Structured Local Optima in Sparse Blind Deconvolution

Yuqian Zhang yz2409@columbia.edu
Han-Wen Kuo hk2673@columbia.edu
John Wright jw2966@columbia.edu
Department of Electrical Engineering & Data Science Institute
Columbia University, New York, 10027

Abstract

Blind deconvolution is a fundamental problem, which asks to recover an unknown kernel $\boldsymbol{a}_0 \in \mathbb{R}^k$ and activation signal $\boldsymbol{x}_0 \in \mathbb{R}^m$ from their convolution $\boldsymbol{y} = \boldsymbol{a}_0 \circledast \boldsymbol{x}_0 \in \mathbb{R}^m$. Unfortunately, this is an ill-posed problem in general. This paper focuses on the *short and sparse* blind deconvolution problem, where the convolutional kernel is short ($k \ll m$) and the activation signal is sparsely and randomly supported ($\|\boldsymbol{x}_0\|_0 \ll m$). This variant captures the structure of the convolutional signals in several important application scenarios. In this paper, we normalize the kernel to have unit Frobenius norm and then cast the blind deconvolution problem as a nonconvex optimization problem over the kernel sphere. We demonstrate that (i) in a certain region of the sphere, each local optimum is close to some shift truncation of the ground truth, and (ii) for a generic $\boldsymbol{a}_0$, when the sparsity rate $\theta \lesssim k^{-2/3}$ and number of measurements $m \gtrsim \mathrm{poly}(k)$, the proposed initialization method together with a proceeding second order algorithm recovers some shifted truncation of the ground truth kernel.

1 Introduction

Blind deconvolution is the problem of recovering two unknown signals $\boldsymbol{a}_0$ and $\boldsymbol{x}_0$ from their convolution $\boldsymbol{y} = \boldsymbol{a}_0 * \boldsymbol{x}_0$. This is a fundamental problem which recurs in several fields. However, it is ill-posed without further knowledge about the unknown signals, as there are infinitely many pairs of signals $(\boldsymbol{a}, \boldsymbol{x})$ whose convolution equals a given observation $\boldsymbol{y}$.

Many practical scenarios, including microscopy data analysis [CLC+17], neural spike sorting [ETS11, Lew98], image deblurring [CL09, KH96], etc., admit a *short-and-sparse* convolution model, in which the observed signal $\boldsymbol{y} \in \mathbb{R}^m$ is the convolution of a *short* kernel $\boldsymbol{a}_0 \in \mathbb{R}^k$ ($k \ll m$) and *random and sparse* activation coefficients $\boldsymbol{x}_0 \in \mathbb{R}^m$ ($\|\boldsymbol{x}_0\|_0 \ll m$). Without loss of generality, we model $\boldsymbol{y}$ as the circular convolution of $\boldsymbol{a}_0$ and $\boldsymbol{x}_0$: $\boldsymbol{y} = \widetilde{\boldsymbol{a}_0} \circledast \boldsymbol{x}_0$. Here, $\widetilde{\boldsymbol{a}_0} = \boldsymbol{\iota}_k \boldsymbol{a}_0 \in \mathbb{R}^m$ denotes the zero padded m-length version of $\boldsymbol{a}_0$; $\boldsymbol{\iota}_k : \mathbb{R}^k \to \mathbb{R}^m$ is a zero padding operator that adds $m-k$ zeros at the end. Its adjoint $\boldsymbol{\iota}_k^* : \mathbb{R}^m \to \mathbb{R}^k$ projects an m-dimensional vector into $\mathbb{R}^k$ by retaining only the first k coordinates.

Because of the basic properties of a convolution operator, the sparse blind deconvolution (SBD) problem exhibits a *scaled-shift symmetry*. Namely, for any nonzero scalar α and integer shift τ,

$$\boldsymbol{y} = (\alpha s_\tau[\widetilde{\boldsymbol{a}_0}]) \circledast \left(\alpha^{-1} s_{-\tau}[\boldsymbol{x}_0]\right). \tag{1}$$

Here, $s_{-\tau}[\boldsymbol{v}]$ denotes the cyclic shift of a vector $\boldsymbol{v}$ by τ entries and can be expressed as

$$s_\tau[\boldsymbol{v}](i) = \boldsymbol{v}\left([i-\tau]_m\right), \quad \forall\, i \in \{1, \cdots, m\}, \tag{2}$$

with $\boldsymbol{v} \in \mathbb{R}^k$ indexed as $\boldsymbol{v} = [v_0, \cdots, v_{k-1}]$ and $[\cdot]_m$ denoting the modulo operator of m.

Clearly, this operation preserves the short-and-sparse structure of $(\boldsymbol{a}_0, \boldsymbol{x}_0)$. This *scaled-shift symmetry* raises challenges for computation, making straightforward convexification approaches ineffective, and leading to a very complicated optimization landscape for nonconvex formulations. [ZLK+17] considers a natural nonconvex formulation of blind deconvolution, in which the kernel $\boldsymbol{a} \in \mathbb{R}^k$ is constrained to have unit ℓ^2 norm. [ZLK+17] argues that this problem has well-structured local optima, in the sense that *every local optimum is close to some shift truncation of the ground truth*. These

OPTML 2017: 10th NIPS Workshop on *Optimization for Machine Learning* (NIPS 2017).

local optima arise as (near) shift-truncations of the ground truth: the shifted and truncated kernel $\boldsymbol{\iota}_k^* s_\tau[\widetilde{\boldsymbol{a}_0}]$ can be convolved with the sparse signal $s_{-\tau}[\boldsymbol{x}_0]$ (shifted in the other direction) to produce a near approximation to $\boldsymbol{y}$ that $(\boldsymbol{\iota}_k^* s_\tau[\widetilde{\boldsymbol{a}_0}]) \circledast s_{-\tau}[\boldsymbol{x}_0] \approx \boldsymbol{y}$. The presence of these local optima can be viewed as a result of the shift symmetry of the convolution operator.[1]

In [ZLK+17], this geometric insight about local solutions is corroborated with experiments, but rigorous proof is only available in the "dilute limit" in which the sparse coefficient signal $\boldsymbol{x}_0$ is a single spike. In this paper, we consider a different objective function over the sphere $\mathbb{S}^{k-1}$ and demonstrate that even when $\boldsymbol{x}_0$ is relatively dense, every local minimum in a certain region of the sphere is still close to a shift truncation $\boldsymbol{\iota}_k^* s_\tau\,[\widetilde{\boldsymbol{a}_0}]$ of the ground truth. Moreover, for a generic kernel $\boldsymbol{a}_0 \in \mathbb{S}^{k-1}$, if the sparsity rate $\theta \lesssim k^{-2/3}$ and the number of measurement $m \gtrsim \mathrm{poly}(k)$, initializing with a k-length window of $\boldsymbol{y}$ and applying any optimization method which (i) is a descent method, and (ii) converges to a local minimizer under a strict saddle hypothesis [ABG07, GWY09], produces a near shift-truncation of the ground truth.

2 Problem Formulation and Main Result

In the "short-and-sparse" deconvolution model, any k consecutive entries in $\boldsymbol{y}$ only depend on corresponding $2k-1$ consecutive entries in $\boldsymbol{x}_0$, i.e.,

$$\boldsymbol{y}_i = \left[y_{i-1}, \cdots, y_{[i+k-2]_m}\right]^T = \sum_{\tau=-(k-1)}^{k-1} x_{[i-1+\tau]_m} \cdot \boldsymbol{\iota}_k^* s_\tau[\boldsymbol{a}_0]. \tag{3}$$

Therefore, the observation $\boldsymbol{y} = \boldsymbol{a}_0 \circledast \boldsymbol{x}_0$ can be equivalently expressed through following matrix multiplication $\boldsymbol{Y} = \boldsymbol{A}_0\boldsymbol{X}_0$. Here, $\boldsymbol{Y} \in \mathbb{R}^{k\times m}$, $\boldsymbol{A}_0 \in \mathbb{R}^{k\times(2k-1)}$, and $\boldsymbol{X}_0 \in \mathbb{R}^{(2k-1)\times m}$ are truncated circulant matrices generated from $\boldsymbol{y}$, $\widetilde{\boldsymbol{a}_0}$, and $\boldsymbol{x}_0$ respectively:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{\iota}_k^* \boldsymbol{C}_{\boldsymbol{y}} = \begin{bmatrix}\boldsymbol{y}_1 & \cdots & \boldsymbol{y}_m\end{bmatrix} & (4)\\ \boldsymbol{A}_0 &= \begin{bmatrix}\boldsymbol{\iota}_k^* s_{-(k-1)}[\widetilde{\boldsymbol{a}_0}] & \cdots & \boldsymbol{a}_0 & \cdots & \boldsymbol{\iota}_k^* s_{k-1}[\widetilde{\boldsymbol{a}_0}]\end{bmatrix}, & (5)\\ \boldsymbol{X}_0 &= \boldsymbol{\iota}_{2k-1}^* \boldsymbol{C}_{s_k[\boldsymbol{x}_0]} = \begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_m\end{bmatrix}. & (6)\end{aligned}$$

with $\boldsymbol{C}_{\boldsymbol{v}} \in \mathbb{R}^{m\times m}$ the circulant matrix generated from vector $\boldsymbol{v}$, whose j-th column is a cyclic shift $s_{j-1}[\boldsymbol{v}]$ of the vector $\boldsymbol{v}$ as defined in Equation 2.

For the randomly and sparsely supported $\boldsymbol{x}_0$, we assume it follows the Bernoulli-Gaussian (BG) model with rate θ: $\boldsymbol{x}_0(i) = \omega_i g_i$ with $\omega_i \sim \mathrm{Ber}(\theta)$ and $g_i \sim \mathcal{N}(0,1)$, with all the random variables jointly independent. For simplicity, we write $\boldsymbol{x}_0 \sim_{i.i.d.} \mathrm{BG}(\theta)$. Note that each column $\boldsymbol{x}_i$ of $\boldsymbol{X}_0$ only contains some $2k-1$ entries of $\boldsymbol{x}_0$, while each row of $\boldsymbol{X}_0$ can be seen as some circularly shifted version of the reversed $\boldsymbol{x}_0$.

2.1 Finding a Shifted Sparse Signal

Up to the shift ambiguity associated with the convolution operator, recovering any row of $\boldsymbol{X}_0$ solves the sparse blind deconvolution problem. Similar idea of casting the bilinear recovery problem as finding sparse vector can be found in a lot of recent works [SWW12, SQW15]. The problem of finding a sparse vector in a subspace has been well studied in several recent works [QSW16, HSSS16]. This motivates us to consider following objective function over the sphere $\mathbb{S}^{k-1}$:

$$\min_{\boldsymbol{q}\in\mathbb{S}^{k-1}} \psi(\boldsymbol{q}) \doteq -\frac{1}{4m}\left\|\boldsymbol{Y}^T\left(\boldsymbol{Y}\boldsymbol{Y}^T\right)^{-1/2}\boldsymbol{q}\right\|_4^4. \tag{7}$$

Since $\mathbb{E}_{\boldsymbol{x}_0\sim_{\text{i.i.d.}}\mathrm{BG}(\theta)}\left[\boldsymbol{Y}\boldsymbol{Y}^T\right] = \theta m \boldsymbol{A}_0\boldsymbol{A}_0^T$, vector $\boldsymbol{Y}^T\left(\boldsymbol{Y}\boldsymbol{Y}^T\right)^{-1/2}\boldsymbol{q}$ is close to the convolution of the observation $\boldsymbol{x}_0$ and $\boldsymbol{A}_0\left(\theta m\boldsymbol{A}_0\boldsymbol{A}_0^T\right)^{-1/2}\boldsymbol{q}$, and maximizing the function $\|\cdot\|_4^4$ encourages the

[1]For certain deconvolution problems in communications, the signals $\boldsymbol{a}_0$ and $\boldsymbol{x}_0$ can be assumed to reside on linear subspaces which are incoherent or random [ARR12, LLJB17]. This model has different properties from the short-and-sparse model; in particular, it does not admit a shift symmetry. Nonconvex optimization approaches to this model have been studied in a number of recent theoretical and algorithmic works – see [LLSW16] and references therein. Because of the very different symmetries (and different applications!) of this model, these results are not directly comparable with ours.

"spikiness" of the vector[2]. Intuitively, the objective function is going to be minimized when $\boldsymbol{Y}^T\left(\boldsymbol{Y}\boldsymbol{Y}^T\right)^{-1/2}\boldsymbol{q}$, or $\boldsymbol{A}_0\left(\boldsymbol{A}_0\boldsymbol{A}_0^T\right)^{-1/2}\boldsymbol{q}$, is close to sparse. Note that the expectation of the objective function $\psi(\boldsymbol{q})$ can be approximated with the following

$$\mathbb{E}_{\boldsymbol{x}_0\sim_{\text{i.i.d.}}\mathrm{BG}(\theta)}\left[\frac{1}{m}\left\|\boldsymbol{Y}^T\left(\boldsymbol{A}_0\boldsymbol{A}_0^T\right)^{-1/2}\boldsymbol{q}\right\|_4^4\right]=3\theta\left(1-\theta\right)\left\|\boldsymbol{A}^T\boldsymbol{q}\right\|_4^4+3\theta^2\left\|\boldsymbol{A}^T\boldsymbol{q}\right\|_2^4, \tag{8}$$

with $\boldsymbol{A}=\left(\boldsymbol{A}_0\boldsymbol{A}_0^T\right)^{-1/2}\boldsymbol{A}_0=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_{2k-1}]$ and $\left\|\boldsymbol{A}^T\boldsymbol{q}\right\|_2=1$. Hence for asymptotic function landscape, we instead consider following surrogate objective function for simplicity:

$$\min_{\boldsymbol{q}\in\mathbb{R}^{k-1}}\varphi(\boldsymbol{q})\doteq-\frac{1}{4}\left\|\boldsymbol{A}^T\boldsymbol{q}\right\|_4^4. \tag{9}$$

2.2 Optimization Function Landscape

Investigating the optimization landscape for nonconvex problems usually involves delicate analysis around stationary points. For our particular problem, in the region where the objective function $\varphi(\boldsymbol{q})$ is small, the objective landscape is particularly regular, and the critical points of φ can be characterized explicitly. This region is given by

$$\mathcal{R}\doteq\left\{\boldsymbol{q}\in\mathbb{S}^{k-1}\mid\left\|\boldsymbol{A}^T\boldsymbol{q}\right\|_4^6\geq C\mu\kappa^2\left\|\boldsymbol{A}^T\boldsymbol{q}\right\|_3^3\right\}. \tag{10}$$

Here, $\kappa\geq1$ is the condition number of $\boldsymbol{A}_0$, and μ is the column incoherence of $\boldsymbol{A}$, given by $\mu\doteq\max_{i\neq j}|\langle\boldsymbol{a}_i,\boldsymbol{a}_j\rangle|$.[3] Hence when restricted on $\mathcal{R}$, the function landscape around a saddle point $\bar{\boldsymbol{q}}$ can be characterized based on the number of nontrivial entries[4] in $\bar{\boldsymbol{\zeta}}=\boldsymbol{A}^T\bar{\boldsymbol{q}}$:

- For any stationary point $\bar{\boldsymbol{q}}\in\mathcal{R}$ as defined, then there must exist some entries of nontrivial magnitude in $\bar{\boldsymbol{\zeta}}=\boldsymbol{A}^T\bar{\boldsymbol{q}}$;
- If $\bar{\boldsymbol{q}}$ is a stationary point and $\bar{\boldsymbol{\zeta}}=\boldsymbol{A}^T\bar{\boldsymbol{q}}$ only has one nontrivial entry ζ_l, then $\bar{\boldsymbol{q}}$ is one local minimum near $\mathcal{P}_{\mathbb{S}}[\boldsymbol{a}_l]$;
- If $\bar{\boldsymbol{q}}$ is a stationary point and $\bar{\boldsymbol{\zeta}}=\boldsymbol{A}^T\bar{\boldsymbol{q}}$ has more than one nontrivial entries, then the Riemannian Hessian at $\bar{\boldsymbol{q}}$ has negative curvature in certain selected direction, then $\bar{\boldsymbol{q}}$ is saddle point.

Therefore, each local optimum $\bar{\boldsymbol{q}}$ in region $\mathcal{R}$ is close to $\mathcal{P}_{\mathbb{S}}[\boldsymbol{a}_l]$ for some l, and $\left(\boldsymbol{A}_0\boldsymbol{A}_0^T\right)^{1/2}\bar{\boldsymbol{q}}$ is close to a signed shift truncation of the ground truth $\boldsymbol{a}_0$ up to scale.[5] For finite sample result, we prove that above characterizations of the function geometry obtain when the observation size m is large enough.

Theorem 2.1 (Main Result) *Suppose observation $\boldsymbol{y}\in\mathbb{R}^m$ is the circulant convolution of $\boldsymbol{a}_0\in\mathbb{S}^{k-1}$ and $\boldsymbol{x}_0\sim_{i.i.d.}\mathrm{BG}(\theta)\in\mathbb{R}^m$. Denote the condition number of $\boldsymbol{A}_0$ with $\kappa\geq1$, and the column incoherence of $\boldsymbol{A}$ with μ. There exist constants $C_1,C_2>0$, whenever $m\geq C_1\left(\theta-\theta^2\right)^{-2}\min\left\{\mu^{-2},\kappa^4k^3\right\}k^6\operatorname{poly}\log k$, then with high probability, any local optima $\bar{\boldsymbol{q}}$ in the sublevel set $\psi(\boldsymbol{q})\leq-C_2\left(\mu\kappa^2\right)^{2/3}$ satisfies $|\langle\bar{\boldsymbol{q}},\mathcal{P}_{\mathbb{S}}[\boldsymbol{a}_i]\rangle|\geq1-C_2^{-1}\kappa^{-2}$ for some integer i.*

Corollary 2.2 *Suppose observation $\boldsymbol{y}\in\mathbb{R}^m$ is the circulant convolution of $\boldsymbol{a}_0\in\mathbb{S}^{k-1}$ and $\boldsymbol{x}_0\sim_{i.i.d.}\mathrm{BG}(\theta)\in\mathbb{R}^m$, with $\theta\leq c_1\mu^{-2/3}\kappa^{-4/3}k^{-1}\left(1+\mu^2k\right)^{-2}$ and signal length $m\geq C\left(\theta-\theta^2\right)^{-2}\min\left\{\mu^{-2},\kappa^4k^3\right\}k^6\operatorname{poly}\log k$, then with high probability, Algorithm 1 recovers $\bar{\boldsymbol{a}}$ such that $\|\bar{\boldsymbol{a}}\pm\mathcal{P}_{\mathbb{S}}[\boldsymbol{\iota}_k s_\tau[\widetilde{\boldsymbol{a}_0}]]\|_2\leq c_2$ for some shift τ. Here, c_1 and c_2 are small positive constants.*

[2] Although the $\|\cdot\|_4^4$ does not lead to strictly "sparse" signal, but serves as a reasonable replacement as well as introduces easier theoretical analysis.

[3] For a generic unit $\boldsymbol{a}_0$, simulation suggests that $\kappa\sim\log^2k$, and $\mu\sim\sqrt{\log k/k}$.

[4] We call any η_i with magnitude smaller than $2\mu\|\boldsymbol{\zeta}\|_3^3/\|\boldsymbol{\zeta}\|_4^4$ to be trivial, and others to be nontrivial.

[5] Both the formulation and the result in this paper share some similarity with [GM17]. Globalizing the result remains an open challenge for both problems. In the tensor decomposition problem, the optimization landscape is highly symmetric as any local solution is equally good, while in blind deconvolution problem, the landscape changes dramatically with respect to the ground truth $\boldsymbol{a}_0$.

Algorithm 1 Short and Sparse Blind Deconvolution

Input: Observations $\boldsymbol{y} \in \mathbb{R}^m$ and kernel size k.
Output: Recovered Kernel $\bar{\boldsymbol{a}}$.

1: Generate random index $i \in [1, m]$ and set $\boldsymbol{q}_{init} = \mathcal{P}_{\mathbb{S}}\left[\left(\boldsymbol{Y}\boldsymbol{Y}^T\right)^{-1/2} \boldsymbol{y}_i\right]$.
2: Solve $\varphi(\boldsymbol{q})$ with a descent method that escapes strict saddle points and set the local optimum $\bar{\boldsymbol{q}} = \arg\min_{\boldsymbol{q}\in\mathbb{S}^{k-1}} \psi(\boldsymbol{q})$
3: Set $\bar{\boldsymbol{a}} = \mathcal{P}_{\mathbb{S}}\left[\left(\boldsymbol{Y}\boldsymbol{Y}^T\right)^{1/2} \bar{\boldsymbol{q}}\right]$.

This result shows that the property *every local solution is close to some shift truncation of the ground truth* holds for relative dense $\boldsymbol{x}_0$. It provides theoretical corroboration for the two stage algorithm proposed in [ZLK+17].

3 Experiment

Phase Transition for Algorithm 1 We present the performance of Algorithm 1 under varying settings. We define the recover error as $err = 1 - \max_i |\langle \bar{\boldsymbol{a}}, \mathcal{P}_{\mathbb{S}}[\boldsymbol{\iota}_k^* s_i[\widetilde{\boldsymbol{a}_0}]]\rangle|$, and calculate the average error from 50 independent experiments. The left figure plots the average error when we fix the kernel size $k = 50$ and change the dimension m and the sparsity θ of $\boldsymbol{x}_0$.[6] The right figure plots the average error when we change the dimensions k, m of both convolution signals, and set sparsity $\theta = k^{-2/3}$.

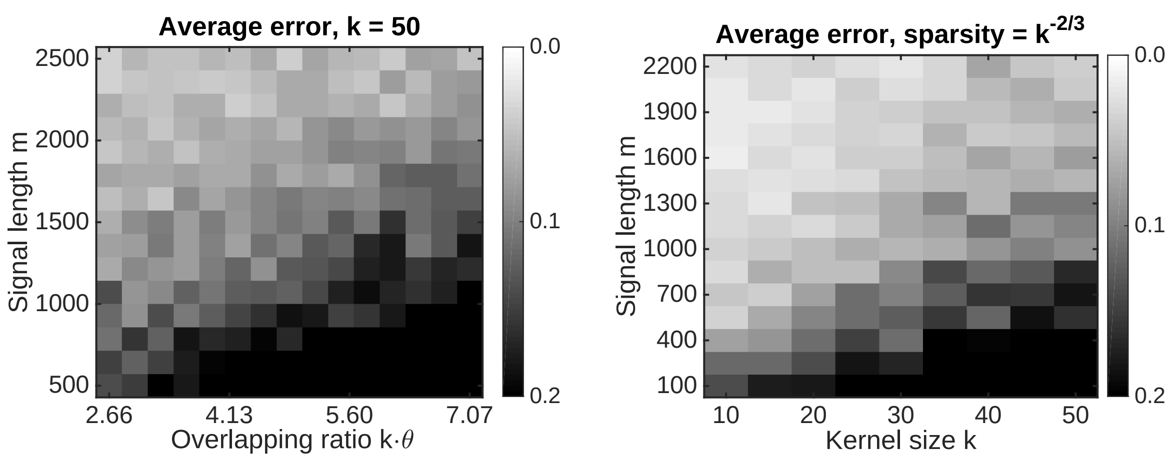

Figure 1: Phase Transition

References

[ABG07] P.-A. Absil, C.G. Baker, and K.A. Gallivan. Trust-region methods on riemannian manifolds. *Foundations of Computational Mathematics*, 7(3):303–330, Jul 2007.

[ARR12] A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programing. *arXiv preprint:1211.5608*, 2012.

[CL09] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. *ACM Transactions on Graphics (SIGGRAPH ASIA 2009)*, 28(5):article no. 145, 2009.

[CLC+17] Sky Cheung, Yenson Lau, Zhengyu Chen, Ju Sun, Yuqian Zhang, John Wright, and Abhay Pasupathy. Beyond the fourier transform: A nonconvex optimization approach to microscopy analysis. *Submitted*, 2017.

[6] Note that the x-axis is indexed with overlapping ratio $k \cdot \theta$, which indicates how many times the kernel $\boldsymbol{a}_0$ present in a k-length window of $\boldsymbol{y}$ on average.

[ETS11] Chaitanya Ekanadham, Daniel Tranchina, and Eero P. Simoncelli. A blind sparse deconvolution method for neural spike identification. In *Advances in Neural Information Processing Systems 24*, pages 1440–1448. 2011.

[GM17] Rong Ge and Tengyu Ma. On the optimization landscape of tensor decompositions. *Advances in Neural Information Processing Systems*, 2017.

[GWY09] Donald Goldfarb, Zaiwen Wen, and Wotao Yin. A curvilinear search method for p-harmonic flows on spheres. *SIAM J. Imaging Sciences*, 2(1):84–109, 2009.

[HSSS16] Samuel B. Hopkins, Tselil Schrammand, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 178–191, 2016.

[KH96] D. Kundur and D. Hatzinakos. Blind image deconvolution. *Signal Processing Magazine, IEEE*, 13(3):43–64, May 1996.

[Lew98] M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):53–78, 1998.

[LLJB17] Kiryung Lee, Yanjun Li, Marius Junge, and Yoram Bresler. Blind recovery of sparse signals from subsampled convolution. *IEEE Transaction of Information Theory*, 63(2):802–821, February 2017.

[LLSW16] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. preprint, 2016.

[QSW16] Qing Qu, Ju Sun, and John Wright. Finding a sparse vector in a subspace: linear sparsity using alternating directions. *IEEE Transactions on Information Theory*, 2016.

[SQW15] Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere. preprint, 2015.

[SWW12] Daniel Spielman, Huan Wang, and John Wright. Exact recovery of sparsely-used dictionaries. preprint, 2012.

[ZLK+17] Yuqian Zhang, Yenson Lau, Han-wen Kuo, Sky Cheung, Abhay Pasupathy, and John Wright. On the global geometry of sphere-constrained sparse blind deconvolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.