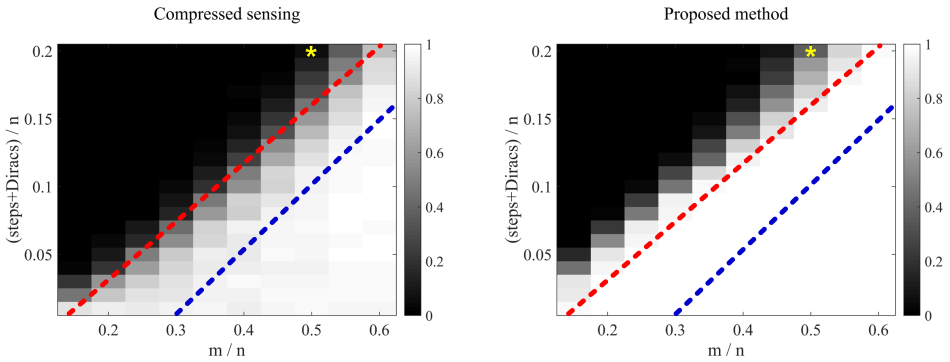

Fig. 7. Phase-transition diagrams for recovering the super-position of the piecewise constant signal and Diracs from m randomly sampled Fourier samples. The size of the target signal (n) is 100 and the annihilating filter size d was set to51. The left and right graphs correspond to the phase-transition diagram of the l_1-TV compressed-sensing approach and the proposed low-rank interpolation approach, respectively. The success ratio is obtained from the success ratio of 300 Monte Carlo runs. Two transition lines from compressed sensing (blue) and the low-rank interpolator (red) are overlaid.

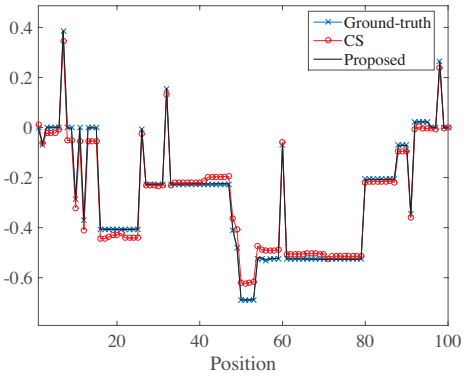

Fig. 8. Sample reconstruction results at the position denoted by the yellow star in Fig. 7. Ground-truth signal (original), l_1-TV (compressed sensing) and the proposed method (low-rank interpolator) are illustrated. The parameters for the experiments are: n = 100, d = 51, and m = 50, with the number of steps and Diracs all being 10.

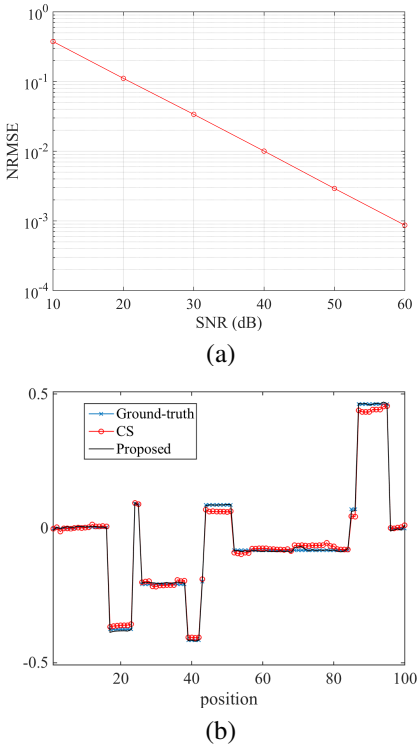

Fig. 9. (a) The reconstructed NMSE plots by the proposed low-rank interpolation scheme at various SNR values. For this simulation, we set the annihilating filter size d = 51, and n = 100. The number of singularity due to steps was 10 and the number of measurements was 50. (b) A reconstruction example from a noisy measurement at 30dB.

weighting, the resulting interpolation was very different from the true measurement (see Fig. 10(a)). These results confirmed our theory.

Fig. 11 illustrates the corresponding reconstruction results from noiseless measurements using the proposed method. We also performed an additional simulation with 40dB of measurement noise. The results clearly showed that the proposed approach accurately reconstructs the underlying piecewise constant signals. Recall that there are no existing off-the-grid reconstruction algorithms for piecewise constant signals. Therefore, the nearly perfect reconstructions by the proposed method clearly show that our theory is quite general and can therefore be used for the recovery of general FRI signals.

VII. Conclusion

While the recent theory of compressed sensing (CS) can overcome the Nyquist limit for recovering sparse signals, standard recovery algorithms are usually im-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

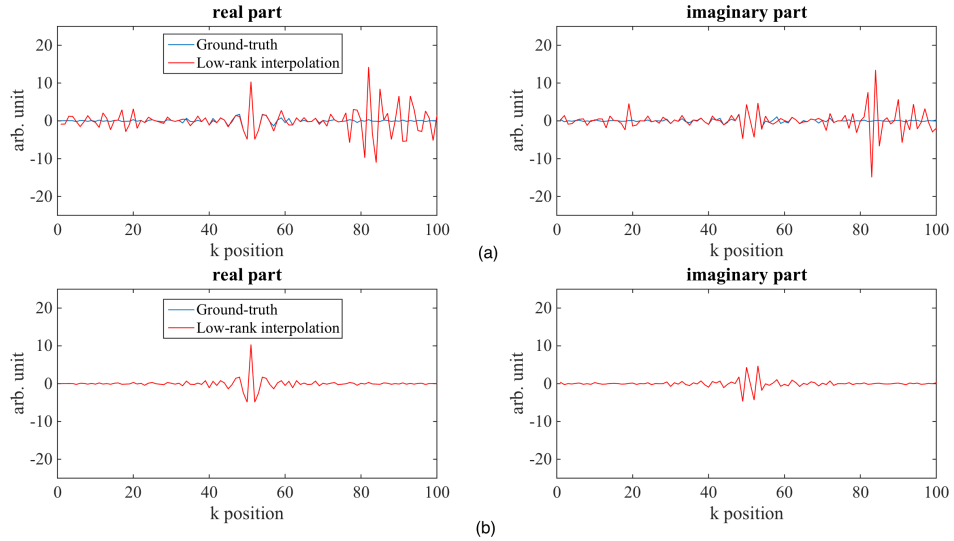

Fig. 10. Fully sampled Fourier measurement and the interpolated data from m = 36 irregularly sampled data. (a) Low-rank interpolation results without spectrum weighting. (b) The proposed low-rank interpolation using optimal weighting $\hat{l}(\omega) = i\omega$. For this simulation, the following parameters were used: d = 51, n = 100 and m = 36.

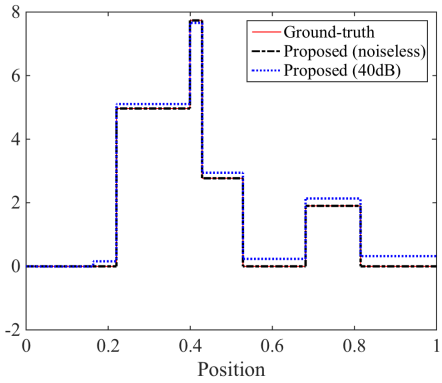

Fig. 11. Proposed reconstruction result of a piecewise constant signal from noiseless and 40dB noisy sparse Fourier samples. For this simulation, the following parameters were used: d = 51, n = 100 and m = 36. The matrix pencil approach was used for signal recovery after the missing Fourier data was interpolated using the proposed low-rank interpolation scheme with optimal weighting $\hat{l}(\omega) = i\omega$.

plemented in the discrete domain as inverse problem approaches that are fully dependent on signal representations. Moreover, existing spectral compressed-sensing algorithms of continuous signals such as off-the-grid spikes are very distinct from their discrete domain counter-parts. To address these issues and unite the theories, this paper developed a near-optimal Fourier CS framework using a structured low-rank interpolator in the measurement domain to use before an analytic reconstruction procedure is applied. This was based on the fundamental duality between the sparsity in the primary space and the low-rankness of the structured matrix in reciprocal spaces. Compared to existing spectral compressed-sensing methods, our theory was generalized to encompass more general signals with a finite rate of innovation such as piecewise polynomials and splines with provable performance guarantees. Numerical results confirmed that the proposed methods exhibited significantly improved phase transition as compared to that by the existing CS approaches.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Michael Unser at EPFL for giving an insight into cardinal spline representations. This study was supported by Korea Science and Engineering Foundation under Grant NRF-2016R1A2B3008104. Kiryung Lee was in part supported in part by the National Science Foundation (NSF) under Grant IIS 14-47879.

APPENDIX

A. *Properties of Linear Difference Equations with Constant Coefficients*

Before proving Theorem II.1, we review some important properties about linear difference equations [50]. In general, the r-th order homogeneous difference equation has the form

$$x[k+r]+a_{r-1}[k]x[k+r-1]+\cdots+a_0[k]x[k] = 0. \quad \text{(A.1)}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

where $\{a_i[k]\}_{0=1}^{r-1}$ are constant coefficients. The functions $f_1[k], f_2[k], \cdots, f_r[k]$ are said to be *linearly independent* for $k \geq k_0$ if there are constants $c_1, c_2, \cdots, c_r$, not all zero, such that

$$c_1 f_1[k] + c_2 f_2[k] + \cdots + c_r f_r[k] = 0, \quad k \geq k_0.$$

And, a set of r linearly independent solutions of (A.1) is called a *fundamental set* of solutions. As in the case of Wronskian in the theory of linear differential equations, we can easily check the linear independence of solution by using Casoratian $W[k]$ which is defined as :

$$W[k] = \begin{vmatrix} x_1[k] & x_2[k] & \cdots & x_r[k] \\ x_1[k+1] & x_2[k+1] & \cdots & x_r[k+1] \\ \vdots & & & \vdots \\ x_1[k+r-1] & x_2[k+r-1] & \cdots & x_r[k+r-1] \end{vmatrix}$$

Lemma A.1. *[50, Lemma 2.13 Corollary 2.14] Let $x_1[k], \cdots, x_r[k]$ be solutions of* (A.1) *and $W[k]$ be their Casoratian. Then,*

1) *For $k \geq k_0$,*

$$W[k] = (-1)^{r(k-k_0)} \left(\prod_{i=k_0}^{k-1} a_0[i] \right) W[k_0];$$

2) *Suppose that $a_0[k] \neq 0$ for all $k \geq k_0$. Then, $W[k] \neq 0$ for all $k \geq k_0$ iff $W[k_0] \neq 0$.*

Now, we state the criterion by which we can easily check the linear independence of solutions of (A.1).

Lemma A.2. *[50, Theorem 2.15, Theorem 2.18],*

1) *The set of solutions $x_1[k], x_2[k], \cdots, x_r[k]$ of* (A.1) *is a fundamental set iff $W[k_0] \neq 0$ for some $k_0 \in \mathbb{N}$.*
2) *If $a_0[k] \neq 0$ for all $k \geq k_0$, then* (A.1) *has a fundamental solutions for $k \geq k_0$.*

Now, for the linear difference equation with constant coefficients, similarly to the case of differential equation with constant coefficients, we have the following result:

Lemma A.3. *[50, Lemma 2.22, Theorem 2.23] Assume that we have a linear difference equation*

$$z_{k+r} + a_{r-1} z_{k+r-1} + \cdots + a_1 z_{k+1} + a_0 z_k = 0, \qquad \text{(A.2)}$$

where the coefficients a_i's are constants and $a_0 \neq 0$, and we are given

$$P(\lambda) := \lambda^r + a_{r-1}\lambda^{r-1} + \cdots + a_1\lambda + a_0 = \prod_{i=0}^{p-1} (\lambda - \lambda_i)^{l_i},$$

as its characteristic equation, where $\lambda_0, \cdots, \lambda_{p-1}$ are all the distinct nonzero complex roots of $P(\lambda)$. Then the set

$$G = \bigcup_{i=0}^{p-1} G_i$$

is a fundamental set of solutions (A.2)*, where*

$$G_i = \left\{ [\lambda_i^k]_k, \left[\binom{k}{1} \lambda_i^{k-1} \right]_k, \left[\binom{k}{2} \lambda_i^{k-2} \right]_k, \cdots, \left[\binom{k}{l_i - 1} \lambda_i^{k-l_i+1} \right]_k \right\}. \qquad \text{(A.3)}$$

Here, we use the notation $[a[k]]_k$ to denote the sequence with k-th term $a[k]$. Now, we have the following results which will be used later.

Lemma A.4. *The sequence set*

$$\left\{ [\lambda_i^k]_k, \left[\binom{k}{1} \lambda_i^{k-1} \right]_k, \left[\binom{k}{2} \lambda_i^{k-2} \right]_k, \cdots, \left[\binom{k}{l_i - 1} \lambda_i^{k-l_i+1} \right]_k \right\}, \qquad \text{(A.4)}$$

and

$$\left\{ [\lambda_i^k]_k, [k\lambda_i^k]_k, \cdots, [k^{l_i-1}\lambda_i^k]_k \right\} \qquad \text{(A.5)}$$

span the same sequence space.

Proof. Note that (A.4) has the same span as

$$\left\{ [\lambda_i^k]_k, \left[\binom{k}{1} \lambda_i^k \right]_k, \cdots, \left[\binom{k}{l_i - 1} \lambda_i^k \right]_k \right\}$$

since $a_0 \neq 0$ and $\lambda_i \neq 0$ for all $0 \leq i \leq p-1$. Moreover, for any given $0 \leq i \leq p-1$ and $0 \leq j \leq l_i - 1$,

$$\binom{k}{j} = \frac{k(k-1)\cdots(k-j+1)}{j!}$$

is a polynomial of k with order j so that each sequence $\left\{ \binom{k}{j} \lambda_i^k \right\}_{k \in \mathbb{N}}$ is a linear combination of

$$\left\{ [\lambda_i^k]_k, [k\lambda_i^k]_k, \cdots, [k^j \lambda_i^k]_k \right\}.$$

On the other hand, every polynomial $P(k)$ on $\mathbb{Z}$ of degree at most j uniquely expressible as a linear combination of

$$\binom{k}{0}, \binom{k}{1}, \cdots, \binom{k}{j}$$

and the explicit representation is given by

$$k^j = \sum_{l=0}^{j} a_l \binom{k}{l} \text{ where } a_l = \sum_{m=0}^{l} (-1)^{l-m} \binom{l}{m} m^j.$$

For the above formula, you may refer [51]. Thus, the sequence $\{k^j \lambda_i^j\}_{k \in \mathbb{N}}$ is given as a linear combination of

$$\left\{ [\lambda_i^k]_k, \left[\binom{k}{1} \lambda_i^k \right]_k, \cdots, \left[\binom{k}{i-1} \lambda_i^k \right]_k \right\}.$$

Thus, (A.4) and the sequence

$$\left\{ [\lambda_i^k]_k, [k\lambda_i^k]_k, \cdots, [k^{l_i-1}\lambda_i^k]_k \right\}$$

spans the same sequence space. Q.E.D. □

B. Proof of Theorem II.1

Proof. First, we will show that $\mathscr{H}(\hat{\mathbf{x}})$ has rank at most r. Let $\hat{\mathbf{h}} \in \mathbb{C}^{r+1}$ be the minimum size annihilating filter.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

Then, (4) can be represented as

$$\hat{\mathbf{h}}_a \quad = \quad \mathscr{C}(\hat{\mathbf{h}})\hat{\mathbf{a}}$$

where $\hat{\mathbf{a}} = \begin{bmatrix}\hat{a}[0] & \cdots & \hat{a}[k_1-1]\end{bmatrix}$ and $\mathscr{C}(\hat{\mathbf{h}}) \in \mathbb{C}^{d\times k_1}$ is a Toeplitz structured convolution matrix from $\hat{\mathbf{h}}$:

$$\mathscr{C}(\hat{\mathbf{h}}) = \begin{bmatrix} \hat{h}[0] & 0 & \cdots & 0 \\ \hat{h}[1] & \hat{h}[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{h}[r] & \hat{h}[r-1] & \cdots & \hat{h}[r-k_1+1] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{h}[r] \end{bmatrix} \in \mathbb{C}^{d\times k_1}$$

where $d = r + k_1$. Since $\mathscr{C}(\hat{\mathbf{h}})$ is a convolution matrix, it is full ranked and we can show that

$$\dim \mathrm{RAN}\mathscr{C}(\hat{\mathbf{h}}) = k_1,$$

where $\dim \mathrm{RAN}(\cdot)$ denotes the dimension of the range space. Moreover, the range space of $\mathscr{C}(\hat{\mathbf{h}})$ now belongs to the null space of the Hankel matrix $\mathscr{H}(\hat{\mathbf{x}})$, so it is easy to show

$$k_1 = \dim \mathrm{RAN}\mathscr{C}(\hat{\mathbf{h}}) \le \dim \mathrm{NUL}\mathscr{H}(\hat{\mathbf{x}}),$$

where $\dim \mathrm{NUL}(\cdot)$ represent the dimension of the null space. Thus,

$$\mathrm{RANK}\mathscr{H}(\hat{\mathbf{x}}) = \min\{d, n-d+1\} - \dim \mathrm{NUL}\mathscr{H}(\hat{\mathbf{x}}) \le d - k_1 = r.$$

Now, we will show by contradiction that the rank of $\mathscr{H}(\hat{\mathbf{x}})$ cannot be smaller than r. Since the rank of the Hankel matrix is at most r, any set of $r+1$ consecutive rows (or columns) of $(n-d+1)\times d$ Hankel matrix with the entries $\hat{x}[k]$ must be linearly dependent. Therefore, $\hat{x}[k]$ should be the solution of the following difference equation:

$$z_{k+r} + a_{r-1}z_{k+r-1} + \cdots + a_1 z_{k+1} + a_0 z_k = 0, \text{ for } 0 \le k \le n-r-1 \tag{A.6}$$

where $\{a_i\}_{i=0}^{r-1}$ are coefficients of the linear difference equation, and

$$P(\lambda) := \lambda^r + a_{r-1}\lambda^{r-1} + \cdots + a_1\lambda + a_0 = \prod_{j=0}^{p-1}(\lambda - \lambda_j)^{l_j}, \tag{A.7}$$

is the characteristic polynomial of the linear difference equation, where $\lambda_0, \cdots, \lambda_{p-1}$ are distinct nonzero complex numbers so that $a_0 \neq 0$ and

$$r = \sum_{j=0}^{p-1} l_j.$$

From Lemma A.3 and A.4, we know that the sequences $\{k^l\lambda_j^k\}_{k\in\mathbb{Z}}$ $(0 \le j \le p-1 \text{ and } 0 \le l \le l_j - 1)$ are the fundamental solutions of the linear difference equation, and $\hat{x}[k]$ can be represented as their linear combination [50]:

$$\hat{x}[k] := \sum_{j=0}^{p-1}\sum_{l=0}^{l_j-1} a_{j,l}k^l\lambda_j{}^k \text{ for } 0 \le k \le n-1, \tag{A.8}$$

where all the leading coefficients a_{j,l_j-1} $(0 \le j \le p-1)$ are nonzero. If we assume that the rank of the Hankel matrix with the sequence as in (A.8) is less than r, then the sequence $\hat{x}[k]$ must satisfy the recurrence relation of order $q < r$, since any collection of $q+1$ consecutive rows (or columns) are linearly dependent. Thus, there exist a recurrence relation for $\hat{x}[k]$ of order $q < r$ such that

$$z_{k+q} + b_{q-1}z_{k+q-1} + \cdots + b_1 z_{k+1} + b_0 z_k = 0, \text{ for } 0 \le k \le n-q-1, \tag{A.9}$$

whose solution is the sequence given by

$$\hat{x}[k] = \sum_{j=0}^{p'-1}\sum_{l=0}^{l'_j-1} a'_{j,l}k^l(\lambda'_j)^k \text{ for } 0 \le k \le n-1, \tag{A.10}$$

where $\sum_{j=0}^{p'-1} l'_j \le r-1$, and

$$P_1(\lambda) = \lambda^q + b_{q-1}\lambda^{q-1} + \cdots + b_1\lambda + b_0 = \prod_{j=0}^{p'-1}(\lambda - \lambda'_j)^{l'_j}$$

is the characteristic polynomial of (A.9). Subtracting (A.10) from (A.8), we have the equation

$$0 = \sum_{j=0}^{P-1}\sum_{l=0}^{L_j-1} c_{j,l}k^l(\Lambda_j)^k, \text{ for } 0 \le k \le n-1, \tag{A.11}$$

where

$$\{\Lambda_j : 1 \le j \le P\} = \{\lambda_j : 0 \le j \le p-1\} \cup \{\lambda'_j : 0 \le j \le p'-1\},$$

and L_j $(1 \le j \le P)$ is the multiplicity of the root of Λ_j for the least common multiple of $P(\lambda)$ and $P_1(\lambda)$. Moreover, we have $N := \sum_{j=0}^{P-1} L_j \le r + (r-1) = 2r-1$.

Now, from Lemma A.3, we know that the sequences $\{\hat{x}_{l,j}[k] = k^l(\Lambda_j)^k : 0 \le j \le P-1, 0 \le l \le L_j - 1\}$ are linearly independent sequences, and from the hypothesis $\min\{n-d+1, d\} > r$, we have $n > 2r-1 \ge N$. Thus, if we write (A.11) as a matrix equation,

$$\mathbf{0} = \Phi\mathbf{c}$$

where

$$\mathbf{c} = \begin{bmatrix} c_{0,0}, \cdots, c_{0,L_0-1}, c_{1,0}, \cdots, \cdots, c_{P-1,0}, \cdots, c_{P-1,L_{P-1}-1} \end{bmatrix}^T$$

is an $N \times 1$ matrix, and Φ is an $n \times N$ matrix which has

$$[\Lambda_0^k, \cdots, k^{L_0-1}(\Lambda_0)^k, \Lambda_1^k, \cdots, \cdots, \Lambda_{P-1}^k, \cdots, k^{L_{P-1}-1}(\Lambda_{P-1})^k]$$

as an $k+1$-th row. By combining Lemma A.1 and A.2, we can conclude that the $N \times N$ principal minor for the matrix Φ must have a nonzero determinant. Thus, the matrix Φ is of full column rank so that all the coefficients

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

vector $\mathbf{c}$ must be zero.

Thus, all the zeros and their multiplicities of zeros for the polynomials $P(\lambda)$, $P_1(\lambda)$ must be identical. That is a contradiction to the hypothesis that the degree of $P_1(\lambda)$ is less than that of $P(\lambda)$. □

C. A basis representation of structured matrices

1) Hankel matrix and variations: The linear space $\mathcal{H}(n,d)$ of $(n-d+1)$-by-d Hankel matrices is spanned by a basis $\{A_k\}_{k=1}^n$ given by

$$A_k = \begin{cases} \dfrac{1}{\sqrt{k}} \displaystyle\sum_{i=1}^{k} \mathbf{e}_i \mathbf{e}_{k-i+1}^*, & 1 \le k \le d, \\ \dfrac{1}{\sqrt{d}} \displaystyle\sum_{i=1}^{d} \mathbf{e}_i \mathbf{e}_{k-i+1}^*, & d+1 \le k \le n-d+1, \\ \dfrac{1}{\sqrt{n-k+1}} \displaystyle\sum_{i=k-n+d}^{d} \mathbf{e}_i \mathbf{e}_{k-i+1}^*, & n-d+2 \le k \le n. \end{cases} \tag{A.12}$$

Note that $\{A_k\}_{k=1}^n$ satisfies the following properties. First, A_k is of unit Frobenius norm and all nonzero entries of A_k are of the same value, i.e.,

$$[A_k]_{i,j} = \begin{cases} \dfrac{1}{\sqrt{\|A_k\|_0}}, & [A_k]_{i,j} \neq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{A.13}$$

for all $k = 1, \ldots, n$. It follows from (A.13) that the spectral norm of A_k is bounded by

$$\|A_k\| \le \|A_k\|_0^{-1/2}. \tag{A.14}$$

Second, each row and column of all A_k's has at most one nonzero element, which implies

$$\sum_{j=1}^{n_2} \left(\sum_{i=1}^{n_1} |[A_k]_{i,j}| \right)^2 = 1, \quad \text{and} \quad \sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} |[A_k]_{i,j}| \right)^2 = 1. \tag{A.15}$$

Last, any two distinct elements of $\{A_k\}_{k=1}^n$ have disjoint supports, which implies that $\{A_k\}_{k=1}^n$ constitutes an orthonormal basis for the subspace spanned by $\{A_k\}_{k=1}^n$. In fact, these properties are satisfied by bases for structured matrices of a similar nature including Toeplitz, Hankel-block-Hankel, and multi-level Toeplitz matrices.

2) Warp-around Hankel matrix and variations: The linear space $\mathcal{H}_c(n,d)$ of n-by-d wrap-around Hankel matrices for $n \ge d$ is spanned by a basis $\{A_k\}_{k=1}^n$ given by

$$A_k = \begin{cases} \dfrac{1}{\sqrt{d}} \left(\displaystyle\sum_{i=1}^{k} \mathbf{e}_i \mathbf{e}_{(k+d-i-1)_n+1}^* + \sum_{j=n-d+k+1}^{n} \mathbf{e}_j \mathbf{e}_{(k+d-j-1)_n+1}^* \right), \\ \qquad 1 \le k \le d-1, \\ \dfrac{1}{\sqrt{d}} \displaystyle\sum_{i=k-d+1}^{k} \mathbf{e}_i \mathbf{e}_{(k+d-i-1)_n+1}^*, \quad d \le k \le n, \end{cases} \tag{A.16}$$

where $(\cdot)_n$ denotes the modulo operation that finds the remainder after division by n. The above basis $\{A_k\}_{k=1}^n$ for $\mathcal{H}_c(n,d)$ also satisfies the aforementioned properties of that for $\mathcal{H}(n,d)$.

Similarly, all elements of a structured matrix with the wrap-around property (e.g., wrap-around Hankel matrix) are repeated by the same number of times. Thus, the corresponding basis $\{A_k\}_{k=1}^n \subset \mathbb{C}^{n_1 \times n_2}$ has an extra property that

$$\|A_k\|_0 = \min\{n_1, n_2\}, \quad \forall k = 1, \ldots, n. \tag{A.17}$$

D. Incoherence conditions

The notion of the incoherence plays a crucial role in matrix completion and structured matrix completion. We recall the definitions using our notations. Suppose that $M \in \mathbb{C}^{n_1 \times n_2}$ is a rank-r matrix whose SVD is $U\Sigma V^*$. It was shown that the standard incoherence condition (9) alone suffices to provide a near optimal sample complexity for matrix completion [52]. For structured matrix completion, Chen and Chi [10] extended the notion of the standard incoherence as follows: M is said to satisfy the *basis incoherence* condition with parameter μ if

$$\begin{aligned} \max_{1 \le k \le n} \|U^* A_k\|_{\mathrm{F}} &\le \sqrt{\frac{\mu r}{n_1}}, \\ \max_{1 \le k \le n} \|V^* A_k^*\|_{\mathrm{F}} &\le \sqrt{\frac{\mu r}{n_2}}. \end{aligned} \tag{A.18}$$

When $A_k = \mathbf{e}_i \mathbf{e}_j^*$, the basis incoherence reduces to the standard incoherence with the same parameter. In general, two incoherence conditions are related as shown in the following lemma.

Lemma A.5. *Let $U \in \mathbb{C}^{n_1 \times r}$ and $V \in \mathbb{C}^{n_2 \times r}$. Let $A_k \in \mathbb{C}^{n_1 \times n_2}$ for $k = 1, \ldots, n$. Then,*

$$\max_{1 \le k \le n} \|U^* A_k\|_{\mathrm{F}} \le \left(\max_{1 \le i' \le n_1} \|U^* \mathbf{e}_{i'}\|_2 \right) \cdot \max_{1 \le k \le n} \left[\sum_{j=1}^{n_2} \left(\sum_{i=1}^{n_1} |[A_k]_{i,j}| \right)^2 \right]^{1/2},$$

$$\max_{1 \le k \le n} \|V^* A_k^*\|_{\mathrm{F}} \le \left(\max_{1 \le j' \le n_2} \|V^* \mathbf{e}_{j'}\|_2 \right) \cdot \max_{1 \le k \le n} \left[\sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} |[A_k]_{i,j}| \right)^2 \right]^{1/2}.$$

Proof of Lemma A.5. Let k be an arbitrary in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

$\{1,\dots,n\}$. Then,

$$\begin{aligned}\|U^*A_k\|_{\mathrm{F}}^2 &= \left\|U^*\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{e}_i\mathbf{e}_j^*[A_k]_{i,j}\right\|_{\mathrm{F}}^2\\ &= \left\|\sum_{j=1}^{n_2}\left(\sum_{i=1}^{n_1}[A_k]_{i,j}U^*\mathbf{e}_i\right)\mathbf{e}_j^*\right\|_{\mathrm{F}}^2\\ &= \sum_{j=1}^{n_2}\left\|\sum_{i=1}^{n_1}[A_k]_{i,j}U^*\mathbf{e}_i\right\|_2^2\\ &\le \sum_{j=1}^{n_2}\left(\sum_{i=1}^{n_1}|[A_k]_{i,j}|\,\|U^*\mathbf{e}_i\|_2\right)^2\\ &\le \left(\max_{1\le i'\le n_1}\|U^*\mathbf{e}_{i'}\|_2\right)^2\sum_{j=1}^{n_2}\left(\sum_{i=1}^{n_1}|[A_k]_{i,j}|\right)^2.\end{aligned}$$

Therefore, the first claim follows by taking maximum over k. The second claim is proved similarly by symmetry. □

By Lemma A.5, if (A.15) is satisfied, then the standard incoherence condition implies the basis incoherence condition with the same parameter μ. However, the converse is not true in general.

E. Proof of Theorem II.2

The previous work by Chen and Chi [10] could have proved the claim (in the case without the wrap-around property) as in their Theorem 4. However, a few steps in their proof depend on a Vandermonde decomposition of $\mathcal{L}(\mathbf{x})$, where the generators of the involved Vandermonde matrices are of unit modulus. Therefore, the original version [10, Theorem 4] only applies to the spectral compressed sensing.

Essentially, their results apply to the setup in this paper with slight modifications. In the below, a summary of the proof in the previous work [10] will be presented with emphasis on necessary changes that enable the extension of the result by Chen and Chi [10] to the setup of this theorem.

We first adopt notations from the previous work [10]. Define $\mathcal{A}_k : \mathbb{C}^{n_1\times n_2}\to\mathbb{C}^{n_1\times n_2}$ by

$$\mathcal{A}_k(M) = A_k\langle A_k, M\rangle$$

for $k = 1,\dots,n$, where $\langle A,B\rangle = \mathrm{Tr}(A^*B)$ and $\mathrm{Tr}(\cdot)$ is the trace of a matrix. Then each $\mathcal{A}_k$ is an orthogonal projection onto the one-dimensional subspace spanned by A_k. The orthogonal projection onto the subspace spanned by $\{A_k\}_{k=1}^n$ is given as $\mathcal{A} = \sum_{k=1}^n \mathcal{A}_k$. The summation of the rank-1 projection operators in $\{\mathcal{A}_k\}_{k\in\Omega}$ is denoted by $\mathcal{A}_\Omega$, i.e., $\mathcal{A}_\Omega := \sum_{k\in\Omega}\mathcal{A}_k$. With repetitions in Ω, $\mathcal{A}_\Omega$ is not a projection operator. The summation of distinct elements in $\{\mathcal{A}_k\}_{k\in\Omega}$ is denoted by $\mathcal{A}'_\Omega$, which is a valid orthogonal projection. Let $\mathcal{L}(\mathbf{x}) = U\Lambda V^*$ denote the singular value decomposition of $\mathcal{L}(\mathbf{x})$. Then the tangent space T with respect to $\mathcal{L}(\mathbf{x})$ is defined as

$$T := \{UM^* + \widetilde{M}V^* :\ M\in\mathbb{C}^{n_2\times r},\ \widetilde{M}\in\mathbb{C}^{n_1\times r}\}.$$

Then the projection onto T and its orthogonal complement will be denoted by $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$, respectively. Let $\mathrm{sgn}(\widetilde{X})$ denote the sign matrix of $\widetilde{X}$, defined by $\widetilde{U}\widetilde{V}^*$, where $\widetilde{X} = \widetilde{U}\widetilde{\Lambda}\widetilde{V}^*$ denotes the SVD of $\widetilde{X}$. For example, $\mathrm{sgn}[\mathcal{L}(\mathbf{x})] = UV^*$. The identity operator for $\mathbb{C}^{n_1\times n_2}$ will be denoted by id.

The proof starts with Lemma A.6, which improves on the corresponding result by Chen and Chi [10, Lemma 1].

Lemma A.6 (Refinement of [10, Lemma 1])**.** *Suppose that $\mathcal{A}_\Omega$ satisfies*

$$\left\|\frac{n}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \le \frac{1}{2}. \tag{A.19}$$

If there exists a matrix $W\in\mathbb{C}^{n_1\times n_2}$ satisfying

$$(\mathcal{A}-\mathcal{A}'_\Omega)(W) = 0, \tag{A.20}$$

$$\|\mathcal{P}_T(W - \mathrm{sgn}[\mathcal{L}(\mathbf{x})])\|_{\mathrm{F}} \le \frac{1}{7n}, \tag{A.21}$$

and

$$\|\mathcal{P}_{T^\perp}(W)\| \le \frac{1}{2}, \tag{A.22}$$

then $\mathbf{x}$ is the unique minimizer to (10).

Proof of Lemma A.6. See Appendix F. □

Lemma A.6, similarly to [10, Lemma 1], claims that if there exists a dual certificate matrix W, which satisfies (A.20) to (A.22), then $\mathbf{x}$ is the unique minimizer to (10). Compared to the previous result [10, Lemma 1], Lemma A.6 allows a larger deviation of the dual certificate W from the sign matrix of $\mathcal{L}(\mathbf{x})$. (Previously, the upper bound was in the order of n^{-2}.)

Remark A.7. *The relaxed condition on W in* (A.21) *provides a performance guarantee at sample complexity of the same order compared to the previous work [10]. However, in the noisy case, this relaxed condition provides an improved performance guarantee with significantly smaller noise amplification factor given in Theorem II.3.*

The next step is to construct a dual certificate W that satisfies (A.20) to (A.22). The version of the golfing scheme by Chen and Chi [10] still works in the setup of this theorem. They construct a dual certificate W as follows: recall that the elements of Ω are i.i.d. following the uniform distribution on $[n] := \{0,\cdots,n-1\}$. The multi-set Ω is partitioned into j_0 multi-sets, $\Omega_1,\dots,\Omega_{j_0}$ so that each Ω_j contains m/j_0 i.i.d. samples. A sequence

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

of matrices $(F_0, \dots, F_{j_0})$ are generated recursively by

$$F_j = \mathcal{P}_T\left(\mathcal{A} - \frac{nj_0}{m}A^*_{\Omega_j}\right)F_{j-1}, \quad j = 1, \dots, j_0,$$

starting from $F_0 = \operatorname{sgn}[\mathcal{L}(\mathbf{x})] = UV^*$. Then, W is obtained by

$$W = \sum_{j=1}^{j_0}\left(\frac{nj_0}{m}\mathcal{A}^*_{\Omega_j} + \mathrm{id} - \mathcal{A}\right)F_{j-1}.$$

Chen and Chi showed that if $j_0 = 3\log_{1/\epsilon} n$ for a small constant $\epsilon < e^{-1}$, then W satisfies (A.20) and (A.21) with high probability [10, Section VI.C]. In fact, they showed that a sufficient condition for (A.21) given by

$$\|\mathcal{P}_T(W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})])\|_{\mathrm{F}} \le \frac{1}{2n^2}$$

is satisfied. Thus, without any modification, their arguments so far apply to completion of structured matrices in the setup of this theorem. Chen and Chi verified that W satisfies (A.20) and (A.21) [10, Section VI.C]. Without any modification, their arguments so far apply to completion of structured matrices in the setup of this theorem.

They verified that W also satisfies the last property in (A.22) with some technical conditions [10, Section VI.D]. Specifically, Chen and Chi verified that W satisfies (A.22) through a sequence of lemmas [10, Lemmas 4,5,6,7] using intermediate quantities given in terms of the following two norms:

$$\|M\|_{\mathcal{A},\infty} := \max_{1\le k\le n} |\langle A_k, M\rangle|\,\|A_k\|, \tag{A.23}$$

and

$$\|M\|_{\mathcal{A},2} := \left(\sum_{k=1}^{n} |\langle A_k, M\rangle|^2\,\|A_k\|^2\right)^{1/2}. \tag{A.24}$$

Since most of their arguments in [10, Sections VI.D, VI.E] generalize to the setup of our theorem, we do not repeat technical details here. However, there was one place where the generalization fails. The results in [10, Lemma 7] provide upper bounds on the initialization of the dual certificate algorithm in the above two norms. We found that Chen and Chi used both the standard incoherence (9) and the basis incoherence (A.18) in this step. In fact, the proof of [10, Lemma 7] depends crucially on a Vandermonde decomposition with generators of unit modulus. In spectral compressed sensing, by controlling the condition number of Vandermonde matrices, both incoherence properties are satisfied with the same parameter. In fact, this is the place where their proof fails to generalize to other structured matrix completion.

In our setup, we assume that (A.15) is satisfied. By Lemma A.5, the standard incoherence property implies the basis incoherence, and then the dependence on the structure due to a Vandermonde decomposition disappears. Thus, only the standard incoherence of $\mathcal{L}(\mathbf{x})$ is included among the hypotheses.

Lemma A.8. *[10, Lemma 7] Let $\|\cdot\|_{\mathcal{A},\infty}$ and $\|\cdot\|_{\mathcal{A},2}$ be defined respectively in* (A.23) *and* (A.24)*. The standard incoherence property with parameter μ implies that there exists an absolute constant c_6 such that*

$$\|UV^*\|_{\mathcal{A},\infty} \le \frac{\mu r}{\min(n_1, n_2)}, \tag{A.25}$$

$$\|UV^*\|^2_{\mathcal{A},2} \le \frac{c_6\mu r\log^2 n}{\min(n_1, n_2)}, \tag{A.26}$$

and

$$\left\|\mathcal{P}_T\left(\|A_k\|_0^{1/2}A_k\right)\right\|^2_{\mathcal{A},2} \le \frac{c_6\mu r\log^2 n}{\min(n_1, n_2)}, \quad \forall k = 1, \dots, n. \tag{A.27}$$

Proof of Lemma A.8. Although [10, Lemma 7] did not assume that $U \in \mathbb{C}^{n_1\times r}$ (resp. $V \in \mathbb{C}^{n_2\times r}$) consists of the left (resp. right) singular vectors of a rank-r Hankel matrix with a Vandermode decomposition with generators of unit modulus, this condition was used in the proof by Chen and Chi [10, Appendix H]. More precisely, they used the Vandermonde decomposition to get the following inequalities:

$$\max_{1\le i\le n_1}\|U^*\mathbf{e}_i\|_2^2 \le \frac{\mu r}{n_1} \quad \text{and} \quad \max_{1\le j\le n_2}\|V^*\mathbf{e}_j\|_2^2 \le \frac{\mu r}{n_2}.$$

These inequalities are exactly the standard incoherence property with parameter μ. Except these inequalities, their proof generalizes without requiring the Vandermonde decomposition. Thus, we slightly modify [10, Lemma 7] by including the standard incoherence as an assumption to the lemma. □

The proof by Chen and Chi [10] focused on the Hankel-block-Hankel matrix where the elements in the basis $\{A_k\}_{k=1}^n$ have varying sparsity levels. In the case of structured matrices with the wrap-around property, the sparsity levels of $\{A_k\}_{k=1}^n$ are the same. Thus, this additional property can be used to tighten the sample complexity by reducing the order of $\log n$ term. More specifically, we improve [10, Lemma 7] with the wrap-around property in the next lemma. (The upper bounds on the terms in $\|\cdot\|_{\mathcal{A},2}$ were larger by factor of $\log^2 n$ in [10, Lemma 7].)

Lemma A.9 (Analog of [10, Lemma 7] with the wrap-around property)**.** *Let $\|\cdot\|_{\mathcal{A},\infty}$ and $\|\cdot\|_{\mathcal{A},2}$ be defined respectively in* (A.23) *and* (A.24)*. The standard*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

incoherence property with parameter μ implies

$$\|UV^*\|_{\mathcal{A},\infty} \le \frac{\mu r}{\min(n_1,n_2)}, \tag{A.28}$$

$$\|UV^*\|_{\mathcal{A},2}^2 \le \frac{\mu r}{\min(n_1,n_2)}, \tag{A.29}$$

and

$$\left\|\mathcal{P}_T\left(\|A_k\|_0^{1/2} A_k\right)\right\|_{\mathcal{A},2}^2 \le \frac{9\mu r}{\min(n_1,n_2)}, \quad \forall k = 1,\dots,n. \tag{A.30}$$

Proof of Lemma A.9. See Appendix G. □

In the wrap-around case, it only remains to verify that we can drop the order of $\log n$ from 4 to 2. In the previous work [10, Section VI.E], the $\log^4 n$ term appears only through the parameter μ_5, which is in the order of $\log^2 n$. Due to Lemma A.9, parameter μ_5 reduces by factor of $\log^2 n$. Thus, the sample complexity reduces by the same factor. This completes the proof.

F. Proof of Lemma A.6

Our proof essentially adapts the arguments of Chen and Chi [10, Appendix B]. The upper bound on the deviation of W from $\operatorname{sgn}[\mathcal{L}(\mathbf{x})]$ in (A.21) is sharpened in order by optimizing parameters.

Let $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{h}$ be the minimizer to (10). We show that $\mathcal{L}(\mathbf{h}) = 0$ in two complementary cases. Then by the injectivity of $\mathcal{L}$, $\mathbf{h} = 0$, or equivalently, $\hat{\mathbf{x}} = \mathbf{x}$.

Case 1: We first consider the case when $\mathcal{L}(\mathbf{h})$ satisfies

$$\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} \le 3n \|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_{\mathrm{F}}. \tag{A.31}$$

Since T is the tangent space of $\mathcal{L}(\mathbf{x})$, $\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{x}) = 0$. Thus $\mathcal{P}_T(\operatorname{sgn}[\mathcal{L}(\mathbf{x})] + \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})]) = \mathcal{P}_T(\operatorname{sgn}[\mathcal{L}(\mathbf{x})])$. Furthermore, $\|\operatorname{sgn}[\mathcal{L}(\mathbf{x})] + \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})]\| \le 1$. Therefore, $\operatorname{sgn}[\mathcal{L}(\mathbf{x})] + \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})]$ is a valid sub-gradient of the nuclear norm at $\mathcal{L}(\mathbf{x})$. Then it follows that

$$\begin{aligned}
&\|\mathcal{L}(\mathbf{x}) + \mathcal{L}(\mathbf{h})\|_* \\
&\ge \|\mathcal{L}(\mathbf{x})\|_* + \langle \operatorname{sgn}[\mathcal{L}(\mathbf{x})] + \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})],\ \mathcal{L}(\mathbf{h})\rangle \\
&= \|\mathcal{L}(\mathbf{x})\|_* + \langle W, \mathcal{L}(\mathbf{h})\rangle + \langle \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})],\ \mathcal{L}(\mathbf{h})\rangle \\
&\quad - \langle W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})], \mathcal{L}(\mathbf{h})\rangle.
\end{aligned} \tag{A.32}$$

In fact, $\langle W, \mathcal{L}(\mathbf{h})\rangle = 0$ as shown below. The inner product of $\mathcal{L}(\mathbf{h})$ and W is decomposed as

$$\begin{aligned}
\langle W, \mathcal{L}(\mathbf{h})\rangle = \ & \langle W, (\mathrm{id} - \mathcal{A})\mathcal{L}(\mathbf{h})\rangle + \langle W, (\mathcal{A} - \mathcal{A}'_\Omega)\mathcal{L}(\mathbf{h})\rangle \\
& + \langle W, \mathcal{A}'_\Omega\mathcal{L}(\mathbf{h})\rangle.
\end{aligned} \tag{A.33}$$

Indeed, all three terms in the right-hand-side of (A.33) are 0. This can be shown as follows. Since $\mathcal{A}$ is the orthogonal projection onto the range space of $\mathcal{L}$, the first term is 0. The second term is 0 by the assumption on W in (A.20). Since $\hat{\mathbf{x}}$ is feasible for (10), $P_\Omega(\hat{\mathbf{x}}) = P_\Omega(\mathbf{x})$. Thus $P_\Omega(\mathbf{h}) = P_\Omega(\hat{\mathbf{x}} - \mathbf{x}) = 0$. Since $\{A_k\}_{k=1}^n$ is an orthonormal basis, we have

$$\mathcal{A}_\omega\mathcal{L}(\mathbf{h}) = \sum_{k\in[n]\setminus\Omega} \langle \mathbf{e}_k, \mathbf{h}\rangle \langle \mathcal{A}_\omega, A_k\rangle = 0, \quad \forall \omega \in \Omega. \tag{A.34}$$

It follows that $\mathcal{A}'_\Omega\mathcal{L}(\mathbf{h}) = 0$. Thus, the third term of the right-hand-side of (A.33) is 0.

Since the $\operatorname{sgn}(\cdot)$ operator commutes with $\mathcal{P}_{T^\perp}$, and $\mathcal{P}_{T^\perp}$ is idempotent, we get

$$\begin{aligned}
\langle \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})],\ \mathcal{L}(\mathbf{h})\rangle &= \langle \mathcal{P}_{T^\perp}\operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})],\ \mathcal{L}(\mathbf{h})\rangle \\
&= \langle \operatorname{sgn}[\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})],\ \mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\rangle \\
&= \|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_*.
\end{aligned}$$

Then (A.32) implies

$$\begin{aligned}
\|\mathcal{L}(\mathbf{x}) + \mathcal{L}(\mathbf{h})\|_* \ \ge \ & \|\mathcal{L}(\mathbf{x})\|_* + \|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_* \\
& - \langle W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})], \mathcal{L}(\mathbf{h})\rangle.
\end{aligned} \tag{A.35}$$

We derive an upper bound on the magnitude of the third term in the right-hand-side of (A.35) given by

$$\begin{aligned}
&|\langle W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})], \mathcal{L}(\mathbf{h})\rangle| \\
&= |\langle \mathcal{P}_T(W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})]), \mathcal{L}(\mathbf{h})\rangle + \langle \mathcal{P}_{T^\perp}(W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})]), \mathcal{L}(\mathbf{h})\rangle| \\
&\le |\langle \mathcal{P}_T(W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})]), \mathcal{L}(\mathbf{h})\rangle| + |\langle \mathcal{P}_{T^\perp}(W), \mathcal{L}(\mathbf{h})\rangle| \qquad \text{(A.36a)} \\
&\le \|\mathcal{P}_T(W - \operatorname{sgn}[\mathcal{L}(\mathbf{x})])\|_{\mathrm{F}} \|\mathcal{P}_T\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} \\
&\quad + \|\mathcal{P}_{T^\perp}(W)\| \|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_* \qquad \text{(A.36b)} \\
&\le \frac{1}{7n}\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} + \frac{1}{2}\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_*, \qquad \text{(A.36c)}
\end{aligned}$$

where (A.36a) holds by the triangle inequality and the fact that $\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{x}) = 0$; (A.36b) by Hölder's inequality; (A.36c) by the assumptions on W in (A.21) and (A.22).

We continue by applying (A.36) to (A.35) and get

$$\begin{aligned}
\|\mathcal{L}(\mathbf{x}) + \mathcal{L}(\mathbf{h})\|_* &\ge \|\mathcal{L}(\mathbf{x})\|_* - \frac{1}{7n}\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} + \frac{1}{2}\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_* \\
&\ge \|\mathcal{L}(\mathbf{x})\|_* - \frac{3}{7}\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} + \frac{1}{2}\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} \\
&= \|\mathcal{L}(\mathbf{x})\|_* + \frac{1}{14}\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_{\mathrm{F}},
\end{aligned}$$

where the second step follows from (A.31).

Then, $\|\mathcal{L}(\hat{\mathbf{x}})\|_* \ge \|\mathcal{L}(\mathbf{x})\|_* \ge \|\mathcal{L}(\hat{\mathbf{x}})\|_*$, which implies $\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h}) = 0$. By (A.31), we also have $\mathcal{P}_T\mathcal{L}(\mathbf{h}) = 0$. Therefore, it follows that $\mathcal{L}(\mathbf{h}) = 0$.

Case 2: Next, we consider the complementary case when $\mathcal{L}(\mathbf{h})$ satisfies

$$\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\|_{\mathrm{F}} \ge 3n \|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\|_{\mathrm{F}}. \tag{A.37}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

Note that (A.34) implies $\mathcal{A}_\Omega\mathcal{L}(\mathbf{h}) = 0$. Then together with $(\mathrm{id} - \mathcal{A})\mathcal{L} = 0$, we get

$$\left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{L}(\mathbf{h}) = 0,$$

which implies

$$\begin{aligned} 0 &= \left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{L}(\mathbf{h})\right\rangle \\ &= \left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\rangle \\ &\quad + \left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\right\rangle. \end{aligned} \tag{A.38}$$

The magnitude of the first term in the right-hand-side of (A.38) is lower-bounded by

$$\begin{aligned} &\left|\left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\rangle\right| \\ &= \left|\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \mathcal{P}_T\mathcal{L}(\mathbf{h})\rangle\right| \\ &\quad - \left|\left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\mathcal{A} - \frac{n}{m}\mathcal{A}_\Omega\right)\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\rangle\right| \\ &\geq \left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2 - \left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{n}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T\right\| \left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2 \\ &\geq \frac{1}{2}\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2, \end{aligned} \tag{A.39}$$

where the last step follows from the assumption in (A.19).

Next, we derive an upper bound on the second term in the right-hand-side of (A.38). Since $\mathcal{A}_{\omega_j}$ is an orthogonal projection for $j \in [m]$, the operator norm of $\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}$ is upper-bounded by

$$\begin{aligned} \left\|\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right\| &\leq \frac{n}{m}\left(\left\|\mathcal{A}_{\omega_1} + \mathrm{id} - \mathcal{A}\right\| + \sum_{j=2}^{m}\left\|\mathcal{A}_{\omega_j}\right\|_{\mathrm{F}}\right) \\ &\leq \frac{n}{m}\left(\max\left(\left\|\mathcal{A}_{\omega_1}\right\|, \left\|\mathrm{id} - \mathcal{A}\right\|\right) + \sum_{j=2}^{m}\left\|\mathcal{A}_{\omega_j}\right\|_{\mathrm{F}}\right) \\ &\leq n, \end{aligned} \tag{A.40}$$

where the second step follows since $\mathcal{A}_{\omega_1}(\mathrm{id} - \mathcal{A}) = 0$.

The second term in the right-hand-side of (A.38) is then upper-bounded by

$$\begin{aligned} &\left|\left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\right\rangle\right| \\ &\leq \left\|\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right\| \left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}} \left\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}} \\ &\leq n\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}} \left\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}, \end{aligned} \tag{A.41}$$

where the last step follows from (A.40).

Applying (A.39) and (A.41) to (A.38) provides

$$\begin{aligned} 0 &= \left|\left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\rangle\right| \\ &\quad - \left|\left\langle \mathcal{P}_T\mathcal{L}(\mathbf{h}), \left(\frac{n}{m}\mathcal{A}_\Omega + \mathrm{id} - \mathcal{A}\right)\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\right\rangle\right| \\ &\geq \frac{1}{2}\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2 - n\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}} \left\|\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}} \\ &\geq \frac{1}{2}\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2 - \frac{1}{3}\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2 \\ &= \frac{1}{6}\left\|\mathcal{P}_T\mathcal{L}(\mathbf{h})\right\|_{\mathrm{F}}^2 \geq 0, \end{aligned}$$

where the second inequality holds by (A.37).

Then, it follows that $\mathcal{P}_T\mathcal{L}(\mathbf{h}) = 0$. By (A.37), we also have $\mathcal{P}_{T^\perp}\mathcal{L}(\mathbf{h}) = 0$. Therefore, $\mathcal{L}(\mathbf{h}) = 0$, which completes the proof.

G. Proof of Lemma A.9

The proof is obtained by slightly modifying that of [10, Lemma 7].

The first upper bound in (A.28) is derived as follows:

$$\begin{aligned} \left\|UV^*\right\|_{\mathcal{A},\infty} &= \max_{1\leq k\leq n} \left|\langle A_k, UV^*\rangle\right| \left\|A_k\right\| \\ &= \max_{1\leq k\leq n} \frac{\left|\sum_{(i,j)\in\mathrm{supp}(A_k)}[UV^*]_{i,j}\right|}{\left\|A_k\right\|_0} \\ &\leq \max_{1\leq k\leq n} \max_{(i,j)\in\mathrm{supp}(A_k)} \left|[UV^*]_{i,j}\right| \\ &= \max_{1\leq i\leq n_1} \max_{1\leq j\leq n_2} \left|[UV^*]_{i,j}\right| \\ &= \max_{1\leq i\leq n_1} \max_{1\leq j\leq n_2} \left|\mathbf{e}_i^* UV^*\mathbf{e}_j\right| \\ &= \max_{1\leq i\leq n_1} \left\|U^*\mathbf{e}_i\right\|_2 \max_{1\leq j\leq n_2} \left\|V^*\mathbf{e}_j\right\|_2 \\ &\leq \frac{\mu r}{\sqrt{n_1 n_2}} \leq \frac{\mu r}{\min(n_1, n_2)}. \end{aligned}$$

This proves (A.28).

Next, to prove (A.29) and (A.30), we use the following lemma.

Lemma A.10. *Let $M \in \mathbb{C}^{n_1\times n_2}$. Then,*

$$\left\|M\right\|_{\mathcal{A},2}^2 \leq \max\left(\max_{1\leq i\leq n_1}\left\|\mathbf{e}_i^* M\right\|_2^2, \max_{1\leq j\leq n_2}\left\|M\mathbf{e}_j\right\|_2^2\right) \tag{A.42}$$

Proof of Lemma A.10. See Appendix H. □

Then, (A.29) is proved as follows: Since U and V are unitary matrices, we have

$$\left\|\mathbf{e}_i^* UV^*\right\|_{\mathrm{F}} = \left\|U^*\mathbf{e}_i\right\|_2 \quad \text{and} \quad \left\|UV^*\mathbf{e}_j\right\|_{\mathrm{F}} = \left\|V^*\mathbf{e}_j\right\|_2$$

for all $1 \leq i \leq n_1$ and for all $1 \leq j \leq n_2$. Thus,

$$\max\left(\max_{1\leq i\leq n_1}\left\|\mathbf{e}_i^* UV^*\right\|_{\mathrm{F}}^2, \max_{1\leq j\leq n_2}\left\|UV^*\mathbf{e}_j\right\|_{\mathrm{F}}^2\right) \leq \frac{\mu r}{\min(n_1, n_2)}. \tag{A.43}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2016.2629078, IEEE Transactions on Information Theory

Then (A.29) follows by applying (A.43) to Lemma A.10 with $M = UV^*$.

Lastly, we prove (A.30). By definition of $\mathcal{P}_T$,

$$\left\|\mathbf{e}_i^*\left[\mathcal{P}_T\left(\|A_k\|_0^{1/2}A_k\right)\right]\right\|_{\mathrm{F}}^2 \le 3\left\|\mathbf{e}_i^*UU^*\|A_k\|_0^{1/2}A_k\right\|_{\mathrm{F}}^2 + 3\left\|\mathbf{e}_i^*\|A_k\|_0^{1/2}A_kVV^*\right\|_{\mathrm{F}}^2 + 3\left\|\mathbf{e}_i^*UU^*\|A_k\|_0^{1/2}A_kVV^*\right\|_{\mathrm{F}}^2, \tag{A.44}$$

for all $i \in \{1,\dots,n_1\}$. The first term in the right-hand-side of (A.44) is upper-bounded by

$$\left\|\mathbf{e}_i^*UU^*\|A_k\|_0^{1/2}A_k\right\|_{\mathrm{F}}^2 \le \|\mathbf{e}_i^*U\|_2^2\left\|\|A_k\|_0^{1/2}A_k\right\|^2 \le \frac{\mu r}{n_1}, \tag{A.45}$$

where the last step follows from $\|A_k\| \le \|A_k\|_0^{-1/2}$. Since $\|VV^*\| \le 1$, the first term dominates the third term in the right-hand-side of (A.44). Note that $\|A_k\|_0^{1/2}A_k$ is a submatrix of a permutation matrix. Therefore, $\mathbf{e}_i^*\|A_k\|_0^{1/2}A_k = \mathbf{e}_j^*$ for some $j \in \{1,\dots,n_1\}$. Then, the second term in the right-hand-side of (A.44) is upper-bounded by

$$\left\|\mathbf{e}_i^*\|A_k\|_0^{1/2}A_kVV^*\right\|_{\mathrm{F}}^2 = \left\|\mathbf{e}_j^*VV^*\right\|_{\mathrm{F}}^2 \le \frac{\mu r}{n_2}. \tag{A.46}$$

Plugging (A.45) and (A.46) to (A.44) provides

$$\max_{1\le i\le n_1}\left\|\mathbf{e}_i^*\left[\mathcal{P}_T\left(\|A_k\|_0^{1/2}A_k\right)\right]\right\|_{\mathrm{F}}^2 \le \frac{9\mu r}{\min(n_1,n_2)}. \tag{A.47}$$

By symmetry, we also get

$$\max_{1\le j\le n_2}\left\|\left[\mathcal{P}_T\left(\|A_k\|_0^{1/2}A_k\right)\right]\mathbf{e}_j\right\|_{\mathrm{F}}^2 \le \frac{9\mu r}{\min(n_1,n_2)}. \tag{A.48}$$

Applying (A.47) and (A.48) to Lemma A.10 with $M = \mathcal{P}_T\left(\|A_k\|_0^{1/2}A_k\right)$ completes the proof.

H. Proof of Lemma A.10

The inequality in (A.42) is proved as follows:

$$\begin{aligned}
\|M\|_{\mathcal{A},2}^2 &= \sum_{k=1}^n |\langle A_k, M\rangle|^2\|A_k\|^2 \\
&= \sum_{k=1}^n \frac{\left|\sum_{(i,j)\in\mathrm{supp}(A_k)}[M]_{i,j}\right|^2}{\|A_k\|_0}\|A_k\|^2 \\
&\le \sum_{k=1}^n \frac{\left(\sum_{(i,j)\in\mathrm{supp}(A_k)}|[M]_{i,j}|\right)^2}{\|A_k\|_0}\|A_k\|^2 \\
&\le \sum_{k=1}^n \sum_{(i,j)\in\mathrm{supp}(A_k)}|[M]_{i,j}|^2\|A_k\|^2 \\
&\le \frac{1}{\min\{n_1,n_2\}}\sum_{k=1}^n\sum_{(i,j)\in\mathrm{supp}(A_k)}|[M]_{i,j}|^2 \\
&= \frac{1}{\min\{n_1,n_2\}}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}|[M]_{i,j}|^2 \\
&\le \max\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\|\mathbf{e}_i^*M\|_2^2,\ \frac{1}{n_2}\sum_{j=1}^{n_2}\|M\mathbf{e}_j\|_2^2\right) \\
&\le \max\left(\max_{1\le i\le n_1}\|\mathbf{e}_i^*M\|_2^2,\ \max_{1\le j\le n_2}\|M\mathbf{e}_j\|_2^2\right),
\end{aligned}$$

where the third inequality follow from (A.17).

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

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

[3] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] R. Prony, "Essai experimentale et analytique," *J. de lEcole Polytechnique*, 1795.

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

[6] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear algebra and its applications*, vol. 18, no. 2, pp. 95–138, 1977.

[7] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

[8] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

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

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

[11] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

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

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

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

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

[16] M. Unser and P. D. Tafti, *An introduction to sparse stochastic processes*. Cambridge University Press, 2010.

[17] M. Unser, "Sampling-50 years after shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[18] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. on Signal Processing*, vol. 42, no. 11, pp. 2915–2925, 1994.

[19] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix," *IEEE Trans. on Computational Imaging (in press)*, 2016.

[20] D. Lee, K. H. Jin, E. Y. Kim, S.-H. Park, and J. C. Ye, "Acceleration of MR parameter mapping using annihilating filter-based low rank Hankel matrix (ALOHA)," *Magn. Reson. Med*, no. 10.1002/mrm.26081, 2016.

[21] J. Lee, K. H. Jin, and J. C. Ye, "Reference-free single-pass EPI Nyquist ghost correction using annihilating filter-based low rank Hankel structured matrix," *Magn. Reson. Med*, no. 10.1002/mrm.26077, 2016.

[22] K. H. Jin, J.-Y. Um, D. Lee, J. Lee, S.-H. Park, and J. C. Ye, "MRI artifact correction using sparse + low-rank decomposition of annihilating flter-based Hankel matrix," *Magn. Reson. Med.*, no. 10.1002/mrm.26330, 2016.

[23] K. H. Jin and J. C. Ye, "Annihilating filter based low rank hankel matrix approach for image inpainting," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 3498 – 3511, 2015.

[24] J. Min, L. Carlini, M. Unser, S. Manley, and J. C. Ye, "Fast live cell imaging at nanometer scale using annihilating filter based low rank Hankel matrix approach," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2015, pp. 95 970V–95 970V.

[25] K. H. Jin and J. C. Ye, "Sparse + low rank decomposition of annihilating filter-based Hankel matrix for impulse noise removal," *arXiv preprint arXiv:1510.05559*, 2015.

[26] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian versus sparse stochastic processes–Part I: Continuous-domain theory," *IEEE Trans. on Information Theory*, vol. 60, no. 3, pp. 1945–1962, 2014.

[27] M. Unser, P. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian versus sparse stochastic processes–Part II: Discrete-domain theory," *IEEE Trans. on Information Theory*, vol. 60, no. 5, pp. 3036–3051, 2014.

[28] D. L. Boley, F. T. Luk, and D. Vandevoorde, "Vandermonde factorization of a Hankel matrix," *Scientific computing (Hong Kong, 1997)*, pp. 27–39, 1998.

[29] R. Badeau, B. David, and G. Richard, "High-resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. on Signal Processing*, vol. 54, no. 4, pp. 1341–1350, 2006.

[30] D. Batenkov and Y. Yomdin, "On the accuracy of solving confluent Prony systems," *SIAM Journal on Applied Mathematics*, vol. 73, no. 1, pp. 134–154, 2013.

[31] D. Batenkov, "Decimated generalized Prony systems," *arXiv preprint arXiv:1308.0753*, 2013.

[32] A. Moitra, "Super-resolution, extremal functions and the condition number of Vandermonde matrices," in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*. ACM, 2015, pp. 821–830.

[33] D. Batenkov, "Stability and super-resolution of generalized spike recovery," *Applied and Computational Harmonic Analysis (in press)*, 2016.

[34] W. Gautschi, "Norm estimates for inverses of Vandermonde matrices," *Numerische Mathematik*, vol. 23, no. 4, pp. 337–347, 1974.

[35] ——, "On inverses of Vandermonde and confluent Vandermonde matrices," *Numerische Mathematik*, vol. 4, no. 1, pp. 117–123, 1962.

[36] ——, "On inverses of Vandermonde and confluent Vandermonde matrices. II," *Numerische Mathematik*, vol. 5, no. 1, pp. 425–430, 1963.

[37] ——, "On inverses of Vandermonde and confluent Vandermonde matrices III," *Numerische Mathematik*, vol. 29, no. 4, pp. 445–450, 1978.

[38] C. Aubel and H. Bölcskei, "Deterministic performance analysis of subspace methods for cisoid parameter estimation," in *IEEE International Symposium on Information Theory, ISIT 2016, Barcelona, Spain, July 10-15, 2016*, 2016, pp. 1551–1555.

[39] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[40] T. K. Sarkar and O. Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," *IEEE Antennas and Propagation Magazine*, vol. 37, no. 1, pp. 48–55, 1995.

[41] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. I. Theory," *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 821–833, 1993.

[42] ——, "B-spline signal processing. II. Efficiency design and applications," *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 834–848, 1993.

[43] M. Signoretto, V. Cevher, and J. A. Suykens, "An SVD-free approach to a class of structured low rank matrix optimization problems with application to system identification," in *IEEE Conf. on Decision and Control*, 2013.

[44] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Math. Prog. Comp.*, vol. 4, no. 4, pp. 333–361, 2012.

[45] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends in Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[46] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *arXiv preprint arXiv:1208.3922*, 2012.

[47] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[48] E. Van Den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[49] T. Goldstein and S. Osher, "The split Bregman method for l_1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[50] S. Elaydi, *An introduction to difference equations*, 3rd ed. Springer Science & Business Media, 2005.

[51] G.-C. Rota and P. Doubilet, *Finite operator calculus*. Academic Press, 1975.

[52] Y. Chen, "Incoherence-optimal matrix completion," *IEEE Trans. on Information Theory*, vol. 61, no. 5, pp. 2909–2923, 2015.