FREQUENCY RECOVERY RESOLUTION ENHANCEMENT USING GENERALIZED REWEIGHTED NONCONVEX RELAXATION

Lama ZIEN ALABIDEEN[1], Oumayma AL-DAKKAK[2], Khaldoun KHORZOM[3]

In this paper, we investigate the problem of joint recovery of frequency-sparse signals sharing common frequency components from the collection of their compressed measurements. Unlike conventional arts in compressed sensing, the frequencies are not assumed to lie on a grid and they are continuously valued in the normalized domain [0 1]. As an extension to the atomic norm minimization (ANM) approach, which states that the frequencies can be recovered only if they are sufficiently separated, we propose a relaxation framework based on Schatten p-norm. This framework is formulated in an iterative weighted minimization approach and solved efficiently via a computationally tractable semidefinite program. Finally, numerical experiments are carried out to illustrate the effectiveness of the proposed approach and its advantages over similarly recent recovery techniques in term of frequency resolution and computational complexity.

Keywords: sparse frequency recovery, Schatten p-norm, super-resolution, atomic norm, semidefinite programming

LIST OF ABBREVIATION

ANM	Atomic Norm Minimization
CCS	Continuous Compressed Sensing
MMV	Multiple Measurement-Vector
SDP	SemiDefinite Programming
RAM	Reweighted ANM
JSFR	Joint Sparse Frequency Recovery
SLA	Sparse Linear Antenna
SMV	Single Measurement Vector
RMP	Rank Minimization Problem
LRR	Low Rank Representation
LRMR	Low Rank Matrix Recovery
SNM	Schatten p-Norm Minimization
NNM	Nuclear Norm Minimization
WNNM	Weighted NNM
MRAM	Modified RAM
SVD	Singular Value Decomposition
MM	Maximization-Minimization

[1] Higher Institute for Applied Sciences and Technology, HIAST, Syrian Arab Republic, e-mail: lamazien79@gmail.com
[2] Higher Institute for Applied Sciences and Technology, HIAST, Syrian Arab Republic
[3] Higher Institute for Applied Sciences and Technology, HIAST, Syrian Arab Republic

1. Introduction

Extracting frequencies from a mixture of super positioned complex exponentials is an opening problem in statistical signal processing. It arises in many applications ranging from communications, radar, array processing to astronomy and seismology. Many methods have been proposed for frequency recovery [1]. The most prominent approaches are grid-based sparse methods which have been in the scope of researches mainly after the development of compressed sensing (CS) concept [2,3]. Principally, it refers to a technique of reconstructing a high dimensional signal from far fewer samples. In this kind of methods, the continuous frequency domain is discretized/gridded into a finite set of grid points. Thus, in order to recover the signal, it must be sparse in a prior known basis, like Discrete Fourier Transform basis; however, no physical field is sparse in a such basis. Consequently, it suffers from basis mismatch problem due to the discretization requirement [4,5]. Many subsequent approaches have been suggested to mitigate this problem [6-8], most of them are still based on gridding the frequency domain.

The gridless or continuous compressed sensing (CCS) method form the most recent class, mainly proposed after the invention of the theory of super-resolution by Cande`s and Fernandes-Granda [9]. This method guarantees fine details recovery of a sparse frequency spectrum from coarse time-domain samples, and bypasses the issues arising from discretization by working directly on the continuous parameter space. Chandrasekaran et al [10] introduced the gridless convex optimization for noiseless full data case based on the atomic norm (or the total variation norm) technique, which can be reformulated as semidefinite programming (SDP) and solved in a polynomial time. They proved that the frequencies could be recovered with infinite accuracy given a set of N regularly spaced samples whenever the frequency components are mutually separated by at least $\Delta_f \geq \frac{4}{N}$.

Motivated by the previous CCS aspects, Tang *et al*. [11] extended the theoretical results to the case of compressive samples with K frequency components, where $K \ll N$. They showed that a number of $M = O(K\log K \log N)$ random samples are sufficient for high probability recovery via ANM under the same previous frequency separation condition.

Recently, many researchers [12-15] investigated the theoretical guarantee of multiple measurement-vector (MMV) in CCS, and provided boosted results verified using extended atomic norm approaches. In this context, Z. Yang and L. Xie [16] proposed a solution based on reweighted atomic norm minimization (RAM) for continuous dictionary, that is able to resolve frequencies with minimum separation as $\Delta_f \geq \frac{0.3}{N}$, for $K \leq 20$, and realized the best enhancement in resolution till this paper.

In this paper, we propose a high resolution gridless sparse approach for joint sparse frequency recovery (JSFR). Our work is inspired by the confident impact of the recent works [12-16] on the optimization relaxation together with MMV to enhance sparsity and resolution. Under this scope, we present a novel Schatten p-norm optimization framework derived to solve the objective function using more efficient sparse metric. We show that the new approach can be solved using a computationally tractable SDP. In fact, the idea of reweighted optimization in the case of sparsity enhancement is not new [16-20]. However, this paper introduces, for the first time, the implementation of ANM using a nonconvex approximation for the main rank function based on Schatten-p norm for JSFR. Numerical results are then provided showing the successful frequency recovery with superior resolution and faster convergence than RAM, under the same technical settings as in [16].

The rest of the paper is organized as follows. Section 2 introduces preliminary mathematical background needed for the problem formulation. Section 3 presents a novel sparse metric for frequency recovery and introduces our approach with theoretical analysis and computational implementation study. Section 4 provides extensive numerical simulations to demonstrate the performance enhancement. Finally, the drawn conclusion and the future aspects are discussed in Section 5. Throughout the paper, bold letters are reserved for matrices and vectors. The transpose is denoted by $(.)^T$, and the complex conjugate or Hermitian is denoted by $(.)^H$. tr(.) and rank(.) represent matrix trace and rank respectively. For an integer N, $[N] := \{1, \dots, N\}$. $\|.\|_0$, $\|.\|_2$ and $\|.\|_F$ refer to l_0, l_2 and Forbenius norms respectively, and $A \geq 0$ means that A is positive semidefinite matrix.

2. Problem Formulation

We study the super-resolution problem of JSFR. Hence, we consider L discrete signals stacked in a full data matrix $\mathbf{Y} \in \mathbb{C}^{N\times L}$. The observed samples of the above matrix are modeled by $\mathbf{Y_\Omega}$ indexed by $\mathbf{\Omega} \subset [N] = \{1, \dots, N\}$, where the size of $\mathbf{\Omega}$ refers to the size of the partial measurements M. The (n, l) indexes the n^{th} entry of the l^{th} measurement vector (snapshot), which is given by:

$$y_{nl} = \sum_{k=1}^{K} \alpha_{kl} e^{i2\pi n f_k}, \quad (1)$$

Where $i = \sqrt{-1}$, $f_k \in [0,1]$ represents the k^{th} normalized frequency, and K is the number of frequency components, which is supposed to be small and unknown, $\alpha_{kl} \in \mathbb{C}$ is the complex amplitude of the k^{th} component at measurement $l \in \{1, \dots . L\}$. From array processing view, the measurement matrix $\mathbf{Y_\Omega}$ represents the output of M sparse linear antenna (SLA) over L measurements ($L > 1$), and each

frequency component refers to individual source. It is worth mentioning, that when $L = 1$, which is known as single measurement vector (SMV), the problem is reduced to line spectrum estimation.

Recall that in JSFR the goal is to recover the whole set of frequency components $\boldsymbol{F} = \{f_1, \dots, f_K\}$ shared among the measurement vectors given in $\mathbf{Y_\Omega}$. For that reason, we seek for the maximally sparse candidate that is composed of the frequency components. To state it properly, let $\mathbf{a}(f)=[1, e^{i2\pi f}, \dots, e^{i2\pi(N-1)f}]^T \in \mathbb{C}^{N\times 1}$ denotes the atom with frequency $f \in [0,1]$, and $\boldsymbol{\alpha_k} = [\alpha_{k1}, \alpha_{k2}, \dots, \alpha_{kL}] \in \mathbb{C}^{1\times L}$ represents the complex amplitude of the k^{th} component through L measurements, it follows that (1) can be written as:

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{a}(f_k)\boldsymbol{\alpha}_k = \sum_{k=1}^{K} s_k \mathbf{a}(f_k)\varphi_k, \tag{2}$$

Where $s_k = \|\boldsymbol{\alpha}_k\|_2 \geq 0$, and $\varphi_k = \frac{\alpha_k}{s_k}$ with $\|\varphi_k\|_2 = 1$. It is evident that, if the frequencies in $\boldsymbol{F}$ are distinct, $\mathbf{Y}$ will be a linear combination of a number of atoms from the continuous dictionary $\mathbf{A}$, which is defined for spectrally-sparse signals as [10]:

$$\mathbf{A} := \{\mathbf{a}(f,\varphi) = \mathbf{a}(f)\varphi : f \in [0,1], \|\varphi\|_2 = 1\}. \tag{3}$$

As a result, the objective of our problem is reduced to look for the atomic decomposition of $\mathbf{Y}$ with smallest number of atoms, where the atomic l_0 norm offers the proper description of this problem, and defined as [11,12]:

$$\|\mathbf{Y}\|_{\mathbf{A},\mathbf{0}} = \inf_{f_k, s_k} \{K : \mathbf{Y} = \sum_{k=1}^{K} \mathbf{a}(f_k,\varphi_k)s_k : \mathbf{a}(f_k,\varphi_k) \in \mathbf{A}, s_k \geq 0\}. \tag{4}$$

According to [13, Theorem 2], $\|\mathbf{Y}\|_{\mathbf{A},\mathbf{0}}$ can be characterized as following constrained rank minimization problem (RMP):

$$\|\mathbf{Y}\|_{\mathbf{A},\mathbf{0}} = \min_{\mathbf{X},\mathbf{u}} \operatorname{rank}(\mathbf{T}(\mathbf{u})), \quad \text{subject to} \quad \begin{bmatrix} \mathbf{X} & \mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y} & \mathbf{T}(\mathbf{u}) \end{bmatrix} \geq 0. \tag{5}$$

Where $\mathbf{T}(\mathbf{u}) \in \mathbb{C}^{N\times N}$ is Hermitian Toeplitz positive semidefinite (PSD) matrix with vector $\mathbf{u}$ being its first row. It can be viewed as the covariance matrix of the full data $\mathbf{Y}$, which is consistent with the observed data $\mathbf{Y_\Omega}^{observed}$.

Consequently, for partially observed signal matrix $\mathbf{Y_\Omega}^{observed} \in \mathbb{C}^{M\times L}$ with the same linear combination form as in (2), and in the absence of noise the RMP in (5) can be expressed as:

$$\min_{\mathbf{X},\mathbf{u},\mathbf{Y}} \operatorname{rank}(\mathbf{T}(\mathbf{u})), \tag{6}$$

$$\text{subject to} \quad \begin{bmatrix} \mathbf{X} & \mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y} & \mathbf{T}(\mathbf{u}) \end{bmatrix} \geq 0, \text{ and } \mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{Y}_{\boldsymbol{\Omega}}{}^{observed}.$$

Where $\mathbf{Y}_{\boldsymbol{\Omega}}$ forms the rows of $\mathbf{Y}$ indexed by $\boldsymbol{\Omega}$.

2.1 Atomic Norm Relaxation in Review

Since RMP in (6) is nonconvex and NP-hard to compute. One way to avoid the nonconvexity is by using convex relaxation through replacing $\|\mathbf{Y}\|_{\mathbf{A},\mathbf{0}}$ by the atomic l_1 norm or briefly atomic norm, which is defined as the gauge function of conv($\mathbf{A}$), the convex hull[4] of $\mathbf{A}$ [10]:

$$\|\mathbf{Y}\|_{\mathbf{A}} = \inf\{t > 0 : \mathbf{Y} \in t.\,\mathrm{conv}(\mathbf{A})\}, \tag{7}$$

$$\inf_{f_k, s_k} \{\textstyle\sum_{k=1} \|s_k\|_2 : \mathbf{Y} = \sum_{k=1}^{K} \mathbf{a}(f_k, \varphi_k) s_k : \mathbf{a}(f_k, \varphi_k) \in \mathbf{A}, s_k \geq 0\}.$$

$\|\mathbf{Y}\|_{\mathbf{A}}$ is a norm and has the following efficient computation SDP formulation [11-13].

$$\|\mathbf{Y}\|_{\mathbf{A}} = \min_{\mathbf{X},\mathbf{u}} \frac{1}{2\sqrt{N}} (\mathrm{tr}(\mathbf{T}(\mathbf{u})) + \mathrm{tr}(\mathbf{X})) \tag{8}$$
$$\text{subject to} \quad \begin{bmatrix} \mathbf{X} & \mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y} & \mathbf{T}(\mathbf{u}) \end{bmatrix} \geq 0.$$

Subsequently, (6) can be casted as the following SDP:

$$\min_{\mathbf{X},\mathbf{u},\mathbf{Y}} \frac{1}{2\sqrt{N}} (\mathrm{tr}(\mathbf{T}(\mathbf{u})) + \mathrm{tr}(\mathbf{X})), \tag{9}$$
$$\text{subject to} \quad \begin{bmatrix} \mathbf{X} & \mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y} & \mathbf{T}(\mathbf{u}) \end{bmatrix} \geq 0, \text{ and } \mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{Y}_{\boldsymbol{\Omega}}{}^{observed}.$$

The frequencies composing $\mathbf{Y}$ are embedded in $\mathbf{T}(\mathbf{u})$. Hence, whenever an optimizer of $\mathbf{u}$ is found the frequencies can be recovered from the Vandermonde decomposition of $\mathbf{T}(\mathbf{u})$ [13,15,22], which states that any PSD Toeplitz matrix can be decomposed as:

$$\mathbf{T}(\mathbf{u}) = \textstyle\sum_{k=1}^{K} p_k\, \mathbf{a}(f_k)\mathbf{a}^H(f_k) = \mathbf{V}(f)\mathbf{P}\mathbf{V}^H(f). \tag{10}$$

Where $\mathbf{V}(f) = [\mathbf{a}(f_1), \mathbf{a}(f_2), \ldots, \mathbf{a}(f_K)]$, and $\mathbf{P} = \mathrm{diag}(p_1, p_2, \ldots, p_K)$ with $p_k > 0$. Generally, the above decomposition is unique if $K < N$ which is fulfilled in JSFR. But to assure that the true frequencies are uniquely obtained, the optimizer of (9) must be global and distinctive.

[4] The convex hull of A is defined as the set of all convex combinations of points in A.

As we have seen, atomic norm is computationally beneficial compared to atomic l_0 norm, but it is constrained by the resolution limit due to the relaxation. In spirit of low rank representation (LRR) as a powerful formulation of our problem due to its strong ability in exploring low-dimensional structures embedded in data, (9) can be interpreted as recovery of low rank matrix **T**(**u**) by relaxing the pseudo rank norm in (6) using the trace or nuclear norm for a PSD matrix. In this context, relaxation in (9) can be achieved in various ways using the promising properties of low rank matrix recovery (LRMR) techniques.

3. Joint Spares Frequency Recovery with Schatten p-Norm

Motivated by applying LRMR techniques to CCS, we propose a new optimization framework to recover low-rank matrix for JSFR problem with Schatten p-norm minimization relaxation, which can be used to solve problems in both (6) and (9). Schatten p-norm generalizes the nuclear norm minimization (NNM), and gives a better approximation to the original RMP.

3.1 Schatten p-Norm Minimization

In this paper, we are interested in the nonconvex Schatten p-norm as a surrogate of rank function. Let $\|.\|_P$ be the Schatten p-norm of the matrix $\mathbf{Y} \in \mathbb{C}^{N\times L}$, which is defined as the l_p norm of its singular values $\sigma_k(\mathbf{Y})$, i.e.

$$\|\mathbf{Y}\|_p = \left(\sum_{k=1}^{\min(N,L)} {\sigma_k}^p(\mathbf{Y})\right)^{\frac{1}{p}} = \left(\mathrm{tr}((\mathbf{Y}^T\mathbf{Y})^{\frac{p}{2}})\right)^{\frac{1}{p}} \text{ with } p \in (0,1]. \quad (11)$$

Recently, Schatten p-norm has received a significant amount of attention from researchers in various domains [21, 23-26] due to its exceptional properties that make it a good choice for rank function approximation. Firstly, it equals to nuclear norm when $p = 1$. Hence, nuclear norm is a special case of Schatten p-norm. Secondly, if p tends to zero, then $\|\mathbf{Y}\|_p \to \mathrm{rank}(\mathbf{Y})$, which indicates that Schatten p-norm can find a low-rank solution when p has a small value [23]. Moreover, [24, 26-28] approved that it will guarantee a more accurate signal recovery while requiring only a weaker restricted isometry property. As well, [29] verified theoretically that Schatten p-norm requires significantly fewer measurements than NNM, and empirically shows superior performance over conventional NNM in many problems.

3.2 Reweighted Iterative Algorithm via Schatten p-Norm

To avoid the non-differentiability of (11), it can be rewritten as follows [25]:

$$\|\mathbf{Y}\|_p = \left(\sum_{k=1}^{\min(N,L)} (\sigma_k(\mathbf{Y}) + \xi)^p\right)^{\frac{1}{p}}. \quad (12)$$

Where $\xi > 0$ is a smoothing parameter. Consequently, we propose the following sparse metric:

$$\mathcal{Q}_p(\mathbf{Y}) = \min_{\mathbf{u}}(\|\mathbf{T}(\mathbf{u})\|_p^p + \mathrm{tr}(\mathbf{X})), \quad \text{subject to} \quad \mathbf{T}(\mathbf{u}) \geq 0. \tag{13}$$

Our iterative algorithm is developed here. Since (13) is not a cost function of a convex optimization problem, we have to linearize it. Thus, we use Taylor expansion for Schatten p-norm term. In the l^{th} iteration, the expansion will be as follows:

$$\left(\sigma_k\left(\mathbf{T}^{\boldsymbol{l}}(\mathbf{u})\right) + \xi\right)^p + \frac{p}{\left(\sigma_k\left(\mathbf{T}^{\boldsymbol{l}}(\mathbf{u})\right)+\xi\right)^{1-p}}\left(\sigma_k(\mathbf{T}(\mathbf{u})) - \sigma_k\left(\mathbf{T}^{\boldsymbol{l}}(\mathbf{u})\right)\right). \tag{14}$$

Where $\{\sigma_k \geq 0\}_1^N$ are descendingly sorted singular values of $\mathbf{T}(\mathbf{u})$. Notice that $\mathbf{T}^{\boldsymbol{l}}(\mathbf{u})$ is fixed, so the associated terms can be removed from the optimization problem.

Additionally, we use the following identity knowing that $\mathbf{T}(\mathbf{u})$ is a PSD matrix [16]:

$$\mathrm{tr}(\mathbf{Y}^{\mathbf{H}}\mathbf{T}(\mathbf{u})^{-1}\mathbf{Y}) = \min_{\mathbf{X}}(\mathrm{tr}(\mathbf{X})), \quad \text{subject to} \quad \begin{bmatrix}\mathbf{X} & \mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y} & \mathbf{T}(\mathbf{u})\end{bmatrix} \geq 0. \tag{15}$$

As a result, (13) can be reformulated as follows:

$$\mathcal{Q}_p(\mathbf{Y}) = \min_{\mathbf{u}}\left(\left\|\boldsymbol{W}^{\boldsymbol{l}}\mathbf{T}(\mathbf{u})\right\|_\star + \mathrm{tr}(\mathbf{Y}^{\mathbf{H}}\mathbf{T}(\mathbf{u})^{-1}\mathbf{Y})\right), \quad \text{subject to} \quad \mathbf{T}(\mathbf{u}) \geq 0. \tag{16}$$

Where $\left\|\boldsymbol{W}^{\boldsymbol{l}}\mathbf{T}(\mathbf{u})\right\|_\star = \mathrm{tr}\left(\boldsymbol{W}^{\boldsymbol{l}}\mathbf{T}(\mathbf{u})\right) = \sum_{k=1}^N (w_k^l \sigma_k(\mathbf{T}(\mathbf{u})))$ is the nuclear norm of the weighted matrix $\boldsymbol{W}^{\boldsymbol{l}}\mathbf{T}(\mathbf{u})$, and $\boldsymbol{W}^{\boldsymbol{l}} = \mathbf{D}\mathrm{diag}\left(w_1^l, w_2^l, \ldots, w_N^l\right)\mathbf{D}^{\mathrm{H}}$ is the reweighting matrix whose elements $w_k^l = \frac{p}{\left(\sigma_k\left(\mathbf{T}^{\boldsymbol{l}}(\mathbf{u})\right)+\xi\right)^{1-p}}$ are updated based on the latest solution of the problem, and $\mathbf{D}$ is obtained using the eigen decomposition of $\mathbf{T}(\mathbf{u})$, so that $\mathbf{D}\mathbf{D}^{\mathrm{H}} = \boldsymbol{I}$.

Obviously, the first term of the proposed sparse metric in (16) is reduced to weighted NNM (WNNM) model [26,29-30], which improves the flexibility of the original NNM problem by assigning different weights to different singular values, namely the success of this model depends on the appropriate setting of the weights. However, since WNNM is not convex in general, it is difficult to analyze the convergence of the approach. Theoretically, no efficient algorithms can guarantee the global optimizer. To work around this, we observe that $\{\sigma_k\}_1^N$ are nonnegative monotonically decreasing, then the non-descending order of the resultant weights with respect to singular values will be kept throughout the reweighting process, and

hence, we always expect to shrink less the larger singular values and keep the major and faithful information of the underneath data. In this context and according to [27], S. Gu et al verified that WNNM converges weakly to the solution if the weights are non-descendingly ordered through a weighted soft-thresholding operator.

Consequently, as Schatten p-norm is concave in general, then at each iteration its value decreases rapidly greater than the decrease of its tangent plane which is approximately equal to the Taylor expansion in (14). It follows that by iteratively solving (16), the objective function in (13) monotonically decreases and converges to a local minimum.

Recall the definitions of both weighted continuous dictionary, and weighted atomic norm as in (17) and (18) respectively [16].

$$\mathbf{A}_{\boldsymbol{\omega}} := \{\mathbf{a}_{\boldsymbol{\omega}}(f) = \omega(f)\mathbf{a}(f): f \in [0,1]\}. \tag{17}$$

$$\|\mathbf{Y}\|_{\mathrm{A},\omega} = \inf_{f_k, s_k} \left\{\sum_{k=1} \frac{\|s_k\|_2}{\omega(f_k)} : \mathbf{Y} = \sum_{k=1}^{K} \mathbf{a}(f_k) s_k\right\}, \tag{18}$$

Or equally,

$$\|\mathbf{Y}\|_{\mathrm{A},\omega} = \min_{\mathbf{u}} \left(\frac{\sqrt{N}}{2}\mathrm{tr}(\boldsymbol{W}\mathbf{T}(\mathbf{u})) + \frac{1}{2\sqrt{N}}\mathrm{tr}(\mathbf{Y}^{\mathbf{H}}\mathbf{T}(\mathbf{u})^{-1}\mathbf{Y})\right), \tag{19}$$
$$\text{subject to} \quad \mathbf{T}(\mathbf{u}) \geq 0.$$

Obviously, by comparing(16) and (19) we can regenerate our problem as RAM by simply implying our weighting matrix to (19) with $\boldsymbol{W^l} = \frac{\mathbf{1}}{\boldsymbol{N}}\mathbf{D}\mathrm{diag}\left(w_1^l, w_2^l, \dots, w_N^l\right)\mathbf{D}^{\mathrm{H}}$ at l iteration, and $\omega(f) = \frac{1}{\sqrt{\boldsymbol{a}^H(f)\,\boldsymbol{W^l}\boldsymbol{a}(f))}}$ by applying [16, Theorem 3]. Therefore, we call our approach a modified RAM (MRAM). In contrast to [16], MRAM defines the reweighting strategy based on iteratively updated singular value decomposition (SVD) of $\mathbf{T}(\mathbf{u})$ matrix. Although, computing SVD can be expensive, but by using techniques such as randomized algorithms [19], which tend to compute the top k singular vectors and singular values of $\mathbf{T}(\mathbf{u})$, the cost of computing weighting matrix can be significantly saved. In RAM the weighting function was chosen more sophisticatedly according to the inverse of $(\mathbf{T}(\mathbf{u}) + \varepsilon\boldsymbol{I})$ matrix, where $\varepsilon > 0$.

3.3 Enhancing Resolution and Sparsity

By using the proposed sparse metric $\mathcal{Q}_p(\mathbf{Y})$, we can cast (9) to the following optimization problem:

$$\min_{\mathbf{Y}} \mathcal{Q}_p(\mathbf{Y}), \tag{20}$$

$$\text{subject to} \quad \mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{Y}_{\boldsymbol{\Omega}}{}^{observed}.$$

Or equally,

$$\min_{\mathbf{u},\mathbf{Y}} \left(\left\| \boldsymbol{W}^{l}\mathbf{T}(\mathbf{u}) \right\|_{\star} + \mathrm{tr}(\mathbf{X}) \right), \tag{21}$$
$$\text{subject to} \quad \begin{bmatrix} \mathbf{X} & \mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y} & \mathbf{T}(\mathbf{u}) \end{bmatrix} \geq 0, \text{ and } \mathbf{Y}_{\boldsymbol{\Omega}} = \mathbf{Y}_{\boldsymbol{\Omega}}{}^{observed}.$$

To illustrate the sparsity property and the promising resolution enhancement of the new metric, we present the properties of $\mathcal{Q}_p(\mathbf{Y})$ in comparison with the log-det sparse metric used in [16]. Since $\mathcal{Q}_p(\mathbf{Y})$ puts penalty on $\sum_{k=1}^{\min(N,L)}(\sigma_k(\mathbf{T}(\mathbf{u})) + \xi)^p$, we plot the function $f_p(\sigma) = (\sigma + \xi)^p$ with p = 0.5, and $h(\sigma) = ln|\sigma + \xi|$ function which prompts the log-det heuristic in Fig.1 for different values for ξ together with l_0, and l_1 norms, where $f_p(\sigma)$ and $h(\sigma)$ are translated and scaled by the same factor as in [16]. Contrary to $h(\sigma)$ curves, $f_p(\sigma)$ gets close to the l_0 norm for large ξ while it approaches the l_1 norm as $\xi \longrightarrow 0$. However, it converges faster toward l_0 whenever $p \longrightarrow 0$ as illustrated in Fig.2.

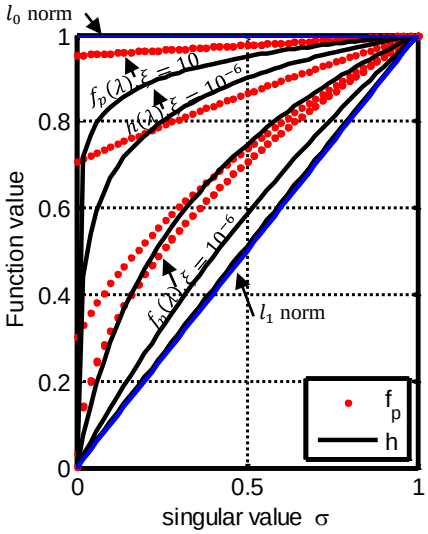

Fig. 1 The sparsity property of $f_p(\sigma)$ for p=0.5 (dot curve) and $h(\sigma)$ for different values of ξ.

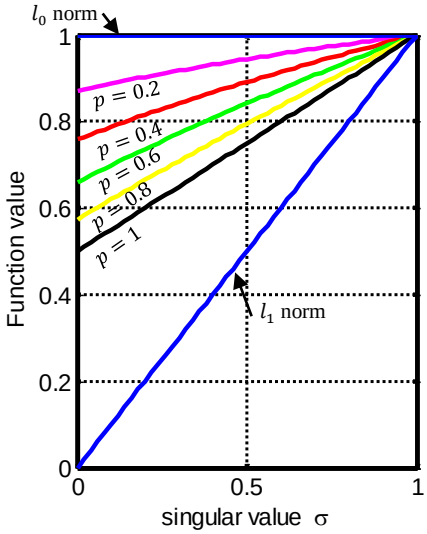

Fig. 2 The sparsity property of $f_p(\sigma)$ with ξ=1, for different values of p.

Hence, we assume that the $\mathcal{Q}_p(\mathbf{Y})$ bridges l_0 and l_1, and consequently $\|\mathbf{Y}\|_{A,0}$ and $\|\mathbf{Y}\|_A$ when ξ diverges from $+\infty$ to 0 with $0 < p \leq 1$. This approach is expected to enhance sparsity and resolution with boosted properties than log-det based sparse metric used in [16]. To validate the convergence of (21), we know that $f_p(\sigma)$ is a concave function in σ, thus the sum of $f_p(\sigma)$ with the trace function is a

concave-convex optimization problem. As a popular solution to this problem we refer to locally convergent maximization-minimization (MM) algorithm [20], which guarantees the monotonically decreasing of the objective function in (21) and its converges to a local minimum. This result coincides with the convergence analysis mentioned in subsection 3.2.

3.4 Efficient Computational Implementations with SDP

Fortunately, as shown in the previous subsections, atomic norm and its weighted version admit SDP formulations, which imply efficient computation using off-the-shelf standard SDP solvers such as SDPT3 [32]. Intuitively, the straightforward implementation is based on the primal SDP formulation. However, according to related literature [11, 15-16], frequency recovery can be done more efficiently through solving dual optimization problem than the primal one. Thus, we cast our analysis and implementation to the dual problem of (21), which can be formulated by standard Lagrangian analysis [10]. We derive the dual problem depending on the dual polynomial, due to its characteristics in frequency recovery problem [11]. Firstly, we define the dual atomic norm $\|\mathbf{x}\|_{A,\omega}^{*}$ as follows:

$$\|\mathbf{x}\|_{A,\omega}^{*} = \max_{\|\mathbf{y}\|_{A,\omega}\leq 1} \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}} = \max_{f} (\omega(f) |\mathbf{a}^{H}(f)\mathbf{x}|). \tag{22}$$

Where, $\langle , \rangle_{\mathbb{R}}$ denotes the real part of the inner product. Inspired by [15], we consider the dual problem of (21) as follows:

$$\begin{aligned} &\max_{\mathbf{x}} \langle \mathbf{x}_{\boldsymbol{\Omega}}, \mathbf{y}_{\boldsymbol{\Omega}}^{observed} \rangle_{\mathbb{R}} \\ &\text{subject to} \quad \|\mathbf{x}\|_{A,\omega}^{*} \leq 1, \text{ and } \mathbf{x}_{\boldsymbol{\Omega}^{c}} = 0. \end{aligned} \tag{23}$$

Where, $\boldsymbol{\Omega}^{\boldsymbol{c}}$ is the complement set of $\boldsymbol{\Omega}$. As a result, the problem is reduced to the following inequality:

$$|\mathbf{a}^{H}(f)\mathbf{x}| \leq \omega^{-1}(f) = \sqrt{\mathbf{a}^{H}(f)\, \boldsymbol{W}\mathbf{a}(f))} \text{ for all } f \in \boldsymbol{\Omega}. \tag{24}$$

In order to characterize (24) as SDP, we reformulate it by using the properties of trigonometric polynomials [33]. Since $\boldsymbol{W}$ is a Hermitian PSD matrix- it can easily be proven- and $\mathbf{a}(\boldsymbol{f})=[1, e^{i2\pi f}, \ldots, e^{i2\pi(\mathrm{N}-1)f}]^{T}$, then $\omega^{-2}(f)$ is a Hermitian nonnegative trigonometric polynomial defined as follows:

$$\omega^{-2}(f) = \mathbf{a}^{H}(f)\, \boldsymbol{W}\mathbf{a}(f)) = \mathrm{tr}(\mathbf{a}(f)\mathbf{a}^{H}(f)\boldsymbol{W}). \tag{25}$$

By applying [33, Theorem 2.3], we have:

$$\omega^{-2}(f) = \mathrm{tr}\left(\left(\sum_{k=-(N-1)}^{N-1} \boldsymbol{\Theta}_{\boldsymbol{k}} e^{i2\pi kf}\right) \boldsymbol{W}\right) = \sum_{k=-(N-1)}^{N-1} \mathrm{tr}(\boldsymbol{\Theta}_{\boldsymbol{k}}\boldsymbol{W})\, e^{i2\pi kf}, \tag{26}$$

Where $\boldsymbol{\Theta_k}$ denotes an $N\times N$ elementary Toeplitz matrix with ones on the k^{th} diagonal and zeros elsewhere. Now, if the inequality (24) is satisfied then by carrying out [33, Theorem 4.24], there exists a matrix $\mathbf{Z} \in C^{N\times N} \geq 0$ realizing

$$\begin{bmatrix} \mathbf{I} & \mathbf{x}^{\mathrm{H}} \\ \mathbf{x} & \mathbf{Z} \end{bmatrix} \geq \mathbf{0} \quad \text{and} \quad \mathrm{tr}(\boldsymbol{\Theta_k}\mathbf{Z}) = \mathrm{tr}(\boldsymbol{\Theta_k}\boldsymbol{W}), \quad k = 1, \dots, N-1, \tag{27}$$

Where $\mathbf{I}$ is $L \times L$ identity matrix. By using the relation between the inner product and matrix trace, the dual problem in (23) can be successfully formulated as the following SDP:

$$\begin{aligned} &\max_{\mathbf{x},\mathbf{Z}} \Re(\mathrm{tr}(\mathbf{x}_{\boldsymbol{\Omega}}^{\boldsymbol{H}}\mathbf{y}_{\boldsymbol{\Omega}}^{observed})) \\ &\text{subject to (27), and } \mathbf{x}_{\boldsymbol{\Omega}^c} = 0, \end{aligned} \tag{28}$$

Where $\Re$ indicates the real part of the argument.

As in [11], once (28) is solved, the frequencies can be localized by identifying the locations where the dual polynomial $|\mathbf{a}^H(f)\mathbf{x}|$ exceeds $|\,\omega^{-1}(f)|$. Note that the optimizer to (21) is given for free via duality when we solve (28). It is worth noting that, in each iteration we indeed reconstruct $\mathbf{T}(\mathbf{u})$ matrix which presents the data covariance of $\mathbf{Y}$ after removing correlations among the sources according to [16]. Hence, the set of frequencies can be recovered via either characterization of the dual polynomial, or by using conventional covariance-based subspace methods of spectrum estimation such as MUSIC [34] and ESPRIT [35].

3.5 Complexity Analysis:

In this subsection, we mainly compare the computational complexity of MRAM and RAM. If we consider the dominant part, i.e., the computation of weighting function. The upper bound of the overall cost for calculating $\boldsymbol{W}$ in MRAM is $O((N+L)^3)$ floating-point operations (flops) due to SVD decomposition, thus it consumes up to $O((N+L)^3)$ flops to compute the inverse of $(\mathbf{T}(\mathbf{u}) + \varepsilon\mathbf{I})$ in RAM. Moreover, we should mention that in general ANM, RAM and MRAM are all based on CVX toolbox [36], where the interior-point method (IPM) such as SDPT3 is implemented to solve the SDP, which requires $O((n+L^2)^2(N+L)^{2.5})$ flops per iteration for both RAM and MRAM, and as total complexity for ANM at best, where n denotes the number of variables [32, 37]. Hence, the overall computational complexity per iteration to extract $\mathbf{T}(\mathbf{u})$ for both RAM and MRAM equals to $O((N+L)^3 + (n+L^2)^2(N+L)^{2.5})$, and the total complexity equals the complexity per iteration times the number of iterations in each approaches plus the complexity of covariance-based algorithm to detect the frequency components.

Although, the complexity of the proposed approach is equivalent to that of RAM, however, empirical results show that the convergence of MRAM is faster

and only several iterations are needed to converge. Therefore, the proposed method performs better in practice. The execution efficiency can be further improved by using alternating direction method of multipliers (ADMM), which requires $O((N+L)^3)$ flops per iteration, and in the case of $L > M$, this complexity can be reduced to $O((N+M)^3)$ by applying the dimensionality reduction technique as presented in [37], but the convergence is much slower than primal-dual IPM.

To demonstrate the effectiveness of our method, we perform extensive numerical simulations in Section 4.

4. Performance Evaluation

In this section, we evaluate the performance of MRAM in noiseless cases. In particular, we examine the effect of suggested sparse metric on JSFR. All simulations are carried out using CVX of MATLAB v.14.a on a PC with Windows 7 system and a 3.3 GHz CPU.

Let $N = 64$, and $M = 30$,[5] in which the observed samples are selected uniformly at random from each measurement vector. Suppose also that we collect L ($< M$) snapshots of uncorrelated random complex amplitudes sinusoidal signals sharing $K = 5$ spikes randomly located in the interval [0,1). When implementing MRAM, we set the total number of iterations to 20, $\xi = 10^{-3}$,[6] and $p = 0.25$. In each iteration p is divided by 2, unless $p \geq 25 \times 10^{-5}$.

4.1 Constructed Dual Polynomial

In [15], it was shown that the reconstructed dual polynomial for ANM in MMV model has a much better localization property than ANM in SMV. Thus, for simplicity, we took L=1 to illustrate the superiority of MRAM dual polynomial characteristics over ANM one. Fig.3 shows the reconstructed dual polynomial of ANM and MRAM results for randomly generated spectrally sparse signals with $f = [0.1, 0.1047, 0.2, 0.2156, 0.3]$. Note that the first two frequencies are mutually separated by only about $\frac{0.3}{N}$ (the best resolution of [16]), while the third and fourth ones are separated by $\frac{1}{N}$ (the best resolution of [15]).

We notice that when the frequency separation condition is violated, ANM localization degrades significantly, while MRAM has superior ability to discriminate the frequencies under the same separation settings.

[5] We choose the value of M according to [11]; $M \approx O(K \log K \log N)$.

[6] We use a small value of ξ due to the reverse relation between the weight and $f_p(\lambda)$.

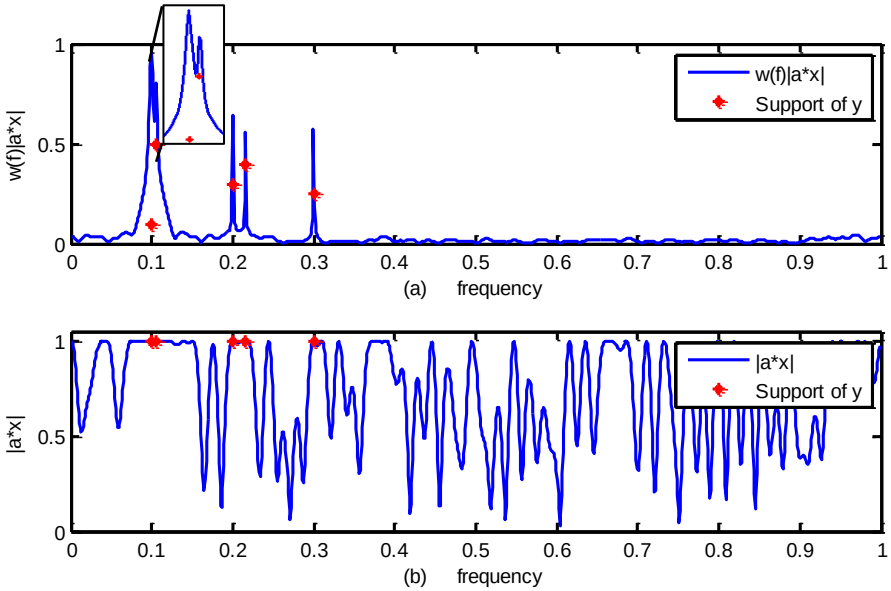

Fig. 3 The reconstructed dual polynomial for: (a) MRAM with true amplitudes, (b) ANM.

4.2 Comparison with Existing Approaches

The following experiment examines the performance of MRAM in reconstructing the frequency spectrum from M picked samples. For better evaluation we implement both MRAM and RAM via SDPT3 solver and compare their performances under the same settings as in 4.1 subsection except for L=5 and $f = [0.1, 0.104, 0.1094, 0.2, 0.5]$. Here the first three frequencies are separated by only $\frac{0.3}{N}$. Whereas, the frequencies for each implementation are considered to be successfully recovered if the root mean squared error (RMSE) is less than 10^{-6}. We plot the simulation results in Fig.4.

The first and third subfigures present the recovered frequencies by using MRAM and RAM respectively, whereas the second and fourth rows plot the weighting functions used during the iterations for MRAM and RAM respectively. Following the reweighting process of MRAM, all the frequencies in this example are correctly recovered from the second iteration and the algorithm tightly converges. While RAM takes three iterations to recover the frequencies properly. This is in agreement with [16] results.

Note that the first iteration in both RAM and MRAM coincide with ANM for constant weighting function. The frequency recovery is coarse; the first three spikes are positioned but cannot precisely be determined. By applying the weighting functions, the five spikes are identified by MRAM from the second iteration, while it needs three iterations to localize precisely via RAM.

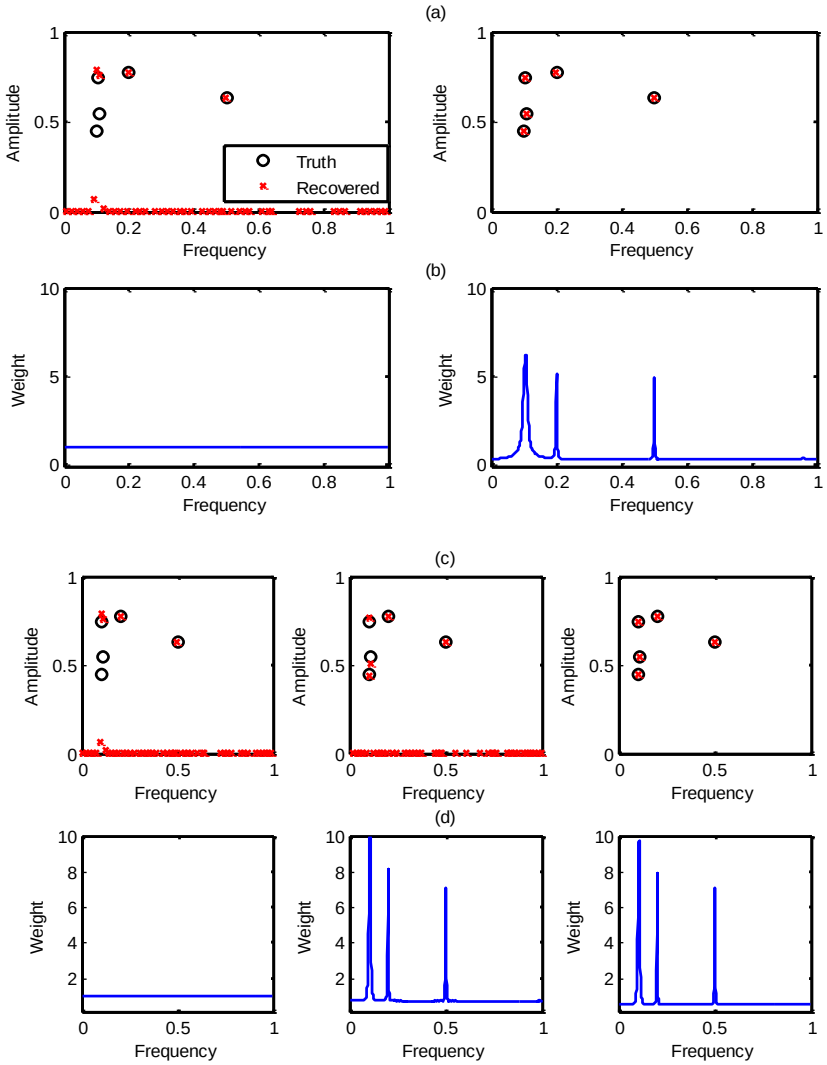

Fig. 4 The recovered frequencies and the associated reweighting functions for MRAM (2 iterations) in (a) and (b), and RAM (3 iterations) in (d) and (c) respectively.

We further compare the MRAM against RAM, and ANM in frequency estimation using the obtained real part of Toeplitz covariance matrix $\mathbf{T}(\mathbf{u})$. This is done by applying rootMUSIC method with a model order estimated iteratively, and under the same prior settings.

Fig.5 shows the power of MRAM in recovering frequency accurately over RAM and ANM. Notice that RAM and MRAM plots are performed for the same number of iterations.

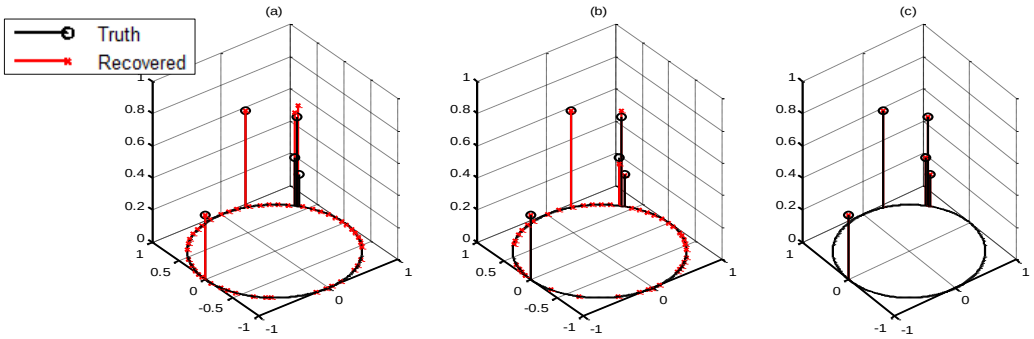

Fig. 5 Frequency estimation using different approaches: (a) ANM, (b) RAM, (c) MRAM.

4.3 Resolution Enhancement

In the third simulation, we show a simple example that provides some evidence for resolution boost of MRAM. Under the same settings, with $f = [0.1, 0.1016, 0.2, 0.3, 0.3016]$, where the frequency separations between the first frequency pair and the last pair are only $\Delta_f = \frac{0.1}{N}$. The simulation results are shown in Fig.6. The first row of subfigures presents the recovered frequencies by MRAM under the proposed frequency separation challenge, while the second row displays the weighting functions used during the iterations. As shown in Fig.6, although the spikes violate the separation conditions of the latest studies in this domain [15, 16], all the frequencies are determined properly in the first three iterations of MRAM. Again, remarkable improvement is obtained by the proposed MRAM compared to RAM and ANM.

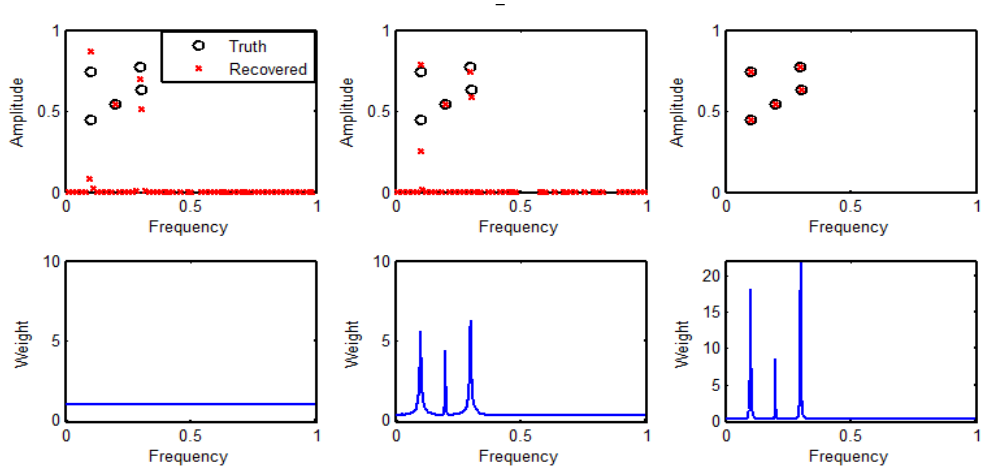

Fig. 6 The recovered frequencies and the associated reweighting functions throughout three first iterations for MRAM with $\Delta_f = \frac{0.1}{N}$.

4.4 Sparsity-Separation Phase Transition

For further clarification, in the following simulation we study the sparse recovery capabilities of MRAM in terms of sparsity separation phase transition[7]. Particularly, we fix $N = 32, M = 15$. We examine the phase transition of frequency recovery for various pairs of (K, Δ_f). For each pair, we randomly generate K frequencies such that they are mutually separated by at least Δ_f. We randomly generate the amplitudes independently and identically from a standard complex normal distribution. Consequently, the frequency recovery is processed using MRAM, whereas the recovery is considered successful if the relative MSE of frequency recovery are less than 10^{-12}. Fig.7 shows the success rates of MRAM with L=1 in (a) and L=5 in (b). The grayscale images present the success rates, where white and black indicate complete success and complete failure, respectively.

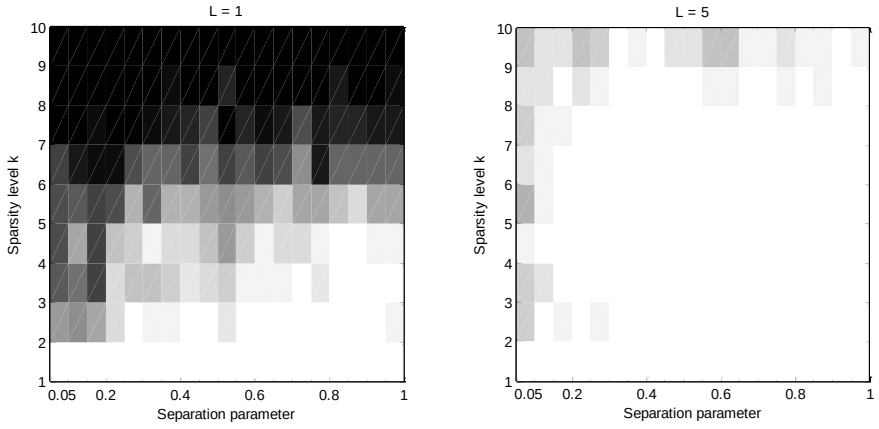

Fig. 7 Sparsity-separation phase transition of MRAM with: (a) L=1, and (b) L=5

Obviously, as L increases the recovered frequencies get more accurate, specifically at close-located frequencies. Practically, under the current simulation settings the successful rate exceeds 95% whenever $K \leq 8$ and frequency separation $\Delta_f \geq \frac{0.1}{N}$. Regardless of the sparsity level, Δ_f of MRAM is about 3 times weaker than the condition in [16], and about 10 times less than the condition in [15]. Moreover, the performance intuitively improves, as more samples per measurement vector are offered.

5. Conclusions

In this paper, we propose a generalized gridless sparse approach for joint sparse frequency recovery model. We present a novel Schatten p-norm optimization

[7] Phase transition diagrams are two-dimensional figures, mainly, the y-axis represents the sparsity, and the x-axis denotes the separation parameters. It provides information about the success of applying compressive measurements under certain conditions.

framework derived to resolve the objective of ANM efficiently via SDP. The effectiveness of the proposed approach is further demonstrated through numerical examples. The simulations results illustrated the excellence of the frequency recovery with superior resolution, and faster convergence, when compared with state-of-the-art approaches. In future studies, we will try different computationally efficient algorithms for matrix rank minimization to push forward the recovery resolution accuracy. In addition, we look forward to examining our approach under noisy measurements scenarios.

R E F E R E N C E S

[1]. *P. Stoica and R. L. Moses,* "Spectral analysis of signals". Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

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

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

[4]. *Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank,* "Sensitivity to basis mismatch in compressed sensing," IEEE Trans. Signal Process., vol. 59, no. 5, pp. 2182–2195, May. 2011.

[5]. *J. M. Nichols, A. K. Oh, and R. M. Willett,* "Reducing basis mismatch in harmonic signal recovery via alternating convex search," IEEE Signal Process. Lett., vol. 21, no. 8, pp. 1007–1011, Aug. 2014.

[6]. *L. Hu, Z. Shi, J. Zhou, and Q. Fu,* "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," IEEE Trans. on Signal Processing, vol. 60, no. 7, pp. 3809–3822, 2012.

[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]. *L. Hu, J. Zhou, Z. Shi, and Q. Fu,* "A fast and accurate reconstruction algorithm for compressed sensing of complex sinusoids," IEEE Trans. on Signal Processing, vol. 61, no. 22, pp. 5744–5754, 2013.

[9]. *E. J. Cande`s 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.

[10]. *V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky,* "The convex geometry of linear inverse problems," Foundations of Computational Mathematics, vol. 12, no. 6, pp. 805–849, 2012.

[11]. *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.

[12]. *Z. Yang and L. Xie,* "Continuous compressed sensing with a single or multiple measurement vectors," in IEEE Workshop on Statistical Signal Processing (SSP), pp. 308–311, 2014.

[13]. *Z. Yang and L. Xie*, "Exact joint sparse frequency recovery via optimization methods," IEEE Trans. on Signal Processing, vol. 64, no. 19, pp. 5145–5157, Oct. 2016.

[14]. *Y. Chi,* "Joint sparsity recovery for spectral compressed sensing," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3938–3942, 2014.

[15]. *Y. Li and Y. Chi,* "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," IEEE Trans. on Signal Processing, vol. 64, pp. 1257 - 1269, 2016.

[16]. *Z. Yang and L. Xie*, "Enhancing sparsity and resolution via reweighted atomic norm minimization," IEEE Trans. on Signal Processing, vol. 64, no. 4, pp. 995–1006, Feb. 2016.

[17]. *E. J. Candes, M. B. Wakin, and S. P. Boyd,* "Enhancing sparsity by reweighted l_1 minimization," Journal of Fourier Analysis and Applications, vol. 14, no. 5-6, pp. 877–905, 2008.

[18]. *D. Wipf and S. Nagarajan,* "Iterative reweighted l_1 and l_2 methods for finding sparse solutions," IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pp. 317–329, 2010.

[19]. *N. Halko, P. G. Martinsson, and J. A. Tropp.* "Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions," SIAM Review, 53(2):217–288, 2011.

[20]. *J. M. Ortega and W. C. Rheinboldt, "*Iterative Solution of Nonlinear Equations in Several Variables," New York, NY, USA: Academic, vol. 30, 1970.

[21]. *K. Mohan and M. Fazel,* "Iterative reweighted algorithms for matrix rank minimization," The Journal of Machine Learning Research, vol. 13, no. 1, pp. 3441–3473, 2012.

[22]. *C. Carathéodory and L. Fejér*, "Uber den zusammenghang der extemen von harmonischen funktionen mit ihren koeffizienten und uber den Picard-Landauschen satz," Rendiconti del Circolo Matematico di Palermo, vol. 32, pp. 218–239, 1911.

[23]. *S. Ji, K.-F. Sze, Z. Zhou, A. M.-C. So, and Y. Ye*, "Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization," in INFOCOM, 2013 Proceedings IEEE, pp. 2499–2507, IEEE, 2013.

[24]. *L. Liu, W. Huang, and D. Chen. "*Exact minimum rank approximation via Schatten p-norm minimization. Journal of Computational and Applied Mathematics, 267(1):218–227, 2014.

[25]. *D. Ge, X. Jiang, and Y. Ye,* "A note on the complexity of lp minimization," Mathematical programming, vol. 129, no. 2, pp. 285–299, 2011.

[26]. *F. Nie, H. Huang, and C. Ding*, "Low-rank matrix recovery via efficient Schatten p-norm minimization," in Proc. 26th AAAI Conf. Artif. Intell., pp. 655–661, 2012.

[27]. *S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, L. Zhang,* "Weighted nuclear norm minimization and Its applications to low level vision," International Journal of Computer Vision, 183-208, 2017.

[28]. *F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding,* "Robust matrix completion via joint Schatten p-norm and Lp-norm minimization," in Proc. IEEE Int. Conf. Data Mining, pp. 566–574, 2012.

[29]. *M. Zhang, Z. Huang, and Y. Zhang,* "Restricted p-isometry properties of nonconvex matrix recovery," IEEE Trans. Inform. Theory, vol. 59, no. 7, pp. 4316–4323, 2013.

[30]. *S. Gu, L. Zhang, W. Zuo, and X. Feng,* "Weighted nuclear norm minimization with application to image denoising," CVPR. 2014.

[31]. *K. Mohan and M. Fazel,* "Reweighted nuclear norm minimization with application to system identification," in Proc. ACC, pp. 2953–2959, 2010.

[32]. *K.-C. Toh, M. J. Todd, and R. H. Tutunu,* "SDPT3–a MATLAB software package for semidefinite programming, version 1.3," Optimization Methods and Software, vol. 11, no. 1-4, pp. 545–581, 1999.

[33]. *B. Dumitrescu*, "Positive trigonometric polynomials and signal processing applications," Springer, 2007.

[34]. *R. Schmidt*, "A signal subspace approach to multiple emitter location spectral estimation," Ph.D. dissertation, Stanford University, 1981.

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

[36]. *M. Grant, S. Boyd, and Y. Ye,* "Cvx: Matlab software for disciplined convex programming," 2008.

[37]. *Z. Yang, J. Li, P. Stoica, and L. Xie*, "Sparse Methods for Direction-of-Arrival Estimation," in Academic Press Library in Signal Processing Volume 7, R. Chellappa and S. Theodoridis, Eds. Academic Press, pp. 509–581, 2017.