

# MULTICHANNEL SPARSE BLIND DECONVOLUTION ON THE SPHERE

Yanjun Li, Yoram Bresler

CSL and Department of ECE, University of Illinois at Urbana-Champaign

## ABSTRACT

Multichannel blind deconvolution is the problem of recovering an unknown signal  $f$  and multiple unknown channels  $x_i$  from convolutional measurements  $y_i = x_i \otimes f$  ( $i = 1, 2, \dots, N$ ). We consider the case where the  $x_i$ 's are sparse, and convolution with  $f$  is invertible. Our nonconvex optimization formulation solves for a filter  $h$  on the unit sphere that produces sparse output  $y_i \otimes h$ . Under some technical assumptions, we show that all local minima of the objective function correspond to the inverse filter of  $f$  up to an inherent sign and shift ambiguity, and all saddle points have strictly negative curvatures. This geometric structure allows successful recovery of  $f$  and  $x_i$  using a simple manifold gradient descent algorithm with random initialization. Our theoretical findings are complemented by numerical experiments, which demonstrate superior performance of the proposed approach over the previous methods.

**Index Terms**— Manifold gradient descent, nonconvex optimization, Riemannian gradient, Riemannian Hessian, strict saddle points, super-resolution fluorescence microscopy

## 1. INTRODUCTION

Blind deconvolution, which aims to recover unknown vectors  $x$  and  $f$  from their convolution  $y = x \otimes f$ , has been extensively studied, especially in the context of image deblurring [1, 2, 3]. Recently, algorithms with theoretical guarantees have been proposed for single channel blind deconvolution [4, 5, 6, 7, 8, 9, 10]. In order for the problem to be well-posed, these previous methods assume that *both*  $x$  and  $f$  are constrained, to either reside in a known subspace or be sparse over a known dictionary. However, these methods cannot be applied if  $f$  (or  $x$ ) is unconstrained, or does not have a subspace or sparsity structure.

In many applications in communications [11], imaging [12], and computer vision [13], convolutional measurements  $y_i = x_i \otimes f$  are taken between a single signal (resp. filter)  $f$  and multiple filters (resp. signals)  $\{x_i\}_{i=1}^N$ . We call such problems multichannel blind deconvolution (MBD). Importantly, in this multichannel setting, one can assume that only  $\{x_i\}_{i=1}^N$  are structured, and  $f$  is unconstrained. While there has been abundant work on single channel blind deconvolution (with both  $f$  and  $x$  constrained), research on MBD (with  $f$  unconstrained) is relatively limited. Traditional MBD works assumed that the channels  $x_i$ 's are FIR filters [14, 15, 16] or IIR filters [17], and proposed to solve MBD using subspace methods. The problem is generally ill-conditioned, and the recovery using the subspace methods is highly sensitive to noise [18].

In this paper, while retaining the unconstrained form of  $f$ , we consider a different structure of the multiple channels  $\{x_i\}_{i=1}^N$ : sparsity. The resulting problem is termed multichannel sparse blind

deconvolution (MSBD). The sparsity structure arises in many real-world applications, including opportunistic underwater acoustics [19], reflection seismology [20], functional MRI [21], and super-resolution fluorescence microscopy [22].

Previous approaches to MSBD have provided efficient iterative algorithms to compute maximum likelihood (ML) estimates of parametric models of the channels  $\{x_i\}_{i=1}^N$  [19], or maximum a posteriori (MAP) estimates in various Bayesian frameworks [20, 13]. However, these algorithms usually do not have theoretical guarantees. Recently, guaranteed algorithms for MSBD have been developed. Wang and Chi [23] proposed a convex formulation of MSBD based on  $\ell_1$  minimization. Li et al. [24] solved a nonconvex formulation using projected gradient descent, and proposed an initialization algorithm to compute a sufficiently good starting point. However, the theoretical guarantees of these algorithms require restrictive assumptions (e.g.,  $f$  has one dominant entry that is significantly larger than other entries [23], or  $f$  has an approximately flat spectrum [24]).

In this paper, we consider MSBD with circular convolution. In addition to the sparsity prior on the channels  $\{x_i\}_{i=1}^N$ , we impose, without loss of generality, the constraint that  $f$  has unit  $\ell_2$  norm, i.e.,  $f$  is on the unit sphere. (This eliminates the scaling ambiguity inherent in the MBD problem.) We show that our sparsity promoting objective function has a nice geometric landscape on the unit sphere: **(S1)** all local minima correspond to signed shifted versions of the desired solution, and **(S2)** the objective function is strongly convex in neighborhoods of the local minima, and has strictly negative curvature directions in neighborhoods of local maxima and saddle points. Similar geometric analysis has been conducted for dictionary learning [25], phase retrieval [26], and single channel sparse blind deconvolution [10]. Recently, Mei et al. [27] analyzed the geometric structure of the empirical risk of a class of machine learning problems (e.g., nonconvex binary classification, robust regression, and Gaussian mixture model). This paper is the first such analysis for MSBD.

Properties **(S1)** and **(S2)** allow simple manifold optimization algorithms to find the ground truth in the nonconvex formulation. Unlike the second order methods in previous works [28, 26], we take advantage of recent advances in the analysis of first-order methods [29, 30], and prove that a simple manifold gradient descent algorithm, with random initialization and a fixed step size, can accurately recover a signed shifted version of the ground truth in polynomial time almost surely. This is the first guaranteed algorithm for MSBD that does *not* rely on restrictive assumptions on  $f$  or  $\{x_i\}_{i=1}^N$ .

## 2. MSBD ON THE SPHERE

### 2.1. Problem Statement

In MSBD, the measurements  $y_1, y_2, \dots, y_N \in \mathbb{R}^n$  are the circular convolutions of unknown sparse vectors  $x_1, x_2, \dots, x_N \in \mathbb{R}^n$  and an unknown vector  $f \in \mathbb{R}^n$ , i.e.,  $y_i = x_i \otimes f$ . In this paper, we solve

This work was supported in part by the National Science Foundation (NSF) under Grant IIS 14-47879.

for  $\{x_i\}_{i=1}^n$  and  $f$  from  $\{y_i\}_{i=1}^N$ . One can rewrite the measurement as  $Y = C_f X$ , where  $C_f$  represents the circulant matrix whose first column is  $f$ , and  $Y = [y_1, y_2, \dots, y_N]$  and  $X = [x_1, x_2, \dots, x_N]$  are  $n \times N$  matrices. Without structures, one can solve the problem by choosing any invertible circulant matrix  $C_f$  and compute  $X = C_f^{-1}Y$ . The fact that  $X$  is sparse narrows down the search space.

Even with sparsity, the problem suffers from inherent scale and shift ambiguities. Suppose  $\mathcal{S}_j : \mathbb{R}^n \rightarrow \mathbb{R}^n$  denotes a circular shift by  $j$  positions, i.e.,  $\mathcal{S}_j(x)_{(k)} = x_{(k-j)}$  for  $j, k \in [n]$ . Here we use  $x_{(j)}$  to denote the  $j$ -th entry of  $x \in \mathbb{R}^n$  (treated as modulo  $n$ ). Note that we have  $y_i = x_i \otimes f = (\alpha \mathcal{S}_j(x_i)) \otimes (\alpha^{-1} \mathcal{S}_{-j}(f))$  for every nonzero  $\alpha \in \mathbb{R}$  and  $j \in [n]$ . Therefore, MSBD has equivalent solutions generated by scaling and circularly shifting  $\{x_i\}_{i=1}^n$  and  $f$ .

Throughout this paper, we assume that the circular convolution with the signal  $f$  is invertible, i.e., there exists a filter  $g$  such that  $f \otimes g = e_1$  (the first standard basis vector). Equivalently,  $C_f$  is an invertible matrix, and the discrete Fourier transform (DFT) of  $f$  is nonzero everywhere. Since  $y_i \otimes g = x_i \otimes f \otimes g = x_i$ , one can find  $g$  by solving the following optimization problem:

$$(P0) \quad \min_{h \in \mathbb{R}^n} \frac{1}{N} \sum_{i=1}^N \|C_{y_i} h\|_0, \quad \text{s.t. } h \neq 0.$$

The constraint eliminates the trivial solution that is 0. If the solution to MSBD is unique up to the aforementioned ambiguities, then the only minimizers of (P0) are  $h = \alpha \mathcal{S}_j$  ( $\alpha \neq 0, j \in [n]$ ).

## 2.2. Smooth Formulation

Minimizing the non-smooth  $\ell_0$  “norm” is usually challenging. Instead, one can choose a smooth surrogate function for sparsity. It is well-known that minimizing the  $\ell_1$  norm can lead to sparse solutions [31]. An intuitive explanation is that the sparse points on the unit  $\ell_2$  sphere (which we call unit sphere from now on) have the smallest  $\ell_1$  norm. As demonstrated in Figure 1, these sparse points also have the largest  $\ell_4$  norm. Therefore, maximizing the  $\ell_4$  norm, a surrogate for the “spikiness” [32] of a vector, is akin to minimizing its sparsity.

Here, we make two observations: (1) one can eliminate the scaling ambiguity by restricting  $h$  to the unit sphere  $S^{n-1}$ ; (2) sparse recovery can be achieved by maximizing  $\|\cdot\|_4^4$ . Based on these observations, we adopt the following optimization problem:

$$(P1) \quad \min_{h \in \mathbb{R}^n} -\frac{1}{4N} \sum_{i=1}^N \|C_{y_i} R h\|_4^4, \quad \text{s.t. } \|h\| = 1.$$

The matrix  $R := (\frac{1}{\theta n N} \sum_{i=1}^N C_{y_i}^T C_{y_i})^{-1/2} \in \mathbb{R}^{n \times n}$  is a preconditioner, where  $\theta$  is a parameter that is proportional to the sparsity level of  $\{x_i\}_{i=1}^N$ . In Section 3, under specific probabilistic assumptions on  $\{x_i\}_{i=1}^N$ , we explain how the preconditioner  $R$  works.

Problem (P1) can be solved using first-order or second-order optimization methods over Riemannian manifolds. The main result of this paper provides a geometric view of the objective function over the sphere  $S^{n-1}$  (see Figure 2). We show that some off-the-shelf optimization methods can be used to obtain a solution  $\hat{h}$  close to a scaled and circularly shifted version of the ground truth. Specifically,

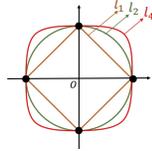


Fig. 1. Unit  $\ell_1$ ,  $\ell_2$ , and  $\ell_4$  spheres in 2-D.

$\hat{h}$  satisfies  $C_f R \hat{h} \approx \pm e_j$  for some  $j \in [n]$ , i.e.,  $R \hat{h}$  is approximately a signed and shifted version of the inverse of  $f$ . Given solution  $\hat{h}$  to (P1), one can recover  $f$  and  $x_i$  ( $i = 1, 2, \dots, N$ ) as follows:

$$\hat{f} = \mathcal{F}^{-1}[\mathcal{F}(R \hat{h})^{\circ -1}], \quad \hat{x}_i = C_{y_i} R \hat{h}. \quad (1)$$

Here, we use  $x^{\circ -1}$  to denote the entrywise inverse of  $x$ .

## 3. GLOBAL GEOMETRIC VIEW

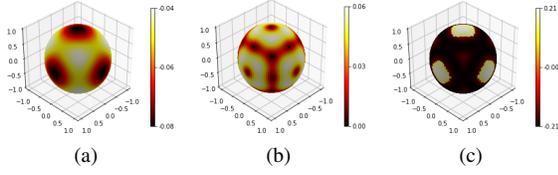
In this paper, we assume that  $\{x_i\}_{i=1}^N$  are random sparse vectors, and  $f$  is invertible:

- (A1) The channels  $\{x_i\}_{i=1}^N$  follow a Bernoulli-Rademacher model. More precisely,  $x_{i(j)} = A_{ij} B_{ij}$  are independent random variables,  $B_{ij}$ 's follow a Bernoulli distribution  $\text{Ber}(\theta)$ , and  $A_{ij}$ 's follow a Rademacher distribution (taking values 1 and  $-1$ , each with probability  $1/2$ ).
- (A2) The circular convolution with the signal  $f$  is invertible. We use  $\kappa$  to denote the condition number of  $f$ , which is defined as  $\kappa := \frac{\max_j |(\mathcal{F}f)_{(j)}|}{\min_k |(\mathcal{F}f)_{(k)}|} = \frac{\sigma_1(C_f)}{\sigma_n(C_f)}$ , i.e., the ratio of the largest and smallest magnitudes of the DFT of  $f$ .

The Bernoulli-Rademacher model is a special case of the Bernoulli-sub-Gaussian models. The derivation in this paper can be repeated for other sub-Gaussian nonzero entries, with different tail bounds. We use the Rademacher distribution for simplicity.

Let  $\phi(x) = -\frac{1}{4} \|x\|_4^4$ . Its gradient and Hessian are defined by  $\nabla \phi(x)_{(j)} = -x_j^3$ , and  $H_{\phi}(x)_{(jk)} = -3x_j^2 \delta_{jk}$ . (We use  $H_{(jk)}$  to denote the entry of  $H \in \mathbb{R}^{n \times n}$  in the  $j$ -th row and  $k$ -th column, and use  $\delta_{jk}$  to denote the Kronecker delta.) Then the objective function in (P1) is  $L(h) = \frac{1}{N} \sum_{i=1}^N \phi(C_{y_i} R h)$ , where  $R = (\frac{1}{\theta n N} \sum_{i=1}^N C_{y_i}^T C_{y_i})^{-1/2}$ . The gradient and Hessian are  $\nabla_L(h) = \frac{1}{N} \sum_{i=1}^N R^T C_{y_i}^T \nabla \phi(C_{y_i} R h)$ , and  $H_L(h) = \frac{1}{N} \sum_{i=1}^N R^T C_{y_i}^T H_{\phi}(C_{y_i} R h) C_{y_i} R$ . Since  $L(h)$  is to be minimized over  $S^{n-1}$ , we use optimization methods over Riemannian manifolds [33]. To this end, we define the tangent space at  $h \in S^{n-1}$  as  $\{z \in \mathbb{R}^n : z \perp h\}$ . We study the Riemannian gradient and Riemannian Hessian of  $L(h)$  (gradient and Hessian along the tangent space at  $h \in S^{n-1}$ ):  $\hat{\nabla}_L(h) = P_{h^\perp} \nabla_L(h)$ , and  $\hat{H}_L(h) = P_{h^\perp} H_L(h) P_{h^\perp} - \langle \nabla_L(h), h \rangle P_{h^\perp}$ , where  $P_{h^\perp} = I - h h^T$  is the projection onto the tangent space at  $h$ . We refer the readers to [33] for a more comprehensive discussion of these concepts.

The toy example in Figure 2 demonstrates the geometric structure of the objective function on  $S^{n-1}$ . (As shown later, the quantity  $\mathbb{E}L''(h)$  is, up to an unimportant rotation of the coordinate system, a good approximation to  $L(h)$ .) The local minima correspond to signed shifted versions of the ground truth (Figure 2(a)). The Riemannian gradient is zero at stationary points, including local minima, saddle points, and local maxima of the objective function when restricted to the sphere  $S^{n-1}$ . (Figure 2(b)). The Riemannian Hessian is positive definite in the neighborhoods of local minima, and has at least one strictly negative eigenvalue in the neighborhoods of local maxima and saddle points (Figure 2(c)). We say that a stationary point is a “strict saddle point” if the Riemannian Hessian has at least one strictly negative eigenvalue. Our main result Theorem 3.1 formalizes the observation that  $L(h)$  only has two types of stationary points: (1) local minima, which are close to signed shifted versions of the ground truth, and (2) strict saddle points. Please refer to the supplementary result for the full proof.



**Fig. 2.** Geometric structure of the objective function over the sphere. For  $n = 3$ , we plot the following quantities on the sphere  $S^2$ : (a)  $\mathbb{E}L''(h)$ , (b)  $\|\mathbb{E}\widehat{\nabla}_{L''}(h)\|$ , and (c)  $\min_{z \perp h, \|z\|=1} z^\top \mathbb{E}\widehat{H}_{L''}(h)z$ .

**Theorem 3.1.** *Suppose Assumptions (A1) and (A2) are satisfied, and the Bernoulli probability satisfies  $\frac{1}{n} \leq \theta < \frac{1}{3}$ . Let  $\kappa$  be the condition number of  $f$ , and let  $\rho < 10^{-3}$  be a small tolerance constant. There exist constants  $c_1, c'_1, c_2, c'_2 > 0$  (depending only on  $\theta$ ), such that: if  $N > \max\{\frac{c_1 n^9}{\rho^4} \log \frac{n}{\rho}, \frac{c_2 \kappa^8 n^8}{\rho^4} \log n\}$ , then with probability at least  $1 - n^{-c'_1} - n^{-c'_2}$ , every local minimum  $h^*$  in (P1) is close to a signed shifted version of the ground truth. I.e., for some  $j \in [n]$ :  $\|C_f R h^* \pm e_j\| \leq 2\kappa\sqrt{\rho}$ . Moreover, one can partition  $S^{n-1}$  into three sets  $\mathcal{H}_1, \mathcal{H}_2$ , and  $\mathcal{H}_3$ , which, for some  $c(n, \theta, \rho) > 0$ , satisfy:*

- $L(h)$  is strongly convex in  $\mathcal{H}_1$ , i.e.,  $\min_{z: \|z\|=1} z^\top \widehat{H}_L(h)z \geq c(n, \theta, \rho) > 0$ .
- $L(h)$  has negative curvature in  $\mathcal{H}_2$ , i.e.,  $\min_{z: \|z\|=1} z^\top \widehat{H}_L(h)z \leq -c(n, \theta, \rho) < 0$ .
- $L(h)$  has a descent direction in  $\mathcal{H}_3$ , i.e.,  $\|\widehat{\nabla}_L(h)\| \geq c(n, \theta, \rho) > 0$ .

Clearly, all the stationary points of  $L(h)$  on  $S^{n-1}$  belong to  $\mathcal{H}_1$  or  $\mathcal{H}_2$ . The stationary points in  $\mathcal{H}_1$  are local minima, and the stationary points in  $\mathcal{H}_2$  are strict saddle points.

*Proof Sketch.* Note that  $R = (\frac{1}{\theta n N} \sum_{i=1}^N C_{y_i}^\top C_{y_i})^{-1/2}$  asymptotically converges to  $(C_f^\top C_f)^{-1/2}$  as  $N$  increases. Therefore,  $L(h)$  can be approximated by

$$\begin{aligned} L'(h) &= \frac{1}{N} \sum_{i=1}^N \phi(C_{y_i} (C_f^\top C_f)^{-1/2} h) \\ &= \frac{1}{N} \sum_{i=1}^N \phi(C_{x_i} C_f (C_f^\top C_f)^{-1/2} h). \end{aligned}$$

Since  $C_f (C_f^\top C_f)^{-1/2}$  is an orthogonal matrix, one can study the objective function  $L''(h') = \frac{1}{N} \sum_{i=1}^N \phi(C_{x_i} h')$  with  $h' = C_f (C_f^\top C_f)^{-1/2} h$ , which is a rotated version of  $L'(h)$  on the sphere. Our analysis consists of three parts:

(1) *Geometric structure of  $\mathbb{E}L''$ :* We first bound the smallest nonzero eigenvalue ( $\min_{z: \|z\|=1, z \perp h} z^\top \mathbb{E}\widehat{H}_{L''}(h)z$ ) of the Riemannian Hessian, which is strictly positive near its local minima, and strictly negative near all other stationary points (the strict saddle points). At the same time, at all other points on  $S^{n-1}$  (the points further away from stationary points), the Riemannian gradient of  $\mathbb{E}L''$  is bounded away from zero.

(2) *Deviation of  $L''$  (or its rotated version  $L'$ ) from  $\mathbb{E}L''$ :* We bound  $\|\widehat{\nabla}_{L''}(h) - \mathbb{E}\widehat{\nabla}_{L''}(h)\|$  and  $\|\widehat{H}_{L''}(h) - \mathbb{E}\widehat{H}_{L''}(h)\|$  using the matrix Bernstein inequality and union bounds.

(3) *Difference between  $L$  and  $L'$ :* We bound  $\|\widehat{\nabla}_L(h) - \widehat{\nabla}_{L'}(h)\|$ , and  $\|\widehat{H}_L(h) - \widehat{H}_{L'}(h)\|$  using the matrix Bernstein inequality and Lipschitz continuity of  $\widehat{\nabla}_L(h)$  and  $\widehat{H}_L(h)$ .

Theorem 3.1 follows by combining the above results.  $\square$

Theorem 3.1 is a sufficient condition for favorable geometric properties, and may not be necessary for successful recovery of MSBD. It holds uniformly for all  $f$ , which is more demanding than the sampling requirement for a specific  $f$ . For these reasons, the sample complexity bound on the number  $N$  of channels is pessimistic, especially when compared to the empirical success of our optimization method introduced in the rest of the paper.

## 4. OPTIMIZATION METHOD

Recently, first-order methods have been shown to escape strict saddle points with random initialization [34, 35]. In this paper, we use the manifold gradient descent algorithm studied by Lee et al. [30]. One can initialize the algorithm with a random  $h^{(0)}$ , and use the following iterative update:

$$h^{(t+1)} = \mathcal{A}(h^{(t)}) := P_{S^{n-1}}(h^{(t)} - \gamma \widehat{\nabla}_L(h^{(t)})). \quad (2)$$

Each iteration takes a Riemannian gradient descent step in the tangent space, and does a retraction by normalizing the iterate (projecting onto  $S^{n-1}$ ). Using the geometric structure introduced in Section 3, and some technical results in [29, 30], the following result gives a theoretical guarantee for manifold gradient descent for our formulation of MSBD: convergence to an accurate estimate (up to the inherent sign and shift ambiguity) of the true solution.

**Theorem 4.1.** *Suppose that the geometric structure in Theorem 3.1 is satisfied. If manifold gradient descent (2) is initialized with a random  $h^{(0)}$  drawn from a uniform distribution on  $S^{n-1}$ , and the step size is chosen as  $\gamma = \frac{1}{128n^3}$ , then (2) converges to a local minimum of  $L(h)$  on  $S^{n-1}$  almost surely. In particular, after at most  $T = \frac{4096n^8}{\theta^2(1-3\theta)^2\rho^4}$  iterations,  $h^{(T)} \in \mathcal{H}_1$ . Moreover, for some  $j \in [n]$*

$$\|C_f R h^{(T)} \pm e_j\| \leq 2\kappa\sqrt{\rho}.$$

**Corollary 4.2.** *If the conditions of Theorem 4.1 are satisfied, then the recovered  $\hat{f}$  and  $\hat{x}_i$  in (1), computed using the output of manifold gradient descent  $\hat{h} = h^{(T)}$ , satisfy (for some  $j \in [n]$ ):*

$$\frac{\|\hat{x}_i \pm \mathcal{S}_j(x_i)\|}{\|\hat{x}_i\|} \leq 2\kappa\sqrt{\rho n}, \quad \frac{\|\hat{f} \pm \mathcal{S}_{-j}(f)\|}{\|\hat{f}\|} \leq \frac{2\kappa\sqrt{\rho n}}{1 - 2\kappa\sqrt{\rho n}}.$$

Theorem 4.1 and Corollary 4.2 show that, with a random initialization and a fixed step size, manifold gradient descent outputs, in polynomial time, a solution that is close to a signed and shifted version of the ground truth.

## 5. NUMERICAL EXPERIMENTS

### 5.1. Deconvolution with Synthetic Data

In this section, we examine the empirical performance of manifold gradient descent (2) in solving MSBD (P1). We synthesize  $\{x_i\}_{i=1}^N$  following the Bernoulli-Rademacher model, and synthesize  $f$  following a Gaussian distribution  $N(\mathbf{0}_{n \times 1}, I_n)$ . In all experiments, we run manifold gradient descent for  $T = 100$  iterations, with a fixed step size of  $\gamma = 0.1$ .

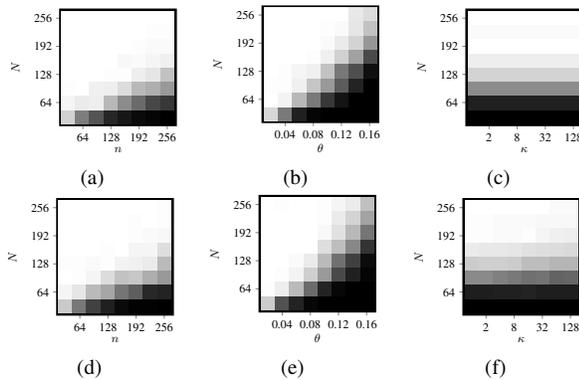
Recall that the desired  $h$  is a signed shifted version of the ground truth, i.e.,  $C_f R h = \pm e_j$  ( $j \in [n]$ ). Therefore, to evaluate the accuracy of the output  $h^{(T)}$ , we compute  $C_f R h^{(T)}$  with the true  $f$ , and

declare successful recovery if  $\|C_f R h^{(T)}\|_\infty / \|C_f R h^{(T)}\| > 0.95$ , or equivalently, if  $\max_{j \in [n]} |\cos \angle(C_f R h^{(T)}, e_j)| > 0.95$ . We compute the success rate based on 100 Monte Carlo instances. In a typical successful instance,  $h^{(t)}$  converges to an accurate estimate of the ground truth after about 50 iterations.

In the first experiment, we fix  $\theta = 0.1$  (sparsity level, mean of the Bernoulli distribution), and run experiments with  $n = 32, 64, \dots, 256$  and  $N = 32, 64, \dots, 256$  (see Figure 3(a)). In the second experiment, we fix  $n = 256$ , and run experiments with  $\theta = 0.02, 0.04, \dots, 0.16$  and  $N = 32, 64, \dots, 256$  (see Figure 3(b)). The empirical phase transitions suggest that, for sparsity level relatively small (e.g.,  $\theta < 0.16$ ), there exist a constant  $c > 0$  such that manifold gradient descent can recover a signed shifted version of the ground truth with  $N \geq cn\theta$ .

In the third experiment, we examine the phase transition with respect to  $N$  and the condition number  $\kappa$  of  $f$ , which is the ratio of the largest and smallest magnitudes of its DFT. To synthesize  $f$  with specific  $\kappa$ , we generate the DFT  $\hat{f}$  of  $f$  that is random with the following distribution: (1) The DFT  $\hat{f}$  is symmetric, i.e.,  $\hat{f}_{(j)} = \hat{f}_{(n+2-j)}$ , so that  $f$  is real. (2) The phase of  $\hat{f}_{(j)}$  follows a uniform distribution on  $[0, 2\pi)$ , except for the phases of  $\hat{f}_{(1)}$  and  $\hat{f}_{(n/2+1)}$  (if  $n$  is even), which are always 0 for symmetry. (3) The gains of  $\hat{f}$  follows a uniform distribution on  $[1, \kappa]$ . We fix  $n = 256$  and  $\theta = 0.1$ , and run experiments with  $\kappa = 1, 2, 4, \dots, 128$  and  $N = 32, 64, \dots, 256$  (see Figure 3(c)). The phase transition suggests that the number  $N$  for successful empirical recovery is not sensitive to the condition number  $\kappa$ .

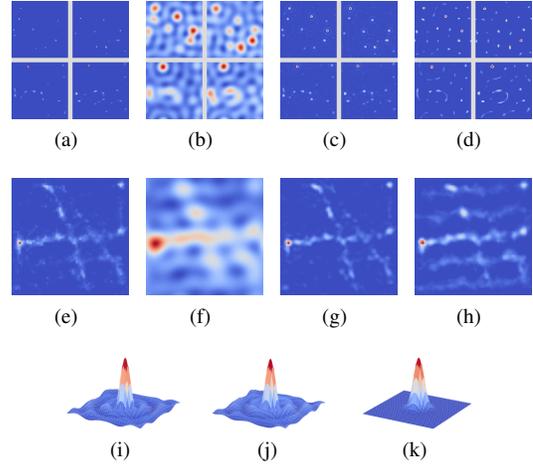
Manifold gradient descent is robust against noise. We repeat the above experiments with noisy measurements:  $y_i = x_i \otimes f + \sigma \varepsilon_i$ , where  $\varepsilon_i$  follows a Gaussian distribution  $N(\mathbf{0}_{n \times 1}, I_n)$ . The phase transitions for  $\sigma = 0.1\sqrt{n\theta}$  (SNR  $\approx 20$  dB) are shown in Figure 3(d), 3(e), and 3(f). For a reasonable noise level, the number  $N$  of noisy measurements we need to accurately recover a signed shifted version of the ground truth is roughly the same as with noiseless measurements.



**Fig. 3.** Empirical phase transition (grayscale values represent success rates). (a) and (d) show the phase transitions of  $N$  versus  $n$ . (b) and (e) show the phase transitions of  $N$  versus  $\theta$ . (c) and (f) show the phase transitions of  $N$  versus  $\kappa$ . (a)(b)(c) are the results for the noiseless case. (d)(e)(f) are the results for SNR  $\approx 20$  dB.

## 5.2. Super-Resolution Fluorescence Microscopy

Manifold gradient descent can be applied to deconvolution of time resolved fluorescence microscopy images. The goal is to recover



**Fig. 4.** Super-resolution fluorescence microscopy experiment using manifold gradient descent. (a) True images. (b) Observed images. (c) Recovered images using blind deconvolution. (d) Recovered images using non-blind deconvolution and a miscalibrated PSF. (e)(f)(g)(h) are average images of (a)(b)(c)(d). (i) True PSF. (j) Recovered PSF using blind deconvolution. (k) Miscalibrated PSF used in non-blind deconvolution. All images in this figure are of the same size ( $64 \times 64$ ).

sharp images  $x_i$ 's from observations  $y_i$ 's that are blurred by an unknown PSF  $f$ .

We use a publicly available microtubule dataset [22], which contains  $N = 626$  images (Figure 4(a)). Since fluorophores are turned on and off stochastically, the images  $x_i$ 's are random sparse samples of the  $64 \times 64$  microtubule image (Figure 4(e)). The observations  $y_i$ 's (Figure 4(b), 4(f)) are synthesized by circular convolutions with the PSF in Figure 4(i). The recovered images (Figure 4(c), 4(g)) and kernel (Figure 4(j)) clearly demonstrate the effectiveness of our approach in this setting.

Blind deconvolution is less sensitive to instrument calibration error than non-blind deconvolution. If the PSF used in a non-blind deconvolution method fails to account for certain optic aberration, the resulting images may suffer from spurious artifacts. For example, if we use a miscalibrated PSF (Figure 4(k)) in non-blind image reconstruction using FISTA [36], then the recovered images (Figure 4(d), 4(h)) suffer from serious spurious artifacts.

## 6. CONCLUSION

In this paper, we study the geometric structure of multichannel sparse blind deconvolution over the unit sphere. Our theoretical analysis reveals that local minima of a sparsity promoting smooth objective function correspond to signed shifted version of the ground truth, and saddle points have strictly negative curvatures. Thanks to the favorable geometric properties of the objective, we can simultaneously recover the unknown signal and unknown channels from convolutional measurements using manifold gradient descent with a random initialization. In practice, many convolutional measurement models are subsampled in the spatial domain (e.g., image super-resolution) or in the frequency domain (e.g., radio astronomy). Studying the effect of subsampling on the geometric structure of multichannel sparse blind deconvolution is an interesting problem for future work.

## 7. REFERENCES

- [1] S. Cho and S. Lee, “Fast motion deblurring,” in *ACM Trans. Graphics (TOG)*. ACM, 2009.
- [2] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding blind deconvolution algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2011.
- [3] L. Xu, S. Zheng, and J. Jia, “Unnatural l0 sparse representation for natural image deblurring,” in *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013.
- [4] A. Ahmed, B. Recht, and J. Romberg, “Blind deconvolution using convex programming,” *IEEE Trans. Information Theory*, 2014.
- [5] S. Ling and T. Strohmer, “Self-calibration and biconvex compressive sensing,” *Inverse Problems*, 2015.
- [6] Y. Chi, “Guaranteed blind sparse spikes deconvolution via lifting and convex optimization,” *IEEE J. Selected Topics in Signal Processing*, 2016.
- [7] X. Li, S. Ling, T. Strohmer, and K. Wei, “Rapid, robust, and reliable blind deconvolution via nonconvex optimization,” *Applied and Computational Harmonic Analysis*, 2018.
- [8] K. Lee, Y. Li, M. Junge, and Y. Bresler, “Blind recovery of sparse signals from subsampled convolution,” *IEEE Trans. Information Theory*, 2017.
- [9] W. Huang and P. Hand, “Blind deconvolution by a steepest descent algorithm on a quotient manifold,” *arXiv preprint arXiv:1710.03309*, 2017.
- [10] Y. Zhang, Y. Lau, H.-W. Kuo, S. Cheung, A. Pasupathy, and J. Wright, “On the global geometry of sphere-constrained sparse blind deconvolution,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [11] L. Tong and S. Perreau, “Multichannel blind identification: from subspace to maximum likelihood methods,” *Proc. IEEE*, 1998.
- [12] H. She, R.-R. Chen, D. Liang, Y. Chang, and L. Ying, “Image reconstruction from phased-array data based on multichannel blind deconvolution,” *Magnetic resonance imaging*, 2015.
- [13] H. Zhang, D. Wipf, and Y. Zhang, “Multi-image blind deblurring using a coupled adaptive sparse prior,” in *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013.
- [14] L. Tong, G. Xu, and T. Kailath, “A new approach to blind identification and equalization of multipath channels,” in *Twenty-Fifth Asilomar Conference on Signals, Systems Computers*, 1991.
- [15] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, “Subspace methods for the blind identification of multichannel fir filters,” *IEEE Trans. Signal Processing*, 1995.
- [16] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approach to blind channel identification,” *IEEE Trans. Signal Processing*, 1995.
- [17] M. I. Gurelli and C. L. Nikias, “EVAM: an eigenvector-based algorithm for multichannel blind deconvolution of input colored signals,” *IEEE Trans. Signal Processing*, 1995.
- [18] K. Lee, F. Kraemer, and J. Romberg, “Spectral methods for passive imaging: Non-asymptotic performance and robustness,” *arXiv preprint arXiv:1708.04343*, 2017.
- [19] N. Tian, S.-H. Byun, K. Sabra, and J. Romberg, “Multichannel myopic deconvolution in underwater acoustic channels via low-rank recovery,” *J. Acoustical Society of America*, 2017.
- [20] K. F. Kaarensen and T. Taxt, “Multichannel blind deconvolution of seismic signals,” *Geophysics*, 1998.
- [21] D. R. Gitelman, W. D. Penny, J. Ashburner, and K.J. Friston, “Modeling regional and psychophysiological interactions in fmri: the importance of hemodynamic deconvolution,” *Neuroimage*, 2003.
- [22] E. A. Mukamel, H. Babcock, and X. Zhuang, “Statistical deconvolution for superresolution fluorescence microscopy,” *Biophysical journal*, 2012.
- [23] L. Wang and Y. Chi, “Blind deconvolution from multiple sparse inputs,” *IEEE Signal Processing Letters*, 2016.
- [24] Y. Li, K. Lee, and Y. Bresler, “Blind gain and phase calibration via sparse spectral methods,” *arXiv preprint arXiv:1712.00111*, 2017.
- [25] J. Sun, Q. Qu, and J. Wright, “Complete dictionary recovery over the sphere i: Overview and the geometric picture,” *IEEE Trans. Information Theory*, 2017.
- [26] J. Sun, Q. Qu, and J. Wright, “A geometric analysis of phase retrieval,” *Foundations of Computational Mathematics*, 2017.
- [27] S. Mei, Y. Bai, and A. Montanari, “The landscape of empirical risk for non-convex losses,” *arXiv preprint arXiv:1607.06534*, 2016.
- [28] J. Sun, Q. Qu, and J. Wright, “Complete dictionary recovery over the sphere ii: Recovery by riemannian trust-region method,” *IEEE Trans. Information Theory*, 2017.
- [29] N. Boumal, P. A. Absil, and C. Cartis, “Global rates of convergence for nonconvex optimization on manifolds,” *arXiv preprint arXiv:1605.08101*, 2016.
- [30] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht, “First-order methods almost always avoid saddle points,” *arXiv preprint arXiv:1710.07406*, 2017.
- [31] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization,” *Proceedings of the National Academy of Sciences*, 2003.
- [32] Y. Zhang, H.-W. Kuo, and J. Wright, “Structured local optima in sparse blind deconvolution,” in *Proc. 10th NIPS Workshop on Optimization for Machine Learning (OPTML)*, 2017.
- [33] P-A Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [34] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, “Gradient descent only converges to minimizers,” in *Conference on Learning Theory*, 2016.
- [35] I. Panageas and G. Piliouras, “Gradient descent only converges to minimizers: Non-isolated critical points and invariant regions,” *arXiv preprint arXiv:1605.00405*, 2016.
- [36] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J Imaging Sciences*, 2009.