

On the Global Geometry of Sphere-Constrained Sparse Blind Deconvolution

Yuqian Zhang[†], Yenson Lau[†], Han-wen Kuo[†], Sky Cheung[◊], Abhay Pasupathy[◊], John Wright^{†*}

[†] Department of Electrical Engineering, Columbia University, New York, USA

[◊] Department of Physics, Columbia University, New York, USA

{yz2409, yl3027, hk2673, scc2134, apn2108, jw2966}@columbia.edu

Abstract

Blind deconvolution is the problem of recovering a convolutional kernel a_0 and an activation signal x_0 from their convolution $y = a_0 \circledast x_0$. This problem is ill-posed without further constraints or priors. This paper studies the situation where the nonzero entries in the activation signal are sparsely and randomly populated. We normalize the convolution kernel to have unit Frobenius norm and cast the sparse blind deconvolution problem as a nonconvex optimization problem over the sphere. With this spherical constraint, every spurious local minimum turns out to be close to some signed shift truncation of the ground truth, under certain hypotheses. This benign property motivates an effective two stage algorithm that recovers the ground truth from the partial information offered by a suboptimal local minimum. This geometry-inspired algorithm recovers the ground truth for certain microscopy problems, also exhibits promising performance in the more challenging image deblurring problem. Our insights into the global geometry and the two stage algorithm extend to the convolutional dictionary learning problem, where a superposition of multiple convolution signals is observed.

1. Introduction

Blind deconvolution aims to recover two unknown signals: a kernel a_0 and some underlying signal x_0 from their convolution $y = a_0 \circledast x_0$. Blind deconvolution is ill-posed in general: there are infinitely many pairs of signals rendering the same convolution. To render the problem well-posed, one may exploit prior knowledge about the structure of a_0 and x_0 . For example, the underlying signal x_0 is *sparse* in many realistic applications:

Microscopy data analysis: In the crystal lattice of nanoscale materials, there exist randomly and sparsely distributed “defects”, whose locations and signatures encode

*The authors gratefully acknowledge support from NSF 1343282, NSF CCF 1527809, and NSF IIS 1546411

crucial information about the electronic structure of the material. Accurate recovery of such information can facilitate investigation of the detailed structure of materials [4].

Neural spike sorting: Neurons communicate by firing brief voltage spikes, whose characteristics reflect important features of the neuron. These spikes occur randomly and sparsely in time. Neurophysiologists are interested in assigning stereotyped spikes to putative cells, as well as in knowing their respective spike times [7, 14].

Image deblurring: Motion blur can be modeled as the convolution of a latent sharp image and a kernel capturing the motion of the camera, usually assumed to be invariant across the image [8]. The inverse process of recovering the original sharp image from a blurry image has been widely studied [6, 12]. Many well-performing approaches leverage the observation that sharp natural images typically have (approximately) sparse gradients [3, 13, 18].

All of these applications lead to instances of the *sparse blind deconvolution* (SBD) problem. The dominant algorithmic approach to SBD involves *nonconvex optimization*¹. Nonconvex formulations for deconvolution can be derived via several probabilistic formalisms (ML/MAP, VB, ect.), or simply from heuristics. For example, in image deblurring, the kernel a can be modeled as residing on a simplex [9, 11, 13, 16]. This is natural from a modeling prospective², but problematic for optimization: natural formulations of deconvolution over the simplex admit trivial global minimizers (corresponding to spiky convolutional kernels $a = \delta$) [2, 18], which provide no information about the ground truth. Practical remedies for this problem include exploiting additional data priors [9, 16, 24] or careful initialization via edge restoration or multi-scale refinement [11, 25], to avoid the trivial spiky global minima.

In contrast, motivated by a careful comparison of MAP

¹In signal processing, a number of elegant *convex* relaxations of the problem have been developed [1, 5, 15]. However, these approaches typically require stronger prior information (subspace constraint rather than sparsity) or exhibit suboptimal scalings.

²Since entries of a roughly represent the fraction of the camera exposure time at a given location.

and VB approaches, [23, 26] propose to instead constrain \mathbf{a} to have unit Frobenius norm – i.e., to reside on a high-dimensional sphere.³ This choice is arguably more appropriate for applications – such as microscopy – in which the kernel \mathbf{a} have negative entries. For image deblurring, \mathbf{a} can be assumed to be nonnegative, and the sphere constraint seems less natural from a modeling perspective.

In this paper, we study the geometry of sphere-constrained sparse blind deconvolution. Our goal is to understand when simple algorithms based on nonconvex optimization can exactly recover the convolutional kernel \mathbf{a} and \mathbf{x} . This goal is motivated by the applications described above – in particular, microscopy data analysis – in which there is a strong, physical sparsity prior and a clear, physical notion of the ground truth. We develop our theory and algorithms under the assumption that \mathbf{a} is a short kernel, and that \mathbf{x} is sparsely and randomly supported. We demonstrate through a theoretical analysis of certain (idealized) cases and lots of numerical experiments that when these assumptions are satisfied, the proposed algorithm correctly recovers \mathbf{a} , and hence \mathbf{x} . These results stem from a striking geometric property of sphere-constrained SBD: although the problem is still nonconvex, every local minimizer $\bar{\mathbf{a}}$ is very close to a signed shift-truncation of the ground truth kernel \mathbf{a}_0 . This observation provides a geometric explanation of how the sphere constraint can facilitate sparse blind deconvolution.

The remainder of this paper is organized as follows. Section 2 discusses the intrinsic symmetry associated to the convolutional operator and its implication on the geometry of sphere-constrained SBD. Section 3 introduces the optimization-based two stage algorithm and some related technical details. Section 4 discusses two other important extensions in image deblurring and convolutional dictionary learning. Section 5 gives experimental corroboration of our theory, and shows promising results on microscopy data analysis, image deblurring, and convolutional dictionary learning. Section 6 discusses directions for future work.

For simplicity, we assume the convolutional signals are one dimensional in both problem formulation and technical proof, while all the results apply to two dimensional signals. Throughout this paper, vectors $\mathbf{v} \in \mathbb{R}^k$ are indexed as $\mathbf{v} = [v_0, v_1, \dots, v_{k-1}]$, and $[\cdot]_m$ denotes the modulo operator of m . We use $\|\cdot\|$ to denote the operator norm, and $\|\cdot\|_p$ to denote the entry wise ℓ^p norm. A projection on the Frobenius sphere is denoted with $\mathcal{P}_{\mathbb{S}}[\cdot] = \frac{\cdot}{\|\cdot\|_F}$, and a projection onto subset with index I is denoted with $(\cdot)_I$.

³[23] contains a wealth of additional ideas about the role of sparsity-promoting priors in obtaining good local minima, and on the probabilistic underpinnings of deconvolution. Our experiments support the viewpoint that the key insight in [23] is the role of the spherical constraint in avoiding bad minimizers.

2. Symmetry and Global Geometry

Without loss of generality, we assume the observation data \mathbf{y} is generated via a circular convolution \circledast of the ground truth $\mathbf{a}_0 \in \mathbb{R}^k$ and $\mathbf{x}_0 \in \mathbb{R}^m$:

$$\mathbf{y}(\mathbf{a}_0, \mathbf{x}_0) = \mathbf{a}_0 \circledast \mathbf{x}_0 = \widetilde{\mathbf{a}_0} \circledast \mathbf{x}_0 \in \mathbb{R}^m. \quad (1)$$

Here, $\widetilde{\mathbf{a}_0} \in \mathbb{R}^m$ denote the zero padded m -length version of \mathbf{a}_0 , which can be expressed as $\widetilde{\mathbf{a}_0} = \iota \mathbf{a}_0$ with $\iota : \mathbb{R}^k \rightarrow \mathbb{R}^m$ be a zero padding operator. Its adjoint $\iota^* : \mathbb{R}^m \rightarrow \mathbb{R}^k$ acts as a projection to lower dimensional space by keeping the first k components. Equivalently, we can write

$$\mathbf{y}(\mathbf{a}_0, \mathbf{x}_0) = \mathbf{C}_{\widetilde{\mathbf{a}_0}} \mathbf{x}_0 = \mathbf{C}_{\mathbf{x}_0} \widetilde{\mathbf{a}_0}. \quad (2)$$

Here, $\mathbf{C}_v \in \mathbb{R}^{m \times m}$ is the circulant matrix generated from vector v , whose j -th column is a cyclic shift $s_{j-1}[v]$ of the vector v :

$$s_\tau[v](i) = v([i - \tau]_m), \quad \forall i \in [0, \dots, m-1]. \quad (3)$$

2.1. Symmetries and Symmetry Breaking

The SBD problem exhibits a *scaled-shift symmetry*, which derives from the symmetries of the convolution operator. Namely, given a pair $(\mathbf{a}_0, \mathbf{x}_0)$ satisfying $\mathbf{y} = \mathbf{a}_0 \circledast \mathbf{x}_0$, for any nonzero scalar α and integer τ

$$\mathbf{y} = (\alpha s_\tau[\widetilde{\mathbf{a}_0}]) \circledast (\alpha^{-1} s_{-\tau}[\mathbf{x}_0]). \quad (4)$$

Note that a scaled shift $\alpha^{-1} s_{-\tau}[\mathbf{x}_0]$ of a sparse signal \mathbf{x}_0 remains sparse, and that a scaled shift $\alpha s_\tau[\widetilde{\mathbf{a}_0}]$ of a length- k kernel \mathbf{a}_0 still has k -nonzero entries. So, these symmetries are intrinsic to the SBD problem, as formulated here. We can only hope to recover $(\mathbf{a}_0, \mathbf{x}_0)$ up to this symmetry.

The presence of nontrivial symmetries is a hallmark of bilinear problems arising in practice – see, e.g., [20, 21] for examples from dictionary learning and generalized phase retrieval. Symmetries render straightforward approaches to convexify the problem ineffective.⁴ They also raise challenges for nonconvex optimization: equivalent symmetric solutions correspond to multiple disconnected global optima. This creates a very complicated objective landscape, which could potentially also contain spurious local optimizers. Certain highly symmetric nonconvex problems arising in signal processing do not exhibit spurious minimizers [20, 21], however, proving this can be challenging.

Symmetry breaking. We employ a weak symmetry breaking mechanism by constraining $\mathbf{a} \in \mathbb{S}^{k-1}$ ⁵: we reduce the

⁴Given any set of points where the convex objective function achieves equal values, the function value will be no larger at any convex combination of them.

⁵This is motivated in part by [20], which demonstrates that a certain formulation of the dictionary learning problem over the sphere has no spurious local minimizers, even for relatively dense target representations. The “simplex constrained” analogue of that work, which optimizes over hyperplanes, requires the target solution to be much sparser [19].

scale ambiguity to a sign ambiguity, by constraining \mathbf{a} to have unit Frobenius norm; we mitigate the shift ambiguity by constraining \mathbf{a} to be supported on the first k entries.

In general, $s_\tau[\widetilde{\mathbf{a}}_0]$ are not be supported on the first k entries, hence constraining \mathbf{a} to be supported on the first k entries removes the shift symmetry. However, effects of such shift symmetry still persist. Since the restriction $\iota^* s_\tau[\widetilde{\mathbf{a}}_0]$ to the first k entries can be convolved with the sparse signal $s_{-\tau}[\mathbf{x}_0]$ to produce a near approximation to \mathbf{y} :

$$(\iota^* s_\tau[\widetilde{\mathbf{a}}_0]) \circledast s_{-\tau}[\mathbf{x}_0] \approx \mathbf{y}, \quad (5)$$

especially when the shift $|\tau|$ is small. We will see that (i) these symmetric solutions $\iota^* s_\tau[\widetilde{\mathbf{a}}_0]$ persist as local minima of a natural optimization formulation of the SBD problem, but that (ii) under conditions, these are the *only* local minima.

2.2. Global Geometry on the Sphere

We study the following objective function, which can be viewed as balancing sparsity of \mathbf{x} with fidelity to the observation \mathbf{y} :⁶

$$\min_{\mathbf{a} \in \mathbb{S}^{k-1}, \mathbf{x}} \psi(\mathbf{a}, \mathbf{x}) \doteq \frac{1}{2} \|\mathbf{y} - \mathbf{a} \circledast \mathbf{x}\|_2^2 + \lambda r(\mathbf{x}). \quad (6)$$

When \mathbf{x}_0 is long and random, it's more convenient to study this function through its “marginalization”

$$\varphi(\mathbf{a}) \doteq \min_{\mathbf{x}} \psi(\mathbf{a}, \mathbf{x}), \quad (7)$$

which is defined over the sphere \mathbb{S}^{k-1} .

In Figure 1, we plot the function value of $\varphi(\mathbf{a})$ on the sphere $\mathbf{a} \in \mathbb{S}^2$: red and blue imply larger and smaller objective value respectively and there are several local minima. For this highly nonconvex function, the ground truth \mathbf{a}_0 achieves the global minimum, while other local minima $\bar{\mathbf{a}}$ are very close to certain signed shift truncations of the ground truth. Figure 1 (right) exhibits an example of a local minimum in a higher-dimensional problem.

Analysis under Restricted Settings. Demonstrating that this observation holds in general is challenging: for most reasonable choices of the regularizer r , there is no closed form expression for the objective $\varphi(\mathbf{a})$. We develop an analysis under several simplifying assumptions. Throughout, we let $r(\mathbf{x})$ be the ℓ^1 norm, although similar conclusions hold for other sparsifying regularizers. With this choice, we can simplify the objective φ by dividing the sphere \mathbb{S}^{k-1} via the sign-support pattern of the minimizing $\mathbf{x}^*(\mathbf{a})$:

$$\mathbf{x}^*(\mathbf{a}) = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{a} \circledast \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (8)$$

Let I and σ denote the support and sign of \mathbf{x}^*

$$I = \text{supp}(\mathbf{x}^*), \quad \sigma = \text{sign}(\mathbf{x}^*), \quad (9)$$

⁶Similar formulation can be found in lot of sparse representation problems [17].

then the whole sphere can be divided via the sign support pattern

$$\mathbb{S}^{k-1} = \bigcup_{\sigma} \mathcal{R}_{\sigma}, \quad \mathcal{R}_{\sigma} = \{\mathbf{a} \mid \text{sign}(\mathbf{x}^*(\mathbf{a})) = \sigma\}. \quad (10)$$

On each \mathcal{R}_{σ} where the sign support pattern σ remains the same, the stationary condition for minimizer $\mathbf{x}^*(\mathbf{a})$ implies

$$\mathbf{x}_I^*(\mathbf{a}) = (\mathbf{C}_a^* \mathbf{C}_a)_I^{-1} (\mathbf{C}_a^* \mathbf{y} - \lambda \sigma)_I. \quad (11)$$

Plugging above expression back to the original objective function $\varphi(\mathbf{a})$ yields

$$\begin{aligned} \varphi_{\sigma}(\mathbf{a}) &= -\frac{1}{2} (\mathbf{C}_a^* \mathbf{y} - \lambda \sigma)_I^* (\mathbf{C}_a^* \mathbf{C}_a)_I^{-1} (\mathbf{C}_a^* \mathbf{y} - \lambda \sigma)_I \\ &\quad + \frac{1}{2} \|\mathbf{y}\|_2^2. \end{aligned} \quad (12)$$

Although the objective function $\varphi(\mathbf{a})$ can be substantially simplified by removing the \mathbf{x} variable in this way, it still maintains a complicated dependence on \mathbf{a} . To obtain some preliminary insights, we make two simplifications for easier calculation whilst preserving similar geometrical properties:

Simplification I: $\mathbf{x}_0 = \delta$. We maximally simplify the underlying sparse signal as a single spike δ and the observation will be $\mathbf{y} = \mathbf{a}_0 \circledast \mathbf{x}_0 = \widetilde{\mathbf{a}}_0$. This case itself is trivial, but its function geometry is a basic but important case to be understood. Also, as the dimension of the sparse signal \mathbf{x}_0 increases, the function geometry will converge to this case.

Simplification II: $\mathbf{C}_a^* \mathbf{C}_a \rightarrow \mathbf{I}$. For a random $\mathbf{a} \in \mathbb{S}^{k-1}$, its expectation satisfies $\mathbb{E} \{ \mathbf{C}_a^* \mathbf{C}_a \} = \mathbf{I}$. Here, we simply use the identity matrix to replace any $\mathbf{C}_a^* \mathbf{C}_a$ and therefore reduce the complexity of Equation 12.

With these two simplifications, the original objective problem can be replaced with the following:

$$\text{minimize } \widehat{\varphi}(\mathbf{a}) \quad \text{subject to } \mathbf{a} \in \mathbb{S}^{k-1}, \quad (13)$$

where

$$\widehat{\varphi}(\mathbf{a}) \doteq \min_{\mathbf{x}} \frac{1}{2} \|\widetilde{\mathbf{a}}_0\|_2^2 + \frac{1}{2} \|\mathbf{x}\|_2^2 - \langle \widetilde{\mathbf{a}} \circledast \mathbf{x}, \widetilde{\mathbf{a}}_0 \rangle + \lambda \|\mathbf{x}\|_1.$$

In this case, the minimizing $\mathbf{x}^*(\mathbf{a})$ has a simple closed form solution:

$$\mathbf{x}^*(\mathbf{a}) = \text{SOFT}_{\lambda} [\mathbf{C}_a^* \widetilde{\mathbf{a}}_0] = \text{SOFT}_{\lambda} [\check{\mathbf{C}}_{\mathbf{a}_0}^* \iota \mathbf{a}], \quad (14)$$

here, $\text{SOFT}_{\lambda}[u] = \text{sign}(u) \max \{|u| - \lambda, 0\}$ is the entry-wise soft-thresholding operator and $\check{\mathbf{C}}_{\mathbf{a}_0} \in \mathbb{R}^{m \times m}$ is the reversed circulant matrix for \mathbf{a}_0 defined via

$$\check{\mathbf{C}}_{\mathbf{a}_0} = [\ s_0 [\widetilde{\mathbf{a}}_0] \ | \ s_{-1} [\widetilde{\mathbf{a}}_0] \ | \ \dots \ | \ s_{-(m-1)} [\widetilde{\mathbf{a}}_0] \]. \quad (15)$$

On a constant sign support pattern σ , $\widehat{\varphi}(\mathbf{a})$ can be written into a simpler quadratic form:

$$\widehat{\varphi}_{\sigma}(\mathbf{a}) = -\frac{1}{2} \|(\mathbf{C}_a^* \widetilde{\mathbf{a}}_0 - \lambda \sigma)_I\|_2^2 + \frac{1}{2} \|\widetilde{\mathbf{a}}_0\|_2^2. \quad (16)$$

For this surrogate $\widehat{\varphi}(\mathbf{a})$, we can show that if λ is sufficiently large compared to the magnitude of \mathbf{x}_0 , every strict local minimizer is a signed shift truncation of the ground truth:

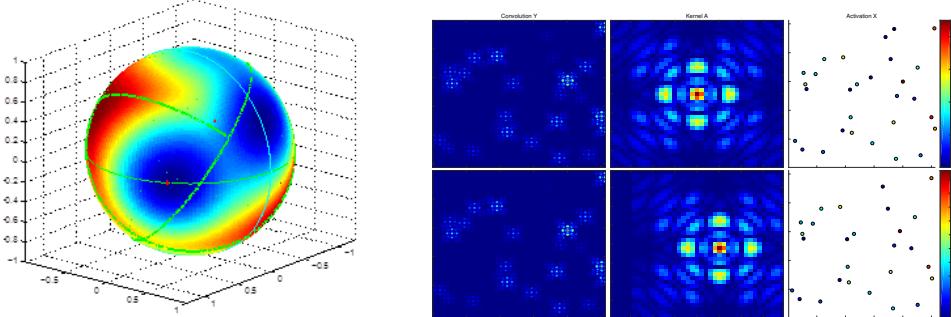


Figure 1: Geometry on the ℓ^2 ball for fixed a_0 and generic x_0 . Left: the objective $\varphi(a)$ in a low dimensional setting $a \in \mathbb{S}^2$ – dark blue represents small values while dark red represents large values. All local minima are close to signed shift truncations of the ground truth a_0 , with a_0 itself achieving global minimum. The green lines indicate regions where a are ill-posed as convolutional kernels. Right: a shift truncation a achieves a local minimum of $\varphi(a)$ in a high dimensional setting. Shown here is the ground truth $y = a_0 \otimes x_0$, a_0 , and x_0 (top right) versus their respective recovered quantities $a \otimes x$, a , and x (bottom right).

Theorem 2.1. Define the set of possible supports of minimizer x^* with

$$\mathcal{I} = \{\text{supp}(\text{SOFT}_\lambda [\check{C}_{a_0}^* \iota a]) \mid a \in \mathbb{S}^{k-1}\}. \quad (17)$$

For each nonempty support $I = \{i_1 < i_2 < \dots < i_{|I|}\}$, let

$$W_I = \left[\frac{\iota^* s_{-i_1} [\widetilde{a}_0]}{\|\iota^* s_{-i_1} [\widetilde{a}_0]\|_F} \middle| \frac{\iota^* s_{-i_2} [\widetilde{a}_0]}{\|\iota^* s_{-i_2} [\widetilde{a}_0]\|_F} \middle| \dots \middle| \frac{\iota^* s_{-i_{|I|}} [\widetilde{a}_0]}{\|\iota^* s_{-i_{|I|}} [\widetilde{a}_0]\|_F} \right]$$

Suppose that $\lambda < 1$ and that for every nonempty $I \in \mathcal{I}$,

$$\|W_I^* W_I - I\| < \frac{\lambda^2}{6}, \quad (18)$$

then every local minimum \bar{a} of $\widehat{\varphi}$ over \mathbb{S}^{k-1} satisfies either $\bar{a} \in \mathcal{R}_0$ (in which case \bar{a} is also a global maximum), or

$$\bar{a} = \pm \frac{\iota^* s_\tau [\widetilde{a}_0]}{\|\iota^* s_\tau [\widetilde{a}_0]\|_F} = \pm \mathcal{P}_{\mathbb{S}} [\iota^* s_{-\tau} [\widetilde{a}_0]] \quad (19)$$

with $x^*(\bar{a}) = \pm \text{SOFT}_\lambda [\|\iota^* s_\tau [\widetilde{a}_0]\|_F] s_{-\tau} [x_0]$ for some shift τ .

Proof. Please refer to the supplement. \square

This theorem says that the only local minima in this ideal case are signed shift truncations of the ground truth a_0 , with certain choice of λ . Moreover, on those local minima \bar{a} , the minimizing sparse $x^*(\bar{a})$ correspond to the soft thresholded, oppositely shifted ground truth x_0 . The quantity $\|W_I^* W_I - I\|$ measures the orthogonality of different shifts of a_0 . In particular, if a_0 is benign enough in the sense that any two different shifts of a_0 are uncorrelated, or $\|W_I^* W_I - I\| \rightarrow 0$ for any I , then any nonzero λ guarantees the desired geometry. On the other hand, for a fixed a_0 , both $|\mathcal{I}|$ and $|I|$ becomes smaller as λ increases, therefore the constraint $\|W_I^* W_I - I\| < \frac{\lambda^2}{6}$ is more likely to be satisfied.

A similar result holds when x_0 is separated enough that copies of the kernel do not overlap. It also holds if x_0 is

a long, sufficiently sparse random vector. For example, if the entries of x_0 satisfy a Bernoulli-Gaussian distribution $x_0(i) = \Omega(i)\mathbf{v}(i)$, with $\Omega(i) \sim \text{Ber}(\theta)$ and $\mathbf{v}(i) \sim \mathcal{N}(0, 1)$, and the probability θ diminishes sufficiently quickly with k . We conjecture that this phenomenon holds much more broadly. In particular, determining how slowly θ can diminish with k is an open problem.

Comparison with the simplex (ℓ^1) constraints. In comparison, the objective value of the same function but with the convolutional kernel a constrained on the ℓ^1 norm ball is shown in Figure 2.

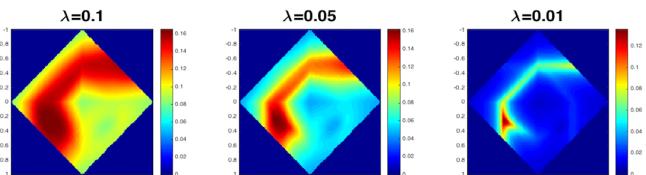


Figure 2: Geometry on the ℓ^1 ball: The trivial spike convolutional kernel achieves global minimum and the ground truth $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$ becomes a local minimum.

There is a significant difference induced by these two constraints: the trivial spike kernel ($a = \delta$) becomes the global minima and other meaningful solutions become local minima with the ℓ^1 norm constraint [2], while the spherical constraint always renders spurious local minima close to some signed shift truncation of the ground truth. This important empirical knowledge of the structure of the local minima enables us to infer the ground truth from any local minima.

3. A Two-Stage Algorithm

Inspired by the geometric property that every local minimum of the simplified problem (2.2) is a signed shift truncation of the ground truth a_0 , we present a two stage algorithm for reliable recovery of the ground truth a_0 in

this section. In the first stage, the algorithm recovers some signed shift truncation of the ground truth, and the following stage infers the ground truth from this partial recovery.

3.1. Stage I: Find the Signed Shift Truncation

Theorem 2.1 suggests that λ needs to be relatively large to guarantee that all local minimizers are signed shift truncations of the ground truth. However, this is not sufficient to guarantee the success of an optimization algorithm due to the non-differentiability of the ℓ^1 regularizer at $\mathbf{x}^* = \mathbf{0}$. Because of this non-differentiability, when λ is too large, there is a nonzero measure set of \mathbf{a} where $\|\mathbf{C}_\mathbf{a}^* \mathbf{y}\|_\infty \leq \lambda$ and therefore $\mathbf{x}^*(\mathbf{a}) = \mathbf{0}$. These \mathbf{a} are not correlated with any signed shift truncation of \mathbf{a}_0 and are the global maxima of φ . The objective function is constant over this region, so there is no way to escape using only local information.

One way to cope with this flat global maxima region is to replace the sparsity penalty function with a differentiable one. A natural choice is the μ -huber loss function, which can be seen as an ℓ^1 penalty but with a rounded bottom for $|x_i| \leq \mu$:

$$h_\mu(\mathbf{x}) = \sum_{|x_i| \leq \mu} \left(\frac{x_i^2}{2\mu} + \frac{\mu}{2} \right) + \sum_{|x_i| > \mu} |x_i| \quad (20)$$

As we choose $\mu \ll \lambda$, the μ -huber function closely approximates the ℓ^1 norm, which still maintaining the effect of “smoothing” the flat region. The flat region for the ℓ^1 penalty objective occurs when $\mathbf{x}^*(\mathbf{a}) = \mathbf{0}$, correspondingly we define a region as $\mathcal{R}_{h,0}$ with small $\mathbf{x}^*(\mathbf{a})$ such that

$$\mathcal{R}_{h,0} := \{\mathbf{a} \in \mathbb{S}^{k-1} : \|\mathbf{x}^*(\mathbf{a})\|_\infty \leq \mu\}. \quad (21)$$

Within the $\mathcal{R}_{h,0}$ region, the original objective function can be rewritten into a simpler form:

$$\varphi_{h_\mu}(\mathbf{a}) = \frac{1}{2} \|\mathbf{y} - \mathbf{a} \otimes \mathbf{x}\|_2^2 + \frac{\lambda}{2\mu} \|\mathbf{x}\|_2^2 + \frac{\mu n}{2}, \quad (22)$$

thus the optimality condition for \mathbf{x}^* implies

$$\mathbf{x}^*(\mathbf{a}) = \left(\mathbf{C}_\mathbf{a}^* \mathbf{C}_\mathbf{a} + \frac{\lambda}{\mu} \mathbf{I} \right)^{-1} \mathbf{C}_\mathbf{a}^* \mathbf{y} \approx \frac{\mu}{\lambda} \mathbf{C}_\mathbf{a}^* \mathbf{y}. \quad (23)$$

Plugging $\mathbf{x}^*(\mathbf{a})$ back to (6) and ignoring the higher order term $O(\frac{\mu^2}{\lambda^2})$ yields

$$\varphi_{h_\mu}(\mathbf{a}) \approx -\frac{\mu}{2\lambda} \|\mathbf{y} \otimes \mathbf{a}\|_2^2 + \frac{1}{2} \|\mathbf{y}\|_2^2 + \frac{\mu n}{2}. \quad (24)$$

In this case, minimization of the objective function $\varphi_{h_\mu}(\mathbf{a})$ within region $\mathcal{R}_{h,0}$ is equivalent to finding the maximum eigenvalue of the matrix $\boldsymbol{\iota}^* \mathbf{C}_\mathbf{y}^* \mathbf{C}_\mathbf{y} \boldsymbol{\iota}$ with the corresponding leading eigenvectors $e_1(\boldsymbol{\iota}^* \mathbf{C}_\mathbf{y}^* \mathbf{C}_\mathbf{y} \boldsymbol{\iota})$ achieving the local minima. However, these points can be excluded from $\mathcal{R}_{h,0}$ by setting $\lambda < \min_{\mathbf{v} \in e_1(\boldsymbol{\iota}^* \mathbf{C}_\mathbf{y}^* \mathbf{C}_\mathbf{y} \boldsymbol{\iota})} \|\mathbf{C}_\mathbf{y}^* \boldsymbol{\iota} \mathbf{v}\|_\infty$.⁷

⁷A computationally easier upper bound would be $\sqrt{\frac{\lambda_1(\boldsymbol{\iota}^* \mathbf{C}_\mathbf{y}^* \mathbf{C}_\mathbf{y} \boldsymbol{\iota})}{k}}$. Note that in some scenario, there exists a local minima appearing in either region $\mathcal{R}_{h,0}$ or $\mathcal{R}_{h,0}^c$ regardless of how we set λ . Such extreme case happens when the ground truth convolutional kernel is only supported on a small consecutive portion of its full size, hence a tight estimate of the kernel size is preferred.

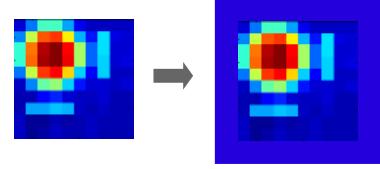


Figure 3: Zero Padding a Signed Shift Truncation The original signed shift truncation (left) and the corresponding zero padded one (right).

With above modifications, the original flat local maxima region $\mathcal{R}_{h,0}$ becomes concave and always have a direction of negative curvature for the algorithm to escape $\mathcal{R}_{h,0}$. Hence, the first stage of the algorithm can find a signed shift-truncation of the ground truth $\bar{\mathbf{a}} = \pm \frac{\boldsymbol{\iota}^* s_\tau[\widetilde{\mathbf{a}}_0]}{\|\boldsymbol{\iota}^* s_\tau[\widetilde{\mathbf{a}}_0]\|_F}$ as desired.

3.2. Stage II: Infer the Ground Truth

The second stage of the algorithm aims to recover the ground truth from its signed shift truncation $\bar{\mathbf{a}}$. To recover the truncated part, we first put $\bar{\mathbf{a}}$ in a higher dimensional sphere by zero padding (Figure 3), and then recover the ground truth \mathbf{a}_0 on a higher dimensional sphere. Intuitively, as $\bar{\mathbf{a}}$ still captures a considerable portion of the ground truth \mathbf{a}_0 (the zero padded $\bar{\mathbf{a}}$ is close to the shifted \mathbf{a}_0 in a higher dimensional space), the zero padded $\bar{\mathbf{a}}$ serves as a good initialization. This intuition is made rigorous in the following lemma:

Lemma 3.1. Let $\lambda_{rel} = \lambda / \|\mathbf{x}_0\|_\infty$, suppose the ground truth \mathbf{a}_0 satisfies

$$|\langle \mathbf{a}_0, \boldsymbol{\iota} s_{\tau \neq 0} [\widetilde{\mathbf{a}}_0] \rangle| < \lambda_{rel}^2 - (2 + 1/\lambda_{rel}^2) \sqrt{1 - \lambda_{rel}^2} \quad (25)$$

for any nonzero shift τ , and \mathbf{x}_0 is separated enough such that any two nonzero components are at least $2k$ entries away from each other. If initialized at some $\mathbf{a} \in \mathbb{S}^{k-1}$ that $|\langle \mathbf{a}, \mathbf{a}_0 \rangle| > \lambda_{rel}$, a gradient descent algorithm minimizing $\varphi(\mathbf{a})$ recovers the signed ground truth $\pm \mathbf{a}_0$.

Proof. Please refer to the supplement. \square

This lemma says when the initial point \mathbf{a} is close enough to the ground truth, the gradient always points to \mathbf{a}_0 as long as $|\langle \mathbf{a}_0, \boldsymbol{\iota} s_{\tau \neq 0} [\widetilde{\mathbf{a}}_0] \rangle|$ is sufficiently small. Theorem 2.1 suggests that the first stage of the algorithm finds one local minimum $\bar{\mathbf{a}}$ that $|\langle \bar{\mathbf{a}}, \mathbf{a}_0 \rangle| \geq \lambda_{rel}$. Hence, the second stage of the algorithm, which minimizes the same objective function but on a higher dimensional sphere, recovers the ground truth up to sign shift ambiguity as desired.

To ensure accurate recovery, it is important to take the effect of λ on the function geometry into consideration. A larger λ encourages a sparser \mathbf{x} and induces a simpler and smoother function landscape, which effectively eliminates undesirable local minima that are not close to any signed

shift truncations, as shown in Figure 4. On the other hand, a smaller λ emphasizes more on the accurate recovery of the signal, therefore the global minima of (6) will be closer to the ground truth when λ decreases.

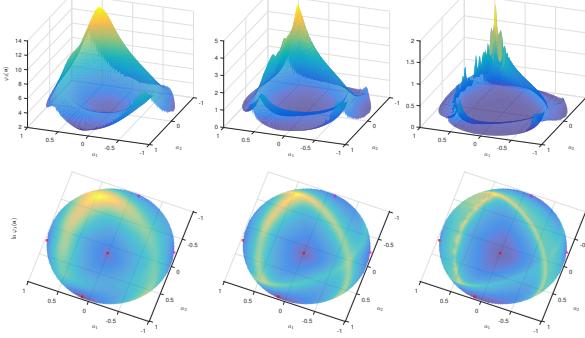


Figure 4: Function Geometry with Varying λ : The objective $\varphi(\mathbf{a})$ over the hemisphere for $\lambda = 10^{-1}, 10^{-3}, 10^{-6}$. Here $\mathbf{a}_0 = \mathcal{P}_{\mathbb{S}^2} [[1, 8, 2]]$ and $\mathbf{x}_0 \sim \text{Ber}(.) \odot \mathcal{N}(0, 1)$. The ground truth kernel \mathbf{a}_0 and its shift-truncations $\mathcal{P}_{\mathbb{S}^2} [[8, 2, 0]], \mathcal{P}_{\mathbb{S}^2} [[0, 1, 8]]$ are shown in red, and sign-flips $\mathcal{P}_{\mathbb{S}^2} [-[8, 2, 0]], \mathcal{P}_{\mathbb{S}^2} [-[0, 1, 8]]$ are shown in magenta. Notice that each signed shift truncation shown on the hemisphere is close to a corresponding local minima, while as the objective landscape becomes less regularized as λ shrinks.

This geometric effect induced by λ suggests a continuation method in the second stage of the algorithm. We start with a relatively big λ for smoother function geometry, which encourages the algorithm to converge to one meaningful local minimum close to some signed shift truncation of the ground truth. Then run the same algorithm with decreasing sequence of λ to produce a finer approximation of the ground truth. The overall algorithm is described in Algorithm 1.

Algorithm 1 Nonconvex Sparse Blind Deconvolution

Ensure: Observation data \mathbf{y} , regularization parameter λ_0 and λ_{\min} , continuation parameter $\beta > 1$

- 1: Solve $\mathbf{a}^{(0)} = \arg \min \varphi_{\lambda_0}(\mathbf{a})$ on \mathbb{S}^{k-1} with random initialization
- 2: Set $\lambda_1 = \lambda_0$, zero pad $\mathbf{a}^{(0)}$ to $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(1)} \in \mathbb{S}^{k'-1}$ ($k' > k$).
- 3: **while** $\lambda_k > \lambda_{\min}$ **do**
- 4: Solve $\mathbf{a}^{(k+1)} = \arg \min \varphi_{\lambda_k}(\mathbf{a})$ on $\mathbb{S}^{k'-1}$ with initialization $\mathbf{a}^{(k)}$.
- 5: $\lambda_{k+1} = \lambda_k / \beta$
- 6: **end while**

We need to note that solving $\mathbf{a} = \arg \min \varphi_\lambda(\mathbf{a})$ in Algorithm 1 involves iteration between (i) finding the marginalization over $\mathbf{x}^*(\mathbf{a})$ step, and (ii) updating \mathbf{a} based on the gradient/Hessian of $\varphi_\lambda(\mathbf{a})$. This could be very computationally consuming, a more efficient variant would be to optimize over the cross space of \mathbf{a} and \mathbf{x} together. The corresponding algorithm can be easily adapted to fit into the same general framework. The only things we want to emphasize in the

proposed algorithm are the dimension lifting of the sphere and the continuation of λ .

4. Further Extensions

In this section, we extend our algorithm to handle two other deconvolution problems of practical interests: image deblurring and convolutional dictionary learning. The proposed two stage algorithm can be modified and applied to these more complicated applications.

4.1. Image Deblurring

Image deblurring aims to recover a sharp natural image from its blurred observation due to unknown photographic processes such as camera shake or defocus. Although the natural images are not necessarily sparse, it is widely acknowledged that their gradients are approximately sparse. Let $\mathbf{y} = \mathbf{a}_0 \circledast \mathbf{x}_0$ denote the observed blurry image, which is the convolution of the original sharp image \mathbf{x}_0 and the blurring kernel \mathbf{a}_0 . Because of the linearity of the convolution operator, the gradient of the blurred image equals the convolution of the kernel and gradient of the original sharp image, which is usually sparse as desired

$$\nabla_x \mathbf{y} = \mathbf{a}_0 \circledast \nabla_x \mathbf{x}_0, \quad \nabla_y \mathbf{y} = \mathbf{a}_0 \circledast \nabla_y \mathbf{x}_0. \quad (26)$$

Here, ∇_x and ∇_y denote derivatives in the x and y directions. In this application, $\nabla_x \mathbf{x}_0$ and $\nabla_y \mathbf{x}_0$ are the underlying sparse signals, and the blind image deblurring problem can be cast as solving:

$$\min_{\mathbf{a} \in \mathbb{S}_+^{k-1}, \mathbf{x}_1, \mathbf{x}_2} \left\{ \begin{array}{l} \frac{1}{2} \|\nabla_x \mathbf{y} - \mathbf{a} \circledast \mathbf{x}_1\|_2^2 + \lambda r(\mathbf{x}_1) \\ + \frac{1}{2} \|\nabla_y \mathbf{y} - \mathbf{a} \circledast \mathbf{x}_2\|_2^2 + \lambda r(\mathbf{x}_2) \end{array} \right\}. \quad (27)$$

Here, \mathbb{S}_+^{k-1} denotes the intersection of the unit sphere and the positive orthant. In this application, the non-negativity of the blurring kernel removes the sign ambiguity. We observe in experiments that local minimizers are all near some shift truncation of the ground truth kernel. The same two stage algorithm can therefore be applied to infer the ground truth.

4.2. Convolutional Dictionary Learning

Convolutional dictionary learning (CDL) is an important problem in machine learning for images, speeches, as well as scientific problems like microscopy data analysis and neural spike sorting. The observation signal \mathbf{y} is the superposition of convolutions of N pairs of kernels \mathbf{a}_{0n} and corresponding coefficients \mathbf{x}_{0n} :

$$\mathbf{y} = \sum_{n=1}^N \mathbf{a}_{0n} \circledast \mathbf{x}_{0n}. \quad (28)$$

Blind deconvolution can be seen as a special case of CDL with $N = 1$. If the coefficients \mathbf{x}_{0n} are sparse, a natural way to extend our knowledge of SBD would be to assume all N

convolutional kernels having unit Frobenius norm and cast it as minimizing following objective function over the product of N spheres:

$$\min_{\mathbf{a}_n \in \mathbb{S}^{k-1}} \min_{\mathbf{x}_n} \frac{1}{2} \|\mathbf{y} - \sum_{n=1}^N \mathbf{a}_n \otimes \mathbf{x}_n\|_2^2 + \lambda \sum_{n=1}^N r(\mathbf{x}_n). \quad (29)$$

We anticipate that all the local minima are near signed shift truncations of the ground truth, provided the target kernels \mathbf{a}_{0n} are sufficiently diverse. The modified two stage algorithm still manages to capture the partial information offered by local minima and hence recovers the ground truth. Experimental results are provided in Section 5.4 to corroborate this claim.

5. Experiments

In this section, we investigate the performance of our algorithm on both synthetic and real data. We first report a systematic investigation, performed in [4], of the performance of our algorithm on synthetic data, which are designed to mimic properties of the microscopy data analysis problem. In Sections 5.2-5.4, we present experiment results showing how our method performs on real data from microscopy and image deblurring.

5.1. Evaluation on Synthetic Data

Noise-free data: we generate the noise-free observation signal of size $m = 256 \times 256$ through circular convolution between a kernel of size k and a random underlying activation signal with a Bernoulli distribution with sparsity θ , i.e. $x_i \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(\theta)$, or $x \sim \text{Ber}(\theta)$. We plot the kernel recovery error for varying kernel size k and sparsity level θ in the left of Figure 5 [4]. Each point on the diagram is the average of 20 independent measurements. The algorithm performs excellently in the blue regions, but begins to fail in the red regions, where either the kernel size is large or the underlying activation signal is dense. The region where typical STM measurements are performed are bounded below by the white dashed line, where the proposed algorithm achieves satisfying performance.

Noisy data: we generate convolutional signals by convolving fixed kernel of dimension k , $k/m = 0.14$ with the random activation map $x \sim \text{Ber}(\theta)$ of dimension m , and applying additive Gaussian noise. We test the performance of our algorithm for varying sparsity θ and noise power. The result is shown in Figure 5 (right): the algorithm achieves noise-robust recovery when the sparsity constraint is satisfied.

5.2. Microscopy Data Analysis

We apply our algorithm on experimental microscopy data obtained from a NaFeCoAs sample. Our results shown in Figure 6 indicate that the proposed algorithm manages to recover the missing details of the ripples in the Fourier domain of the defect, which encode the physical scattering processes of electrons at work.

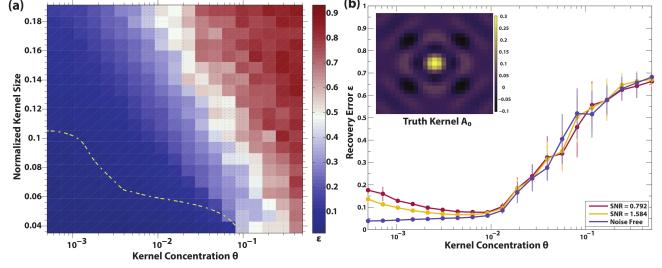


Figure 5: Recovery accuracy [4]. Left: phase transition diagram from noise-free simulated results. Right: performance of algorithm 1 in the presence of additive noise in the measurement; the error increases for small θ due to a lack of samples, whereas extremely large θ leads to algorithmic failure.

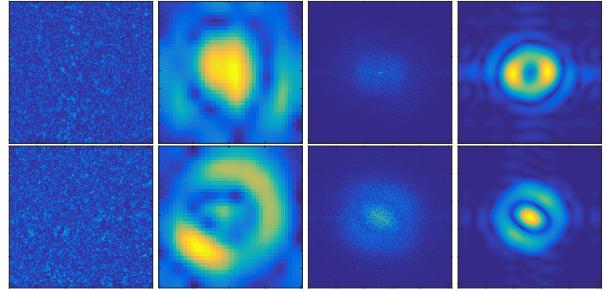


Figure 6: STM Data Analysis. From left to right: the microscopy images, extracted convolutional kernels (defect patterns), and their respective Fourier magnitude images.

5.3. Image Deblurring

We test our algorithm on the image deblurring dataset from [13], solving (28) to recover the convolutional kernel. To clearly separate the inaccuracy of the algorithm and the universal blurring kernel model, all the experiments are done on three kinds of blurred images: (i) synthetic blurred images generated by the convolution of sharp images and blurry kernels; (ii) noisy blurred images generated by adding Gaussian noise to the clean synthetic blurred images (SNR=100); and (iii) real blurry images taken with camera shakes [13].

We compare with algorithms by Zhang et al.[26], Krishnan et al.[11], Sun et al.[22], and Liu et al.[16].⁸ Because of the shift ambiguity, we evaluate the accuracy of the recovered blurring kernel considering all possible shifts. The kernel recovery error is defined as $\min_{\tilde{\mathbf{a}}} \|\mathbf{t}^* s_\tau [\tilde{\mathbf{a}}] / \|\tilde{\mathbf{a}}\|_1 - \mathbf{a}_0 / \|\mathbf{a}_0\|_1\|_F$, and the cumulative distribution is shown in Figure 7.

We use the same non-blind deblurring algorithm from [10], with the same parameter. We consider the blurred image using the ground truth kernel to be the bench mark, and evaluate the quality of the deblurred image by calculating the Frobenius norm of its difference to such bench mark.

⁸We use the default parameters for these algorithms. It's possible that better performance could be obtained by tuning the parameters more carefully. In our algorithm, we fix the λ 's to be 0.1, 0.01, 0.001, 0.001 for all the instances.

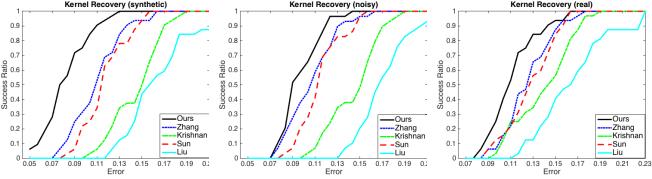


Figure 7: Blur Kernel Recovery Error: Cumulative distributions of recovered blur kernel error from synthetic (left), noised(middle) and real (right) blurry images.

Results are shown are in Figure 8.

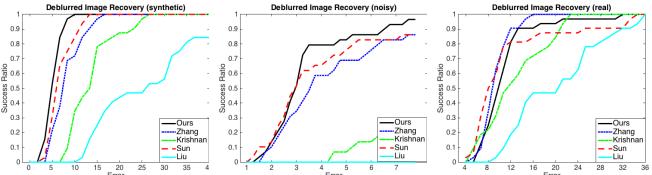


Figure 8: Non-blind Restoration Error: Cumulative distributions of deblurred image error from synthetic (left), noised(middle) and real (right) blurry images.

Our algorithm achieves better convolutional kernel recovery for all three types of images, but its improvement on deblurred image is less obvious, especially for real images. This could be due to (i) the convolutional kernel in this dataset is not strictly uniform across the image, and (ii) the non-blind deconvolution algorithm exploits the heavy-tailed distribution of a natural image’s gradient and becomes less sensitive to the accuracy of the recovered convolutional kernel.

5.4. Convolutional Dictionary Learning

We show results of recovering multiple convolutional kernels on both synthetic data (Figure 9) and real STM data (Figure 10). In the synthetic data, the three convolutional kernels are of size 16×16 and their corresponding activation signals are generated through a Bernoulli model of sparsity 0.005. Results of both stages of the algorithm are shown in Figure 9: the first stage returns kernels close to some shift truncations of the ground truth, and the second stage recovers the ground truth on a higher dimensional space.

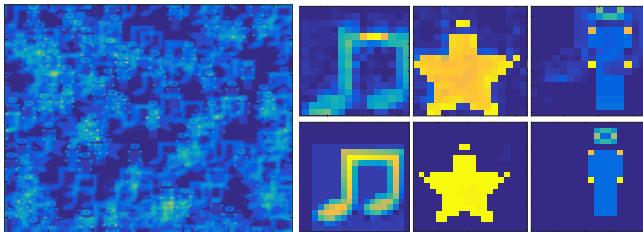


Figure 9: Multi Kernel Blind Deconvolution on Synthetic Data: Input image (left) and the recovered convolutional kernels of Stage I and Stage II of the algorithm (right).

We repeat this experiment with microscopy data obtained

from a NaFeAs sample. The algorithm manages to differentiate the two convolutional kernels (defect patterns), as shown in Figure 10. For this material, the kernel orientations depend on the history of the material (stress, temperature, etc.), and using convolutional dictionary learning can be used to automatically detect these features.

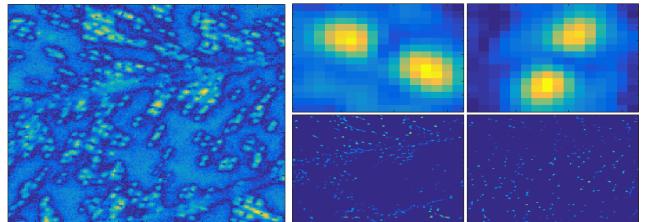


Figure 10: Multi Kernel Blind Deconvolution on Real STM Image: Input image (left) and recovered convolutional kernels and their corresponding activation signals (right).

6. Discussions

This work studies the global geometry of a nonconvex optimization problem for SBD when the kernel is assumed to have unit Frobenius norm. In this setting, we find that all the local minima are benign, in the sense that they are close to some signed shift truncation of the ground truth. With this insight, we propose a two stage algorithm that recovers the ground truth by exploiting the information hidden in local minima.

This problem reveals the challenges faced when analyzing the SBD problem via a geometrical approach. For problems enjoying stronger symmetry properties [20, 21], similar approaches yield a global understanding of the function geometry and recovery guarantees. We expect that the weak symmetries in SBD make a major contribution to the difficulties encountered for this problem.

There are lots of additional further directions could be of great interests for both theory and application: Our empirical results show that our characterization of local minima carries through to the convolutional dictionary learning problem, which can also be efficiently solved by slight adaptation of the proposed algorithm. However, the theory part is open, it would be interesting to know how many kinds of kernels, or what kinds of kernels are recoverable, probably by some measures of incoherence, which is a common assumption in dictionary learning problem.

Two other imperfections we encounter in scientific measurement are resolution limit and measurement error, which inspire us to consider (i) if it’s possible to integrate blind deconvolution and super-resolution process together; (ii) if we can come up with a robust blind deconvolution algorithm to automatically rule out noisy entries.

References

- [1] A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programing. *arXiv preprint:1211.5608*, 2012. 1
- [2] A. Benichoux, E. Vincent, and R. Gribonval. A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors. *38th International Conference on Acoustics, Speech, and Signal Processing*, May 2013. 1, 4
- [3] T. Chan and C. Wong. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375, Mar 1998. 1
- [4] S. Cheung, Y. Lau, Z. Chen, J. Sun, Y. Zhang, J. Wright, and A. Pasupathy. Beyond the fourier transform: A nonconvex optimization approach to microscopy analysis. *Submitted*, 2017. 1, 7
- [5] Y. Chi. Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):782–794, June 2016. 1
- [6] S. Cho and S. Lee. Fast motion deblurring. *ACM Transactions on Graphics (SIGGRAPH ASIA 2009)*, 28(5):article no. 145, 2009. 1
- [7] C. Ekanadham, D. Tranchina, and E. P. Simoncelli. A blind sparse deconvolution method for neural spike identification. In *Advances in Neural Information Processing Systems 24*, pages 1440–1448. 2011. 1
- [8] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM Trans. Graph.*, 25(3):787–794, June 2006. 1
- [9] D. Gong, M. Tan, Y. Zhang, A. van den Hengel, and Q. Shi. Blind image deconvolution by automatic gradient activation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1
- [10] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *Advances in Neural Information Processing Systems 22*, pages 1033–1041. 2009. 7
- [11] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 233–240, June 2011. 1, 7
- [12] D. Kundur and D. Hatzinakos. Blind image deconvolution. *Signal Processing Magazine, IEEE*, 13(3):43–64, May 1996. 1
- [13] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2354–2367, Dec 2011. 1, 8
- [14] 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. 1
- [15] X. Li, S. Ling, T. Strohmer, and K. Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *preprint*, 2016. 1
- [16] G. Liu, S. Chang, and Y. Ma. Blind image deblurring using spectral properties of convolution operators. *IEEE Transactions on Image Processing*, 23(12):5047–5056, Dec 2014. 1, 7
- [17] J. Mairal, F. Bach, and J. Ponce. Sparse modeling for image and vision processing. *Found. Trends. Comput. Graph. Vis.*, 8(2-3), Dec. 2014. 3
- [18] D. Perrone and P. Favaro. Total variation blind deconvolution: The devil is in the details. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1
- [19] D. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. *preprint*, 2012. 2
- [20] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere. *preprint*, 2015. 2, 8
- [21] J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. *preprint*, 2016. 2, 8
- [22] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based blur kernel estimation using patch priors. In *Proc. IEEE International Conference on Computational Photography*, 2013. 7
- [23] D. Wipf and H. Zhang. Revisiting bayesian blind deconvolution. *arXiv preprint:1305.2362*, 2013. 2
- [24] L. Xiao, J. Wang, W. Heidrich, and M. Hirsch. Learning high-order filters for efficient blind deconvolution of document photographs. In *14th European Conference on Computer Vision (ECCV)*, pages 734–749, 2016. 1
- [25] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *11th European Conference on Computer Vision (ECCV)*, pages 157–170, 2010. 1
- [26] H. Zhang, D. Wipf, and Y. Zhang. Multi-image blind deblurring using a coupled adaptive sparse prior. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, January 2013. 2, 7