Parameter Estimation for Mixture Models via Convex Optimization

Yuanxin Li
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, Ohio, 43210
Email: li.3822@osu.edu

Yuejie Chi
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, Ohio, 43210
Email: chi.97@osu.edu

***Abstract*—Many applications encounter signals that are a linear combination of multiple components, where each component represents a low-resolution observation of a point source model captured through a low-pass point spread function. This paper proposes a convex optimization algorithm to simultaneously separate and identify the point source models of each component from a noisy observation corrupted by possibly adversarial noise, by leveraging the recently proposed atomic norm framework. The proposed algorithm can be solved efficiently via semidefinite programming, where locations of the point sources can be identified via the constructed dual polynomials without estimating the model orders a priori. Stability of the proposed algorithm is established under certain conditions of the point source models and the point spread functions in the presence of bounded noise. Furthermore, numerical examples are provided to corroborate the theoretical analysis, with comparisons against the Cramèr-Rao bound for parameter estimation.**

I. Introduction

Many applications encounter signals that are a noisy linear combination of multiple components, where each component represents a low-resolution observation of a point source model captured through a low-pass point spread function. Specifically, consider the following *mixture model*, where the acquired signal, $y(t)$, is given as

$$\begin{aligned} y(t) &= \sum_{i=1}^{I} x_i(t) * g_i(t) + w(t) \\ &= \sum_{i=1}^{I} \left(\sum_{k=1}^{K_i} a_{ik} g_i(t-\tau_{ik}) \right) + w(t), \end{aligned} \tag{1}$$

where $*$ denotes convolution, $w(t)$ is the additive noise, and I is the total number of components. Moreover,

$$x_i(t) = \sum_{k=1}^{K_i} a_{ik}\delta(t-\tau_{ik})$$

is the point source model of the ith component, with $\tau_{ik} \in [0,1)$ and $a_{ik} \in \mathbb{C}$ denoting the location and the amplitude of the kth point source, $1 \le k \le K_i$, and $g_i(t)$ is the corresponding point spread function, respectively. The point source model can be used to model a variety of physical phenomena, such as the activation pattern of fluorescence in single-molecule imaging [1], sparse channel impulse response in multi-path fading environments, the locations of pollution plants in urban areas, firing times of neurons, and many more.

978-1-4673-7353-1/15/$31.00 ©2015 IEEE

A. Motivations

The mixture model shows up in the modeling and analysis of many practical problems, such as spike sorting in neural recording [2], [3], three-dimensional super-resolution single-molecule imaging [4], multi-user multi-path channel identification [5], [6], and blind calibration of time-interleaved analog-to-digital converters (ADCs) [7]. The goal therein is to stably *invert* for the parameters of the point source models of each component from the acquired signal in the presence of noise. Moreover, typically we are interested in resolving the locations of the point sources at a resolution much higher than that of the acquired signal $y(t)$, determined by the Rayleigh limit, or in other words, the bandwidth of the point spread functions. We provide two example applications below, highlighting the difference in the generating mechanisms of the $g_i(t)$'s.

- **Spike sorting:** It is known that the action potentials, i.e. spikes, fired by different neurons, have its own stereotyped shapes [8]. Neural recordings can be modeled as a *superposition* of returns from multiple neurons, as in (1), where the return of each neuron corresponds to the sum of its characteristic spike shape delayed by the sequence of its firing times. It is of great interest to separate and estimate the firing times of each neuron from the neural recording, in order to understand the mechanisms of the brain. In this application, the point spread functions are the spike shapes of each neuron, which are determined by the physics and estimated from the neural recording.
- **Multi-user channel estimation:** In multi-user multiple access model, each active user transmits a signature waveform modulated via a signature sequence, which can be designed to optimize performance, and then the base station receives a *superposition* of returns from active users, where the received component for each active user corresponds to a sparse linear combination of delayed and attenuated copies of its signature waveform, determined by the unknown *sparse* multi-path channel from the user to the base station. The goal is to identify the channel

state information of each active user from the received signal at the base station.

B. Existing Approaches and Our Contribution

Conventional approaches for parameter estimation such as matched filtering and subspace methods [9] yield suboptimal outcomes or cannot be applied directly to the mixture model due to the mutual interference between different components in the signal observation. Sparse recovery algorithms have been proposed to estimate the mixture model in [5], [6] with a discretized set of delays, but the performance is at stake when actual delays do not belong to the discrete grid [10]. More recently, [2], [3] have proposed heuristic sparse recovery algorithms to estimate the continuous-valued delays in the mixture model for neural spike sorting, however no performance guarantees are available. Finally, an algebraic root-finding approach has been proposed in [7], but it is rather sensitive to noise and does not extend well to a large number of components due to the prohibitive sample complexity.

In this paper, we start by recognizing that in the frequency domain, the mixture model can be regarded as a linear combination of spectrally-sparse signals. The atomic norm [11] of spectrally-sparse signals is developed in [12], [13] as an efficient convex surrogate, which can be computed efficiently via semidefinite programming. Inspired by [12], [13], we apply atomic norm minimization to parameter estimation of mixture models, by seeking to simultaneously recover the set of spectrally-sparse signals via minimizing their respective atomic norms, in addition to satisfying the observation constraints. The locations of each of the point sources can be identified via the constructed dual polynomials without estimating the model orders a priori. We establish the stability of the proposed algorithm under certain conditions of the point source models and the point spread functions, as well as robustness in the presence of bounded noise from a small number of measurements. Furthermore, numerical examples are provided to corroborate the theoretical analysis, with comparisons against the Cramér-Rao bound for parameter estimation.

II. Problem Formulation and Background

A. Problem Formulation

Due to hardware and physical limits, the resolution of the sensor suite is limited by the diffraction limit, so it is reasonable to assume $g_i(t)$'s are band-limited with cut-off frequency $2M$. Correspondingly, in the Fourier domain, we have

$$g_{in} = \left\langle g_i(t), e^{j2\pi nt} \right\rangle = 0$$

when $n \notin \Omega_M = \{-2M, \ldots, 2M\}$, where $2M$ is inverse proportional to the diffraction limit. In this paper, we consider randomly generated point spread functions in frequency, with g_{in}'s uniformly at random drawn from a complex unit circle.

Taking the discrete Fourier transform of (1) and specializing to the case with $I = 2$, the measurements can be represented as, in the Fourier domain,

$$y_n = \sum_{i=1}^{2} g_{in} \cdot \left(\sum_{k=1}^{K_i} a_{ik} e^{-j2\pi n \tau_{ik}} \right) + w_n, \quad n \in \Omega_M, \quad (2)$$

where the noise w_n can be written as

$$w_n = \int_{-\infty}^{\infty} w(t) e^{j2\pi nt} dt.$$

Equivalently, the measurements y_n's in (2) can be regarded as a combination of two spectrally-sparse signals, each composed of a few distinct complex harmonics. With slight abuse of notation, the measurement vector $\boldsymbol{y} = [y_{-2M}, \ldots, y_0, \ldots, y_{2M}]^T$ can be written as

$$\boldsymbol{y} = \boldsymbol{x}_1^\star + \boldsymbol{g} \odot \boldsymbol{x}_2^\star + \boldsymbol{w} \in \mathbb{C}^{4M+1}, \quad (3)$$

where $\odot$ is the Hadamard element-wise product operation, and each entry in $\boldsymbol{g}$ is $g_n = g_{2n}/g_{1n}$. The ground truth signals $\boldsymbol{x}_1^\star$ and $\boldsymbol{x}_2^\star$ are spectrally-sparse and can be formulated as

$$\boldsymbol{x}_1^\star = \sum_{k=1}^{K_1} a_{1k} \boldsymbol{c}(\tau_{1k}), \quad (4)$$

$$\boldsymbol{x}_2^\star = \sum_{k=1}^{K_2} a_{2k} \boldsymbol{c}(\tau_{2k}), \quad (5)$$

where each atom $\boldsymbol{c}(\tau)$ is defined as

$$\boldsymbol{c}(\tau) = \left[e^{-j2\pi(-2M)\tau}, \ldots, 1, \ldots, e^{-j2\pi(2M)\tau} \right]^T.$$

In this paper, we consider the case when $\boldsymbol{w}$ is bounded, given by $\|\boldsymbol{w}\|_2^2 \le \sigma_w^2$.

B. Atomic Norm

The atomic norm is proposed as a unifying framework for developing convex relaxations to find parsimonious representations under different geometric signal models [11]. In particular, one can define the atomic norm of a signal $\boldsymbol{x}$ with respect to the atomic set $\mathcal{A} = \{\boldsymbol{c}(\tau), \tau \in [0,1)\}$ as [13]

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf_{a_k \in \mathbb{C}, \tau_k \in [0,1)} \left\{ \sum_k |a_k| \mid \boldsymbol{x} = \sum_k a_k \boldsymbol{c}(\tau_k) \right\},$$

which has been shown to be an efficient convex relaxation to promote spectral sparsity in the signal representation, i.e. finding a superposition of complex harmonics with the frequencies from a continuous-valued space. Interestingly, the atomic norm $\|\boldsymbol{x}\|_{\mathcal{A}}$ can be equivalently rewritten using the following semidefinite program characterization,

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf_{\boldsymbol{u}, t} \frac{1}{2} \left(\frac{1}{(4M+1)} \mathrm{Tr}(\mathrm{toep}(\boldsymbol{u})) + t \right) \quad \text{s.t.} \quad \begin{bmatrix} \mathrm{toep}(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^H & t \end{bmatrix} \succeq \boldsymbol{0},$$

making its efficient computation possible. Here, $\mathrm{toep}(\boldsymbol{u})$ represents a Hermitian Toeplitz matrix with $\boldsymbol{u}$ as its first column and $(\cdot)^H$ represents Hermitian transpose. Furthermore, define the real-valued inner product as $\langle \boldsymbol{p}, \boldsymbol{x} \rangle_{\mathbb{R}} = \mathrm{Re}(\boldsymbol{x}^H \boldsymbol{p})$, the dual norm of $\|\cdot\|_{\mathcal{A}}$ can be represented as

$$\|\boldsymbol{p}\|_{\mathcal{A}}^{\star} = \sup_{\|\boldsymbol{x}\|_{\mathcal{A}} \leq 1} \langle \boldsymbol{p}, \boldsymbol{x} \rangle_{\mathbb{R}} = \sup_{\tau \in [0,1)} \left| \sum_{n=-2M}^{2M} p_n e^{j2\pi n\tau} \right|.$$

III. Super-resolution of Mixture Models via Atomic Norm Minimization

A. Noise-free Case

We start by considering the noise-free case with $\boldsymbol{w} = \boldsymbol{0}$, then the signal model can be simplified as $\boldsymbol{y} = \boldsymbol{x}_1^{\star} + \boldsymbol{g} \odot \boldsymbol{x}_2^{\star}$. Leveraging the atomic norm framework, we propose a convex demixing algorithm as

$$\begin{aligned} \{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2\} = \underset{\boldsymbol{x}_1, \boldsymbol{x}_2}{\operatorname{argmin}} & \|\boldsymbol{x}_1\|_{\mathcal{A}} + \|\boldsymbol{x}_2\|_{\mathcal{A}}, \\ \text{s.t.} \quad & \boldsymbol{y} = \boldsymbol{x}_1 + \boldsymbol{g} \odot \boldsymbol{x}_2. \end{aligned} \tag{6}$$

By standard Lagrangian calculation the dual problem of (6) can be written as

$$\begin{aligned} \max_{\boldsymbol{p}} \ & \langle \boldsymbol{p}, \boldsymbol{y} \rangle_{\mathbb{R}}, \\ \text{s.t.} \quad & \|\boldsymbol{p}\|_{\mathcal{A}}^{\star} \leq 1, \ \|\bar{\boldsymbol{g}} \odot \boldsymbol{p}\|_{\mathcal{A}}^{\star} \leq 1, \end{aligned} \tag{7}$$

where $\bar{(\cdot)}$ denotes element-wise conjugate. Let $\hat{\boldsymbol{p}}$ be the dual solution of (7), then the dual polynomials $\hat{P}(\tau)$ and $\hat{Q}(\tau)$ constructed from $\hat{\boldsymbol{p}}$ can be defined as

$$\hat{P}(\tau) = \sum_{n=-2M}^{2M} \hat{p}_n e^{j2\pi n\tau}, \quad \hat{Q}(\tau) = \sum_{n=-2M}^{2M} \hat{p}_n \bar{g}_n e^{j2\pi n\tau}.$$

With these dual polynomials, the corresponding point sources in each signal can be identified as

$$\hat{\Upsilon}_1 = \left\{ \tau \in [0,1) : \left| \hat{P}(\tau) \right| = 1 \right\}$$

and

$$\hat{\Upsilon}_2 = \left\{ \tau \in [0,1) : \left| \hat{Q}(\tau) \right| = 1 \right\}$$

without a priori knowledge of model orders.

The performance guarantee for the noise-free case has been presented in details in the companion paper [14], and we provide here for completeness. Define the separation condition of point sources in each signal as

$$\Delta_i = \min_{k \neq j} |\tau_{ik} - \tau_{ij}|,$$

which is the wrapped-around distance on $[0,1)$, and the minimum separation of all signals as $\Delta = \min_i \Delta_i$. We have the following theorem whose proof can be found in the companion paper [14].

Theorem 1. *Let $M \geq 4$. Assume that $g_n = e^{j2\pi\phi_n}$'s are i.i.d. randomly generated from a uniform distribution on the complex unit circle with $\phi_n \sim \mathcal{U}[0,1]$, and that the signs of the coefficients a_{ik}'s are i.i.d. generated from a symmetric distribution on the complex unit circle. Provided that $\Delta \geq 1/M$, there exists a numerical constant C such that*

$$\begin{aligned} M \geq C \max \Bigg\{ & \log^2 \left(\frac{M(K_1+K_2)}{\eta} \right), \\ & \max\{K_1, K_2\} \log \left(\frac{K_1+K_2}{\eta} \right) \log \left(\frac{M(K_1+K_2)}{\eta} \right) \Bigg\} \end{aligned}$$

is sufficient to guarantee that $\boldsymbol{x}_1^{\star}$ and $\boldsymbol{x}_2^{\star}$ are the unique solutions of (6) with probability at least $1-\eta$.

Theorem 1 indicates that as soon as the number of measurements is on the order $M = O(\max\{K_1, K_2\} \log(K_1 + K_2) \log M)$, the proposed convex demixing algorithm recovers the locations of the point sources exactly with high probability. This suggests that the performance of the convex demixing algorithm is near optimal in terms of the sample complexity.

B. Noisy Case

We proceed to consider the case where the measurements are corrupted by noise as in (3). We modify the convex denoising algorithm as a regularized atomic norm minimization algorithm as

$$\begin{aligned} \{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2\} = \underset{\boldsymbol{x}_1, \boldsymbol{x}_2}{\operatorname{argmin}} & \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}_1 - \boldsymbol{g} \odot \boldsymbol{x}_2\|_2^2 \\ & + \lambda_w (\|\boldsymbol{x}_1\|_{\mathcal{A}} + \|\boldsymbol{x}_2\|_{\mathcal{A}}), \end{aligned} \tag{8}$$

where λ_w is the regularization parameter to balance the data fitting term and the structural promoting term. The dual problem of (8) can be obtained as

$$\begin{aligned} \max_{\boldsymbol{q}} \ & \frac{1}{2} \left(\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{y} - \lambda_w \boldsymbol{q}\|_2^2 \right), \\ \text{s.t.} \quad & \|\boldsymbol{q}\|_{\mathcal{A}}^{\star} \leq 1, \ \|\bar{\boldsymbol{g}} \odot \boldsymbol{q}\|_{\mathcal{A}}^{\star} \leq 1. \end{aligned} \tag{9}$$

Proposition 1 provides the optimality condition of (8).

Proposition 1. *$\{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2\}$ is the minimizer of (8) if and only if*

$$\begin{aligned} \|\boldsymbol{y} - (\hat{\boldsymbol{x}}_1 + \boldsymbol{g} \odot \hat{\boldsymbol{x}}_2)\|_{\mathcal{A}}^{\star} &\leq \lambda_w, \\ \|\bar{\boldsymbol{g}} \odot (\boldsymbol{y} - (\hat{\boldsymbol{x}}_1 + \boldsymbol{g} \odot \hat{\boldsymbol{x}}_2))\|_{\mathcal{A}}^{\star} &\leq \lambda_w, \\ \langle \boldsymbol{y} - (\hat{\boldsymbol{x}}_1 + \boldsymbol{g} \odot \hat{\boldsymbol{x}}_2), \hat{\boldsymbol{x}}_1 + \boldsymbol{g} \odot \hat{\boldsymbol{x}}_2 \rangle_{\mathbb{R}} &= \lambda_w (\|\hat{\boldsymbol{x}}_1\|_{\mathcal{A}} + \|\hat{\boldsymbol{x}}_2\|_{\mathcal{A}}). \end{aligned}$$

Let $\{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2\}$ be the solution of primal problem (8), and $\hat{\boldsymbol{q}}$ be the solution of dual problem (9) respectively. Then there is no dual gap between (8) and (9) as $\boldsymbol{y} = \hat{\boldsymbol{x}}_1 + \boldsymbol{g} \odot \hat{\boldsymbol{x}}_2 + \lambda_w \hat{\boldsymbol{q}}$, which provides an easy way to obtain $\hat{\boldsymbol{q}}$ from the primal solution. With the dual solution $\hat{\boldsymbol{q}}$, we construct the dual polynomials as

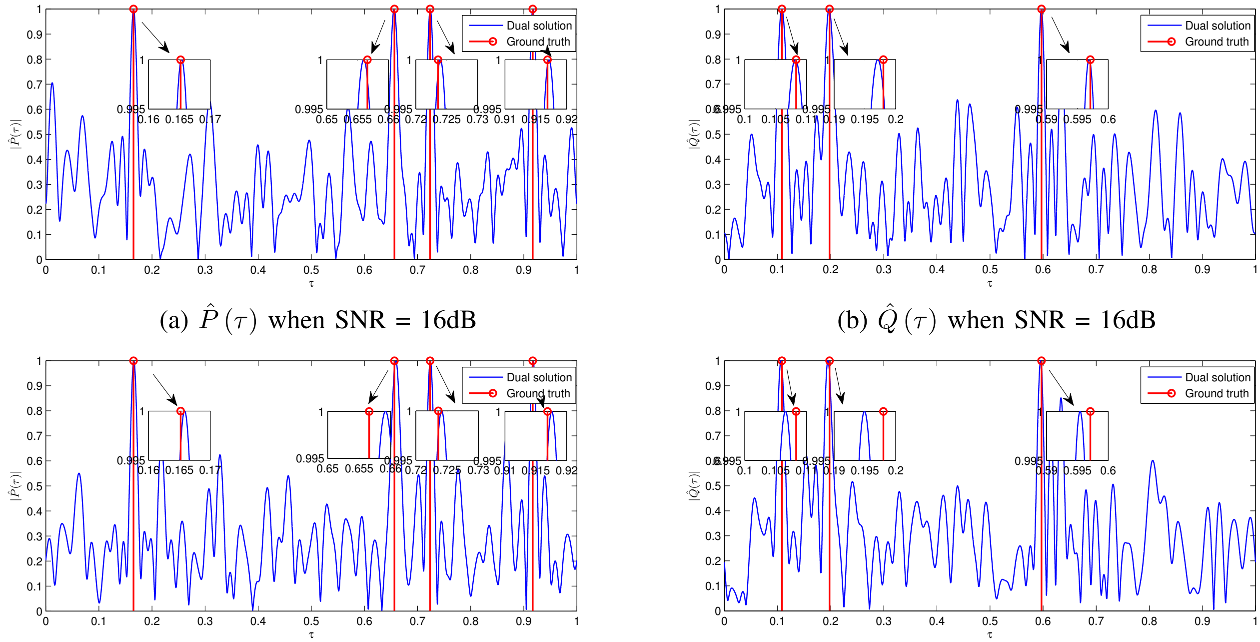
$$\hat{P}(\tau) = \sum_{n=-2M}^{2M} \hat{q}_n e^{j2\pi n\tau}, \quad \hat{Q}(\tau) = \sum_{n=-2M}^{2M} \hat{q}_n \bar{g}_n e^{j2\pi n\tau}.$$

Then the corresponding point sources in each signal can be determined similarly as

$$\hat{\Upsilon}_1 = \left\{ \tau \in [0,1) : \left| \hat{P}(\tau) \right| = 1 \right\}$$

and

$$\hat{\Upsilon}_2 = \left\{ \tau \in [0,1) : \left| \hat{Q}(\tau) \right| = 1 \right\}$$



(a) $\hat{P}(\tau)$ when SNR = 16dB

(b) $\hat{Q}(\tau)$ when SNR = 16dB

(c) $\hat{P}(\tau)$ when SNR = 5dB

(d) $\hat{Q}(\tau)$ when SNR = 5dB

Fig. 1: Point source localization from dual polynomials when the SNR =16dB in (a) and (b), and the SNR = 5dB in (c) and (d), for $M = 16$, $K_1 = 4$ and $K_2 = 3$.

respectively.

The following theorem characterizes the algorithm performance when the noise is bounded as $\|\boldsymbol{w}\|_2^2 \le \sigma_w^2$.

Theorem 2. *Let $M \ge 4$ and $\lambda_w = \sigma_w\sqrt{4M+1}$. Assume that $g_n = e^{j2\pi\phi_n}$'s are i.i.d. randomly generated from a uniform distribution on the complex unit circle with $\phi_n \sim \mathcal{U}[0,1]$. Provided that the separation $\Delta \ge 1/M$, then as long as the number of measurements satisfies*

$$M \ge C \max\left\{ \log^2\left(\frac{M}{\eta}\right), \max\{K_1, K_2\} \log\left(\frac{M}{\eta}\right), \max\{K_1^2, K_2^2\} \log\left(\frac{K_1+K_2}{\eta}\right) \right\}$$

for some constant C, the solution to (8) *satisfies*

$$\max\{\|\hat{\boldsymbol{x}}_1 - \boldsymbol{x}_1^\star\|_2^2, \|\hat{\boldsymbol{x}}_2 - \boldsymbol{x}_2^\star\|_2^2\} \le C'\sigma_w^2,$$

with high probability at least $1-\eta$, where C' is a constant that depends on M, K_1 and K_2.

Compared with Theorem 1, we no longer require the signs of the spikes to be random, at a price of more measurements (note the quadratic dependence with $\max\{K_1, K_2\}$ in the last term). When $\sigma_w^2 = 0$, Theorem 2 degenerate to the noise-free case, providing a performance guarantee of convex demixing algorithm when the support has deterministic signs. Theorem 2 guarantees the stability for inversion in presence of bounded noise, even when the noise is adversarially generated.

IV. Numerical Experiments

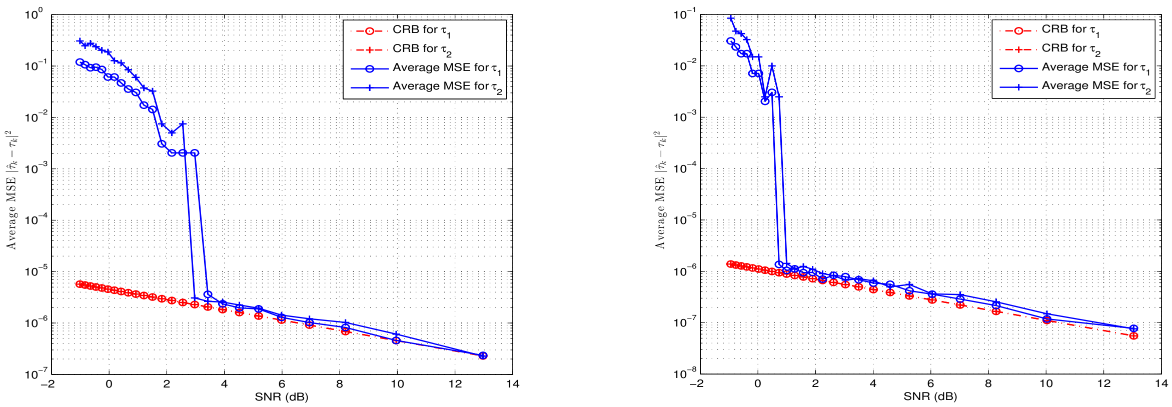
We perform a series of numerical experiments to validate the performance of the proposed algorithms in the noisy case.

A. Point Source Localization by Dual Polynomials

Fix $M = 16$, $K_1 = 4$ and $K_2 = 3$. We first randomly generate a pair of point sources that satisfy a separation condition $\Delta \ge 1/(2M)$, which is in fact a little smaller than the theoretical constraint, with the coefficients of the point sources i.i.d. drawn from the complex standard Gaussian distribution. The noise is generated with i.i.d. complex Gaussian entries $\mathcal{CN}(0,\sigma^2)$. We solve the proposed algorithm by CVX [15], which also returns the dual solution simultaneously. The amplitudes of constructed dual polynomials $\hat{P}(\tau)$ and $\hat{Q}(\tau)$ are shown in Fig. 1 for SNR = 16 dB and SNR = 5dB, respectively, where the Signal-to-Noise Ratio (SNR) is defined as $\mathrm{SNR} = 10\log_{10}\left(\frac{\|\boldsymbol{x}_1^\star + \boldsymbol{g}\odot\boldsymbol{x}_2^\star\|_2^2/(4M+1)}{\sigma^2}\right)$. It is clear that the source locations can be estimated stably from the dual solutions, and the performance degenerates gracefully with the increase of the noise level.

B. Comparison with CRB for Parameter Estimation

We further examine the performance of (8) on estimating the locations of the point sources from noisy measurements by comparing it against the Cramér-Rao bound (CRB). Specifically, consider the special case with a single point source for each component, by letting $K_1 = K_2 = 1$. Denote the point source location in $\boldsymbol{x}_1^\star$ and $\boldsymbol{x}_2^\star$ as τ_1 and τ_2 respectively. We assume the corresponding amplitude of each point source is



(a) $M = 10$ (b) $M = 16$

Fig. 2: The comparison between the average MSE of point source localization and the corresponding CRB with respect to SNR, when (a) $M = 10$, (b) $M = 16$.

known and unity when computing the CRB for estimating τ_1 and τ_2, which can be found by inverting the diagonal entries of the following Fisher information matrix as

$$\boldsymbol{J}(\tau_1, \tau_2) = \frac{8\pi^2}{\sigma^2} \begin{bmatrix} \sum_{n=-2M}^{2M} n^2 & \mathrm{Re}\left(\sum_{n=-2M}^{2M} C_n\right) \\ \mathrm{Re}\left(\sum_{n=-2M}^{2M} C_n\right) & \sum_{n=-2M}^{2M} n^2 \end{bmatrix},$$

where $C_n = n^2 \bar{g}_n e^{-j2\pi n(\tau_1 - \tau_2)}$. We randomly generate 200 noise realizations and compute the average Mean Square Error (MSE) of point source localization using the dual solution of (8), defined as $(\hat{\tau}_k - \tau_k)^2$, as a function of SNR. Fig. 2 shows the average MSE in comparison with the CRB against the SNR when $M = 10$ in (a) and $M = 16$ in (b). The performance of parameter estimation shows a similar "thresholding effect" as for conventional spectrum estimation algorithms, where the average MSE approaches the CRB as soon as SNR is large enough. Moreover, as we increase M, the threshold SNR becomes smaller.

V. Conclusion

We propose a convex optimization algorithm based on atomic norm minimization to simultaneously estimate the parameters from mixture models, where the point sources can be localized via the dual polynomials without priori model order knowledge. We demonstrate the stability of the proposed algorithm for inversion in the presence of bounded noise, under certain conditions on the point source separation and the generation of the point spread functions. In the future work, we plan to implement the proposed algorithms to practical applications outlined in the introduction.

Acknowledgment

This work is supported in part by the Ralph E. Powe Junior Faculty Enhancement Award from the Oak Ridge Associated Universities.

References

[1] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.

[2] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "A unified framework and method for automatic neural spike identification," *Journal of neuroscience methods*, vol. 222, pp. 47–55, 2014.

[3] K. C. Knudson, J. Yates, A. Huk, and J. W. Pillow, "Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit," in *Advances in Neural Information Processing Systems*, 2014, pp. 1215–1223.

[4] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, February 2008.

[5] Y. Chi, Y. Xie, and R. Calderbank, "Compressive demodulation of mutually interfering signals," *submitted to IEEE Transactions on Information Theory*, 2013. [Online]. Available: http://arxiv.org/abs/1303.3904

[6] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Asynchronous code-division random access using convex optimization," *Physical Communication*, vol. 5, no. 2, pp. 129–147, 2012.

[7] Y. M. Lu and M. Vetterli, "Multichannel sampling with unknown gains and offsets: A fast reconstruction algorithm," in *Proc. Allerton Conference on Communication, Control and Computing, Monticello*, 2010.

[8] G. Gerstein and W. Clark, "Simultaneous studies of firing patterns in several neurons," *Science*, vol. 143, no. 3612, pp. 1325–1327, 1964.

[9] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, 1991, vol. 98.

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

[11] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex algebraic geometry of linear inverse problems," *48th Annual Allerton Conference on Communication, Control, and Computing*, pp. 699–703, 2010.

[12] 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.

[13] 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.

[14] Y. Li and Y. Chi, "Super-resolution of mutually interfering signals," in *Proceeding of IEEE International Symposium on Information Theory*, Hong Kong, June 2015.

[15] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," *Online accessiable: http://stanford. edu/~boyd/cvx*, 2008.