

Spectrally Sparse Signal Recovery via Hankel Matrix Completion with Prior Information

Xu Zhang, Yulong Liu and Wei Cui

Abstract—This paper studies the problem of reconstructing spectrally sparse signals from a small random subset of time domain samples via low-rank Hankel matrix completion with the aid of prior information. By leveraging the low-rank structure of spectrally sparse signals in the lifting domain and the similarity between the signals and their prior information, we propose a convex method to recover the undersampled spectrally sparse signals. The proposed approach integrates the inner product of the desired signal and its prior information in the lift domain into vanilla Hankel matrix completion, which maximizes the correlation between the signals and their prior information. Theoretical analysis indicates that when the prior information is reliable, the proposed method has a better performance than vanilla Hankel matrix completion, which reduces the number of measurements by a logarithmic factor. We also develop an ADMM algorithm to solve the corresponding optimization problem. Numerical results are provided to verify the performance of proposed method and corresponding algorithm.

Index Terms—Maximizing correlation, Hankel matrix completion, spectrally sparse signals, prior information.

I. INTRODUCTION

Spectrally sparse signal recovery refers to recovering a spectrally sparse signal from a small number of time domain samples, which is fundamental in various applications, such as medical imaging [1], radar imaging [2], analog-to-digital conversion [3] and channel estimation [4]. Let $\mathbf{x} = [x_0, \dots, x_{n-1}]^T \in \mathbb{C}^n$ denote the one-dimensional spectrally sparse signal to be estimated. Each entry of the desired signal \mathbf{x} is a weighted superposition of r complex sinusoids

$$x_k = \sum_{l=1}^r w_l e^{i2\pi k f_l},$$

where $k = 0, \dots, n-1$, $\{f_1, \dots, f_r\}$ and $\{w_1, \dots, w_r\}$ denote the normalized frequencies and amplitudes for the r sinusoids, respectively, and $f_l \in [0, 1)$ for $l = 1, \dots, r$.

In many practical applications, we only have access to a small subset of signal samples. For example, in the field of computed tomography (CT), only part of the desired signals can be observed to protect the patients from high-dose radiation [5]; in wideband signal sampling, it's challenging to build analog-to-digital converter according to Shannon sampling theorem, and hence only a small number of samples of the wideband signals can be acquired for reconstruction [3].

X. Zhang and W. Cui are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: connorzx@bit.edu.cn; cuiwei@bit.edu.cn).

Y. Liu is with the School of Physics, Beijing Institute of Technology, Beijing 100081, China (e-mail: yulongliu@bit.edu.cn).

Therefore, we have to figure out a way to recover the original signal \mathbf{x} from its random undersampled observations

$$\mathcal{P}_\Omega(\mathbf{x}) = \sum_{k \in \Omega} x_k \mathbf{e}_k,$$

where $\Omega \in \{0, \dots, n-1\}$ denote the index set of the entries we observe, \mathbf{e}_k denotes the k -th canonical basis of \mathbb{R}^n , and $\mathcal{P}_\Omega(\cdot)$ denotes the projection operator on the sampling index set Ω , i.e., $\mathcal{P}_\Omega(\mathbf{z}) = \sum_{k \in \Omega} \langle \mathbf{z}, \mathbf{e}_k \rangle \mathbf{e}_k$ for $\mathbf{z} \in \mathbb{C}^n$.

In order to reconstruct \mathbf{x} , structured low-rank completion methods have been proposed by using the low-rank Hankel structure of the spectrally sparse signals in the lifting domain

$$\begin{aligned} \min_{\mathbf{z}} \quad & \text{Rank}(\mathcal{H}(\mathbf{z})) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{z}) = \mathcal{P}_\Omega(\mathbf{x}), \end{aligned} \quad (1)$$

where $\text{Rank}(\cdot)$ returns the rank of matrix, and $\mathcal{H} : \mathbb{C}^n \rightarrow \mathbb{C}^{n_1 \times n_2}$ is a linear lifting operator to generate the Hankel low-rank structure. In particular, for a vector $\mathbf{x} = [x_0, x_1, \dots, x_{n-1}]^T \in \mathbb{C}^n$, the Hankel matrix $\mathcal{H}(\mathbf{x})$ is defined as

$$\mathcal{H}(\mathbf{x}) \triangleq \begin{bmatrix} x_0 & x_1 & \dots & x_{d-1} \\ x_1 & x_2 & \dots & x_d \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-d} & x_{n-d+1} & \dots & x_{n-1} \end{bmatrix},$$

where d denotes the matrix pencil parameter, $n_1 = n - d + 1$ and $n_2 = d$.

By using Vandermonde decomposition, the Hankel matrix $\mathcal{H}(\mathbf{x})$ can be decomposed as

$$\mathcal{H}(\mathbf{x}) = \sum_{l=1}^r w_l \mathbf{y}_l \mathbf{z}_l^H,$$

where $\mathbf{y}_l = [1, e^{i2\pi f_l}, \dots, e^{i2\pi(n_1-1)f_l}]^T$ and $\mathbf{z}_l = [1, e^{i2\pi f_l}, \dots, e^{i2\pi(n_2-1)f_l}]^T$, $l = 1, \dots, r$. When the frequencies are all distinct and $r \ll \min\{n_1, n_2\}$, $\mathcal{H}(\mathbf{x})$ is a low-rank matrix with $\text{Rank}(\mathcal{H}(\mathbf{x})) \leq r$.

Since Eq. (1) is a non-convex problem and solving it is NP-hard, an alternative approach based on convex relaxation is proposed to complete the low rank matrix, that is, Hankel matrix completion program [6], [7]

$$\begin{aligned} \min_{\mathbf{z}} \quad & \|\mathcal{H}(\mathbf{z})\|_* \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{z}) = \mathcal{P}_\Omega(\mathbf{x}), \end{aligned} \quad (2)$$

where $\|\cdot\|_*$ denotes the nuclear norm. Theoretical analysis was given to show that $O(r \log^4 n)$ samples are enough to recover the desired signal with high probability [6].

Apart from the sparsity constraint in the spectral domain, a reference signal ϕ that is similar to the original signal x sometimes is available to us. There are two main sources of this kind of prior information. The first source comes from natural (non-self-constructed) signals. In high resolution MRI [8]–[10], adjacent slices show good similarity with each other; in multiple-contrast MRI [11]–[13], different contrasts in the same scan are similar in structure; in dynamic CT [14], the scans for the same slice in different time have similar characteristics. The second source comes from self-constructed signals. One way is to use classical method to construct a similar signal. For example, filtered backprojection reconstruction algorithm from the dynamic scans was used to construct the prior information in dynamic CT [15]; smooth method was used to generate prior information in sparse-view CT [16]; the standard spectrum of dot object was used as the prior information in radar imaging [17]. The other way is to use machine learning to generate a similar signal. In [18], the authors generated the reference image by using a CNN network; similarly, other algorithms from deep learning can be used to create reference signals, see e.g. in [19], [20].

In this paper, we propose a convex approach to integrate prior information into the reconstruction of spectrally sparse signals by maximizing the correlation between signal z and prior information ϕ in the lifting domain

$$\begin{aligned} \min_z \quad & \|\mathcal{H}(z)\|_* - 2\lambda \operatorname{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(z) \rangle) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(z) = \mathcal{P}_\Omega(x), \end{aligned} \quad (3)$$

where $\lambda > 0$ is a tradeoff parameter, $\mathcal{G}(\cdot) = \mathcal{F}(\mathcal{H}(\cdot))$ is a composition operator, $\langle X, Y \rangle = (\operatorname{vec}(Y))^H \operatorname{vec}(X)$ is the inner product and $\operatorname{Re}(\cdot)$ returns the real part of a complex number. Here, $\mathcal{F} : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^{n_1 \times n_2}$ is a suitable operator to be designed in the sequel. Theoretical guarantees are provided to show that our method has better performance than vanilla Hankel matrix completion when the prior information is reliable. In addition, we propose an Alternating Direction Method of Multipliers (ADMM)-based optimization algorithm for efficient reconstruction of the desired signals.

A. Related Literature

Recovery of spectrally sparse signals has attracted great attentions in the past years. Conventional compressed sensing [21], [22] was used to estimate the spectrally sparse signals when the frequencies are located on a grid. In many practical applications, however, the frequencies lie off the grid, leading to the mismatch for conventional compressed sensing.

To recover the signals with off-the-grid frequencies, two kinds of methods are proposed: atomic norm minimization and low-rank structured matrix completion. By promoting the sparsity in a continuous frequency domain, atomic norm minimization [23], [24] demonstrated that $r \log(r) \log(n)$ random samples are sufficient to recover the desired signals exactly with high probability when the frequencies are well separated. Due to the fact that the sparsity in frequency domain leads to the low-rankness in the lifting time domain, low rank structured matrix completion [6], [7] was proposed to promote the low-rank structure in the lifting time domain. Their results

showed that $O(r \log^4(n))$ random samples are enough to correctly estimate the original signals with high probability when some incoherence conditions are satisfied.

Besides the sparse prior knowledge, other kinds of prior information are used to further improve the recovery performance. By using the similarity between original signal and reference signal, an adaptive weighted compressed sensing approach was considered in [25], which presented a better performance than conventional approach. Assuming that some frequency intervals or likelihood of each frequency of the desired signal is known a priori, a weighted atomic norm method was studied in [26], [27], which outperforms standard atomic norm approach.

While the above work considered spectrally sparse signal recovery with prior information based on conventional compressed sensing or atomic norm minimization, little work incorporates the prior information into low-rank structured matrix completion.

Recently, we proposed a novel method to recover structured signals by using the prior information via maximizing correlation [28], [29]. By introducing a negative inner product of the prior information and the desired signal into the objective function, theoretical guarantees and numerical results illustrated that the matrix completion approach proposed in [29] outperforms standard matrix completion procedure in [30]–[33] when the prior information is reliable.

Inspired by [29], this paper leverages the transform low-rank information in the lifting domain to recover the undersampled spectrally sparse signals with the help of the prior information. Different from [29], this paper studies the low-rank property in the lifting domain while the previous approach studies the low-rank property in original domain, leading to the change of the desired matrix from random matrix to Hankel random matrix. Accordingly, the sampling operator changes from sampling random entries to sampling random skew-diagonal. Therefore, different theoretical guarantees should be given to analyze the proposed approach. The analysis also should be extended from real number domain to complex number domain since the spectrally sparse signals are complex.

B. Paper Organization

The structures of this paper are arranged as follows. Preliminaries are provided in Section II. Performance guarantees are given in Section III. An extension to multi-dimensional models is provided in Section IV. The ADMM optimization algorithm is presented in Section V. Simulations are included in Section VI, and the conclusion is drawn in Section VII.

II. PRELIMINARIES

In this section, we introduce some important notation and definitions, which will be used in the sequel.

Let $\{\mathbf{A}_k\}_{k=0}^{n-1} \in \mathbb{C}^{n_1 \times n_2}$ be an orthonormal basis of Hankel matrices [7], [34], which is defined as

$$\mathbf{A}_k = \frac{1}{\sqrt{w_k}} \mathcal{H}(e_k), \quad k \in \{0, \dots, n-1\}$$

where

$$w_k = |\{(i, j) | i + j = k, 0 \leq i \leq n_1 - 1, 0 \leq j \leq n_2 - 1\}|,$$

and $|S|$ returns the cardinality of the set S . Then $\mathcal{H}(\mathbf{x})$ can be expressed as

$$\mathcal{H}(\mathbf{x}) = \sum_{k=0}^{n-1} \mathcal{A}_k(\mathcal{H}(\mathbf{x})) = \sum_{k=0}^{n-1} \langle \mathcal{H}(\mathbf{x}), \mathbf{A}_k \rangle \mathbf{A}_k, \quad (4)$$

where $\mathcal{A}_k(\mathbf{X}) \triangleq \langle \mathbf{X}, \mathbf{A}_k \rangle \mathbf{A}_k$ for $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$.

Let $\mathcal{H}(\mathbf{x}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ denote the compact singular value decomposition (SVD) of $\mathcal{H}(\mathbf{x})$ with $\mathbf{U} \in \mathbb{C}^{n_1 \times r}$, $\mathbf{\Sigma} \in \mathbb{R}^{r \times r}$ and $\mathbf{V} \in \mathbb{C}^{r \times n_2}$. Let the subspace \mathcal{T} denote the support of $\mathcal{H}(\mathbf{x})$ and \mathcal{T}^\perp be its orthogonal complement. Let $\text{sgn}(\widetilde{\mathbf{X}}) = \widetilde{\mathbf{U}}\mathbf{\Sigma}^H$ denote the sign matrix of $\widetilde{\mathbf{X}}$, where $\widetilde{\mathbf{X}} = \widetilde{\mathbf{U}}\mathbf{\Sigma}\mathbf{V}^H$ denotes the compact SVD of $\widetilde{\mathbf{X}}$.

In order to analyse the matrix completion problem theoretically, we need to introduce the standard incoherence condition as follows

$$\begin{aligned} \max_{1 \leq i \leq n_1} \|\mathbf{U}^H \mathbf{e}_i\|_2^2 &\leq \frac{\mu r}{n_1}, \\ \max_{1 \leq j \leq n_2} \|\mathbf{V}^H \mathbf{e}_j\|_2^2 &\leq \frac{\mu r}{n_2}. \end{aligned} \quad (5)$$

We also need to introduce the following norms which, respectively, measure the largest spectral norm among matrices $\{\mathcal{A}_k(\mathbf{X})\}_{k=1}^n$ and the ℓ_2 norm of $\{\|\mathcal{A}_k(\mathbf{X})\|\}_{k=1}^n$ [6], [7]

$$\|\mathbf{X}\|_{\mathcal{A}, \infty} \triangleq \max_{1 \leq k \leq n} |\langle \mathbf{X}, \mathbf{A}_k \rangle| \|\mathbf{A}_k\|, \quad (6)$$

and

$$\|\mathbf{X}\|_{\mathcal{A}, 2} \triangleq \left(\sum_{k=1}^n |\langle \mathbf{X}, \mathbf{A}_k \rangle|^2 \|\mathbf{A}_k\|^2 \right)^{1/2}. \quad (7)$$

III. THEORETICAL GUARANTEES

In this section, we start by giving the theoretical guarantees for the proposed method. Then we extend the analysis to noisy circumstance. Our main result shows that when the prior information is reliable, the proposed approach (3) can outperform previous approach (2) by a logarithmic factor.

Theorem 1. *Let $\mathcal{H}(\mathbf{x})$ be a rank- r matrix and satisfy the standard incoherence condition in Eq. (5) with parameter μ . Consider a multi-set $\Omega = \{j_1, \dots, j_m\}$ whose indicies $\{j_k\}_{k=1}^m$ are i.i.d. and follow the uniform distribution on $\{0, \dots, n-1\}$. If the sample size satisfies*

$$m \geq \max\{\Delta^2, 1\} c \mu c_s r \log^3 n \max\{\log(7n \|\mathbf{F}_0\|_F), 1\},$$

and the prior information satisfies

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\| < \frac{1}{2},$$

where $c > 0$ is an absolute constant,

$$c_s \triangleq \max\left\{\frac{n}{n_1}, \frac{n}{n_2}\right\}, \quad \mathbf{F}_0 \triangleq \mathcal{P}_{\mathcal{T}}(\text{sgn}[\mathcal{H}(\mathbf{x})] - \lambda \mathcal{G}(\phi)),$$

and

$$\Delta \triangleq \frac{4(\|\mathbf{F}_0\|_{\mathcal{A}, 2} + \|\mathbf{F}_0\|_{\mathcal{A}, \infty})}{1 - 2\|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\|},$$

then \mathbf{x} is the unique minimizer for the approach (3) with high probability.

Remark 1 (Comparison with [7, Theorem 1]). When there is no prior information or no reliable prior information, ϕ is set to be $\mathbf{0}$ and the program (3) would reduce to (2). However, when the prior information is reliable, the proposed approach can reduce the sampling size by $O(\log n)$ compared with the results [7, Theorem 1].

Remark 2 (The choice of operator \mathcal{F}). It should be noted that the choice of operator \mathcal{F} will influence the performance of the proposed program. According to the definition of \mathbf{F}_0 , it's not hard to see that $\mathcal{F}(\cdot) = \text{sgn}(\cdot)$ is a suitable choice to improve the sampling bound. In this case, the value of $\|\mathbf{F}_0\|_F$ will be very small when the subspace information of $\mathcal{H}(\phi)$ is very similar to that of $\mathcal{H}(\mathbf{x})$ and $\lambda = 1$. Accordingly, the program becomes

$$\begin{aligned} \min_{\mathbf{z}} \quad & \|\mathcal{H}(\mathbf{z})\|_* - 2\lambda \text{Re}(\langle \text{sgn}(\mathcal{H}(\phi)), \mathcal{H}(\mathbf{z}) \rangle) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{z}) = \mathcal{P}_\Omega(\mathbf{x}). \end{aligned}$$

Remark 3 (The choice of weight λ). Note that the sampling lower bound is determined by the value of $\|\mathbf{F}_0\|_F$ and the best choice of λ is the one that minimizes $\|\mathbf{F}_0\|_F$. The expression of $\|\mathbf{F}_0\|_F^2$ can be rewritten as

$$\begin{aligned} \|\mathbf{F}_0\|_F^2 &= \lambda^2 \|\mathcal{P}_{\mathcal{T}}(\mathcal{G}(\phi))\|_F^2 + \|\mathcal{P}_{\mathcal{T}}(\text{sgn}[\mathcal{H}(\mathbf{x})])\|_F^2 \\ &\quad - 2\lambda \text{Re}(\langle \mathcal{P}_{\mathcal{T}}(\text{sgn}[\mathcal{H}(\mathbf{x})]), \mathcal{G}(\phi) \rangle). \end{aligned}$$

So the optimal weight is

$$\lambda^* = \frac{\text{Re}(\langle \mathcal{P}_{\mathcal{T}}(\text{sgn}[\mathcal{H}(\mathbf{x})]), \mathcal{G}(\phi) \rangle)}{\|\mathcal{P}_{\mathcal{T}}(\mathcal{G}(\phi))\|_F^2}.$$

Let $\mathcal{G}(\phi) = \text{sgn}(\mathcal{H}(\phi))$ as Remark 2. When the prior information is close to the desired signal, λ should be around 1. On the contrary, when the prior information is extremely different from the desired signal, λ should be around 0.

Remark 4 (The wrap-around operator). When $\mathcal{H}(\cdot)$ is replaced with the following operator $\mathcal{H}_c(\cdot)$ with the wrap-around property

$$\mathcal{H}_c(\mathbf{x}) \triangleq \begin{bmatrix} x_0 & x_1 & \dots & x_{d-1} \\ x_1 & x_2 & \dots & x_d \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-d} & x_{n-d+1} & \dots & x_{n-1} \\ x_{n-d+1} & x_{n-d+2} & \dots & x_0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & x_0 & \dots & x_{d-2} \end{bmatrix},$$

where $\mathcal{H}_c(\mathbf{x}) \in \mathbb{C}^{n \times d}$, it is straightforward to obtain the lower bound for sample size by following the proof in [7]

$$m \geq \max\{\Delta^2, 1\} c \mu c_s r \log n \max\{\log(7n \|\mathbf{F}_0\|_F), 1\}.$$

In this case, $O(r \log n)$ samples are enough to exactly reconstruct the original signals when the prior information is reliable, which outperforms the atomic norm minimization in [23], [24].

$$\mathcal{H}^d(\mathcal{X}^d) = \begin{bmatrix} \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(0)) & \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(1)) & \cdots & \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(n_d-1)) \\ \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(1)) & \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(2)) & \cdots & \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(n_d)) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(N_d-n_d)) & \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(N_d-n_d+1)) & \cdots & \mathcal{H}^{d-1}(\mathcal{X}^{d-1}(N_d-1)) \end{bmatrix} \quad (8)$$

By straightly following [35, Theorem 7], an extension to the noisy version with bounded noise can be shown as follows.

Corollary 1. *Let $\mathcal{H}(\mathbf{x})$ be a rank- r matrix and satisfy the standard incoherence condition in Eq. (5) with parameter μ . Consider a multi-set $\Omega = \{j_1, \dots, j_m\}$ whose indicies $\{j_k\}_{k=1}^m$ are i.i.d. and follow the uniform distribution on $\{0, \dots, n-1\}$. Suppose the noisy observation $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{n} denotes bounded noise. Let \mathbf{x}^\dagger be the solution of the noisy version program*

$$\begin{aligned} \min_{\mathbf{z}} \quad & \|\mathcal{H}(\mathbf{z})\|_* - 2\lambda \operatorname{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{z}) \rangle) \\ \text{s.t.} \quad & \|\mathcal{P}_\Omega(\mathbf{z}) - \mathcal{P}_\Omega(\mathbf{y})\|_2 \leq \delta. \end{aligned}$$

If the sample size satisfies

$$m \geq \max\{\Delta^2, 1\} \mu c_s r \log^3 n \max\{\log(7n \|\mathbf{F}_0\|_F), 1\},$$

and the prior information satisfies

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\| < \frac{1}{2},$$

then the solution \mathbf{x}^\dagger satisfies that

$$\|\mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{x}^\dagger)\|_F \leq c\delta \left(n^2 + n^{\frac{3}{2}} \|\lambda \mathcal{G}(\phi)\|_F \right)$$

with high probability.

IV. EXTENSIONS TO MULTI-DIMENSIONAL MODELS

In this section, we extend the analysis from one-dimensional signal to multi-dimensional signal. Consider a d -way tensor $\mathcal{X}^d \in \mathbb{C}^{N_1 \times \dots \times N_d}$, each of whose entries can be denoted as

$$\mathcal{X}^d(k_1, \dots, k_d) = \sum_{l=1}^r w_l e^{j2\pi \sum_{j=1}^d k_j f_{j,l}},$$

for $(k_1, \dots, k_d) \in \{1, \dots, N_1\} \times \dots \times \{1, \dots, N_d\}$. Denote $\mathbf{f}_l = (f_{1,l}, \dots, f_{d,l}) \in [0, 1]^d$ as the frequency vector for $l = 1, \dots, r$.

Similar to the one-dimensional case, we use multi-level Hankel operator to lift \mathcal{X}^d to a low-rank matrix. See Eq. (8), where $\mathcal{X}^{d-1}(i) = \mathcal{X}^d(:, \dots, :, i)$, $i = 0, \dots, \max\{n_d - 1, N_d - n_d\}$. When $d = 1$, the above operator degrades to normal Hankel operator. According to [6], the rank of the lifted matrix satisfies $\operatorname{Rank}(\mathcal{H}^d(\mathcal{X}^d)) \leq r$ by using high-dimensional Vandermonde decomposition. Let Φ^d denote the prior information of \mathcal{X}^d . We can complete the d -way tensor by using the following nuclear norm minimization

$$\begin{aligned} \min_{\mathcal{Z}^d} \quad & \|\mathcal{H}^d(\mathcal{Z}^d)\|_* - 2\lambda \operatorname{Re}(\langle \mathcal{G}(\Phi^d), \mathcal{H}^d(\mathcal{Z}^d) \rangle) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathcal{Z}^d) = \mathcal{P}_\Omega(\mathcal{X}^d), \end{aligned} \quad (9)$$

where $\Omega = \{(k_1, \dots, k_d) \in \{0, \dots, N_1 - 1\} \times \dots \times \{0, \dots, N_d - 1\}\}$ denotes the index set of known entries of

\mathcal{X}^d and $\langle \mathcal{X}^d, \mathcal{Y}^d \rangle = (\operatorname{vec}(\mathcal{Y}^d))^H \operatorname{vec}(\mathcal{X}^d)$ denotes the inner product of the d -way tensors.

Let $\mathbf{E}(k_1, \dots, k_d)$ be the canonical basis in the domain $\mathbb{C}^{N_1 \times \dots \times N_d}$, and define the following orthonormal basis,

$$\mathbf{A}(k_1, \dots, k_d) = \frac{1}{\sqrt{w(k_1, \dots, k_d)}} \mathcal{H}(\mathbf{E}(k_1, \dots, k_d)),$$

where

$$w(k_1, \dots, k_d) = \prod_{l=1}^d w_l$$

and

$$w_l = |\{(i_l, j_l) | i_l + j_l = k_l, 0 \leq i_l \leq n_l - 1, 0 \leq j_l \leq N_l - n_l + 1\}|.$$

So each d -way tensor \mathcal{X}^d can be rewritten as

$$\begin{aligned} \mathcal{H}(\mathcal{X}^d) &= \sum_{k_1=0}^{N_1-1} \cdots \sum_{k_d=0}^{N_d-1} \langle \mathcal{H}(\mathcal{X}^d), \mathbf{A}(k_1, \dots, k_d) \rangle \mathbf{A}(k_1, \dots, k_d). \end{aligned}$$

It's straightforward to extend the theoretical guarantee from one-dimensional case to multi-dimensional case.

Theorem 2. *Let $\Omega = \{j_1, \dots, j_m\}$ be a multi-set consisting of random indices where $\{j_k\}_{k=1}^m \in \mathbb{R}^d$ are i.i.d. and follow the uniform distribution on $[N_1] \times \dots \times [N_d]$. Suppose, furthermore, that $\mathcal{H}(\mathbf{x})$ is of rank- r and satisfies the standard incoherence condition in (5) with parameter μ . Then there exists an absolute constant c_1 such that \mathbf{x} is the unique minimizer to (3) with high probability, provided that*

$$\begin{aligned} m \geq \max\{\Delta^2, 1\} c \mu c_s r \log^\alpha \left(\prod_{k=1}^d N_k \right) \\ \cdot \max \left\{ \log \left(7 \prod_{k=1}^d N_k \|\mathbf{F}_0\|_F \right), 1 \right\}, \end{aligned}$$

and

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\| < \frac{1}{2},$$

where

$$c_s \triangleq \max \left\{ \prod_{k=1}^d \frac{N_k}{n_k}, \prod_{k=1}^d \frac{N_k}{N_k - n_k + 1} \right\},$$

$$\mathbf{F}_0 \triangleq \mathcal{P}_{\mathcal{T}}(\operatorname{sgn}[\mathcal{H}(\mathbf{x}^*)] - \lambda \mathcal{G}(\phi)),$$

and

$$\Delta \triangleq \frac{4(\|\mathbf{F}_0\|_{\mathcal{A},2} + \|\mathbf{F}_0\|_{\mathcal{A},\infty})}{1 - 2\|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\|}.$$

Here, $\alpha = 1$ if the lifting operator has the wrap-around property; $\alpha = 3$ if the lifting operator doesn't have the wrap-around property.

V. OPTIMIZATION ALGORITHM

Due to the high computational complexity of low rank methods, we decide to use the non-convex method to solve the problem. First we use matrix factorization to decompose $\mathcal{H}(z)$ to two low complexity matrices, i.e. $\mathcal{H}(z) = \mathbf{U}\mathbf{V}^H$ with $\mathbf{U} \in \mathbb{C}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{C}^{n_2 \times r}$. Then we use Alternating Direction Method of Multipliers (ADMM) [36] to solve the problem.

First of all, we denote the nuclear norm as follows [37, Lemma 8]

$$\|\mathcal{H}(z)\|_* = \min_{\mathbf{U}, \mathbf{V}: \mathcal{H}(z) = \mathbf{U}\mathbf{V}^H} \frac{1}{2} (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2). \quad (10)$$

Incorporating (10) into the problem (3) yields

$$\begin{aligned} \min_{\mathbf{z}, \mathbf{U}, \mathbf{V}} \quad & \frac{1}{2} (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) - \lambda \text{Re} (\langle \mathcal{G}(\phi), \mathbf{U}\mathbf{V}^H \rangle) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{z}) = \mathcal{P}_\Omega(\mathbf{x}), \quad \mathcal{H}(z) = \mathbf{U}\mathbf{V}^H. \end{aligned}$$

Then, we start ADMM by the following argueded Lagrange function

$$\begin{aligned} L(\mathbf{U}, \mathbf{V}, \mathbf{z}, \mathbf{\Lambda}) = & \Pi(\mathbf{z}) - 2\lambda \text{Re} (\langle \mathcal{G}(\phi), \mathbf{U}\mathbf{V}^H \rangle) \\ & + \frac{1}{2} (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \frac{\mu}{2} \|\mathcal{H}(z) - \mathbf{U}\mathbf{V}^H + \mathbf{\Lambda}\|_F^2, \end{aligned} \quad (11)$$

where $\mu > 0$ is an absolute constant, and $\Pi(\mathbf{z})$ is an indicator function

$$\Pi(\mathbf{z}) = \begin{cases} 0, & \text{if } \mathcal{P}_\Omega(\mathbf{z}) = \mathcal{P}_\Omega(\mathbf{x}), \\ \infty, & \text{otherwise.} \end{cases}$$

Next, we decompose (11) into three subproblems to get $\mathbf{z}^{(n+1)}$, $\mathbf{U}^{(n+1)}$ and $\mathbf{V}^{(n+1)}$

$$\begin{aligned} \mathbf{z}^{(n+1)} &= \arg \min_{\mathbf{z}} \Pi(\mathbf{z}) + \frac{\mu}{2} \|\mathcal{H}(z) - \mathbf{U}^{(n)}\mathbf{V}^{(n)H} + \mathbf{\Lambda}^{(n)}\|_F^2 \\ \mathbf{U}^{(n+1)} &= \arg \min_{\mathbf{U}} \frac{1}{2} \|\mathbf{U}\|_F^2 - 2\lambda \text{Re} (\langle \mathcal{G}(\phi), \mathbf{U}\mathbf{V}^{(n)H} \rangle) \\ &\quad + \frac{\mu}{2} \|\mathcal{H}(z^{(n+1)}) - \mathbf{U}\mathbf{V}^{(n)H} + \mathbf{\Lambda}^{(n)}\|_F^2 \\ \mathbf{V}^{(n+1)} &= \arg \min_{\mathbf{V}} \frac{1}{2} \|\mathbf{V}\|_F^2 - 2\lambda \text{Re} (\langle \mathcal{G}(\phi), \mathbf{U}^{(n+1)}\mathbf{V}^H \rangle) \\ &\quad + \frac{\mu}{2} \|\mathcal{H}(z^{(n+1)}) - \mathbf{U}^{(n+1)}\mathbf{V}^H + \mathbf{\Lambda}^{(n)}\|_F^2 \end{aligned}$$

and the Lagrangian update is

$$\mathbf{\Lambda}^{(n+1)} = \mathcal{H}(z^{(n+1)}) - \mathbf{U}^{(n+1)}\mathbf{V}^{(n+1)H} + \mathbf{\Lambda}^{(n)}.$$

By simple calculations, we can obtain

$$\mathbf{z}^{(n+1)} = \mathcal{P}_{\Omega^c} \mathcal{H}^\dagger(\mathbf{U}^{(n)}\mathbf{V}^{(n)H} - \mathbf{\Lambda}^{(n)}) + \mathcal{P}_\Omega(\mathbf{x}),$$

where \mathcal{P}_{Ω^c} denotes the projection operator on Ω^c and $\mathcal{H}^\dagger(\cdot)$ denotes the Penrose-Moore pseudo-inverse mapping corresponding to $\mathcal{H}(\cdot)$.

Then, by taking the derivative of the other two problems and setting them to zero, we can obtain

$$\begin{aligned} \mathbf{U}^{(n+1)} = & \left[\mu (\mathcal{H}(z^{(n+1)}) + \mathbf{\Lambda}^{(n)}) + \lambda \mathcal{G}(\phi) \right] \\ & \cdot \mathbf{V}^{(n)} (\mathbf{I} + \mu \mathbf{V}^{(n)H} \mathbf{V}^{(n)})^{-1} \end{aligned}$$

Algorithm 1 Reference-based Structured Matrix Completion

Input: Sampling index set Ω , measurements $\mathcal{P}_\Omega(\mathbf{x})$, prior information ϕ

Output: Estimated result \mathbf{z}

- 1: Initialize $k = 0$, ε , tol , and K
 - 2: **if** $\|\mathcal{P}_\Omega(\mathbf{x} - \phi)\| \leq \varepsilon$ **then**
 - 3: $[\mathbf{U}_{(r)}, \mathbf{\Sigma}_{(r)}, \mathbf{V}_{(r)}] = r\text{-SVD}(\mathcal{H}(\phi)); \mathbf{\Lambda}^{(0)} = \mathbf{0}; \mathbf{U}_0 = \mathbf{U}_{(r)}\mathbf{\Sigma}_{(r)}; \mathbf{V}_0 = \mathbf{V}_{(r)}; G(\phi) = \mathbf{U}_{(r)} * \mathbf{V}_{(r)}^H$
 - 4: **else**
 - 5: $[\mathbf{U}_0, \mathbf{V}_0] = \text{LMaFit}(\mathcal{H}, \mathcal{P}_\Omega(\mathbf{x})); \mathbf{\Lambda}^{(0)} = \mathbf{0}; G(\phi) = \mathbf{0}$
 - 6: **end if**
 - 7: **repeat**
 - 8: $k = k + 1$
 - 9: $\mathbf{z}_k = \mathcal{P}_{\Omega^c} \mathcal{H}^\dagger(\mathbf{U}_{k-1}\mathbf{V}_{k-1}^H - \mathbf{\Lambda}_{k-1}) + \mathcal{P}_\Omega(\mathbf{x})$
 - 10: $\mathbf{U}_k = [\mu (\mathcal{H}(\mathbf{z}_k) + \mathbf{\Lambda}_{k-1}) + \lambda \mathcal{G}(\phi)] \cdot \mathbf{V}_{k-1} \cdot (\mathbf{I} + \mu \mathbf{V}_{k-1}^H \mathbf{V}_{k-1})^{-1}$
 - 11: $\mathbf{V}_k = [\mu (\mathcal{H}(\mathbf{z}_k) + \mathbf{\Lambda}_{k-1}) + \lambda \mathcal{G}(\phi)]^H \cdot \mathbf{U}_k \cdot (\mathbf{I} + \mu \mathbf{U}_k^H \mathbf{U}_k)^{-1}$
 - 12: $\mathbf{\Lambda}_k = \mathcal{H}(\mathbf{z}_k) - \mathbf{U}_k \mathbf{V}_k^H + \mathbf{\Lambda}_{k-1}$
 - 13: **until** $k > K$ or $\|\mathbf{z}_k - \mathbf{z}_{k-1}\|_F < tol$
-

and

$$\begin{aligned} \mathbf{V}^{(n+1)} = & \left[\mu (\mathcal{H}(z^{(n+1)}) + \mathbf{\Lambda}^{(n)}) + \lambda \mathcal{G}(\phi) \right]^H \\ & \cdot \mathbf{U}^{(n+1)} (\mathbf{I} + \mu \mathbf{U}^{(n+1)H} \mathbf{U}^{(n+1)})^{-1}. \end{aligned}$$

The last question is how to initialize \mathbf{U} and \mathbf{V} . In order to converge quickly, the authors in [38] uses an algorithm named LMaFit [39], which is

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{Z}} \frac{1}{2} \|\mathbf{U}\mathbf{V}^H - \mathbf{Z}\|_F^2 \quad \text{s.t. } \mathcal{P}_\Omega(\mathcal{H}^\dagger(\mathbf{Z})) = \mathcal{P}_\Omega(\mathbf{x}). \quad (12)$$

However, LMaFit only uses the undersampled measurements and cannot guarantee that \mathbf{Z} has the lifting matrix structure. Instead, we can take advantage of the reference image to initialize \mathbf{U} and \mathbf{V} by using truncated SVD when the reference image is reliable. Here, we use the value of $\|\mathcal{P}_\Omega(\mathbf{x} - \phi)\|$ as a criterion to choose the suitable initialization strategy.

Then we can give the corresponding algorithm in Algorithm 1. Here, $r\text{-SVD}(\cdot)$ returns the results of truncated SVD. And $\text{LMaFit}(\mathcal{H}, \mathcal{P}_\Omega(\mathbf{x}))$ denotes the algorithm in (12).

VI. SIMULATIONS

In this section, we carry on numerical simulations to show the improvement of the proposed method (3) compared to standard Hankel matrix completion (2). Besides, we compare the performance under two different solvers: CVX solver and ADMM-solver. Here, we use CVX package [40], [41] to get the convex results and use Algorithm 1 to get the ADMM results.

A. Simulations for 1-D signals

We begin by giving the numerical results for one-dimensional signals.

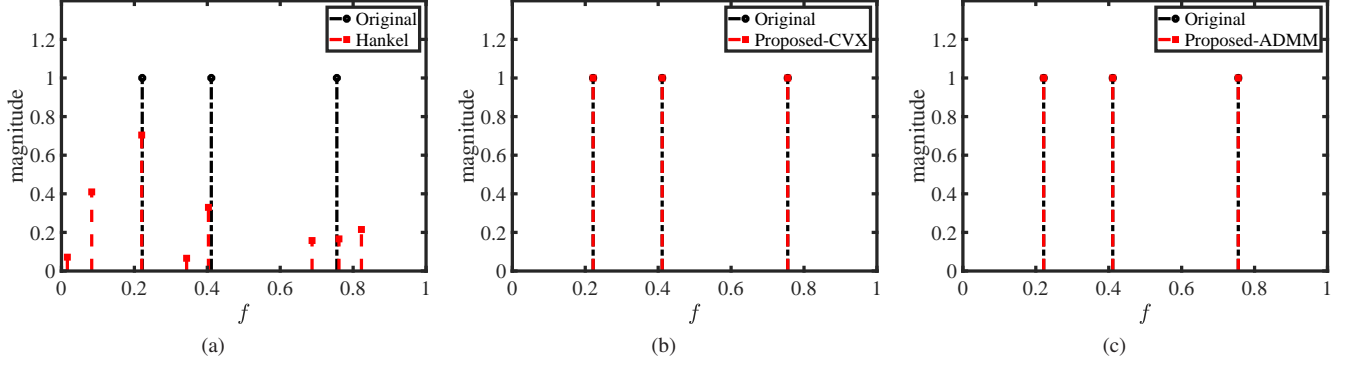


Fig. 1: Performance Comparisons for one-dimensional signals when $n = 32$, $r = 3$ and $m = 10$. (a) Hankel matrix completion; (b) Proposed method (CVX); (c) Proposed method (ADMM).

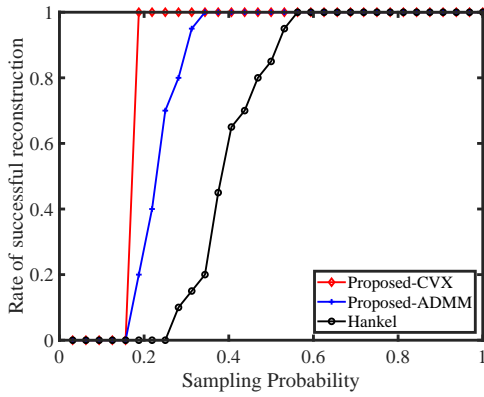


Fig. 2: Rate of successful reconstruction v.s. sampling probability for Hankel matrix completion and reference based Hankel matrix completion.

Consider a one-dimensional spectrally sparse signal $\mathbf{x}^* \in \mathbb{C}^n$ and the signal is a weighted superposition of r complex sinusoids with unit amplitudes. The reference signal is created by $\phi = \mathbf{x}^* + \sigma \mathbf{n} \in \mathbb{C}^n$, where the entries of the real and imaginary part of \mathbf{n} follow i.i.d. standard Gaussian distribution, i.e., $\text{Re}(n_i), \text{Im}(n_i) \sim \mathcal{N}(0, 1)$ for $i = 1, \dots, n$.

We first show the reconstruction results for standard Hankel matrix completion, the proposed method with CVX solver and the proposed method with ADMM solver. We set $n = 32$, $r = 3$, $m = 10$ and $\sigma = 0.5$. The matrix pencil method is used to estimate the location and amplitude of frequencies [42]. The frequency estimation results are shown in Fig. 1. As expected, with the reliable reference signal, the proposed scheme with different solvers exactly reconstructs the original signal, which has a better performance than standard Hankel matrix completion.

We next provide the successful reconstruction rate as a function of sampling probability standard Hankel matrix completion, the proposed method with CVX solver and the proposed method with ADMM solver. We set $n = 32$, $r = 3$, $\sigma = 0.1$. We set $\eta = 10^{-4}$ for CVX solver and $\eta = 10^{-2}$ for ADMM solver since CVX solver gets the exact solution while ADMM solver has performance degradation due to finite iteration. For each sampling probability, we sample the desired signals in

TABLE I: Running time comparison for 1-D signals

Methods	16	64	96
Proposed-CVX	0.483s	3.185s	21.456s
Proposed-ADMM	0.003s	0.017s	0.028s

time domain randomly and the results are averaged over 300 independent trials. Then we count the number of successful trials, and calculate the related probability. Here, we claim a trial as a successful trial if the solution \mathbf{x}^\dagger satisfies

$$\frac{\|\mathbf{x}^* - \mathbf{x}^\dagger\|_2}{\|\mathbf{x}^*\|_2} < \eta.$$

The results are presented in Fig. 2. The results indicate that the proposed approach (3) outperforms the standard Hankel matrix completion with reliable reference.

We then compare the running time for the proposed method with different solvers when the dimension of signals is 16, 64 and 96. The numerical simulations are carried on an Intel desktop with 2.5 GHz CPU and 8 GB RAM. The results in Table I show that ADMM solver can dramatically improve the running time. Besides, the proposed scheme with ADMM solver has a much better performance compared with the standard Hankel matrix completion.

B. Simulations for 2-D signals

We proceed by giving the numerical results for two-dimensional signals.

Consider a two-dimensional spectrally sparse signal $\mathbf{X}^* \in \mathbb{R}^{N_1 \times N_2}$ and the signal is a weighted superposition of r complex sinusoids with unit amplitudes. The reference signal is created by $\Phi = \mathbf{X}^* + \sigma \mathbf{N}$, where the entries of the real and imaginary part of \mathbf{N} follow i.i.d. standard Gaussian distribution, i.e., $\text{Re}(N_{ij}), \text{Im}(N_{ij}) \sim \mathcal{N}(0, 1)$ for $i = 1, \dots, N_1, j = 1, \dots, N_2$.

We first show the recovery results for the proposed method and standard Hankel matrix completion. We set $N_1 = 10$, $N_2 = 10$, $r = 3$, $m = 20$ and $\sigma = 0.1$. 2D-MUSIC is applied to obtain the location and amplitude of frequencies [43]. The results are presented in Fig. 3. The results show that

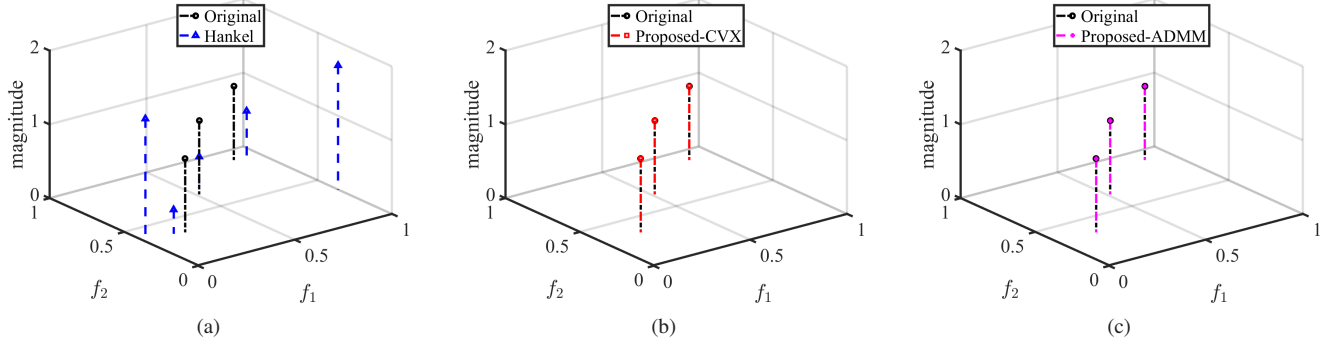


Fig. 3: Performance Comparisons for two-dimensional signals when $N_1 = 10, N_2 = 10, r = 3, m = 20$. (a) Hankel matrix completion; (b) Proposed method (CVX); (c) Proposed method (ADMM).

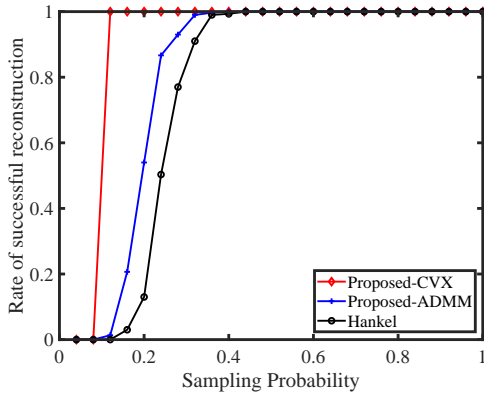


Fig. 4: Phase transitions for Hankel matrix completion and reference based Hankel matrix completion.

TABLE II: Running time comparison for 2-D signals

Methods	8×8	10×10	12×12
Proposed-CVX	1.420s	3.449s	12.847s
Proposed-ADMM	0.399s	0.529s	0.786s

the proposed method with different solvers can exactly recover the desired signals while Hankel matrix cannot.

We next present the successful reconstruction rate as a function of sampling probability standard Hankel matrix completion, the proposed method with CVX solver and the proposed method with ADMM solver. We set $N_1 = 10, N_2 = 10, r = 3$ and $\sigma = 0.1$. We increase the number of samples m from 1 to 100. We Fig. 4 gives the simulation results. As expected, the proposed scheme with the CVX solver performs the best, followed by the proposed scheme with ADMM solver and standard Hankel matrix completion.

Finally, Table II compares the running time for the proposed method with different solvers when $N_1 = N_2 = 8, N_1 = N_2 = 10$ and $N_1 = N_2 = 12$. The results present that the proposed scheme with ADMM solver has smaller running time than that with CVX solver, especially when the dimension of signals is large.

VII. CONCLUSION

In this paper, we have integrated prior information to improve the performance of spectrally sparse signal recovery via structured matrix completion problem and have provided the related performance guarantees. Furthermore, we have designed corresponding ADMM algorithm to reduce the computational complexity. Both the theoretical and experimental results show that the proposed scheme outperforms standard Hankel matrix completion.

APPENDIX A PROOF OF THEOREM 1

Dual certification is used to deviate the theoretical results. In particular, we use the golfing method from [33] to proceed the process. And we adjust the methods from [6], [7] to suit our model.

Recall the definition of the operator $\mathcal{A}_k : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^{n_1 \times n_2}$ by

$$\mathcal{A}_k(\mathbf{X}) = \langle \mathbf{X}, \mathbf{A}_k \rangle \mathbf{A}_k, k = 1, \dots, n.$$

Then each \mathcal{A}_k is an orthogonal projection onto the one-dimensional subspace spanned by \mathbf{A}_k . The orthogonal projection onto the subspace spanned by $\{\mathbf{A}_k\}_{k=1}^n$ is given as $\mathcal{A} = \sum_{k=1}^n \mathcal{A}_k$. Let \mathcal{A}^\perp denote the orthogonal complement of \mathcal{A} . The summation of the rank-1 projection operators in $\{\mathcal{A}_k\}_{k \in \Omega}$ is denoted by \mathcal{A}_Ω , i.e., $\mathcal{A}_\Omega \triangleq \sum_{k \in \Omega} \mathcal{A}_k$. Since Ω is a multi-set and there may exist repetitions in Ω , \mathcal{A}_Ω may be not a projection operator. The summation of distinct elements in $\{\mathcal{A}_k\}_{k \in \Omega}$ is denoted by \mathcal{A}'_Ω , which is a valid orthogonal projection.

Before proving the theorem, let's review the proposed program

$$\begin{aligned} \min_{\mathbf{z}} \quad & \|\mathcal{H}(\mathbf{z})\|_* - 2\lambda \operatorname{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{z}) \rangle) \\ \text{s.t.} \quad & \mathcal{P}_\Omega(\mathbf{z}) = \mathcal{P}_\Omega(\mathbf{x}). \end{aligned}$$

We begin by presenting two lemmas, which are necessary for the proof.

Lemma 1 ([7, Lemma 19] & [6, Lemma 3]). Suppose that

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

So we have

$$\begin{aligned} \max_{1 \leq k \leq n} \|\mathbf{U}^H \mathbf{A}_k\|_F^2 &\leq \frac{\mu r}{n_1}, \\ \max_{1 \leq k \leq n} \|\mathbf{V}^H \mathbf{A}_k^H\|_F^2 &\leq \frac{\mu r}{n_2}. \end{aligned} \quad (13)$$

Then for any small constant $0 < \epsilon \leq \frac{1}{2}$, one has

$$\|\mathcal{P}_\mathcal{T} \mathcal{A} \mathcal{P}_\mathcal{T} - \frac{n}{m} \mathcal{P}_\mathcal{T} \mathcal{A}_\Omega \mathcal{P}_\mathcal{T}\| \leq \epsilon \quad (14)$$

with probability exceeding $1 - n^{-4}$, provided that $m > c\mu c_s r \log n$ for some universal constant $c > 0$ and $c_s \triangleq \max\{\frac{n}{n_1}, \frac{n}{n_2}\}$.

Lemma 2. Consider a multi-set Ω that contains m random indices. Suppose that the sampling operator \mathcal{A}_Ω obeys

$$\|\mathcal{P}_\mathcal{T} \mathcal{A} \mathcal{P}_\mathcal{T} - \frac{n}{m} \mathcal{P}_\mathcal{T} \mathcal{A}_\Omega \mathcal{P}_\mathcal{T}\| \leq \frac{1}{2}. \quad (15)$$

If there exists a matrix \mathbf{W} satisfying

$$\mathcal{A}'_{\Omega^\perp}(\mathbf{W}) = 0, \quad (16)$$

$$\|\mathcal{P}_\mathcal{T}(\text{sgn}[\mathcal{H}(\mathbf{x}^*)] - \mathbf{W} - \lambda \mathcal{G}(\phi))\|_F \leq \frac{1}{7n}, \quad (17)$$

and

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W} + \lambda \mathcal{G}(\phi))\| \leq \frac{1}{2}, \quad (18)$$

then the program (3) can achieve exact recovery, i.e., \mathbf{x} is the unique minimizer.

Proof: See Appendix B. ■

As shown in [6], we generate j_0 independent random multi-sets $\{\Omega_i\}_{i=1}^{j_0}$ and each set contains $\frac{m}{j_0}$ entries. Note that the distribution of Ω and $\cup_{i=1}^{j_0} \Omega_i$ is the same. Then we construct of a dual certificate \mathbf{W} via the golfing scheme:

- 1) Define $\mathbf{F}_0 \triangleq \mathcal{P}_\mathcal{T}(\text{sgn}[\mathcal{H}(\mathbf{x})] - \lambda \mathcal{G}(\phi))$;
- 2) For every i ($1 \leq i \leq j_0$), set

$$\mathbf{F}_i \triangleq \mathcal{P}_\mathcal{T} \left(\mathcal{A} - \frac{n j_0}{m} \mathcal{A}_{\Omega_i} \right) \mathcal{P}_\mathcal{T}(\mathbf{F}_{i-1});$$

- 3) Define $\mathbf{W} \triangleq \sum_{i=1}^{j_0} \left(\frac{n j_0}{m} \mathcal{A}_{\Omega_i} + \mathcal{A}^\perp \right) (\mathbf{F}_{i-1})$.

By the construction, it's easy to see that \mathbf{W} is in the range space of $\mathcal{A}_\Omega \cup \mathcal{A}^\perp$, then

$$\mathcal{A}'_{\Omega^\perp}(\mathbf{W}) = 0. \quad (19)$$

By recursive calculation as [6, Eq. (40)], we can obtain

$$-\mathcal{P}_\mathcal{T}(\mathbf{W} - \mathbf{F}_0) = \mathcal{P}_\mathcal{T}(\mathbf{F}_{j_0}). \quad (20)$$

Using Lemma 1 yields

$$\begin{aligned} \|\mathcal{P}_\mathcal{T}(\mathbf{W} - \mathbf{F}_0)\|_F &= \|\mathcal{P}_\mathcal{T}(\mathbf{F}_{j_0})\|_F \\ &\leq \left\| \mathcal{P}_\mathcal{T} \left(\mathcal{A} - \frac{n j_0}{m} \mathcal{A}_{\Omega_i} \right) \mathcal{P}_\mathcal{T} \right\|^{j_0} \|\mathbf{F}_0\|_F \\ &\leq \epsilon^{j_0} \|\mathbf{F}_0\|_F \leq \frac{1}{2^{j_0}} \|\mathbf{F}_0\|_F. \end{aligned} \quad (21)$$

Let

$$j_0 = \max \{ \log(7n \|\mathbf{F}_0\|_F), 1 \}, \quad (22)$$

then we have

$$\|\mathcal{P}_\mathcal{T}(\mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x}^*)] + \lambda \mathcal{G}(\phi))\|_F \leq \frac{1}{7n} \quad (23)$$

except with a probability at most $j_0 n^{-4} = o(n^{-3})$, as long as

$$m > c\mu c_s r \log n \max \{ \log(7n \|\mathbf{F}_0\|_F), 1 \}.$$

For the last condition, using triangle's inequality yields

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W} + \lambda \mathcal{G}(\phi))\| \leq \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W})\| + \|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\|.$$

According to the result of [6, VI. E], we have

$$\begin{aligned} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W})\| &\leq \sum_{l=1}^{j_0} \left\| \mathcal{P}_{\mathcal{T}^\perp} \left(\frac{n j_0}{m} \mathcal{A}_{\Omega_l} + \mathcal{A}^\perp \right) \mathcal{P}_\mathcal{T}(\mathbf{F}_{l-1}) \right\| \\ &\leq \sum_{l=1}^{j_0} \left(\frac{1}{2} \right)^{l-1} \left(\sqrt{\frac{n j_0 \log n}{m}} \|\mathbf{F}_0\|_{\mathcal{A},2} + \frac{n j_0 \log n}{m} \|\mathbf{F}_0\|_{\mathcal{A},\infty} \right) \\ &< \frac{2}{\Delta} (\|\mathbf{F}_0\|_{\mathcal{A},2} + \|\mathbf{F}_0\|_{\mathcal{A},\infty}), \end{aligned}$$

as long as

$$\begin{aligned} m &\geq c \max\{\mu c_s, \nu\} \\ &\quad \cdot \max\{\Delta^2, 1\} r \log n \max \{ \log(7n \|\mathbf{F}_0\|_F), 1 \}, \end{aligned}$$

where $\nu = o(\mu c_s \log^2 n)$ from [7, Appendix E]. Set

$$\Delta = \frac{4(\|\mathbf{F}_0\|_{\mathcal{A},2} + \|\mathbf{F}_0\|_{\mathcal{A},\infty})}{1 - 2 \|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\|}.$$

If $\|\mathcal{P}_{\mathcal{T}^\perp}(\lambda \mathcal{G}(\phi))\| < \frac{1}{2}$, we have

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W} + \lambda \mathcal{G}(\phi))\| \leq \frac{1}{2}.$$

Therefore, we conclude, if

$$m \geq \max\{\Delta^2, 1\} c\mu c_s r \log^3 n \max \{ \log(7n \|\mathbf{F}_0\|_F), 1 \},$$

then with high probability, we can achieve the unique minimum.

Remark 5. Form the operator $\mathcal{H}_c(\mathbf{x})$ with wrap-around property, $\nu = o(\mu c_s)$ according to [7, Appendix E]. Therefore, we can get the following bound of sample size

$$m \geq \max\{\Delta^2, 1\} c\mu c_s r \log n \max \{ \log(7n \|\mathbf{F}_0\|_F), 1 \}.$$

APPENDIX B PROOF OF LEMMA 2

Let $\mathbf{z} = \mathbf{x} + \mathbf{h}$ be the minimizer to (3). We will show that $\mathcal{H}(\mathbf{h}) = 0$. Then by the injectivity of the operator \mathcal{H} , we achieve $\mathbf{h} = 0$, so we have $\mathbf{z} = \mathbf{x}$.

According to case 2 in the proof of [7, Lemma 20], $\|\mathcal{P}_\mathcal{T}(\mathcal{H}(\mathbf{h}))\|_F \geq 3n \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_F$ leads to $\mathcal{H}(\mathbf{h}) = 0$. So we only need to prove that when

$$\|\mathcal{P}_\mathcal{T}(\mathcal{H}(\mathbf{h}))\|_F \leq 3n \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_F, \quad (24)$$

we also have $\mathcal{H}(\mathbf{h}) = 0$. In the subsequent analysis, we assume that the condition (24) is correct.

According to the definition of nuclear norm, there exists \mathbf{B} such that $\langle \mathbf{B}, \mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h})) \rangle = \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_*$ and $\|\mathbf{B}\| \leq 1$. Then $\text{sgn}(\mathcal{H}(\mathbf{x})) + \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{B})$ is a sub-gradient of the nuclear norm at $\mathcal{H}(\mathbf{x})$. Then it follows that

$$\begin{aligned} & \|\mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h}) \rangle) \\ & \geq \|\mathcal{H}(\mathbf{x})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) \rangle) \\ & \quad + 2\text{Re}(\langle \text{sgn}[\mathcal{H}(\mathbf{x})] + \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{B}) - \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle) \\ & = \|\mathcal{H}(\mathbf{x})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) \rangle) + 2\text{Re}(\langle \mathbf{W}, \mathcal{H}(\mathbf{h}) \rangle) \\ & \quad + 2\text{Re}(\langle \text{sgn}[\mathcal{H}(\mathbf{x})] + \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{B}) - \lambda \mathcal{G}(\phi) - \mathbf{W}, \mathcal{H}(\mathbf{h}) \rangle). \end{aligned} \quad (25)$$

We can get $\text{Re}(\langle \mathbf{W}, \mathcal{H}(\mathbf{h}) \rangle) = 0$ as shown in [7, A.33-A.34]. In addition, we have

$$\begin{aligned} \langle \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{B}), \mathcal{H}(\mathbf{h}) \rangle & = \langle \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{B}), \mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h})) \rangle \\ & = \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_*. \end{aligned}$$

Then the inequality (25) becomes

$$\begin{aligned} & \|\mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h}) \rangle) \\ & \geq \|\mathcal{H}(\mathbf{x})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) \rangle) + 2\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* \\ & \quad - 2\text{Re}(\langle \mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle). \end{aligned} \quad (26)$$

Next, we are going to derive

$$\begin{aligned} & \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* \\ & \quad - \text{Re}(\langle \mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle) \geq 0. \end{aligned}$$

Noting that $\text{Re}(x) \leq |x|$ for $x \in \mathbb{C}$, it's enough to prove

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - |\langle \mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle| \geq 0.$$

By using the triangle inequality, we have

$$\begin{aligned} & \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - |\langle \mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle| \\ & \geq \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - |\langle \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W} + \lambda \mathcal{G}(\phi)), \mathcal{H}(\mathbf{h}) \rangle| \\ & \quad - |\langle \mathcal{P}_{\mathcal{T}}(\mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi)), \mathcal{H}(\mathbf{h}) \rangle|. \end{aligned} \quad (27)$$

Using Holder's inequality and the properties of \mathbf{W} (Eqs. (17) and (18)) yields

$$\begin{aligned} & \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - |\langle \mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle| \\ & \geq \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{W} + \lambda \mathcal{G}(\phi))\| \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* \\ & \quad - \|\mathcal{P}_{\mathcal{T}}(\mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi))\|_F \|\mathcal{P}_{\mathcal{T}}(\mathcal{H}(\mathbf{h}))\|_F \\ & \geq \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - \frac{1}{2} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - \frac{1}{7n} \|\mathcal{P}_{\mathcal{T}}(\mathcal{H}(\mathbf{h}))\|_F \\ & \geq \frac{1}{2} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - \frac{1}{7n} \|\mathcal{P}_{\mathcal{T}}(\mathcal{H}(\mathbf{h}))\|_F. \end{aligned} \quad (28)$$

By using Eq. (24), we obtain

$$\begin{aligned} & \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - |\langle \mathbf{W} - \text{sgn}[\mathcal{H}(\mathbf{x})] + \lambda \mathcal{G}(\phi), \mathcal{H}(\mathbf{h}) \rangle| \\ & \geq \frac{1}{2} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* - \frac{3}{7} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* \\ & = \frac{1}{14} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_* \geq 0. \end{aligned} \quad (29)$$

Therefore, we get

$$\begin{aligned} & \|\mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h}) \rangle) \\ & \geq \|\mathcal{H}(\mathbf{x})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) \rangle) + \frac{1}{14} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h}))\|_*. \end{aligned} \quad (30)$$

Since $\mathbf{z} = \mathbf{x} + \mathbf{h}$ be the minimizer to (3), we also have

$$\begin{aligned} & \|\mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) + \mathcal{H}(\mathbf{h}) \rangle) \\ & \leq \|\mathcal{H}(\mathbf{x})\|_* - 2\lambda \text{Re}(\langle \mathcal{G}(\phi), \mathcal{H}(\mathbf{x}) \rangle). \end{aligned} \quad (31)$$

Combining Eqs. (30) and (31), we get $\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{H}(\mathbf{h})) = 0$. By using (24), we also have $\mathcal{P}_{\mathcal{T}}(\mathcal{H}(\mathbf{h})) = 0$. So we conclude $\mathcal{H}(\mathbf{h}) = 0$ and $\mathbf{h} = 0$.

REFERENCES

- [1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [2] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proc. IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.
- [3] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *arXiv preprint arXiv:0902.0026*, 2009.
- [4] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 383–395, 2007.
- [5] D. J. Brenner and E. J. Hall, "Computed tomography an increasing source of radiation exposure," *N. Engl. J. Med.*, vol. 357, no. 22, pp. 2277–2284, 2007.
- [6] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, Oct 2014.
- [7] J. C. Ye, J. M. Kim, K. H. Jin, and K. Lee, "Compressive sampling using annihilating filter-based low-rank interpolation," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 777–801, 2017.
- [8] J. P. Haldar, D. Hernando, S.-K. Song, and Z.-P. Liang, "Anatomically constrained reconstruction from noisy data," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 59, no. 4, pp. 810–818, 2008.
- [9] X. Peng, H.-Q. Du, F. Lam, S. D. Babacan, and Z.-P. Liang, "Reference-driven MR image reconstruction with sparsity and support constraints," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 89–92.
- [10] B. Wu, R. P. Millane, R. Watts, and P. J. Bones, "Prior estimate-based compressed sensing in parallel MRI," *Magn. Reson. Med.*, vol. 65, no. 1, pp. 83–95, 2011.
- [11] B. Bilgic, V. K. Goyal, and E. Adalsteinsson, "Multi-contrast reconstruction with bayesian compressed sensing," *Magn. Reson. Med.*, vol. 66, no. 6, pp. 1601–1615, 2011.
- [12] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator," *Med. Image Anal.*, vol. 18, no. 6, pp. 843–856, 2014.
- [13] J. Huang, C. Chen, and L. Axel, "Fast multi-contrast MRI reconstruction," *Magn. Reson. Imaging*, vol. 32, no. 10, pp. 1344–1352, 2014.
- [14] M. Selim, E. A. Rashed, and H. Kudo, "Low-dose multiphase abdominal ct reconstruction with phase-induced swap prior," in *Developments in X-Ray Tomography X*, vol. 9967. International Society for Optics and Photonics, 2016, p. 99671P.
- [15] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.
- [16] M. Selim, E. A. Rashed, M. A. Atia, and H. Kudo, "Image reconstruction using self-prior information for sparse-view computed tomography," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*. IEEE, 2018, pp. 146–149.
- [17] Y. Li, X. Wang, Z. Ding, X. Zhang, Y. Xiang, and X. Yang, "Spectrum recovery for clutter removal in penetrating radar imaging," *IEEE Trans. Geosci. Remote Sens.*, 2019.

- [18] M. Vella and J. F. Mota, "Single Image Super-Resolution via CNN Architectures and TV-TV Minimization," *arXiv preprint arXiv:1907.05380*, 2019.
- [19] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2015.
- [20] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.
- [21] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [22] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [23] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.
- [24] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499–512, 2014.
- [25] L. Weizman, Y. C. Eldar, and D. Ben Bashat, "Reference-based MRI," *Med. Phys.*, vol. 43, no. 10, pp. 5357–5369, 2016.
- [26] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Spectral super-resolution with prior knowledge," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5342–5357, 2015.
- [27] Y. Li, X. Zhang, Z. Ding, and X. Wang, "Compressive multidimensional harmonic retrieval with prior knowledge," *arXiv preprint arXiv:1904.11404*, 2019.
- [28] X. Zhang, W. Cui, and Y. Liu, "Recovery of structured signals with prior information via maximizing correlation," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3296–3310, June 2018.
- [29] —, "Matrix completion with prior subspace information via maximizing correlation," *arXiv preprint arXiv:2001.01152*, 2020.
- [30] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.
- [31] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [32] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, p. 717, 2009.
- [33] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [34] J.-F. Cai, T. Wang, and K. Wei, "Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion," *Appl. Comput. Harmon. Anal.*, vol. 46, no. 1, pp. 94–121, 2019.
- [35] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [36] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [37] N. Srebro, "Learning with matrix factorizations," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.
- [38] 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. Comput. Imaging*, vol. 2, no. 4, pp. 480–495, Dec 2016.
- [39] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.
- [40] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [41] —, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.
- [42] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, 1990.
- [43] C. R. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, "Signal processing for passive radar using ofdm waveforms," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 226–238, 2010.