# PRIOR KNOWLEDGE AIDED SUPER-RESOLUTION LINE SPECTRAL ESTIMATION: AN ITERATIVE REWEIGHTED ALGORITHM

*Feiyu Wang**, *Jun Fang**, and *Hongbin Li*‡

*National Key Laboratory on Communications, University of Electronic Science and Technology of China
‡Department of Electrical and Computer Engineering, Stevens Institute of Technology

### ABSTRACT

This paper concerns detecting the frequency components from a spectral sparse, undersampled signal. This problem is also called super-resolution line spectral estimation because the frequencies can take arbitrary continuous values. The prior knowledge of the frequency distribution is often available in many applications. To exploit the prior knowledge, a weighting function $w(f)$ designed according to the frequency distribution $p(f)$ is introduced. The prior information can be harnessed through minimizing the corresponding weighted log-sum penalty function. We solve the optimization problem through iteratively decreasing a surrogate function majorizing the original penalty function. Simulation results show that the proposed algorithm outperforms other methods both in noiseless and noisy case, and it also presents superior performance in resolving closely-spaced frequency components.

***Index Terms*—** Compressed sensing, super-resolution, iterative reweighted method, probabilistic prior

## I. INTRODUCTION

Line spectral estimation aims to infer the spectral components of the observed signal and is an extensively studied problem in communications, radar, sonar and so on. Conventional methods like MUSIC [1] and ESPRIT [2] are based on subspace and thus require an estimate of the data covariance. Such an estimate, however, is always difficult to obtain in compressive data. There are some state-of-the-art spectrum estimation techniques based on sparse representation of the under-sampled signal over the past few years. Candès and Fernandez-Granda proposed the breakthrough technique of super-resolution [3] that recover frequencies in a continuous domain through total variation minimization. Tang *et al*. proposed another grid-free method based on atomic norm minimization [4]. Both of the two approaches have the theoretical guarantee that the continuous-valued frequencies can be exactly recovered provided the frequencies are well separated. Apart from harnessing the spectral sparsity in the continuous domain directly, another line of research [5]–[7] combined sparse signal recovery and dictionary refinement. In [5], the perturbation of the dictionary caused by gird mismatch is modeled as a structured parameterized matrix through first-order Taylor expansion. The frequencies are obtained from both the support of the sparse signal and the estimate of the off-grid difference. In [6], the Fourier dictionary is parameterized by unknown frequencies and both the frequency grids and the corresponding complex amplitudes are estimated under Bayesian framework. In [7], we proposed an iterative reweighted approach to jointly update the sparse signal and parameterized dictionary.

Nevertheless, except for spectral sparsity, these super-resolution techniques assume no other prior information of the signal. While in many applications, the prior knowledge of the frequency distribution is always available from previous observations. There are some works [8]–[10] that handle the prior knowledge by weighted atomic norm minimization. The weighting scheme is similar to those methods in conventional compressed sensing with partially known support [11], [12]. In [8], [9], the authors construct a piecewise constant weighting function to exploit the block prior and transform the problem to a semidefinite program (SDP) with several linear matrix inequalities (LMI). In [10], a Capon's power spectrum like weighting function is proposed leading to an SDP formulation with only a single LMI that can be solved efficiently.

In this paper, we propose an algorithm for super-resolution line spectral estimation when prior knowledge of the frequency $f$ is available. We extend our previous work [7] to improve the performance of frequency detection by harnessing the prior information. The weighting function $w(f)$ designed according to the frequency distribution $p(f)$ is introduced and the prior information can be exploited through minimizing the corresponding weighted log-sum penalty function. We solve the optimization problem through iteratively decreasing a surrogate function majorizing the original penalty function. Simulation results show that the proposed algorithm outperforms other methods both in noiseless and noisy case, and it also presents superior performance in resolving closely-spaced frequency components.

## II. PROBLEM FORMULATION

The frequency sparse signal in line spectral estimation or direction-of-arrival (DOA) estimation could be formulated as a summation of a number of complex sinusoids:

$$y_l = \sum_{k=1}^{K} z_k^\star e^{-j2\pi f_k^\star l} + \varepsilon_l, \qquad l \in \mathcal{L} \tag{1}$$

where $\mathcal{L} = \{1, \dots, L\}$, $f_k^\star \in [0, 1)$ denotes the frequency of the $k$-th component, $z_k^\star$ is the corresponding complex amplitude and $\varepsilon_l$ represents the observation noise. In many practical applications, the original signal is observed on an index set

$\mathcal{M} \subset \mathcal{L}$, $|\mathcal{M}| = M \leq L$. Denote $\mathcal{M} \triangleq \{s_1, \ldots, s_M\}$ and define $\boldsymbol{a}(f) \triangleq [e^{-j2\pi f s_1} \; e^{-j2\pi f s_2} \; \ldots \; e^{-j2\pi f s_M}]^T$. The observed signal can be rewritten in a vector-matrix form as

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{f}^\star)\boldsymbol{z}^\star + \boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{y} \triangleq [y_{s_1} \; \ldots \; y_{s_M}]^T$, $\boldsymbol{\varepsilon} \triangleq [\varepsilon_{s_1} \; \ldots \; \varepsilon_{s_M}]^T$, $\boldsymbol{z}^\star \triangleq [z_1^\star \; \ldots \; z_K^\star]^T$, and $\boldsymbol{A}(\boldsymbol{f}^\star) \triangleq [\boldsymbol{a}(f_1^\star) \; \ldots \; \boldsymbol{a}(f_K^\star)]$. Suppose the prior knowledge of the probability distribution of the frequencies $\{f_k^\star\}$ is obtained. Let $F \in [0, 1]$ be a random variable that describes the signal frequencies. The probability density function (pdf) of $F$ is $p(f)$. Our objective is to recover the continuous-valued frequencies $\{f_k^\star\}$ given the observed signal $\boldsymbol{y}$ and the prior distribution $p(f)$.

### A. Weighted log-sum penalty function

To exploit the sparsity of the frequency components in the original signal, in our previous work [7], we construct a parametric dictionary $\boldsymbol{A}(\boldsymbol{f}) \triangleq [\boldsymbol{a}(f_1) \; \ldots \; \boldsymbol{a}(f_N)]$, $M \ll N$ and formulate the problem as:

$$\begin{aligned} \min_{\boldsymbol{z}, \boldsymbol{f}} \quad & \sum_{n=1}^{N} \log(|z_n|^2 + \epsilon) \\ \text{s.t.} \quad & \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{f})\boldsymbol{z}\|_2 \leq \eta \end{aligned} \tag{3}$$

where $z_n$ denotes the $n$th component of the vector $\boldsymbol{z}$, $\epsilon > 0$ is a positive parameter to ensure that the function is well-defined, and $\eta$ is the error tolerance parameter depended on the noise variance. The log-sum penalty function, which was proved theoretically [13] to be more sparsity-encouraging than the $\ell_1$-norm, is used and both the parameters $\{f_n\}$ and $\{z_n\}$ are optimized in the continuous domain.

When the prior information about the frequency distribution is available, inspired by some prior-knowledge aided approaches in compressed sensing [11], [12] and spectral super-resolution [8]–[10], we replace the objective in (3) with its weighted counterpart to improve the recovery performance. The new optimization problem is formulated as:

$$\begin{aligned} \min_{\boldsymbol{z}, \boldsymbol{f}} \quad & L(\boldsymbol{z}, \boldsymbol{f}) = \sum_{n=1}^{N} \log(w(f_n)|z_n|^2 + \epsilon) \\ \text{s.t.} \quad & \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{f})\boldsymbol{z}\|_2 \leq \eta \end{aligned} \tag{4}$$

where $w(f) > 0$ is a weighting function to penalize the frequency $f$ and the formation of it would be discussed later. By removing the constraint and adding a data fitting penalty, the optimization (4) can be formulated as an unconstraint problem as:

$$\min_{\boldsymbol{z}, \boldsymbol{f}} \quad G(\boldsymbol{z}, \boldsymbol{f}) \triangleq L(\boldsymbol{z}, \boldsymbol{f}) + \lambda \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{f})\boldsymbol{z}\|_2^2 \tag{5}$$

To gain some insights into our proposed model, denote $v(f) \triangleq w(f)/\epsilon$. We have an equivalent penalty of the weighted log function:

$$\log(w(f)|z|^2 + \epsilon) \propto \log(v(f)|z|^2 + 1) \tag{6}$$

Consider two penalty functions defined as $h_1(z) \triangleq |z|$ and $h_{2,v}(z) \triangleq \log(v|z|^2 + 1)/\log(v + 1)$, $v > 0$, while both functions satisfy $h_1(\pm 1) = h_{2,v}(\pm 1) = 1$. The plots of $h_1(z)$ and $h_{2,v}(z)$ with two different values of $v$ are shown in Fig. 1. It can be observed that $h_{2,v}(z)$ has a steeper slope than $h_1(z)$ has at the origin, which indicates that the log penalty function has the potential to be much more sparsity-encouraging than the $\ell_1$ norm. Furthermore, to $h_{2,v}(z)$, the capability in promoting sparsity increases with the value of $v$.

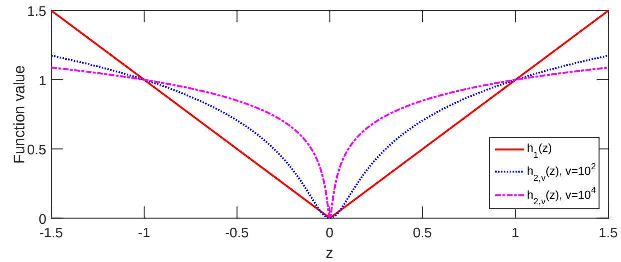


Fig. 1. The plots of three penalty functions: $h_1(z)$, $h_{2,10^2}(z)$ and $h_{2,10^4}(z)$

Guided by the insights discussed above, it is reasonable for $w(f)$ be a relatively large value when there is a small likelihood of occurrence of frequency $f$. Conversely, those frequency components with large $p(f)$ should be weighted lightly in order to express a preference of them. The rule to choose $w(f)$ is flexible. For example, for any kind of prior distribution, the inverse of Capon's power spectrum computed by $p(f)$ [10] could be used as the weighting function.

The most common knowledge available in super-resolution is the block prior, where all the frequencies are known to lie in the union of certain frequency bands. Let $f_{L_i}$ and $f_{H_i}$, $i = 1, \ldots, I$ are the lower and upper cut-off frequencies of each bands respectively. Therefore, we have the union be $\mathcal{B} \triangleq \bigcup_{i=1}^{I} \mathcal{B}_i$ where $\mathcal{B}_i \triangleq [f_{L_i}, f_{H_i}]$, $i = 1, \ldots, I$, are disjoint interval. Suppose the frequencies are uniformly distributed on $\mathcal{B}$. Then we have $p(f) = B^{-1}$ on $\mathcal{B}$ and $p(f) = 0$ on $[0, 1) \backslash \mathcal{B}$ where $B = \sum_{i=1}^{I} (f_{H_i} - f_{L_i})$. In the rest of this paper we consider only the block prior and in the next subsection we proposed an applicable weighting function for it.

### B. Weighting function for block prior

The ideal weighting function for block prior used in [8], [9] is the piecewise constant function in the domain $[0, 1)$:

$$\bar{w}_{\mathcal{B}}(f) = \begin{cases} d_1, & f \in \mathcal{B} \\ d_2, & f \notin \mathcal{B} \end{cases} \tag{7}$$

where $d_1, d_2 \in \mathbb{R}^+$ and $d_1 \ll d_2$. However, this function is non-differentiable at the boundary points of each $\mathcal{B}_i$. It would make the optimization problem (5) intractable. To circumvent this problem, a differentiable weighting function is proposed empirically to fit $\bar{w}(f)$ and exploit the block prior information, which is given by:

$$\widetilde{w}_{\mathcal{B}}(f) = d_2 \left( 1 - \sum_{i=1}^{I} e^{-\frac{1}{2}((f - \mu(\mathcal{B}_i))/\sigma(\mathcal{B}_i))^{2p}} \right) + d_1 \tag{8}$$

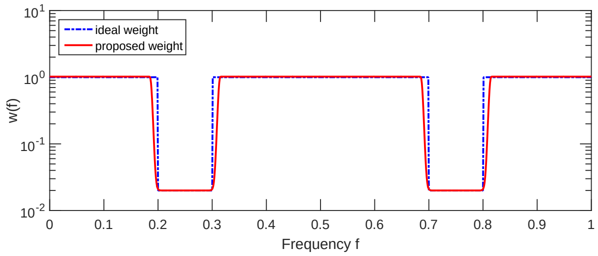


Fig. 2. The plots of weighting functions of $\bar{w}_{\mathcal{B}}(f)$ and $\widetilde{w}_{\mathcal{B}}(f)$ with $\mathcal{B} = [0.2, 0.3] \cup [0.7, 0.8]$, $p = 16$, $d_1 = 0.02$ and $d_2 = 1$.

where $p \in \mathbb{Z}^+$ is a constant. $\mu(\mathcal{B}_i) \triangleq (f_{L_i} + f_{H_i})/2$ is the midpoint of $\mathcal{B}_i$ and $\sigma(\mathcal{B}_i) \triangleq (f_{L_i} - f_{H_i})/2$ is half the width of $\mathcal{B}_i$. The plots of $\bar{w}_{\mathcal{B}}(f)$ and $\widetilde{w}_{\mathcal{B}}(f)$ are shown in Fig. 2 for $\mathcal{B} = [0.2, 0.3] \cup [0.7, 0.8]$. We can see that $\bar{w}_{\mathcal{B}}(f)$ is well fitted by $\widetilde{w}_{\mathcal{B}}(f)$ while it has nearly constant values at $\mathcal{B}$ and a sharp slope at the bound of $\mathcal{B}_i$. Furthermore, the differentiability of $\widetilde{w}_{\mathcal{B}}(f)$ can make the problem (5) tractable.

## III. Proposed Algorithm

To solve the optimization (5), we now develop an iterative reweighted algorithm for joint dictionary parameter learning and sparse signal recovery based on majorization-minimization (MM) approach [14], [15]. We first construct a differentiable and convex surrogate function majorizing $L(\boldsymbol{z}, \boldsymbol{f})$ which is given by

$$Q(\boldsymbol{z}, \boldsymbol{f}|\hat{\boldsymbol{z}}^{(t)}, \hat{\boldsymbol{f}}^{(t)}) \triangleq \sum_{n=1}^{N} \Bigg( \frac{w(f_n)|z_n|^2 + \epsilon}{w(\hat{f}_n^{(t)})|\hat{z}_n^{(t)}|^2 + \epsilon} - 1 + \log(w(\hat{f}_n^{(t)})|\hat{z}_n^{(t)}|^2 + \epsilon) \Bigg) \quad (9)$$

where $\hat{\boldsymbol{z}}^{(t)} \triangleq [\hat{z}_1^{(t)} \ \ldots \ \hat{z}_N^{(t)}]^T$, $\hat{\boldsymbol{f}}^{(t)} \triangleq [\hat{f}_1^{(t)} \ \ldots \ \hat{f}_N^{(t)}]^T$ denotes an estimate of $\boldsymbol{z}$ and $\boldsymbol{f}$ at iteration $t$. The inequality $Q(\boldsymbol{z}, \boldsymbol{f}|\hat{\boldsymbol{z}}^{(t)}, \hat{\boldsymbol{f}}^{(t)}) - L(\boldsymbol{z}, \boldsymbol{f}) \geq 0$ could be easily verified and the equality hold when $\boldsymbol{z} = \hat{\boldsymbol{z}}^{(t)}$, $\boldsymbol{f} = \hat{\boldsymbol{f}}^{(t)}$. Consequently the surrogate function for $G(\boldsymbol{z}, \boldsymbol{\theta})$ is

$$S(\boldsymbol{z}, \boldsymbol{f}|\hat{\boldsymbol{z}}^{(t)}, \hat{\boldsymbol{f}}^{(t)}) \triangleq Q(\boldsymbol{z}, \boldsymbol{f}|\hat{\boldsymbol{z}}^{(t)}, \hat{\boldsymbol{f}}^{(t)}) + \lambda \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{f})\boldsymbol{z}\|_2^2 \quad (10)$$

We now minimize the surrogate function iteratively to solve (5). With the terms independent of $\{\boldsymbol{z}, \boldsymbol{f}\}$ ignored, the minimization of the surrogate function (10) is simplified as

$$\min_{\boldsymbol{z}, \boldsymbol{f}} \quad \boldsymbol{z}^H \boldsymbol{D}^{(t)}(\boldsymbol{f})\boldsymbol{z} + \lambda \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{z}\|_2^2 \quad (11)$$

where $[\cdot]^H$ denotes the conjugate transpose, and $\boldsymbol{D}^{(t)}(\boldsymbol{f})$ is a diagonal matrix given as

$$\boldsymbol{D}^{(t)}(\boldsymbol{f}) \triangleq \operatorname{diag}\left\{ \frac{w(f_1)}{w(f_1^{(t)})|\hat{z}_1^{(t)}|^2 + \epsilon}, \ldots, \frac{w(f_N)}{w(f_N^{(t)})|\hat{z}_N^{(t)}|^2 + \epsilon} \right\}$$

Conditioned on $\boldsymbol{f}$, the optimal $\boldsymbol{z}$ of (11) can be readily obtained as

$$\boldsymbol{z}^*(\boldsymbol{f}) = \boldsymbol{C}^{-1}(\boldsymbol{f})\boldsymbol{A}^H(\boldsymbol{f})\boldsymbol{y} \quad (12)$$

where $\boldsymbol{C}(\boldsymbol{f}) \triangleq \boldsymbol{A}^H(\boldsymbol{f})\boldsymbol{A}(\boldsymbol{f}) + \lambda^{-1}\boldsymbol{D}^{(t)}(\boldsymbol{f})$. Substituting (12) back into (11), then the optimization becomes searching for the unknown parameter $\boldsymbol{f}$:

$$\min_{\boldsymbol{f}} \quad g(\boldsymbol{f}) \triangleq -\boldsymbol{y}^H \boldsymbol{A}(\boldsymbol{f})\boldsymbol{C}^{-1}(\boldsymbol{f})\boldsymbol{A}^H(\boldsymbol{f})\boldsymbol{y} \quad (13)$$

It is difficult to obtain an analytical solution of the above optimization (13). In our algorithm, however, we only need to search for a new estimate $\hat{\boldsymbol{f}}^{(t+1)}$ satisfying the following inequality

$$g(\hat{\boldsymbol{f}}^{(t+1)}) \leq g(\hat{\boldsymbol{f}}^{(t)}) \quad (14)$$

Since $g(\boldsymbol{f})$ is differentiable for our case, such an estimate can be easily obtained by using a gradient descent method. Given $\hat{\boldsymbol{f}}^{(t+1)}$, $\hat{\boldsymbol{z}}^{(t+1)}$ can be obtained via (12), with $\boldsymbol{f}$ replaced by $\hat{\boldsymbol{f}}^{(t+1)}$, i.e.

$$\hat{\boldsymbol{z}}^{(t+1)} = \boldsymbol{z}^*(\hat{\boldsymbol{f}}^{(t+1)}) \quad (15)$$

The new estimate $\{\hat{\boldsymbol{z}}^{(t+1)}, \hat{\boldsymbol{f}}^{(t+1)}\}$ could result in a non-increasing objective function value, that is,

$$G(\hat{\boldsymbol{z}}^{(t+1)}, \hat{\boldsymbol{f}}^{(t+1)}) \leq G(\hat{\boldsymbol{z}}^{(t)}, \hat{\boldsymbol{f}}^{(t)}) \quad (16)$$

The proof of (16) is the same as that in [7] and we omit it due to space restrictions.

$\lambda$ is the regularization parameter to control the tradeoff between the data fitting error and the sparsity of the solution. We have proposed a strategy to update $\lambda$ in [7] that $\lambda^{(t)}$ can be updated as

$$\lambda^{(t)} = \frac{M}{\|\boldsymbol{y} - \boldsymbol{A}(\hat{\boldsymbol{f}}^{(t)})\hat{\boldsymbol{z}}^{(t)}\|_2^2} \quad (17)$$

The iterative update of $\lambda$ can also be seamlessly integrated into our algorithm here. We summarize our algorithm as follows.

**Prior-Knowledge Aided Iterative Reweighted Algorithm**

1. Given an initialization $\hat{\boldsymbol{z}}^{(0)}, \hat{\boldsymbol{f}}^{(0)}$, and $\lambda^{(0)}$.
2. At iteration $t = 0, 1, \ldots$: Based on $\hat{\boldsymbol{z}}^{(t)}$, $\hat{\boldsymbol{f}}^{(t)}$ and $\lambda^{(t)}$, construct the surrogate function as depicted in (10). Search for a new estimate of the frequency vector $\hat{\boldsymbol{f}}^{(t+1)}$ through gradient descent method while keeping the inequality (14) satisfied. Compute a new estimate of the sparse signal, denoted as $\hat{\boldsymbol{z}}^{(t+1)}$, via (15). Compute a new regularization parameter $\lambda^{(t+1)}$ according to (17).
3. Go to Step 2 if $\|\hat{\boldsymbol{z}}^{(t+1)} - \hat{\boldsymbol{z}}^{(t)}\|_2 > \nu$, where $\nu$ is a prescribed tolerance value; otherwise stop.

## IV. Simulation Results

We now carry out experiments to illustrate the performance of the proposed prior-knowledge aided super-resolution iterative reweighted algorithm (denoted as KA-SURE-IR). The recovery performance is evaluated by success rate which is computed as the ratio of the number of successful trials to

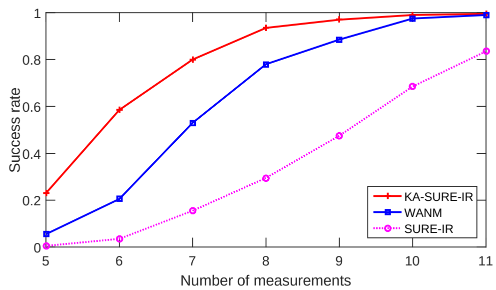


Fig. 3. Success rates of respective algorithms vs. $M$ (without noise), $N = 64$, $K = 3$.

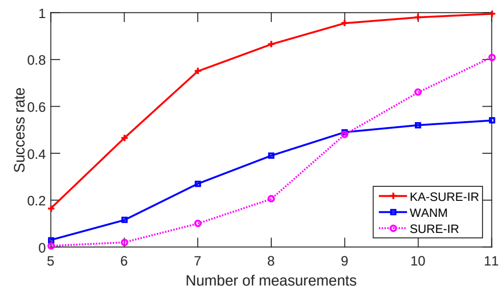


Fig. 4. Success rates of respective algorithms vs. $M$, $N = 64$, $K = 3$, and PSNR $= 25$dB.

the total amount of experiments. A single trial is considered successful if the error between the frequency estimate $\{\hat{f}_k\}$ and the groundtruth $\{f_k^\star\}$ is smaller than $10^{-3}$, i.e. $\|\boldsymbol{f}^\star - \hat{\boldsymbol{f}}\|_2 \leq 10^{-3}$. The standard super-resolution iterative reweighted algorithm (SURE-IR) [7] without exploiting any prior information is compared in the experiments. We also compare our proposed algorithm with weighted atomic norm minimization (WANM) [9] which is another prior knowledge aided method. Here SDPT3 [16] is used to solve WANM.

In the first simulation the respective methods are examined in the noiseless scenario. The observed signal is obtained via (2) with the amplitudes $\{z_k^\star\}$ uniformly distributed on the unit circle and the frequencies $\{f_k^\star\}$ uniformly generated over $\mathcal{B}$ where $\mathcal{B} = [0.2, 0.3] \cup [0.7, 0.8]$. Get $\widetilde{w}_{\mathcal{B}}(f)$ from $\mathcal{B}$ according to (8) where $p = 16$, $d_1 = 0.02$ and $d_2 = 1$. We do not control the minimum separation of the frequencies. In Fig. 3 we plot the success rates of these algorithms as a function of the number of measurements $M$ with $N = 64$ and $K = 3$. Each points are computed by $10^3$ independent runs. Compared with the performance of SURE-IR, we see that the proposed method significantly improves the success rate by exploiting prior knowledge. While our algorithm also outperforms WANM which is guaranteed to attain the global optimum. This is because the log-sum penalty is more sparsity-encouraging than the atomic norm which is the continuous counterpart of $\ell_1$-norm in the discrete domain.

In the second experiment we test the anti-noise capability of each method. The observed signal is corrupted by independent and identically distributed (i.i.d.) zero-mean complex Gaussian noise $\boldsymbol{\varepsilon}$. The effect of noise is evaluated by the peak-signal-to-noise ratio (PSNR) which is defined as PSNR $\triangleq 10\log_{10}(1/\sigma^2)$, where $\sigma^2$ is the noise variance. We set PSNR $= 25$dB and the other setup are the same as those in the first simulation. The performance is presented in Fig. 4. It can be observed that our proposed method is superior to other methods. Furthermore, the success rates of SURE-IR and KA-SURE-IR decrease slightly compared with the noiseless case while WANM suffers from a serious performance degradation.

At last we study the ability of our algorithm in resolving closely-spaced frequency components. The signal $\boldsymbol{y}$ is a mixture of three complex sinusoids with their frequencies given by $\{f_0 - \mu/N, f_0, f_0 + \mu/N\}$ where $f_0$ is randomly generated over

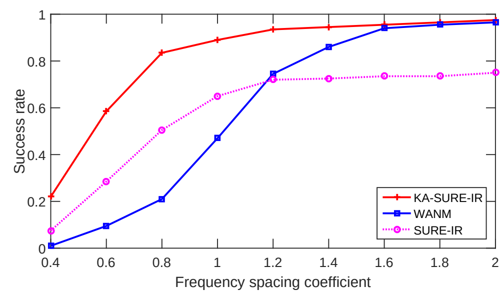


Fig. 5. Success rates of respective algorithms vs. the frequency spacing coefficient $\mu$, $N = 64$, $M = 10$, $K = 3$.

TABLE I

RUN TIMES OF RESPECTIVE ALGORITHMS

| Algorithm | KA-SURE-IR | WANM | SURE-IR |
|---|---|---|---|
| Running Time (sec) | 0.71 | 25.67 | 1.93 |
| Success rate | 0.89 | 0.47 | 0.65 |

$\mathcal{B}$ and $\mu$ is the frequency spacing coefficient ranging from 0.4 to 2. Fig. 5 depicts the success rates of respective algorithms vs. the frequency spacing coefficient $\mu$ with $N = 64$ and $M = 10$. Each point is obtained based on $10^3$ Monte Carlo runs. When the frequency components are very close to each other, (say, $\mu \leq 1$), the proposed method still have a good performance whereas WANM can hardly identify the true frequency parameters and its performance is even worse than SURE-IR which is developed without prior information. This is possibly because the atomic norm based methods have limited capability in distinguishing closely located complex sinusoids. The average running times of respective algorithms are also provided (Table I), where $\mu$ is set to 1.

## V. CONCLUSIONS

In this paper we proposed an algorithm for super-resolution line spectral estimation when the prior knowledge of frequency $f$ is available. The weighting function $w(f)$ depending on the frequency distribution $p(f)$ is introduced and the prior information can be exploited through minimizing the corresponding weighted log-sum penalty function. Simulation results show that the proposed algorithm outperforms other methods both in noiseless and noisy case, and it also has a good performance in resolving closely-spaced frequency components.

## REFERENCES


[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

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

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

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

[5] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Processing*, vol. 61, no. 1, pp. 38–42, Jan. 2013.

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

[7] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Processing*, vol. 64, no. 18, pp. 4649–4662, Sept. 2016.

[8] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Super-resolution line spectrum estimation with block priors," in *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2014, pp. 1211–1215.

[9] ——, "Spectral super-resolution with prior knowledge," *IEEE Trans. Signal Processing*, vol. 63, no. 20, pp. 5342–5357, Oct. 2015.

[10] Z. Yang and L. Xie, "A weighted atomic norm approach to spectral super-resolution with probabilistic priors," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings*, Shanghai, China, 2016, pp. 4598–4602.

[11] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted $\ell_1$ minimization for sparse recovery with prior information," in *IEEE International Symposium on Information Theory*, Seoul, Korea, 2009, pp. 483–487.

[12] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4595–4607, Sept. 2010.

[13] Y. Shen, J. Fang, and H. Li, "Exact reconstruction analysis of log-sum minimization for compressed sensing," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1223–1226, Dec. 2013.

[14] K. Lange, D. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, Mar. 2000.

[15] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, Dec. 2008.

[16] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming*, vol. 95, no. 2, pp. 189–217, 2003.