A WEIGHTED ATOMIC NORM APPROACH TO SPECTRAL SUPER-RESOLUTION WITH PROBABILISTIC PRIORS

Zai Yang[*†], *Member, IEEE, and Lihua Xie*[*], *Fellow, IEEE*

[*]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
[†]School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China

ABSTRACT

This paper concerns the line spectral estimation problem within the recent super-resolution framework. The frequencies of interest are assumed to follow a prior probability distribution. To effectively and efficiently exploit the prior information, we devise a weighted atomic norm approach that is physically sound and can be formulated as convex programming like the standard atomic norm method. Numerical simulations are provided to demonstrate the superior performance of the proposed approach in accuracy and speed compared to the state-of-the-art.

Index Terms— Spectral super-resolution, compressed sensing, weighted atomic norm, probabilistic prior.

1. INTRODUCTION

This paper concerns the line spectral estimation problem that has wide applications in communications, radar, sonar, seismology, astronomy and so on. In particular, let $f_k \in [0,1]$, $k = 1,\ldots,K$ be the unknown frequencies of interest (K is unknown as well). We have the data model:

$$\boldsymbol{y}^o = \sum_{k=1}^{K} \boldsymbol{a}\left(f_k\right) s_k \tag{1}$$

and observe part of the entries of $\boldsymbol{y}^o \in \mathbb{C}^N$ that form a subvector $\boldsymbol{y}^o_{\boldsymbol{\Omega}} \in \mathbb{C}^M$, where $\boldsymbol{\Omega} \subset \{1,\ldots,N\}$ denotes the index set with $M = |\boldsymbol{\Omega}| \leq N$. In (1), $\boldsymbol{a}\left(f\right) = \left[1, e^{i2\pi f}, \ldots, e^{i2\pi(N-1)f}\right]^T \in \mathbb{C}^N$ denotes a complex sinusoid with frequency f, and $s_k \in \mathbb{C}$ is the amplitude of the kth sinusoid. Given $\boldsymbol{y}^o_{\boldsymbol{\Omega}}$ we want to recover the frequencies f_k, $k = 1,\ldots,K$. This is known as the compressive data case since only part of the full data $\boldsymbol{y}^o$ is observed.

The study of line spectral estimation has a long history. Well known conventional methods include beamforming, Capon's beamforming and subspace methods such as MUSIC and ESPRIT. Readers are referred to [1] for a complete review. Since Capon's beamforming and subspace methods require a data covariance estimate, their application is challenging in the compressive data case in which this estimate is difficult to obtain.

With the development of sparse representation and compressed sensing, sparse methods have been successfully applied in line spectral estimation in the past decade by using the fact that the number of sinusoids K is small. In early sparse methods, however, the continuous frequency domain has to be gridded/discretized, resulting in several drawbacks such as grid mismatch and weak theoretical guarantees [2, 3]. These drawbacks have recently been resolved by using *gridless* sparse methods within the recent super-resolution and continuous compressed sensing framework [4–7]. In this framework, the frequencies are dealt with directly in the continuous domain; convex optimization methods are proposed by using the atomic norm—a continuous counterpart of the ℓ_1 norm [8]; and they are guaranteed to recover the frequencies provided the frequencies are sufficiently separated.

In this paper, besides sparsity, we assume that the frequencies of interest follow a prior probability distribution that can be obtained in practical applications, e.g., based on historical data or certain physical constraints (note that K is still unknown). The use of prior distribution for improving performance is common in the literature on statistical estimation and is typically accomplished by using statistical inference that however needs the value of K and often requires nonconvex optimization. Different from the existing methods, we solve this frequency recovery problem by using deterministic convex optimization within the super-resolution framework. In particular, we devise a *weighted* atomic norm approach in which the prior information is used to define a weighting function that is further used to penalize the frequencies. We cast the weighted approach as convex programming and solve it using off-the-shelf solvers.

Before proceeding to the main context, we recall some related works. The paper [9] assumes that the number of frequencies is known and each follows a prior distribution. The maximum a posterior (MAP) estimator was derived by using nonconvex optimization that suffers from local convergence. Weighted optimization is common in the literature on compressed sensing (in the discrete setting) for exploiting prior support information of the sparse signal. A subset of the sup-

The research of the project was supported by Ministry of Education, Republic of Singapore, under grant AcRF TIER 1 RG78/15.

port is typically assumed known [10]. But results are rare on probabilistic priors. Related papers include [11, 12], in which weighted ℓ_1 norm methods were studied given the probability of each entry of the sparse signal being nonzero (note that the number of nonzeros is thus approximately known). To date, weighted atomic norm has been studied for frequency recovery by two research groups. In our previous work [13, 14], weighted atomic norm was iteratively implemented for enhancing sparsity and resolution without any prior information. In [15,16], a different weighted method was used to deal with a class of prior distributions that are piecewise constant. While the approach of this paper can also deal with such specialized prior information, numerical simulations show that it can be a magnitude faster with comparable accuracy.

2. PROPOSED SOLUTION

2.1. Problem Statement

Let $F \in [0,1]$ be a random variable that describes the frequencies $\{f_k\}$ and let the probability density function (pdf) of F be $p(f)$. The objective is to recover the frequencies $\{f_k\}$ given the data model in (1), the observed data $\boldsymbol{y}^o_{\boldsymbol{\Omega}}$ and the prior distribution $p(f)$.

2.2. Preliminaries

We first recall the atomic norm method without any prior information except sparsity. Note that $\boldsymbol{y}^o$ in (1) is a linear combination of K atoms in the set

$$\mathcal{A} = \left\{ \boldsymbol{a}\left(f,\phi\right) = \boldsymbol{a}(f)\phi : \ f \in [0,1], \phi \in \mathbb{S}^1 \right\} \tag{2}$$

that is called the set of atoms, where $\mathbb{S}^1 = \{\phi \in \mathbb{C} : \ |\phi| = 1\}$ is the unit circle. Motivated by the literature on sparse representation, we seek the sparsest solution by solving the optimization problem:

$$\min_{\boldsymbol{y}} \|\boldsymbol{y}\|_{\mathcal{A},0}, \text{ subject to } \boldsymbol{y}_{\boldsymbol{\Omega}} = \boldsymbol{y}^o_{\boldsymbol{\Omega}}, \tag{3}$$

where $\|\boldsymbol{y}\|_{\mathcal{A},0}$ denotes the atomic ℓ_0 norm and is defined as the smallest number of atoms composing $\boldsymbol{y}$ (a continuous counterpart of the ℓ_0 norm). While (3) is in fact a rank minimization problem and cannot be practically solved, we turn to its convex relaxation—the atomic norm problem [4, 5]:

$$\min_{\boldsymbol{y}} \|\boldsymbol{y}\|_{\mathcal{A}}, \text{ subject to } \boldsymbol{y}_{\boldsymbol{\Omega}} = \boldsymbol{y}^o_{\boldsymbol{\Omega}}, \tag{4}$$

where $\|\boldsymbol{y}\|_{\mathcal{A}}$ denotes the atomic norm and is defined as:

$$\|\boldsymbol{y}\|_{\mathcal{A}} = \inf_{f_k,\phi_k,c_k>0} \left\{ \sum_k c_k : \ \boldsymbol{y} = \sum_k c_k \boldsymbol{a}\left(f_k,\phi_k\right) \right\}. \tag{5}$$

It follows from [4,5] that $\|\boldsymbol{y}\|_{\mathcal{A}}$ can be formulated as semidefinite programming (SDP).

2.3. Proposed Weighted Atomic Norm Approach

To exploit the prior information, we propose a weighted atomic norm approach in this paper. In particular, we transform the prior distribution $p(f)$ into a weighting function $w(f) \geq 0$ that, instead of $p(f)$ in statistical inference methods, is used to indicate the preference of the atoms. To be more specific, $\mathcal{A}$ is modified into the set of weighted atoms:

$$\mathcal{A}^w = \left\{ w(f)\boldsymbol{a}\left(f,\phi\right) : \ f \in [0,1], \phi \in \mathbb{S}^1 \right\}. \tag{6}$$

The weighted atomic norm is the atomic norm induced by $\mathcal{A}^w$ that can be written as:

$$\|\boldsymbol{y}\|_{\mathcal{A}^w} = \inf_{f_k,\phi_k,c_k>0} \left\{ \sum_k \frac{c_k}{w(f)} : \ \boldsymbol{y} = \sum_k c_k \boldsymbol{a}\left(f_k,\phi_k\right) \right\}. \tag{7}$$

Then we solve the weighted atomic norm problem:

$$\min_{\boldsymbol{y}} \|\boldsymbol{y}\|_{\mathcal{A}^w}, \text{ subject to } \boldsymbol{y}_{\boldsymbol{\Omega}} = \boldsymbol{y}^o_{\boldsymbol{\Omega}}. \tag{8}$$

Inspired by the weighted ℓ_1 technique in [17], we attempt to let $w(f)$ be approximately the magnitude of the potential frequency component at f so that the weighted atomic norm can approximate the atomic ℓ_0 norm in (3). On the other hand, $w(f)$ must be chosen such that $\|\boldsymbol{y}\|_{\mathcal{A}^w}$ can be efficiently computed. This is the main challenge of the weighted atomic norm approach, making it substantially different from the discrete setting in compressed sensing.

We first study how to choose $w(f)$ to meet the computational criterion. Since strong duality holds [18], we consider the dual problem of (8):

$$\max_{\boldsymbol{x}} \langle \boldsymbol{x}_{\boldsymbol{\Omega}}, \boldsymbol{y}^o_{\boldsymbol{\Omega}} \rangle_{\mathbb{R}}, \text{ subject to } \|\boldsymbol{x}\|^*_{\mathcal{A}^w} \leq 1 \text{ and } \boldsymbol{x}_{\boldsymbol{\Omega}^c} = \boldsymbol{0}, \tag{9}$$

where $\langle \cdot,\cdot \rangle_{\mathbb{R}}$ denotes the inner product and $\boldsymbol{\Omega}^c$ is the complement of $\boldsymbol{\Omega}$. In (9), the dual weighted norm $\|\boldsymbol{x}\|^*_{\mathcal{A}^w}$ is defined as:

$$\|\boldsymbol{x}\|^*_{\mathcal{A}^w} = \max_{\|\boldsymbol{y}\|_{\mathcal{A}^w} \leq 1} \langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{R}} = \max_f w(f) \left| \boldsymbol{a}^H(f)\boldsymbol{x} \right|. \tag{10}$$

So it suffices to characterize the inequality

$$\left| \boldsymbol{a}^H(f)\boldsymbol{x} \right| \leq w^{-1}(f), \quad f \in [0,1]. \tag{11}$$

Motivated by [4] that deals with the standard atomic norm for which $w(f) \equiv 1$, we attempt to characterize (11) as a linear matrix inequality (LMI) by applying the bounded real lemma (BRL) for trigonometric polynomials [19]. It follows from [19, Theorem 4.24] that this can be done if $w^{-1}(f) = \left| \boldsymbol{a}^H(f)\boldsymbol{h} \right|$ for some $\boldsymbol{h} \in \mathbb{C}^N$. Further, it follows from the Riesz-Fejér theorem (see, e.g., [19, Theorem 1.1]) that the above condition is satisfied if $w^{-2}(f)$ is a positive (Hermitian) trigonometric polynomial.

Based on the arguments above, we let $w^2(f)$ be the Capon's power spectrum:

$$w^2(f) = \left[\boldsymbol{a}^H(f)\boldsymbol{C}^{-1}\boldsymbol{a}(f) \right]^{-1} \tag{12}$$

that is well known to estimate the sinusoid power as a function of the frequency f if the data covariance matrix $\boldsymbol{C}$ is given. Applying [19, Theorem 4.24], we have that (11) holds if and only if there exists $\boldsymbol{H} \in \mathbb{C}^{N\times N}$ satisfying that

$$\begin{bmatrix} 1 & \boldsymbol{x}^H \\ \boldsymbol{x} & \boldsymbol{H} \end{bmatrix} \geq \boldsymbol{0} \text{ and } \sum_{n=1}^{N-j} \boldsymbol{H}_{n,n+j} = \sum_{n=1}^{N-j} \left[\boldsymbol{C}^{-1}\right]_{n,n+j}, \tag{13}$$

$j = 0, \ldots, N-1$, where $\boldsymbol{H}_{n,n+j}$ denotes the $(n, n+j)$th entry of $\boldsymbol{H}$. So we successfully cast the dual problem (9) as SDP that generalizes the case of standard atomic norm in [4] where $\boldsymbol{C}$ is the identity matrix. A straightforward consequence of the strong duality is that, once (9) is solved, the frequencies of interest can be located at those f's satisfying that $|\boldsymbol{q}(f)| = w^{-1}(f)$, where, as in [4, 5],

$$\boldsymbol{q}(f) = \boldsymbol{a}^H(f)\boldsymbol{x} \tag{14}$$

is called the dual polynomial. It follows from similar arguments as in [4] that the frequencies can be obtained by solving the roots of the polynomial $\boldsymbol{a}^H(f)\left(\boldsymbol{C}^{-1} - \boldsymbol{x}\boldsymbol{x}^H\right)\boldsymbol{a}(f)$.

The remaining task is to compute $\boldsymbol{C}$ based on the given prior distribution $p(f)$. Note first that we need only to determine $\boldsymbol{C}$ up to a positive scaling factor. We make a common assumption that the complex amplitudes $\{s_k\}$ in (1) have random phases, leading to a common expression of $\boldsymbol{C}$ (as a function of $\left\{|s_k|^2\right\}$ and $\{f_k\}$): $\boldsymbol{C} = \sum_{k=1}^{K} |s_k|^2 \boldsymbol{a}(f_k)\boldsymbol{a}^H(f_k)$. We further assume that the sinusoid power is independent of the frequency. As a result, the expectation of $\boldsymbol{C}$ is

$$\mathbb{E}\boldsymbol{C} \propto \mathbb{E}\boldsymbol{a}(f)\boldsymbol{a}^H(f) = \int_0^1 \boldsymbol{a}(f)\boldsymbol{a}^H(f)p(f)\mathrm{d}f \tag{15}$$

that can be explicitly given or at least numerically computed and will be used as the estimate of $\boldsymbol{C}$ in the proposed approach (note that in the case when $\boldsymbol{C}$ is ill-conditioned, it can be modified to $\boldsymbol{C} + \epsilon\boldsymbol{I}$ where $\epsilon > 0$ is a small regularization parameter).

To illustrate the proposed weighted approach, we provide an example in Fig. 1. There the prior distribution is Gaussian, with mean 0.5 and variance 0.1^2, truncated on the interval $[0, 1]$. The proposed weighting function $w(f)$ is computed with $N = 64$. It is interesting to note that $w(f)$ (solid red) is very close to the squared root of the prior pdf (dashed blue).

2.4. The Perspective of Covariance Fitting

While we have cast the dual problem (9) as SDP, a standard Lagrangian analysis allows us to derive the SDP formulation of the primal problem (8):

$$\begin{aligned} &\min_{t,\boldsymbol{u}} \frac{1}{2}t + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{C}^{-1}T(\boldsymbol{u})\right), \\ &\text{subject to } \begin{bmatrix} t & \boldsymbol{y}^H \\ \boldsymbol{y} & T(\boldsymbol{u}) \end{bmatrix} \geq \boldsymbol{0} \text{ and } \boldsymbol{y}_{\boldsymbol{\Omega}} = \boldsymbol{y}^o_{\boldsymbol{\Omega}}, \end{aligned} \tag{16}$$

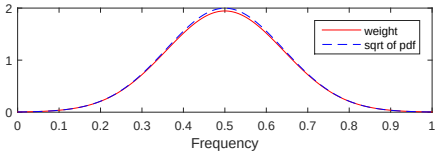

Fig. 1. Weighting function (solid red) of the proposed approach under a truncated Gaussian prior. The dashed blue line plots the squared root of the prior pdf.

where $T(\boldsymbol{u})$ denotes a (Hermitian) Toeplitz matrix the first row of which is given by the transpose of $\boldsymbol{u} \in \mathbb{C}^N$. Without surprise, (16) recovers the formulation of the weighed atomic norm that was obtained in [14] by using the Vandermonde decomposition of Toeplitz matrices (see, e.g., [1]). While the atomic norm method can be viewed as a covariance fitting method [7], it is interesting to note that the weighted atomic norm approach corresponds to weighted covariance fitting. In (16), $T(\boldsymbol{u})$ plays the role of the data covariance of interest and $\boldsymbol{C}$ is its prior estimate. Therefore, the matrix weighting strategy in (16) helps to shape the covariance estimate.

Moreover, the primal SDP formulation (16) provides another way to frequency retrieval, instead of using the dual polynomial. In particular, given $\boldsymbol{u}$ that can be obtained for free when we solve the dual problem (9), we can compute the Vandermonde decomposition of $T(\boldsymbol{u})$ that provides the solution of $\{f_k\}$. Readers are referred to [14] for the details. This method will be used in the present paper since we find that it results in higher numerical accuracy.

3. THE CASE OF BLOCK PRIORS

To compare with [16], we specialize in this section the proposed weighted approach to the case of block priors where all the frequencies are known to lie in certain frequency bands.[1] Let the union of the frequency bands be

$$\mathcal{D} = \cup_{j=1}^{J} \left[f_{L_j}, f_{H_j}\right], \tag{17}$$

where $\left[f_{L_j}, f_{H_j}\right]$, $j = 1, \ldots, J$ are disjoint intervals in $[0, 1]$. In [16], a piecewise constant weighting function with $w(f) = 1$, if $f \in \mathcal{D}$, and $w(f) = 0$, otherwise was used to implement the weighted atomic norm (or constrained atomic norm in the language of [16]). In the resulting dual problem, the inequality $\left|\boldsymbol{a}^H(f)\boldsymbol{x}\right| \leq 1$ was characterized on each frequency band by using two LMIs, leading to an SDP formulation consisting of $2J$ LMIs. We next provide a different yet more computationally efficient approach that contains only a single LMI.

To apply the proposed weighted approach, we naturally assume that the frequencies are uniformly distributed on $\mathcal{D}$. This means that $p(f) = B^{-1}$ on $\mathcal{D}$ and $p(f) = 0$ on $[0, 1]\backslash\mathcal{D}$,

[1] Though the paper [16] can deal with all piecewise constant prior distributions, it does not provide the optimal weight in its method in general.

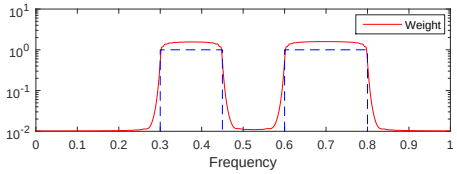

Fig. 2. Weighting function (solid red) of the proposed approach under the block prior with $\mathcal{D} = [0.3, 0.45] \cup [0.6, 0.8]$, $N = 64$ and a regularization parameter $\epsilon = 10^{-4}$.

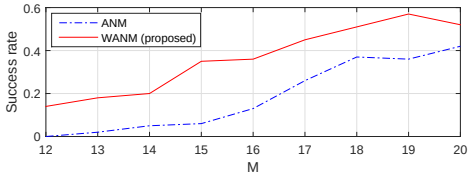

Fig. 3. Results of success rates under the truncated Gaussian prior as presented in Fig. 1.

where $B = \sum_{j=1}^{J} \left(f_{H_j} - f_{L_j}\right)$. It follows from (15) that $\mathbb{E}\boldsymbol{C} \propto T(\boldsymbol{v})$, where $v_1 = 1$ and for $n = 2, \ldots, N$, $v_n = \frac{i}{2\pi(n-1)B} \sum_{j=1}^{J} \left(e^{-i2\pi(n-1)f_{H_j}} - e^{-i2\pi(n-1)f_{L_j}}\right)$. Therefore, the weighted atomic norm approach can be implemented by using this data covariance estimate. As an example, we plot in Fig. 2 the weighting function $w(f)$ in the case of $\mathcal{D} = [0.3, 0.45] \cup [0.6, 0.8]$ and $N = 64$. It is interesting to note that it is very close to the weight used in [16] (up to a scaling factor).

4. NUMERICAL SIMULATIONS

In this section, we numerically study the performances of the proposed weighted atomic norm minimization (WANM) approach, that is implemented in Matlab and solved by using SDPT3 [20]. In particular, we set $N = 64$, $K = 5$ and vary

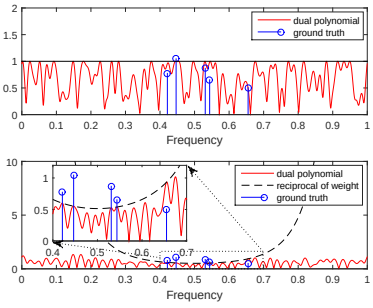

Fig. 4. Results of dual polynomials of ANM (upper) and the proposed WANM (lower) under the truncated Gaussian prior in a single run.

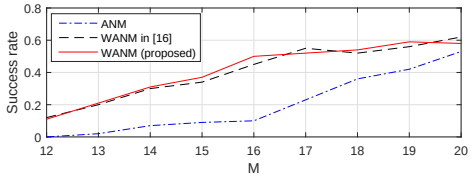

Fig. 5. Results of success rates under the block prior as presented in Fig. 2.

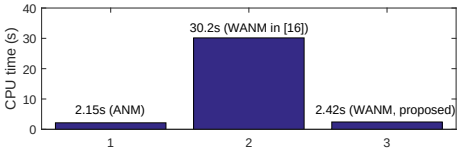

Fig. 6. Results of averaged CPU times under the block prior.

M. The frequencies are independently generated from a prior distribution $p(f)$. Note that we do not control the minimum separation of the frequencies.

In the first simulation, we consider the truncated Gaussian prior as presented in Fig. 1. While no existing methods can deal with this prior (note that [9] needs to know K), we only compare with the standard atomic norm minimization (ANM) in [5]. We plot in Fig. 3 the success rates of the two methods in recovering the frequencies (based on 100 Monte Carlo runs for each M). It is shown that the proposed approach significantly improves the successful recovery rate by exploiting the prior information. Moreover, we plot in Fig. 4 the results of a single run at $M = 16$, in which the proposed WANM (lower) succeeds while ANM (upper) fails. Note that the absolute value of the dual polynomial of WANM intersects with $w^{-1}(f)$ at the true frequencies.

In the second simulation, we consider the block prior as presented in Fig. 2 and compare the proposed WANM with ANM and the method in [16]. It is shown in Fig. 5 that the two weighted approaches ([16] and ours) have comparable frequency recovery capability and both of them outperform ANM. Moreover, it is shown in Fig. 6 that the proposed WANM is comparable with ANM in computational speed and is more than a magnitude faster than the method in [16].

5. CONCLUSION

In this paper, we proposed a weighted atomic norm approach to spectral super-resolution to exploit prior information of the frequencies. By the proposed weighting strategy, the proposed approach can effectively exploit general probabilistic priors without increasing the computational complexity. In the case of block priors it can be a magnitude faster than the state-of-the-art with comparable accuracy.

6. REFERENCES

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

[2] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[3] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.

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

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

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

[7] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.

[8] 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.

[9] D. Zachariah, P. Wirfält, M. Jansson, and S. Chatterjee, "Line spectrum estimation with probabilistic priors," *Signal Processing*, vol. 93, no. 11, pp. 2969–2974, 2013.

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

[11] M. A. Khajehnejad, W. Xu, B. Hassibi *et al.*, "Analyzing weighted minimization for sparse recovery with nonuniform sparse models," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1985–2001, 2011.

[12] J. Scarlett, J. S. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 427–439, 2013.

[13] Z. Yang and L. Xie, "Achieving high resolution for super-resolution via reweighted atomic norm minimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 3646–3650.

[14] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Transactions on Signal Processing, DOI: 10.1109/TSP.2015.2493987*, 2015.

[15] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Off-the-grid spectral compressed sensing with prior information," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1010–1014.

[16] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Spectral super-resolution with prior knowledge," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5342–5357, 2015.

[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] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, London, 2004.

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

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