Approximate Support Recovery of Atomic Line Spectral Estimation: A Tale of Resolution and Precision

Qiuwei Li and Gongguo Tang
Department of Electrical Engineering and Computer Science,
Colorado School of Mines, Golden, CO 80401 Email: qiuli@mymail.mines.edu, gtang@mines.edu

***Abstract*—This work investigates the parameter estimation performance of line spectral estimation/super-resolution using atomic norm minimization. The focus is on analyzing the algorithm's accuracy of inferring the frequencies and complex magnitudes from noisy observations. When the Signal-to-Noise Ratio is reasonably high and the true frequencies are separated by $O(\frac{1}{n})$, the atomic norm estimator is shown to localize the correct number of frequencies, each within a neighborhood of size $O(\sqrt{\frac{\log n}{n^3}}\sigma)$ of one of the true frequencies. Here n is half the number of temporal samples and σ^2 is the Gaussian noise variance. The analysis is based on a primal-dual witness construction procedure. The error bound obtained matches the Cramér-Rao lower bound up to a logarithmic factor. The relationship between resolution (separation of frequencies) and precision/accuracy of the estimator is highlighted.**

I. Introduction

Line spectral estimation, which aims at approximately inferring the frequency and coefficient parameters from a superposition of complex sinusoids embedded in white noise, is one of the fundamental problems in statistical signal processing. When the temporal and frequency domains are exchanged, this classical problem was reinterpreted as the problem of mathematical super-resolution recently [1], [2]. This new line of works also promote the use of a convex sparse regularizer to solve inverse problems involving spectrally sparse signals, distinguishing them from classical methods based on root finding and singular value decompositions (e.g., Prony's method, MUSIC, ESPIRIT, Matrix Pencil, etc.). The convex regularizer, a special instance of the general atomic norms, has been shown to achieve optimal performance in signal completion [3], denoising [4], and outlier removal [5], [6]. However, the most relevant question of noisy parameter estimation performance has been elusive. We address the question by deriving conditions under which the atomic norm formulation will return the correct number of frequencies, and establishing bounds on the estimation errors. These bounds match the one given by the Cramér-Rao bound (CRB), the best possible for any unbiased estimator, thus establishing the optimality of atomic norm minimization in performing spectral parameter estimation. We remark that our results hold for finite-length, single-snapshot signals, while classical methods such as MUSIC are inefficient (i.e., not approaching CRB) even with infinite number of snapshots, as long as the signal length is finite.

Q. Li and G. Tang were supported by the NSF Grant CCF-1464205.

II. Signal Model and Problem Setup

This paper considers the spectral estimation problem: given noisy temporal samples, how well can we estimate the locations and determine the amplitudes of spectral lines? The signal of interest $x^\star(t)$ is composed of only a small number of spectral spikes located in a normalized interval $\mathbb{T} = [0, 1]$, that is,

$$x_t^\star \quad = \quad \sum_{\ell=1}^{k} c_\ell^\star \exp(i2\pi f_\ell^\star t) \tag{1}$$

where the set of frequencies $T^\star = \{f_\ell^\star\}_{\ell=1}^k \subset \mathbb{T}$ and the complex amplitudes $\{c_\ell^\star\}_{\ell=1}^k$ are unknown. In this paper, we abuse notation and call $T^\star$ the support of $\mathbf{x}^\star$. The goal is to approximately localize these parameters from a small number of equispaced noisy samples

$$y_t \quad = \quad x_t^\star + w_t, \ t = -n, \dots, n.$$

For technical simplicity, we sample the signal symetrically and assume that $n = 2M$ is an even number. The noise components w_t are i.i.d. centrally symmetric complex Gaussian variables with variance σ^2. We stack the temporal samples into column vectors and write the observation model as

$$\mathbf{y} \quad = \quad \mathbf{x}^\star + \mathbf{w} \tag{2}$$

where $\mathbf{x}^\star := [x_{-n}^\star, \dots, x_n^\star]^T$, $\mathbf{y} := [y_{-n}, \dots, y_n]^T$ and $\mathbf{w}^\star := [w_{-n}^\star, \dots, w_n^\star]^T$.

To estimate the frequency vector $\mathbf{f}^\star := [f_1^\star, \dots, f_k^\star]^T$ and the complex coefficient vector $\mathbf{c}^\star := [c_1^\star, \dots, c_k^\star]^T$, we assume k is small and treat $\mathbf{x}^\star$ as a sparse combination of atoms $\mathbf{a}(f) := [e^{i2\pi(-n)f}, \dots, e^{i2\pi(n)f}]^T$ parameterized by the frequency $f \in \mathbb{T}$, that is,

$$\mathbf{x}^\star = \sum_{\ell=1}^{k} c_\ell^\star \mathbf{a}(f_\ell^\star). \tag{3}$$

To exploit the structure of $\mathbf{x}^\star$ encoded in the set of atoms $\mathcal{A} := \{\mathbf{a}(f), f \in \mathbb{T}\}$, we follow [7], [3] and define the associated atomic norm as

$$\begin{aligned} \|\mathbf{x}\|_{\mathcal{A}} \quad &= \quad \inf\{t : \mathbf{x} \in t\,\mathrm{conv}(\mathcal{A})\} \\ &= \quad \inf\left\{\sum_{\ell} |c_l| : \mathbf{x} = \sum_{\ell=1} c_\ell \mathbf{a}(f_\ell), \forall f_\ell \in \mathbb{T}\right\}. \end{aligned} \tag{4}$$

The dual norm of the atomic norm, which is useful both algorithmically and theoretically, is defined for a vector $\mathbf{z}$ as $\|\mathbf{z}\|_{\mathcal{A}}^* = \sup_{f\in\mathbb{T}} |\langle \mathbf{a}(f), \mathbf{z}\rangle|$.

Given the noisy observation model (2), it is natural to denoise $\mathbf{x}^\star$ by solving an atomic norm regularized program [8], [4]:

$$\hat{\mathbf{x}}^{glob} = \underset{\mathbf{x}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_Z^2 + \lambda\|\mathbf{x}\|_{\mathcal{A}}. \tag{5}$$

For technical reasons, we used a weighted ℓ_2 norm, $\|\mathbf{z}\|_Z := \sqrt{\mathbf{z}^H Z \mathbf{z}}$, to measure data fidelity. Here $Z = \operatorname{diag}(\frac{g_M(\ell)}{M}) \in \mathbb{R}^{4M+1}$ with $g_M(\ell), \ell = -2M, \ldots, 2M$ defined in [3] as the discrete convolution of two triangular functions. When we exchange the frequency and temporal domains, this weighting scheme trusts low-frequency samples more than high-frequency ones, even though the noise levels are the same.

The second term is a regularization term which penalizes solutions with large atomic norms. Such solutions typically correspond to spectrally dense signals. The regularization parameter λ, whose value will be given later, controls the trade-off between data fidelity and spectral sparsity.

Once $\hat{\mathbf{x}}^{glob}$ was solved, we can extract estimates of the frequencies either from the primal optimal solution $\hat{\mathbf{x}}^{glob}$ or from the dual optimal solution. Our goal is to characterize conditions such that i) we obtain exactly k estimated frequencies; ii) there is a natural correspondence between the estimated frequencies and the true frequencies, whose distances can be explicitly controlled; iii) the distances between the corresponding coefficients can also be explicitly bounded.

III. Main Theorem

It is known that there is a resolution limit of the atomic norm approach in resolving the atoms, or the frequency parameter $\mathbf{f}^\star$, even from the noiseless data [9]. Therefore, to recover the support of the line spectral signal $\mathbf{x}^\star$, we need to impose certain separation condition on the distances of the true frequencies. For this purpose, we define $\Delta(T) = \min_{(f_\ell, f_m)\subset T: f_\ell \neq f_m} |f_\ell - f_m|$, where $|\cdot|$ is understood as the wrap-around distance. To present our main result, we need several other quantities: 1) the dynamic range of the coefficients $B^\star := \frac{c^\star_{\max}}{c^\star_{\min}}$, where $c^\star_{\max}$ and $c^\star_{\min}$ denote the maximal and minimal modules of $\{c^\star_\ell\}_{\ell=1}^k$; 2) the normalized noise level $\gamma_0 := \sigma\sqrt{\frac{\log n}{n}}$; and 3) the Noise-to-Signal Ratio $\gamma := \gamma_0/c^\star_{\min}$. Now we are ready to present our main result.

Theorem 3.1: Suppose we observe $2n+1$ noisy consecutive samples $y_\ell = x^\star_\ell + w_\ell$ of the signal (1) or (3) with w_ℓ being i.i.d. complex Gaussians of mean zero and variance σ^2. Set the regularization parameter $\lambda = 0.658X^\star\gamma_0$ in (5). For some large $\eta \in [1,\infty)$, we assume $n \geq 128$ and

$$\Delta(T^\star) \geq 2.5009/n \tag{6}$$

$$X^\star B^\star \gamma \leq 0.001 \text{ and } \eta B^\star/X^\star \leq 0.001 \tag{7}$$

Then with high probability, the optimal solution of (5) has a decomposition $\hat{\mathbf{x}}^{glob} = \sum_{\ell=1}^k \hat{c}^{glob}_\ell \mathbf{a}(\hat{f}^{glob}_\ell)$ involving exactly k atoms, whose frequencies and coefficients, when properly ordered, satisfy

$$\max_{1\leq\ell\leq k} |c^\star_\ell||\hat{f}^{glob}_\ell - f^\star_\ell| \leq 0.4(X^\star + 14\eta)\gamma_0/n, \tag{8}$$

$$\max_{1\leq\ell\leq k} |\hat{c}^{glob}_\ell - c^\star_\ell| \leq (X^\star + 14\eta)\gamma_0. \tag{9}$$

Several remarks on the conditions follow. Because of the weighting scheme we use in (5), our choice of λ differs from the standard one in [10] by a factor $1/n$ and ensures that the weighted dual atomic norm of the noise, $\|Z\mathbf{w}\|_{\mathcal{A}}^*$, is less than λ with high probability. For technical reasons, our separation condition (6) is stronger than the one in previous work [1], [3], [4]. It is possible to use the standard separation condition at the expense of strengthened conditions (7). The conditions (7) wrap several requirements on the problem parameters for the conclusions to hold: the dynamic range of the coefficients $B^\star$, the Noise-to-Signal Ratio γ, and the normalized noise γ_0 should all be small while the regularization parameter λ should be large enough as measured by $X^\star$.

The error bounds (8) and (9) warrant more discussions. The error bound on estimating the coefficients that is proportional to λ, is expected as a consequence of soft thresholding. The error bound (8) on frequencies can be rewritten as $O(\frac{\sqrt{\log n}}{n^{3/2}}\sigma)$, which matches the CRB (for estimating a single frequency or in the large n and large number of snapshots scenario) $O(\frac{1}{n^{3/2}}\sigma)$ up to a logarithmic factor. We also note that the MUSIC and the Maximum Likelihood Estimator (MLE) have asymptotic standard deviation scales as $O(\frac{\sigma^2}{T|c|^2n^3} + \frac{\sigma^4}{T|c|^4n^4})$, which is worse than the CRB. Here T represents the number of snapshots.

IV. Proof Outline: Primal-Dual Witness Construction

Duality plays an important role in understanding atomic norm regularized line spectral estimation. Standard Lagrangian analysis shows that the dual problem of (5) reads:

$$\hat{\mathbf{q}}^{glob} = \underset{\mathbf{q}}{\operatorname{argmax}} \ \frac{1}{2}\|\mathbf{y}\|_Z^2 - \frac{1}{2}\|\mathbf{y}-\lambda\mathbf{q}\|_Z^2$$
$$\text{subject to } \ \|Z\mathbf{q}\|_{\mathcal{A}}^* \leq 1. \tag{10}$$

The polynomial $Q(f) := \mathbf{a}(f)^H Z\mathbf{q}$ corresponding to a dual feasible solution $\mathbf{q}$ is called a dual polynomial. The dual polynomial associated with the unique dual optimal solution $\hat{Q}^{glob}(f) := \mathbf{a}(f)^H Z\hat{\mathbf{q}}^{glob}$ certifies the optimality of the unique primal solution $\hat{\mathbf{x}}^{glob}$, and vice versa. The uniqueness comes from the strongly convex quadratic term in (5). In particular, the primal-dual optimal solutions are related by $\hat{\mathbf{q}}^{glob} = (\mathbf{y} - \hat{\mathbf{x}}^{glob})/\lambda$. We summarize this in the follow proposition:

Proposition 4.1: Let $\hat{\mathbf{x}} = \sum_{l=1}^{\hat{k}} \hat{c}_l \mathbf{a}(\hat{f}_l)$ with distinct $\hat{f}_\ell$ and $\hat{\mathbf{q}} = \frac{\mathbf{y}-\hat{\mathbf{x}}}{\lambda}$. Suppose the corresponding dual polynomial $\hat{Q}(f) = \hat{\mathbf{q}}^H Z\mathbf{a}(f)$ satisfies the following *Bounded Interpolation Property (BIP)*:

$$\hat{Q}(\hat{f}_l) = \operatorname{sign}(\hat{c}_l), l = 1, \ldots, \hat{k} \ \text{ (Interpolation)};$$
$$|\hat{Q}(f)| < 1, f \in \mathbb{T}/\{\hat{f}_1, \ldots, \hat{f}_{\hat{k}}\} \ \text{ (Boundedness)};$$

then $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$ are the unique primal-dual optimal solutions to (5) and (10), that is, $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{glob}, \hat{\mathbf{q}} = \hat{\mathbf{q}}^{glob}$.

The uniqueness of the dual solution for (5) makes the construction of a dual certificate much harder compared with the line spectral signal completion problem [3] and demixing problem [5], [6]. For the later two problems, while the primal optimal solution is unique, the dual optimal solutions are non-unique. One usually choose one dual solution that is easier to analyze (e.g., the one with minimal energy). For the support recovery problem, we need to simultaneously construct the primal and dual solutions, which witness the optimality of each other. In the compressive sensing literature, this construction process is called the *primal-dual witness construction* [11]. For sparse recovery problems, a candidate primal solution is relatively easy to find, since when the noise is relatively small, the support of the recovered signal would not change. So one only needs to solve a LASSO problem restricted to the true support to determine the candidate coefficients, as was done in [11]. For the optimization (5), due to the continuous nature of the atoms, even a bit of noise would drive the support away from the true one. So to construct a candidate primal solution (hence a dual candidate solution), we need to simultaneously seek for the candidate support $\{\hat{f}_\ell\}$ and the candidate coefficients $\{\hat{c}_\ell\}$.

We solve the following nonlinear, nonconvex program to find plausible candidates for $\{\hat{f}_\ell\}$ and $\{\hat{c}_\ell\}$

$$\underset{\mathbf{f},\mathbf{c}}{\text{minimize}}\, \frac{1}{2}\|A(\mathbf{f})\mathbf{c}-\mathbf{y}\|_Z^2+\lambda\|\mathbf{c}\|_1 \tag{11}$$

where $\mathbf{f}:=[f_1,\ldots,f_k]^T$, $\mathbf{c}:=[c_1,\ldots,c_k]^T$ and $A(\mathbf{f}):=[\mathbf{a}(f_1),\ldots,\mathbf{a}(f_k)]$. Again, unlike in compressive sensing, we cannot fix $\mathbf{f}=\mathbf{f}^\star$ to solve for $\mathbf{c}$ only as in [11]. The program (11) is high nonconvex, and solving it to global optimality is hard even in theory. We are primarily interested in its local minimum $(\{\hat{f}_\ell\},\{\hat{c}_\ell\})$ in a neighborhood of the true frequencies and coefficients $(\mathbf{f}^\star,\mathbf{c}^\star)$. To find this local minimum, we run gradient descent to (11) using $(\mathbf{f}^\star,\mathbf{c}^\star)$ as initialization. We argue that under conditions presented in Theorem 3.1, each $\hat{f}_\ell$ and $\hat{c}_\ell$ stay close to $f_\ell^\star$ and $c_\ell^\star$ as given in (8) and (9), respectively. The major tool we use is the contraction mapping theorem.

The rest of arguments would consist of showing that $\hat{\mathbf{x}}=\sum_{l=1}^{\hat{k}}\hat{c}_\ell\mathbf{a}(\hat{f}_\ell)$ with $\{\hat{f}_\ell\}$ and $\{\hat{c}_\ell\}$ constructed as described above satisfy the Bounded Interpolation Property of Proposition 4.1. The Interpolation property is automatically satisfied due to the construction process and the main challenge is to show the boundedness property $|\hat{Q}(f)|\le 1$, or dual feasibility of $\hat{\mathbf{q}}$. It turns out that directly showing dual feasibility is difficult. So we modify the construction process to a two-step procedure. We first find local minimum of $(\mathbf{f}^\lambda,\mathbf{c}^\lambda)$ of $\frac{1}{2}\|A(\mathbf{f})\mathbf{c}-\mathbf{x}^\star\|_Z^2+\lambda\|\mathbf{c}\|_1$ around $(\mathbf{f}^\star,\mathbf{c}^\star)$, where one should note we replaced $\mathbf{y}$ in (11) with $\mathbf{x}^\star$. We then run gradient descent to (11) using $(\mathbf{f}^\lambda,\mathbf{c}^\lambda)$ as initialization. The intermediate quantities $(\mathbf{f}^\lambda,\mathbf{c}^\lambda)$ will serve as a bridge between $(\mathbf{f}^\star,\mathbf{c}^\star)$ and $(\hat{\mathbf{f}},\hat{\mathbf{c}})$ to make the proof easier. The key observations are that $\hat{Q}(f)=\mathbf{a}(f)^HZ\hat{\mathbf{q}}$ is close to $Q^\lambda(f)=\mathbf{a}(f)^HZ\mathbf{q}^\lambda$, where $\mathbf{q}^\lambda=(\mathbf{x}^\star-\mathbf{x}^\lambda)/\lambda$ and $\mathbf{x}^\lambda=\sum_{\ell=1}^k c_\ell^\lambda\mathbf{a}(f_\ell^\lambda)$, and $Q^\lambda(f)$ is close to $Q^\star(f)=\mathbf{a}(f)^HZ\mathbf{q}^\star$. Here $\mathbf{q}^\star$ is a dual certificate used to certify the atomic decomposition of $\mathbf{x}^\star$. The former claim can be showed using the closeness of $(\mathbf{f}^\lambda,\mathbf{c}^\lambda)$ and $(\hat{\mathbf{f}},\hat{\mathbf{c}})$. The later claim, however, must take advantage of the fact that $\mathbf{q}^\star=\lim_{\lambda\to 0}\mathbf{q}^\lambda=-\frac{\partial}{\partial\lambda}\mathbf{x}^\lambda|_{\lambda=0}$ and applying triangle inequality to

$$Q^\lambda(f)-Q^\star(f)=\frac{1}{\lambda}\int_0^\lambda \mathbf{a}(f)^HZ\left(\frac{\partial}{\partial t}\mathbf{x}^0-\frac{\partial}{\partial t}\mathbf{x}^t\right)dt.$$

The closeness of $(\mathbf{f}^\lambda,\mathbf{c}^\lambda)$ and $(\mathbf{f}^\star,\mathbf{c}^\star)$ ensures that the partial derivatives are also close. Finally, we exploit the nice properties of $Q^\star(f)$ which are similar to those established in [1] to complete the proof. Once we proved Theorem 3.1 using this argument and showed $\hat{\mathbf{x}}^{glob}=\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}^{glob}=\hat{\mathbf{q}}$, we have also showed that $(\hat{\mathbf{f}},\hat{\mathbf{c}})$ obtained in the proof using local gradient descent is the global optimum of (11). Therefore, we can view the atomic regularized least-squares problem (5) as a convex substitute for the nonconvex ℓ_1 regularized least-squares problem (11) – both programs share the same global optimum under the conditions of Theorem 3.1.

V. Numerical Experiments

We present numerical results to support our theoretical findings. In particular, we first examine the phase transition curve of the rate of success in Figure 1. In preparing Figure 1, the k complex coefficients $c_1^\star,\ldots,c_k^\star$ were generated uniformly from the unit complex circle such that $c_{\min}^\star=c_{\max}^\star=1$, implying $B^\star=1$. We also generated k normalized frequencies $f_1^\star,\ldots,f_k^\star$ uniformly chosen from $[0,1]$ such that every pair of frequencies are separated at least $2.5/n$. Then the signal $\mathbf{x}^\star$ was formed according to (3). We created our observation $\mathbf{y}$ by adding Gaussian noise of mean zero and variance σ^2 to the target signal $\mathbf{x}^\star$. Let $\lambda=x\gamma_0$. We varied x and the Noise-to-Signal Ratio γ. For each fixed (x,γ) pair, 20 instances of the spectral line signals were generated. We then solved (5) for each instance and extracted the frequencies and coefficients. We declared success for an instance if i) the correct number of frequencies were esitmated, ii) the recovered frequency vector is within $\gamma/2n$ distance of the true frequency vector $\mathbf{f}^\star$, and iii) the recovered coefficient vector is within 2λ distance of the true frequency vector $\mathbf{c}^\star$. The rate of success for each configuration is the proportion of successful instances.

From Figure 1, we observe that solving (5) is unable to identify the sinusoidal parameters if $x\le 1$ and the performance of the method is unstable when x is around 1. When x is set to be slightly larger than 1, however, we almost always succeed in finding good estimates of the sinusoidal parameters as long as $x\gamma\le c$ for some small constant c. This matches the findings in Theorem 3.1. Figure 1 also shows the required constant 0.001 in Theorem 3.1 is too conservative.

We also ran simulations to compare the mean-squared error for our frequency estimate with those for MUSIC and the MLE, as well as the CRB. We note that the MLE is initialized using the true frequencies and coefficients, which are not available in practice. We focus on the case of two unknown frequencies and examine the effect of separation. We observe that the atomic norm minimization method always outperforms MUSIC, with increased performance gap when the frequencies become closer. While the MLE performs the best, its initialization is not practical.

VI. Conclusions

This work considers the problem of approximately estimating the frequencies and coefficients of a superposition of

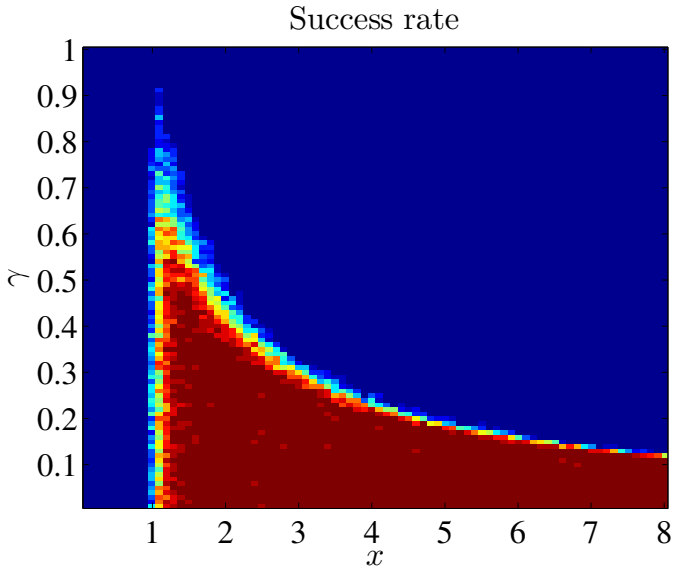

Fig. 1. Rate of success for line spectral estimation by solving the atomic norm regularized program (5).

complex sinusoids in white noise. By using a primal-dual witness construction, we established theoretical performance guarantees for atomic norm regularized Lasso method in line spectral parameter estimation.

REFERENCES

[1] E. J. Candès and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, pp. 906–956, June 2014.

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

[3] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed Sensing Off the Grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, 2013.

[4] G. Tang, B. N. Bhaskar, and B. Recht, "Near Minimax Line Spectral Estimation.," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, 2015.

[5] G. Tang, P. Shah, B. N. Bhaskar, and B. Recht, "Robust line spectral estimation," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 301–305, Nov 2014.

[6] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *arXiv preprint arXiv:1609.02247*, 2016.

[7] 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.

[8] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic Norm Denoising With Applications to Line Spectral Estimation," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5987–5999, 2013.

[9] G. Tang, "Resolution limits for atomic decompositions via markov-bernstein type inequalities," in *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pp. 548–552, IEEE, 2015.

[10] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5987–5999, 2013.

[11] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (lasso)," *IEEE transactions on information theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

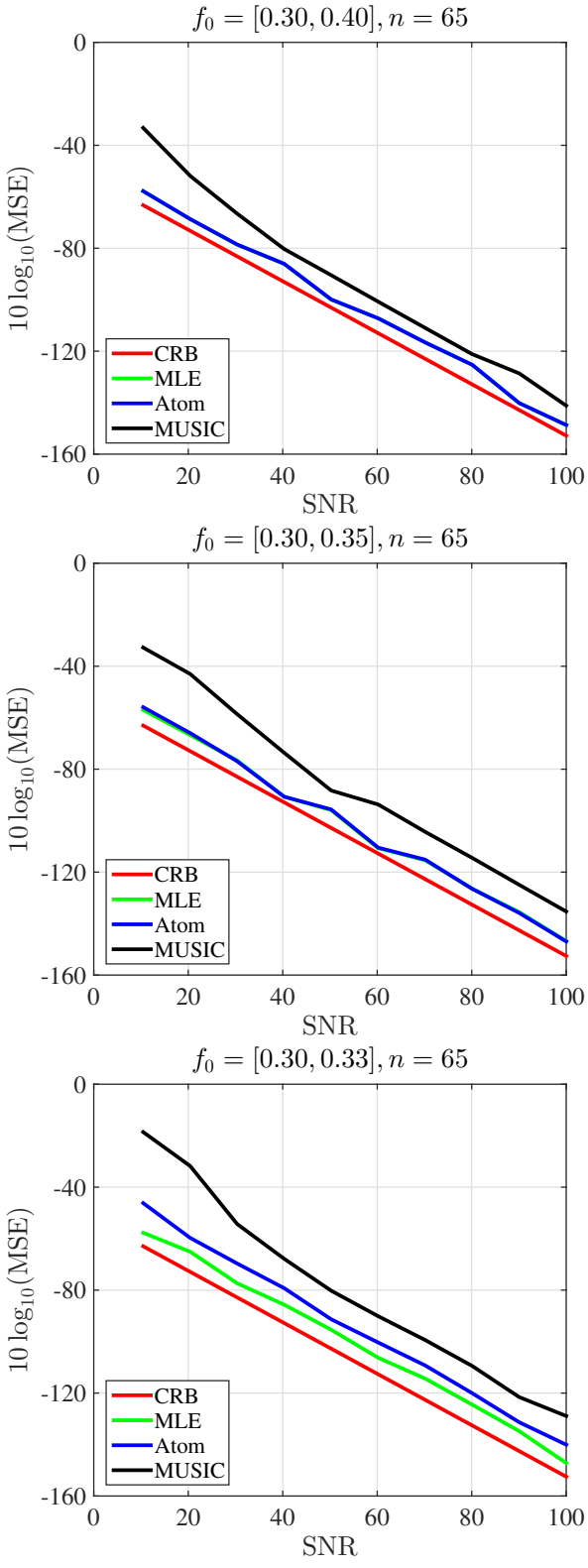

Fig. 2. Performance comparison: Atomic norm minimization (5) (labeled as "Atom"), MUSIC, MLE initialized by the true parameters, and the CRB.