Robust Line Spectral Estimation

Gongguo Tang*, Parikshit Shah†, Badri Narayan Bhaskar‡, Benjamin Recht§
* Colorado School of Mines
gtang@mines.edu
† Wisconsin Institutes for Discovery
pshah@discovery.wisc.edu
‡ University of Wisconsin-Madison
bnbhaskar@wisc.edu
§ University of California, Berkeley
brecht@berkeley.edu

***Abstract*—Line spectral estimation is a classical signal processing problem that finds numerous applications in array signal processing and speech analysis. We propose a robust approach for line spectral estimation based on atomic norm minimization that is able to recover the spectrum exactly even when the observations are corrupted by arbitrary but sparse outliers. The resulting optimization problem is reformulated as a semidefinite program. Our work extends previous work on robust uncertainty principles by allowing the frequencies to assume values in a continuum rather than a discrete set.**

I. Introduction

Line spectral estimation aims to infer the distinct frequency components of a signal from a finite number of observations. Quite a few methods have been developed to solve this fundamental signal processing problem [1], with the earliest dating back to a 1795 article by Prony [2]. While several of these approaches, most notably MUSIC and ESPRIT, exhibit superior statistical performance when the noise is Gaussian, none of them can handle gross corruptions with guaranteed success. Such gross corruptions might be a result of sensor failure, device overflow, missing observation, or data transmission error, which become more common nowadays with the deployment of massive but cheap microsensors and the use of low-power wireless communications in hostile environments.

The following measurement model captures the signal and noise characteristics for robust line spectral estimation:

$$\begin{aligned} y(t) &= x^\star(t) + z^\star(t) \\ &:= \sum_{j=1}^{k} c_j e^{\mathrm{i}2\pi f_j t} + z^\star(t), t = 0, 1, 2, \ldots, n-1. \quad (1) \end{aligned}$$

Here $z^\star \in \mathbb{C}^n$ is a sparse signal with at most s non-zeros, $x^\star \in \mathbb{C}^n$ is the desired line spectral signal, and f_is are a set of k arbitrary frequencies in $[0, 1)$. Estimating both $x^\star$ and $z^\star$ is a special demixing problem [3].

When the frequencies $\{f_i\}$ lie in the discrete set $\{\frac{l}{n}, l = 0, \ldots, n-1\}$, conditions for successful demixing are specified by uncertainty principles [4], [5], [6]. For example, Candés and Romberg showed that, if the support of $z^\star$ and the set of frequencies $\{f_j, j = 1, \ldots, k\} \subset \{\frac{l}{n}, l = 0, \ldots, n-1\}$ are random, and

$$k + s \leq C \frac{n}{\log n}, \quad (2)$$

then with high probability, $x^\star$ and $z^\star$ can be recovered exactly from y by solving an ℓ_1 minimization problem [6]. Unfortunately, in line spectral estimation applications, the frequencies almost never lie in a pre-specified set of discrete frequencies.

In this work, we allow the frequencies $\{f_i\}$ to take values in the continuum $[0, 1)$ and solve the robust line spectral estimation problem using atomic norm minimization [7]. When applied to line spectral estimation, atomic norm minimization has led to new formulations that allow the recovery of continuous frequency parameters from incomplete and noisy data [8], [9], [10]. We further extend this framework to handle sparse outliers in this work. We remark that another approach for line spectral estimation based on matrix completion ideas is also able to handle continuous parameters, incomplete and noisy measurements, and sparse outliers [11]. Due to space limitations, we leave the comparison of these two methods to future work.

The paper is organized as follows. In Section II we describe the atomic norm approach for solving line spectral estimation with gross corruptions. Section III is devoted to the main theorem. Numerical experiments

are performed to support the theorem in Section IV. The paper is concluded in Section V.

II. Atomic Norm Minimization for Line Spectral Estimation

The line spectral signal $x^\star$ is a linear combination of a small number of atoms from an *infinite* continuous dictionary $\mathcal{A}$ corresponding to all possible complex sinusoids. Define the complex sinusoid atom $a(f)$ parameterized by the frequency $f \in [0,1)$ as a vector whose jth component for $j \in \{0,\dots,n-1\}$ is

$$[a(f)]_j = \frac{1}{\sqrt{n}} e^{\mathrm{i}2\pi jf}, \mathrm{i} = \sqrt{-1}. \tag{3}$$

The atomic norm corresponding to the set of atoms $\mathcal{A} = \{a(f) : f \in [0,1)\}$ is defined as

$$\|x\|_{\mathcal{A}} := \inf \Big\{ \sum_j |c_j| : x = \sum_j c_j a(f_j), c_j \in \mathbb{C}, f_j \in [0,1) \Big\} \tag{4}$$

which can be thought of as an extension of the ℓ_1 regularizer to the infinite dictionary $\mathcal{A}$.

The atomic norm $\|x\|_{\mathcal{A}}$ admits an exact semidefinite reformulation:

Proposition 1 ([8], [9]): For any $x \in \mathbb{C}^n$, $\|x\|_{\mathcal{A}}$ is equal to the optimal value of

$$\begin{aligned} \underset{u,t}{\text{minimize}} \ & \frac{1}{2}\operatorname{trace}(\operatorname{Toep}(u)) + \frac{1}{2}t \\ \text{subject to} \ & \begin{bmatrix} \operatorname{Toep}(u) & x \\ x^* & t \end{bmatrix} \succeq 0. \end{aligned} \tag{5}$$

Here the superscript * denotes conjugate transpose, and $\operatorname{Toep}(\cdot)$ maps a complex vector onto the corresponding Hermitian Toeplitz matrix.

We propose to recover $x^\star$ and $z^\star$ as well as to estimate $\{f_j\}$ by solving the following optimization problem:

$$\underset{x,z}{\text{minimize}} \ \|x\|_{\mathcal{A}} + \|z\|_1 \quad \text{subject to} \ \ y = x + z. \tag{6}$$

In light of Proposition 1, we can rewrite (6) into a semidefinite program:

$$\begin{aligned} \underset{u,t,x,z}{\text{minimize}} \ & \frac{1}{2}\operatorname{trace}(\operatorname{Toep}(u)) + \frac{1}{2}t + \|z\|_1 \\ \text{subject to} \ & \begin{bmatrix} \operatorname{Toep}(u) & x \\ x^* & t \end{bmatrix} \succeq 0 \\ & y = x + z \end{aligned} \tag{7}$$

III. Main Result

The main result of this paper is that the optimization (6), or equivalently, the semidefinite program (7) almost always separates $x^\star$ and $z^\star$, provided the number of frequencies and the size of corruptions are small, and additionally the frequencies are well-separated. We formalize this statement in the following theorem.

Theorem 1: Suppose we observe

$$y = x^\star + z^\star = \sum_{f_j \in \Omega} c_j^\star a(f_j) + \sum_{j \in S} z_j^\star e_j \tag{8}$$

where $\Omega \subset [0,1)$ is the set of k unknown frequencies, and S is selected uniformly among all subsets of $\{0,\dots,n-1\}$ of size s. Additionally, assume both $\{\operatorname{sign}(c_j^\star)\}$ and $\{\operatorname{sign}(z_j^\star)\}$ are drawn i.i.d. from the uniform distribution on the complex unit circle and the minimum separation between frequencies in Ω

$$\Delta_f = \min_{f_l, f_j \in \Omega} |f_l - f_j| > \frac{1}{\lfloor (n-1)/4 \rfloor}, \tag{9}$$

where the distance $|f_l - f_j|$ is understood as the wrap-around distance on the unit circle. Then there exists a numerical constant C such that

$$n - s \geq C \max\left\{ \log^2 \frac{n}{\delta}, k \log \frac{k}{\delta} \log \frac{n}{\delta} \right\}, \tag{10}$$

$$n \geq C s \log k \log \frac{n}{\delta}, \tag{11}$$

is sufficient to guarantee the exact recovery of $x^\star$ and $z^\star$ via the semidefinite program (7) with probability at least $1 - \delta$.

Corollary 1: Under the same setup as Theorem 1, if

$$s + k \leq C \frac{n}{\log^2 \frac{n}{\delta}} \tag{12}$$

for some numerical constant C, then that we can recover $x^\star$ and $z^\star$ exactly via the semidefinite program (7) with probability at least $1 - \delta$.

Once $x^\star$ and $z^\star$ are separated exactly, the frequencies can be identified by Prony's method [2], a matrix pencil approach [12], or other linear prediction methods [13]. These frequencies can also be recovered efficiently and accurately from the dual optimal solution [8]. After identifying the frequencies, the coefficients $\{c_j^\star\}$ can be obtained by solving a linear system.

The key to show that the optimization (6) succeeds is to construct a dual solution that certifies the optimality of $(x^\star, z^\star)$:

Proposition 2: $(x^\star, z^\star)$ is the unique optimizer to (6) if there exists a dual polynomial

$$Q(f) \ = \ \frac{1}{\sqrt{n}} \sum_{j \in J} q_j e^{-i2\pi jf} = \langle q, a(f) \rangle \tag{13}$$

satisfying

$$
\begin{aligned}
Q\left(f_j\right) &= \operatorname{sign}\left(c_j^\star\right), \forall f_j \in \Omega && (14)\\
\left|Q\left(f\right)\right| &< 1, \forall f \notin \Omega && (15)\\
q_j &= \operatorname{sign}\left(z_j^\star\right), \forall j \in S && (16)\\
\left|q_j\right| &< 1, \forall j \notin S && (17)
\end{aligned}
$$

The dual certificate can be interpreted as a polynomial with bounded modulus on the unit circle. The coefficients of the polynomial are required to equal to the complex signs of the sparse corruption vector on its support. In the case that there are no corruptions at all, the polynomial constructed by Candès and Fernandez-Granda [14] suffices to guarantee optimality. Indeed they write the certificate polynomial via a kernel expansion and show that one can explicitly find appropriate kernel coefficients that certify optimality.

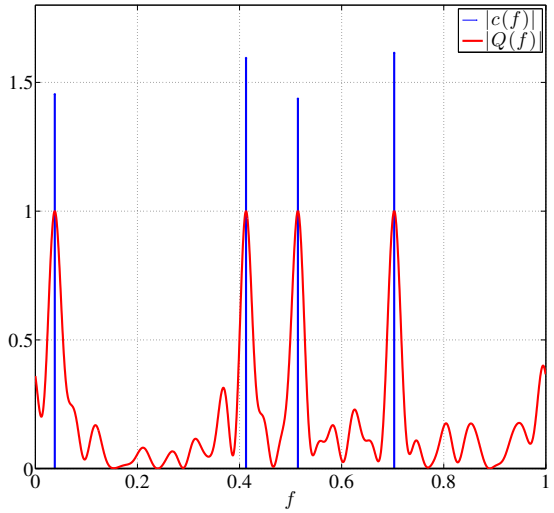
In another case that only part of the signal components are observed, the missing components can be recovered also using atomic norm minimization. To certify the correctness of the completed signal, the authors of [8] constructed a dual polynomial using a random kernel. This random kernel has nonzero coefficients only in the indices corresponding to observed locations (the randomness enters because the samples are observed at random). This random-kernel-based dual polynomial also show up in the construction the dual certificate in Proposition 2, and the results of [8] plays an important role in our proofs.

The requirements of the certifying polynomial in this work are more stringent and require introducing a new term composed of complex monomials, in addition to the term expanded using squared Fejer kernels. Furthermore, the coefficient vector is required to have its ℓ_∞ norm bounded by 1, which needs additional care. We leave details of the construction to the full version of this paper.

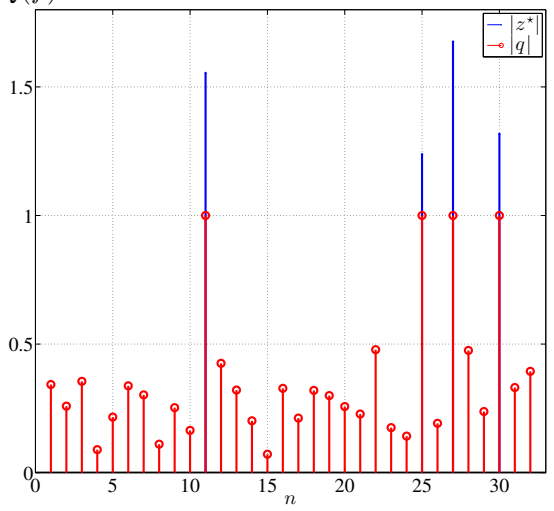
IV. Numerical Experiments

We conducted two sets of numerical experiments to test the performance of (6) under various parameter settings. We solved the semidefinite program using the SDPT3 solver [15] with default parameters.

To generate signal instances of the form (8), we sampled $k = \rho_k n$ normalized frequencies from $[0, 1)$ randomly with an additional constraint on the minimal separation Δ_f. The signal coefficient magnitudes $|c_1|, \cdots, |c_k|$ are equal to $.5 + w^2$ with w a zero mean unit variance Gaussian random variable. The signs $\{\operatorname{sign}(c_j), j = 1, \cdots, k\}$ follow the uniform distribution



(a) Line spectrum of x and the dual polynomial $Q(f)$.



(b) The sparse signal z and the dual vector q.

Fig. 1: Exact demixing for a signal with $n = 32, k = 4$, and $s = 4$.

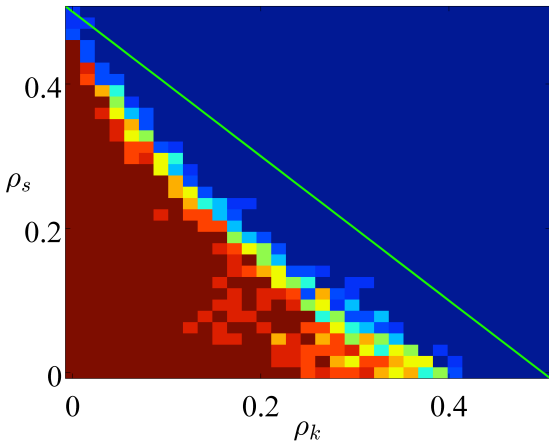
on the complex unit circle. To generate the sparse signal z, a support of size $s = \rho_s n$ is generated uniform randomly, and the coefficients have random phases and magnitudes also of the form $.5 + w^2$. A length n signal was then formed according to model (8).

In the first experiment, we visually illustrate the result of our robust line spectral estimation method on a randomly generated small instance with $n = 32, k = 4$ and $s = 4$. Figures 1a and 1b show exact demixing and the associated dual polynomial and coefficient vector. We observe that the dual vector q and the corresponding dual polynomial $Q(f) = \langle q, a(f) \rangle$ satisfy the conditions prescribed in Proposition 2.

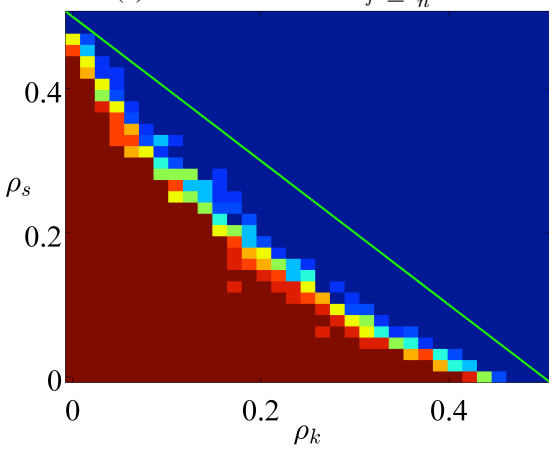
In the second set of experiments, we compiled two phase transition plots. To prepare Figure 2a, we pick $n = 128$ and vary $\rho_k = 0 : \frac{2}{n} : \frac{1}{2}$ and $\rho_s = 0 : \frac{2}{n} : \frac{1}{2}$. For each fixed (ρ_k, ρ_s), we randomly generated $k = n\rho_k$ frequencies while maintaining a frequency separation $\Delta_f \geq \frac{1}{n}$, and uniform randomly generated $s = n\rho_s$ support locations for the sparse corruptions. The coefficients for the complex exponentials and the sparse vectors were both generated with random magnitudes and random phases. We then ran the SDPT3-SDP algorithm to separate $x^\star$ and $z^\star$. The recovery is considered successful if the relative error $\|\hat{x} - x^\star\|_2 / \|x^\star\|_2 \leq 10^{-6}$. This process was repeated 10 times and the rate of success was recorded. Figure 2a shows the phase transition results. The x-axis indicates the fraction of corrupted entries ρ_k, while the y-axis is ρ_s. The color represents the rate of success with red corresponding to perfect recovery and blue corresponding to complete failure. We also plot the line $\rho_s + \rho_k \leq 1/2$.

From Figure 2a, we see that there is a transition from perfect recovery to complete failure. However, the transition boundary is not very sharp. In particular, we notice failures below the boundary of the transition where complete success should happen. Examination of the failures show that they correspond to instances with minimal frequency separations marginally exceeding $\frac{1}{n}$. We expect to get cleaner phase transitions if the frequency separation is increased.

To prepare Figure 2b, we repeated the same process in preparing Figure 2a except that the frequency separation was increased from $\frac{1}{n}$ to $\frac{1.5}{n}$. We now see a much sharper phase transition.



(a) Phase transition: $\Delta_f \geq \frac{1}{n}$



(b) Phase transition: $\Delta_f \geq \frac{1.5}{n}$

Fig. 2: **Phase transition:** The phase transition plots were prepared with $n = 128$, $\rho_k, \rho_s = 0 : 2/n : 1/2$. The frequencies were generated randomly with minimal separation Δ_f. Both signs and magnitudes of the coefficients for the complex exponentials and the non-zero components of the sparse vector are random. In Figure 2a, the separation $\Delta_f \geq 1/n$ while in Figure 2b, the separation $\Delta_f \geq 1.5/n$.

V. Conclusions

By leveraging the framework of atomic norm minimization, we were able to demix line spectral signals from gross corruptions. For signals with well-separated frequencies, we showed that exact demixing is possible as long as the number of frequencies and the number of outliers are small, reminiscent of the Uncertainty Principle developed in a line of work [4], [5], [6].

References

[1] Petre Stoica and R. L. Moses, *Spectral analysis of signals*, Prentice Hall, Upper Saddle River, New Jersey, 1 edition, 2005.

[2] Baron Gaspard Riche de Prony, "Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures," *Journal de l'école Polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[3] Michael McCoy and Joel Tropp, "Sharp recovery bounds for convex demixing, with applications," *Foundations of Computational Mathematics*, pp. 1–65, 2014.

[4] David L Donoho and Philip B Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[5] David L Donoho and Xiaoming Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, 2001.

[6] Emmanuel J Candès and Justin Romberg, "Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions," *Found. Comput. Math.*, vol. 6, no. 2, Apr. 2006.

[7] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[8] Gongguo Tang, B.N. Bhaskar, P. Shah, and B. Recht, "Com-

pressed sensing off the grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.

[9] B.N. Bhaskar, Gongguo 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, Dec 2013.

[10] G. Tang, B.N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *Information Theory, IEEE Transactions on*, vol. 61, no. 1, pp. 499–512, Jan 2015.

[11] Yuxin Chen and Yuejie Chi, "Robust spectral compressed sensing via structured matrix completion," *Information Theory, IEEE Transactions on*, vol. 60, no. 10, pp. 6576–6601, Oct 2014.

[12] Y Hua and T.K Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[13] Petre Stoica, "List of references on spectral line analysis," *Signal Process.*, vol. 31, no. 3, pp. 329–340, Apr. 1993.

[14] Emmanuel J Candès and Carlos Fernandez-Granda, "Towards a mathematical theory of super-resolution," *arXiv.org*, vol. cs.IT, Mar. 2012.

[15] K. C. Toh, M.J. Todd, and R. H. Tütüncü, *SDPT3: A MATLAB software package for semidefinite-quadratic-linear programming*, Available from `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`.