
Short-and-Sparse Deconvolution — A Geometric Approach

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Short-and-sparse deconvolution (SaSD) is the problem of extracting localized,
2 recurring motifs in signals with spatial or temporal structure. Variants of this problem
3 arise in applications such as image deblurring, microscopy, neural spike sorting,
4 and more. SaSD is challenging in both theory and practice, as natural optimization
5 formulations are nonconvex. Moreover, practical deconvolution problems involve
6 smooth motifs (kernels) whose spectra decay rapidly, resulting in poor conditioning
7 and numerical challenges. This paper is motivated by recent theoretical advances
8 [1, 2], which characterize the optimization landscape of a particular nonconvex formulation
9 of SaSD. This is used to derive a *provable* algorithm which exactly solves
10 certain non-practical instances of the SaSD problem. We leverage the key ideas
11 from this theory (sphere constraints, data-driven initialization) to develop a *practical*
12 algorithm, which performs well on data arising from a range of application areas.
13 We highlight key additional challenges posed by the ill-conditioning of real
14 SaSD problems, and suggest heuristics (acceleration, continuation, reweighting) to
15 mitigate them. Experiments demonstrate both the performance and generality of
16 the proposed method.

17

1 Introduction

18 Many signals arising in science and engineering can be modeled as superpositions of basic, recurring
19 motifs, which encode critical information about a physical process of interest. Signals of this type
20 can be modeled as the convolution of a zero-padded short kernel $\mathbf{a}_0 \in \mathbb{R}^{p_0}$ (the motif) with a longer
21 sparse signal $\mathbf{x}_0 \in \mathbb{R}^m$ ($m \gg p_0$) which encodes the locations of the motifs in the sample¹:

$$\mathbf{y} = \iota \mathbf{a}_0 \circledast \mathbf{x}_0. \quad (1)$$

22 We term this a short-and-sparse (SaS) model. In practice, often only \mathbf{y} is directly observed. *Short-and-sparse deconvolution* (SaSD) is the problem of recovering both \mathbf{a}_0 and \mathbf{x}_0 from \mathbf{y} . Variants of
23 this problem arise in areas such as microscopy [3], astronomy [4], and neuroscience [5]. SaSD is a
24 challenging inverse problem in both theory and practice. Natural formulations are nonconvex, and
25 until recently very little algorithmic theory was available. Moreover, practical instances are typically
26 ill-conditioned, due to the spectral decay of the kernel \mathbf{a}_0 .

28 This paper is motivated by recent theoretical advances in nonconvex optimization – and in particular,
29 on the geometry of SaSD. [1, 2] study particular optimization formulations for SaSD and show that
30 the landscape is largely driven by the *problem symmetries* of SaSD. They derive provable methods
31 for idealized problem instances, which exactly recover $(\mathbf{a}_0, \mathbf{x}_0)$ up to trivial ambiguities. While the
32 inspiring, these methods are *not practical* and perform poorly on real problem instances. Where the

¹For simplicity, (1) uses cyclic convolution; algorithms are results also apply to linear convolution with minor modifications. Here ι denotes the zero padding operator.

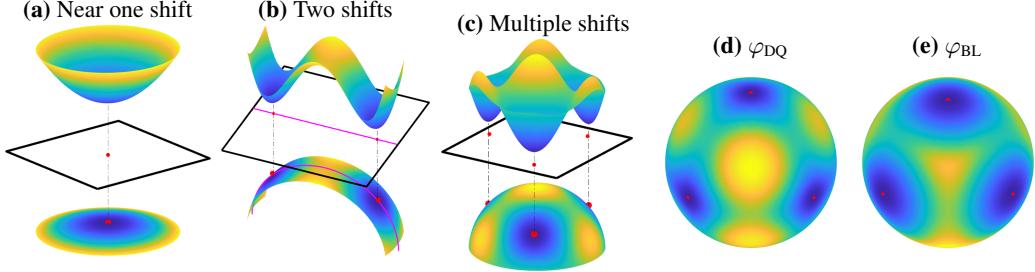


Figure 1: **Geometry of φ_{DQ} near superpositions of shifts of \mathbf{a}_0** [2]. (a) Regions near single shifts are strongly convex. (b) Regions between two shifts contain a saddle-point, with negative curvature pointing towards each shift and positive curvature orthogonally. (c) The span of three shifts. When $\mu_s(\mathbf{a}_0) \approx 0$, marginalizations of the (d) Dropped Quadratic and (e) Bilinear Lasso ($\varphi_{\text{BL}} \doteq \min_{\mathbf{x}} \Psi_{\text{BL}}(\mathbf{a}, \mathbf{x})$) are similar empirically.

emphasis of [1, 2] is on theoretical guarantees, here we focus on practical computation. We show how to combine ideas from this theory with heuristics that better address the properties of practical deconvolution problems, to build a novel method that performs well on data arising in a range of application areas. A critical issue in moving from theory to practice is the poor conditioning of naturally-occurring deconvolution problems: we show how to address this with a combination of ideas from sparse optimization, such as momentum, continuation, and reweighting. The end result is a general purpose method, which we demonstrate on data from neural spike sorting, calcium imaging and fluorescence microscopy.

Notation The zero-padding operator is denoted by $\iota : \mathbb{R}^p \rightarrow \mathbb{R}^m$. Projection of a vector $\mathbf{v} \in \mathbb{R}^p$ onto the sphere is denoted by $\mathcal{P}_{\mathbb{S}^{p-1}}(\mathbf{v}) \doteq \mathbf{v}/\|\mathbf{v}\|_2$, and $\mathcal{P}_{\mathbf{z}}(\mathbf{v}) \doteq \mathbf{v} - \langle \mathbf{v}, \mathbf{z} \rangle \mathbf{z}$ denotes projection onto the tangent space of $\mathbf{z} \in \mathbb{S}^{p-1}$. The Riemannian gradient of a function $f : \mathbb{S}^{p-1} \rightarrow \mathbb{R}$ on the sphere is given by $\text{grad } f \doteq \mathcal{P}_{\mathbb{S}^{p-1}} \circ \nabla f$.

2 The role of symmetry in SaSD

2.1 Symmetry and shift-coherence

An important observation of the SaSD problem is that it admits multiple equivalent solutions. This is purely due to the cyclic convolution between \mathbf{a}_0 and \mathbf{x}_0 , which exhibits the trivial ambiguity²

$$\mathbf{y} = \iota \mathbf{a}_0 \circledast \mathbf{x}_0 = (\alpha s_\ell[\iota \mathbf{a}_0]) \circledast (\frac{1}{\alpha} s_{-\ell}[\mathbf{x}_0]),$$

for any nonzero scalar α and cyclic shift s_ℓ [1]. Since these *scale and shift symmetries* create several acceptable candidates for \mathbf{a}_0 and \mathbf{x}_0 , they largely drive the behavior of certain nonconvex optimization problems formulated for SaSD. Another important aspect of SaSD is the *shift-coherence* of its kernel,

$$\mu(\mathbf{a}_0) \doteq \max_{\ell \neq 0} |\langle \iota \mathbf{a}_0, s_\ell[\iota \mathbf{a}_0] \rangle| \in [0, 1]. \quad (2)$$

Geometrically, the shifts of \mathbf{a}_0 grow further apart on the sphere as $\mu(\mathbf{a}_0)$ diminishes. SaSD problems are also “easier” when $\mu(\mathbf{a}_0)$ is small, in the sense that they can be solved with denser \mathbf{x}_0 as overlapping shifts are easier to distinguish.

2.2 Landscape geometry under shift-incoherence

A natural approach to solving SaSD is to formulate it as a suitable optimization problem. For instance, consider the *Bilinear Lasso* (BL) problem, which minimizes the squared error between the observation \mathbf{y} and its reconstruction $\mathbf{a} \circledast \mathbf{x}$, plus a ℓ_1 -norm sparsity penalty on \mathbf{x} ,

$$\min_{\mathbf{a} \in \mathbb{S}^{p-1}, \mathbf{x} \in \mathbb{R}^m} \left[\Psi_{\text{BL}}(\mathbf{a}, \mathbf{x}) \doteq \frac{1}{2} \|\mathbf{y} - \iota \mathbf{a} \circledast \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right]. \quad (\text{BL})$$

We will later see that the recovered kernel length p should be set slightly larger than p_0 .

²We therefore assume w.l.o.g. that $\|\mathbf{a}_0\|_2 = 1$ in this paper.

60 The Bilinear Lasso is a *nonconvex* optimization problem, as the shift symmetries of SaSD create
 61 discrete local minimizers in the objective landscape. The regularizing effect created by problem
 62 symmetries is a fairly general phenomenon [6] and, as [2] shows, its influence extends far beyond
 63 local minimizers. The authors there analyze the *Dropped Quadratic* (DQ) objective³

$$\Psi_{\text{DQ}}(\mathbf{a}, \mathbf{x}) \simeq \Psi_{\text{BL}}(\mathbf{a}, \mathbf{x}), \quad \text{when } \mu(\mathbf{a}) \simeq 0.$$

64 This non-practical objective is a valid simplification of the Bilinear Lasso (BL) when the true kernel
 65 is itself incoherent, i.e. $\mu(\mathbf{a}_0) \simeq 0$ (see Figures 1d and 1e or Section A in the appendix). We are
 66 particularly interested in the objective of the *marginalized objective*⁴

$$\min_{\mathbf{a} \in \mathbb{S}^{p-1}} \left[\varphi_{\text{DQ}}(\mathbf{a}) \doteq \min_{\mathbf{x} \in \mathbb{R}^m} \Psi_{\text{DQ}}(\mathbf{a}, \mathbf{x}) \right], \quad (3)$$

67 which is greatly simplified when \mathbf{x} is generic due to concentration of measure, whilst also reducing
 68 the space of optimization to a significantly smaller dimension $p \ll m$.

69 **Regularity in the span of a few shifts.** Under suitable conditions on \mathbf{a}_0 and \mathbf{x}_0 , φ_{DQ} enjoys a
 70 number of nice properties on the sphere. Suppose $\mathbf{a} \simeq \alpha_1 s_{\ell_1}[\mathbf{a}_0] + \alpha_2 s_{\ell_2}[\mathbf{a}_0] \in \mathbb{S}^{p-1}$ is near the
 71 span of two shifts⁵ of \mathbf{a}_0 . If $\alpha_1 \simeq 1$ (or $\alpha_2 \simeq 0$), [2] asserts that \mathbf{a} is in a strongly convex region of
 72 φ_{DQ} , containing a single minimizer near $s_{\ell_1}[\mathbf{a}_0]$, and vice versa (Figure 1a). Near the *balanced point*
 73 $\alpha_1 \simeq \alpha_2$, the influence $s_{\ell_1}[\mathbf{a}_0]$ and $s_{\ell_2}[\mathbf{a}_0]$ on φ_{DQ} creates a saddle-point, characterized by large
 74 negative curvature along the two shifts and positive curvature in orthogonal directions (Figure 1b).
 75 Between these two cases, large negative gradients point towards individual shifts.

76 This characterization of φ_{DQ} — strong convexity near single shifts, and saddle-points near balanced
 77 points — extends to regions of the sphere spanned by *several* shifts (Figure 1c); we elaborate more
 78 on multiple shifts in Section A.1 of the supplementary material. This regional landscape guarantees
 79 that \mathbf{a}_0 can be efficiently recovered up to a signed shift using methods for first and second-order
 80 descent, as soon as \mathbf{a} can be brought sufficiently close to the span of a few shifts.

81 **Optimization on the sphere.** These nice properties of φ_{DQ} depend strongly on its restriction to
 82 the sphere, which creates a small but uniform increase in the *Riemannian* curvature of φ_{DQ} [7]. As a
 83 result, the sphere prevents the creation of spurious local minimizers from being created as a result of
 84 any particular points of the constraint surface possessing nonsmoothness or large positive curvature⁶.

85 **Initializing near a few shifts.** The landscape structure of φ_{DQ} makes single shifts of \mathbf{a}_0 easy to
 86 locate, if \mathbf{a} is initialized near a span of a few shifts. Fortunately, this is a relatively simple matter in
 87 SaSD, as \mathbf{y} is itself a sparse superposition of shifts. Setting $p = 3p_0 - 2$, we initialize \mathbf{a} by randomly
 88 choosing a length- p_0 window $\tilde{\mathbf{y}}_i \doteq [y_i \ y_{i+1} \dots \ y_{i+p_0-1}]^T$ and setting

$$\mathbf{a}^{(0)} \doteq \mathcal{P}_{\mathbb{S}^{p-1}} \left([\mathbf{0}_{p-1}; \tilde{\mathbf{y}}_i; \mathbf{0}_{p-1}] \right). \quad (4)$$

89 This brings $\mathbf{a}^{(0)}$ suitably close to the sum of a few shifts of \mathbf{a}_0 ; any border-truncation effects are
 90 absorbed by padding the ends of $\tilde{\mathbf{y}}_i$ ⁷.

91 **Implications for practical computation.** The Dropped Quadratic problem (3) shows us an ex-
 92 ample of a nice regional landscape for SaSD where efficient recovery of \mathbf{a}_0 is guaranteed when
 93 $\mu(\mathbf{a}_0) \simeq 0$. As SaS applications are often motivated by sharpening or resolution tasks [10, 11, 12],
 94 however, a practical algorithm must be able to handle cases where motifs are smooth and shift-
 95 coherent (i.e., $\mu(\mathbf{a}_0) \approx 1$). The Dropped Quadratic is therefore a poor approximation to the Bilinear
 96 Lasso for practical purposes, yet the two problems share qualitatively similar landscapes (Figures 1d
 97 and 1e). This suggests that the algorithmic implications discussed in Section 2 — namely optimization
 98 on the sphere and data initialization — are also applicable in practical settings.

³As the intention here is to apply some key messages from the Dropped Quadratic towards the Bilinear Lasso, we intentionally omit the concrete form of $\Psi_{\text{DQ}}(\mathbf{a})$. Readers may refer to Section A for more details.

⁴ \mathbf{a}_0 can be recovered via convex optimization once \mathbf{a}_0 is found.

⁵Setting $p > p_0$ ensures that \mathbb{S}^{p-1} contains at least two shifts.

⁶Conversely, the popular ℓ_1 -norm constraint set tends to create trivial sparse minimizers w.r.t. \mathbf{a} [8, 9, 1].

⁷This initialization strategy is improved and made rigorous in [2] — readers may refer to Section A.2 in the supplementary material for more details.

Algorithm 1 Inertial Alternating Descent Method (iADM)

Input: Initializations $\mathbf{a}^{(0)} \in \mathbb{S}^{p-1}$, $\mathbf{x} \in \mathbb{R}^m$; observation $\mathbf{y} \in \mathbb{R}^m$; penalty $\lambda \geq 0$; momentum $\alpha \in [0, 1]$.

Output: $(\mathbf{a}^{(k)}, \mathbf{x}^{(k)})$, a local minimizer of Ψ_{BL} .

Initialize $\mathbf{a}^{(1)} = \mathbf{a}^{(0)}$, $\mathbf{x}^{(1)} = \mathbf{x}^{(0)}$.

for $k = 1, 2, \dots$ until converged **do**

Update \mathbf{x} with accelerated proximal gradient step:

$$\mathbf{w}^{(k)} \leftarrow \mathbf{x}^{(k)} + \alpha \cdot (\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$$

$$\mathbf{x}^{(k+1)} \leftarrow \text{soft}_{\lambda t_k} [\mathbf{w}^{(k)} - t_k \cdot \nabla_{\mathbf{x}} \psi_{\lambda}(\mathbf{a}^{(k)}, \mathbf{w}^{(k)})],$$

where $\text{soft}_{\lambda}(\mathbf{v}) \doteq \text{sign}(\mathbf{v}) \odot \max(|\mathbf{v}| - \lambda, 0)$ denotes the soft-thresholding operator.

Update \mathbf{a} with accelerated Riemannian gradient step:

$$\mathbf{z}^{(k)} \leftarrow \mathcal{P}_{\mathbb{S}^{p-1}}(\mathbf{a}^{(k)} + \frac{\alpha}{\langle \mathbf{a}^{(k)}, \mathbf{a}^{(k-1)} \rangle} \cdot \mathcal{P}_{\mathbf{a}^{(k-1)}}(\mathbf{a}^{(k)}))$$

$$\mathbf{a}^{(k+1)} \leftarrow \mathcal{P}_{\mathbb{S}^{p-1}}(\mathbf{z}^{(k)} - \tau_k \cdot \text{grad}_{\mathbf{a}} \psi_{\lambda}(\mathbf{z}^{(k)}, \mathbf{x}^{(k+1)})).$$

end for

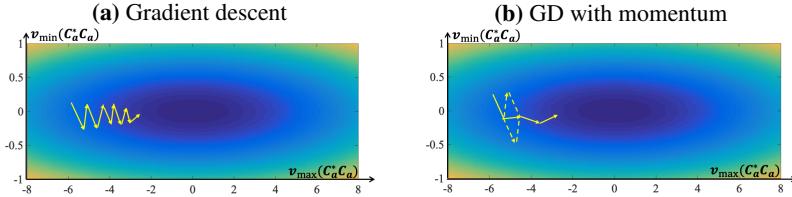


Figure 2: **Momentum acceleration.** a) Iterates of gradient descent oscillate on ill-conditioned functions. b) Momentum dampens oscillation and speeds up convergence.

99 **3 Designing a practical SaSD algorithm**

100 In this section, we borrow the algorithmic implications from the Dropped Quadratic (3) and build an
 101 algorithm based on the Bilinear Lasso (BL), which more accurately accounts for interactions between
 102 (highly coherent) shifts of the ground truth. We show how to address the negative effects of large
 103 coherence using a number of heuristics, leading to an efficient algorithm for SaSD.

104 Several algorithms for SaSD-type problems have been developed for specific applications, such as
 105 image deblurring [8, 4, 12] and neuroscience [13, 14, 5], and image super-resolution [15, 16, 17], or
 106 are augmented with additional structure [18, 19, 20]. Here we will instead attempt to leverage recent
 107 developments in algorithmic theory in SaSD (Section 2) to build an algorithm that performs well in
 108 general practical settings.

109 **3.1 Solving Bilinear Lasso with accelerated alternating descent**

110 When \mathbf{a}_0 is shift-coherent, the Hessian of Ψ_{BL} becomes ill-conditioned as \mathbf{a} converges to single
 111 shifts. Such situations are known to cause slow convergence for first-order methods [21]. One remedy
 112 is to add *momentum* [22, 23] to standard first-order iterations. For instance, consider augmenting
 113 gradient descent, on some smooth $f(\mathbf{z})$ with stepsize τ , using the term \mathbf{w} ,

$$\mathbf{w}^{(k)} \leftarrow \mathbf{z}^{(k)} + \alpha \cdot (\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}) \quad (5)$$

$$\mathbf{z}^{(k+1)} \leftarrow \mathbf{w}^{(k)} - \tau \cdot \nabla f(\mathbf{w}^{(k)}). \quad (6)$$

114 Here, α controls the momentum added⁸. As illustrated in Figure 2, this additional term improves
 115 convergence by reducing oscillations of the iterates for ill-conditioned problems. Momentum has
 116 been shown to improve convergence for nonconvex and nonsmooth problems [24, 25]. In Algorithm 1,
 117 we provide an inertial alternating descent method (iADM) for finding local minimizers of Ψ_{BL} . It
 118 modifies iPALM [24] to perform updates on \mathbf{a} via retraction on the sphere [7]⁹.

⁸Setting $\alpha = 0$ removes momentum and reverts to standard gradient descent.

⁹The stepsizes t_k and τ_k are obtained by backtracking [26, 24] to ensure sufficient decrease for $\Psi_{BL}(\mathbf{a}^{(k)}, \mathbf{w}^{(k)}) - \Psi_{BL}(\mathbf{a}^{(k)}, \mathbf{x}^{(k+1)})$ and $\Psi_{BL}(\mathbf{z}^{(k)}, \mathbf{x}^{(k+1)}) - \Psi_{BL}(\mathbf{a}^{(k+1)}, \mathbf{w}^{(k+1)})$.

Algorithm 2 SaS-BD with homotopy continuation

Input: Observation $\mathbf{y} \in \mathbb{R}^m$, motif size p_0 ; momentum $\alpha \in [0, 1]$; initial $\lambda^{(1)}$ final λ^* , penalty decrease $\eta \in (0, 1)$; precision factor $\delta \in (0, 1)$.

Output: Solution path $\{(\hat{\mathbf{a}}^{(n)}, \hat{\mathbf{x}}^{(n)}; \lambda^{(n)})\}$ for SaS-BD.

Set number of iterations $N \leftarrow \lfloor \log(\lambda^*/\lambda^{(1)}) / \log \eta \rfloor$.

Initialize $\hat{\mathbf{a}}^{(0)} \in \mathbb{R}^{3p_0-2}$ using (4), $\hat{\mathbf{x}}^{(0)} = \mathbf{0} \in \mathbb{R}^m$.

for $n = 1, \dots, N$ **do**

Minimize $\Psi_{\lambda^{(n)}}$ to precision $\delta \lambda^{(n)}$ with Algorithm 1:

$(\hat{\mathbf{a}}^{(n)}, \hat{\mathbf{x}}^{(n)}) \leftarrow \text{iADM}(\hat{\mathbf{a}}^{(n-1)}, \hat{\mathbf{x}}^{(n-1)}; \mathbf{y}, \lambda^{(n)}, \alpha)$.

Update $\lambda^{(n+1)} \leftarrow \eta \lambda^{(n)}$.

end for

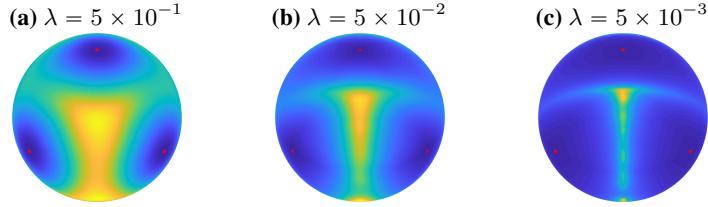


Figure 3: **Bilinear-lasso objective φ_λ on the sphere \mathbb{S}^{p-1}** , for $p = 3$ and varying λ . The function landscape of φ_λ flattens as sparse penalty λ decreases from left to right.

119 **3.2 A SaSD algorithm using homotopy continuation**

120 It is also possible to improve optimization by modifying the objective Ψ_{BL} directly through the
121 sparsity penalty λ . Variations of this idea appear in both [1] and [2], and can also help to mitigate the
122 effects of large shift-coherence in practical problems.

123 When solving (BL) in the noise-free case, it is clear that larger choices of λ encourage sparser
124 solutions for \mathbf{x} . Conversely, smaller choices of λ place local minimizers of the marginal objective
125 $\varphi_{BL}(\mathbf{a}) \doteq \min_{\mathbf{x}} \Psi_{BL}(\mathbf{a}, \mathbf{x})$ closer to signed-shifts of \mathbf{a}_0 by emphasizing reconstruction quality.
126 When $\mu(\mathbf{a}_0)$ is large, however, φ_{BL} becomes ill-conditioned as $\lambda \rightarrow 0$ due to the poor spectral
127 conditioning of \mathbf{a}_0 , leading to severe flatness near local minimizers (Figure 3) and the creation
128 of spurious local minimizers when noise is present. At the expense of precision, larger values of λ limit
129 \mathbf{x} to a small set of support patterns and simplify the landscape of φ_{BL} . It is therefore important both
130 for fast convergence and accurate recovery for λ to be chosen appropriately.

131 When problem parameters – such as the severity of noise, or p_0 and θ – are not known a priori, a
132 *homotopy continuation method* [27, 28, 29] can be used to obtain a *range* of solutions for SaSD.
133 Using the initialization (4), a rough estimate $(\hat{\mathbf{a}}^{(1)}, \hat{\mathbf{x}}^{(1)})$ is first obtained by solving (BL) with iADM
134 using a large choice for $\lambda^{(1)}$; this estimate is refined by gradually decreasing $\lambda^{(n)}$ to produce the
135 *solution path* $\{(\hat{\mathbf{a}}^{(n)}, \hat{\mathbf{x}}^{(n)}; \lambda^{(n)})\}$. Homotopy also ensures that \mathbf{x} remains sparse along the solution
136 path, effectively providing the objective Ψ_{BL} with (restricted) strong convexity w.r.t. both \mathbf{a} and \mathbf{x}
137 throughout optimization [30]. As a result, homotopy achieves linear convergence for SaSD where
138 sublinear convergence is expected otherwise (Figures 4c and 4d). In Algorithm 2, we provide a
139 complete algorithm for SaSD combining Bilinear Lasso and homotopy continuation.

140 **4 Experiments**

141 **4.1 Synthetic experiments**

142 We begin with simulations of SaSD in both coherent and incoherent settings. For coherent settings
143 we use a descretized Gaussian kernel $\mathbf{a}_0 = \mathbf{g}_{n_0, 2}$, where $\mathbf{g}_{p, \sigma} \doteq \mathcal{P}_{\mathbb{S}^{p-1}}\left(\left[\exp\left(-\frac{(2i-p-1)^2}{\sigma^2(p-1)^2}\right)\right]_{i=1}^p\right)$.
144 Incoherent kernels are simulated by sampling $\mathbf{a}_0 \sim \text{Unif}(\mathbb{S}^{n_0-1})$ uniformly on the sphere.

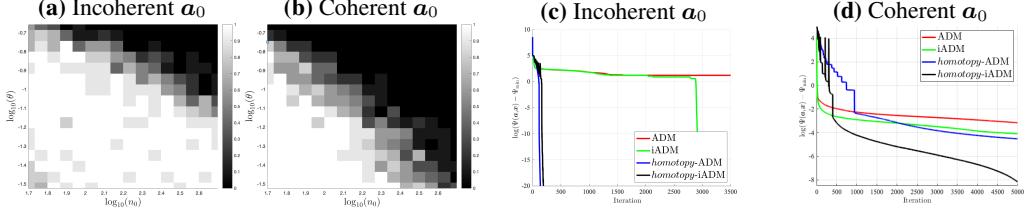


Figure 4: **Synthetic experiments for Bilinear Lasso.** Success probability (a, b): $\mathbf{x}_0 \sim_{\text{i.i.d.}} \mathcal{BR}(\theta)$, the success probability of SaS-BD by solving (BL), shown by increasing brightness, is large when the sparsity rate θ is sufficiently small compared to the length of \mathbf{a}_0 , and vice versa. Success with a fixed sparsity rate is more likely when \mathbf{a}_0 is incoherent. Algorithmic convergence (c, d): convergence of function value for iADM with $\alpha_k = (k-1)/(k+1)$ vs. $\alpha_k = 0$ (ADM); with and without homotopy. Homotopy significantly improves convergence rate, and momentum improves convergence when \mathbf{a}_0 is coherent.

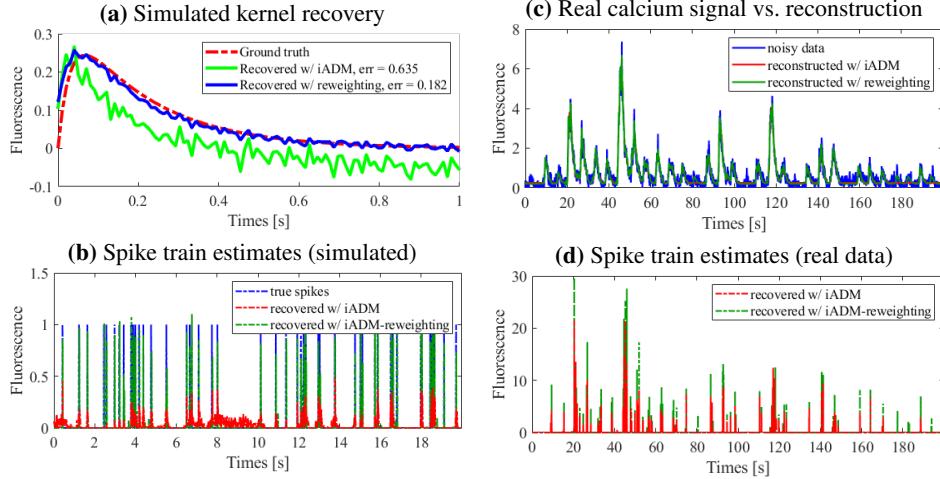


Figure 5: **Deconvolution for calcium imaging** using Algorithm 2 with iADM and with reweighting (Section B, supplementary material). *Simulated data:* (a) recovered AR2 kernel; (b) estimate of spike train. *Real data:* (c) reconstructed calcium signal (d) estimate of spike train. Reweighting improves estimation quality in each case.

145 4.1.1 Recovery performance

146 We test recovery probability for varying kernel lengths n_0 and sparsity rates θ . To ensure the
147 problem size is sufficiently large, we set $m = 100n_0$. For each n_0 and θ , we randomly generate¹⁰
148 data $\mathbf{a}_0 \circledast \mathbf{x}_0$ using iADM with $\lambda = \frac{10^{-2}}{\sqrt{n_0\theta}}$. The probability of recovering a signed shift of \mathbf{a}_0 is
149 shown in Figure 4. Recovery is likely when sparsity is low compared to the kernel length. The
150 coherent problem setting has a smaller success region compared to the incoherent setting.
151

152 4.1.2 Momentum and homotopy

153 We demonstrate the effects of momentum acceleration and homotopy on convergence of the objective
154 Ψ_λ . We deconvolve clean observations with $n_0 = 10^2$, $m = 10^4$, and $\theta = n_0^{-3/4}$ for both coherent
155 and incoherent \mathbf{a}_0 . Algorithm 1 with data initialization is used to solve (BL) with $\lambda = \frac{0.3}{\sqrt{n_0\theta}}$,
156 with and without momentum ($\alpha = 0$) and homotopy. For iADM with momentum, we use an
157 iteration dependent $\alpha_k = \frac{k-1}{k+2}$ [24]. With homotopy, we apply Algorithm 2 with the hyperparameters
158 $\lambda^{(1)} = \max_\ell |\langle s_\ell[\mathbf{a}^{(0)}], \mathbf{y} \rangle|$ [29], $\lambda^* = \frac{0.3}{\sqrt{n_0\lambda}}$, $\eta = 0.8$, and $\delta = 0.1$. The final solve of (BL),
159 regardless of method, uses a precision of $\varepsilon^* = 10^{-6}$. The results show the effectiveness of momentum
160 and homotopy on coherent problem settings, see Figures 4c and 4d.

¹⁰ $\mathcal{BR}(\theta)$ denotes the Bernoulli-Rademacher distribution, which has values ± 1 w.p. $\theta/2$ and zero w.p. $1 - \theta$.

161 **4.2 Imaging applications**

162 We demonstrate the performance and generality of the proposed method with experiments on calcium
 163 fluorescence imaging, a popular modality for studying spiking activity in large neuronal populations
 164 [31], and stochastic optical reconstruction microscopy (STORM) [32, 33, 34], a superresolution
 165 microscopy modality used to image structures within living cells.¹¹

166 **4.2.1 Sparse deconvolution of calcium signals**

167 Neural spike trains are temporal signals created by action potentials, which induce a transient change
 168 in the amount of calcium present in the surrounding environment. The observed calcium concentration
 169 can be modeled as a convolution $a_0 \circledast x_0$ of the transient response a_0 and the spike train x_0 . Typically
 170 neither a_0 nor x_0 is perfectly known ahead of time.

171 We first test our method on synthetic data generated according to an AR2 model for a_0 , which
 172 was used for deconvolution in [14]. This kernel is highly shift-coherent and poses a challenging
 173 problem setting for SaSD. Here, $x_0 \sim_{\text{i.i.d.}} \text{Bernoulli}(n_0^{-4/5}) \in \mathbb{R}^{10^4}$ and additive noise is generated
 174 as $n \sim_{\text{i.i.d.}} \mathcal{N}(0, 5 \cdot 10^{-2})$. Figures 5a and 5b demonstrate accurate recovery of both a_0 and x_0 in
 175 this synthetic setting. We next test our method on real data¹²; Figures 5c and 5d demonstrate recovery
 176 of spike locations in the real setting. Although iADM provides decent performance in each case,
 177 noise suppression and estimation quality can be improved by stronger sparsification methods, such as
 178 the reweighting technique [37] — see Section B of the supplementary material.

179 **4.2.2 Sparse blind deconvolution for super-resolution fluorescence microscopy**

180 The spatial resolution of fluorescence microscopy is often limited by the diffraction of light: its
 181 wavelength (i.e. several hundred nanometers) is often larger than typical molecular length-scales in
 182 cells, preventing a detailed characterization of most subcellular structures.

183 The STORM technique is developed to overcome this resolution limit. Instead of activating all the
 184 fluorophores at the same time, STORM multiplexes the image by randomly activating photoswitchable
 185 fluorescent probes over multiple frames, each containing a subset of the molecules present (Figure 6).
 186 If the location of these molecules can be precisely determined for each frame, synthesizing all
 187 deconvolved frames will produce a super-resolution microscopy image with nanoscale resolution.

188 For each image frame, the localization task can be formulated via the SaS model

$$\underbrace{\mathbf{Y}_t}_{\text{STORM frame}} = \underbrace{\iota \mathbf{A}_0}_{\text{point spread function}} \circledast \underbrace{\mathbf{X}_{0,t}}_{\text{sparse point sources}} + \underbrace{\mathbf{N}_t}_{\text{noise}}. \quad (7)$$

189 Here we will solve the localization task on the single-molecule localization microscopy (SMLM)
 190 benchmarking dataset¹³ via SaSD, recovering both the PSF \mathbf{A}_0 and the point source map $\mathbf{X}_{0,t}$
 191 simultaneously. We apply iADM with reweighting (see Section B of the supplementary material)
 192 on frames from the video sequence “Tubulin” containing 500 frames of size 128×128 , where each
 193 pixel is of 100nm^2 resolution¹⁴; the fluorescence wavelength is 690nm and the imaging frequency is
 194 $f = 25\text{Hz}$. The recovered activation maps individual time frames and the aggregated super-resolution
 195 image is shown in Figure 6. These results demonstrate that our approach can accurately predict the
 196 PSF and the activation map for each video frame, producing higher resolution microscopy images¹⁵.

197 **4.2.3 Localizing neurons in calcium images**

198 Our methods are easily extended to handle superpositions of multiple SaS signals. In calcium imaging,
 199 this can potentially be used to track the neurons in video sequences, a challenging task due to (non-)
 200 rigid motion, overlapping sources, and irregular background noise [38, 39]. We consider frames

¹¹Other such methods developed for similar modalities include photoactivated localization microscopy (PALM) [35], and fluorescence photoactivation localization microscopy (fPALM) [36].

¹²Obtained at <http://spikefinder.codeneuro.org>.

¹³Data can be accessed at <http://bigwww.epfl.ch/smlm/datasets/index.html>.

¹⁴Here we solve SaSD on the same 128×128 grid. In practice, the localization problem is solved on a finer grid, so that the resulting resolution can reach $20 - 30\text{ nm}$.

¹⁵The recovered PSF is provided in Section C in the supplementary material.

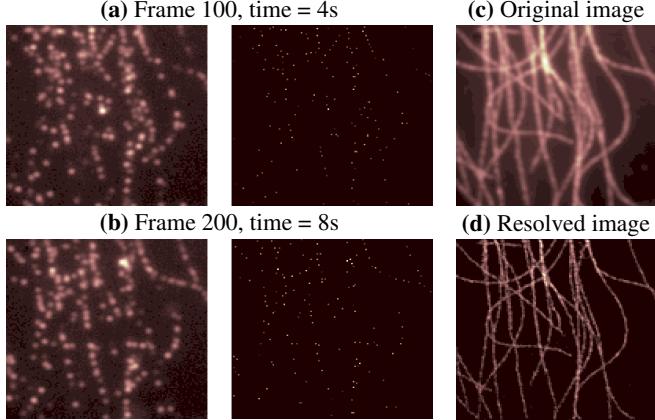


Figure 6: **SaSD for STORM imaging.** (a, b) Individual frames (left) and predicted point process map using SaSD (right). (c, d) shows the original microscopy and the super-resolved image obtained by our method.

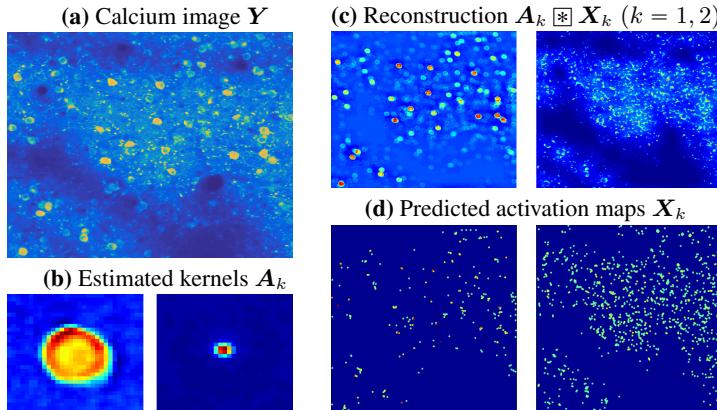


Figure 7: **Classification of calcium images.** (a) Original calcium image; (b) respective kernel estimates; (c) reconstructed images with the (left) neuron and (right) dendrite kernels; (d) respective occurrence map estimates.

video obtained via the two-photon calcium microscopy dataset from the Allen Institute for Brain Science¹⁶, shown in Figure 7. Each frame contains the cross section of several neurons and dendrites, which have distinct sizes. We model this as the SaS signal $\mathbf{Y}_t = \iota \mathbf{A}_1 \boxtimes \mathbf{X}_{1,t} + \iota \mathbf{A}_2 \boxtimes \mathbf{X}_{2,t}$, where each summand consists of neurons or dendrites exclusively. By extending Algorithm 2 to recover each of the kernels \mathbf{A}_k and maps \mathbf{X}_k , we can solve this *convolutional dictionary learning* (SaS-CDL) problem which allows us to separate the dendritic and neuronal components from this image for localization of firing activity, etc. As a result, the application of SaS-CDL as a denoising or analysis tool for calcium imaging videos provides a very promising direction for future research.

209 5 Discussion

Many nonconvex inverse problems — including SaSD — are strongly regulated by their problem symmetries. Understanding this regularity (and when and how it breaks down) can provide a strong basis for developing effective algorithms. In this paper, we have attempted to illustrate this point by combining geometric intuition with practical heuristics motivated by common challenges in real deconvolution to produce an efficient, general purpose method that performs well on data arising from a range of application areas. Our approach, therefore, can serve as a general baseline for studying and developing extensions to SaSD, such as SaS-CDL [40, 41, 42] and Bayesian models [43, 18].

¹⁶Obtained at <http://observatory.brain-map.org/visualcoding/>.

217 **References**

- 218 [1] Y. Zhang, Y. Lau, H.-w. Kuo, S. Cheung, A. Pasupathy, and J. Wright, “On the global geometry of
219 sphere-constrained sparse blind deconvolution,” in *Proceedings of the IEEE Conference on Computer
220 Vision and Pattern Recognition*, 2017, pp. 4894–4902.
- 221 [2] H.-W. Kuo, Y. Zhang, Y. Lau, and J. Wright, “Geometry and symmetry in short-and-sparse deconvolution,”
222 in *International Conference on Machine Learning (ICML)*, June 2019.
- 223 [3] S. Cheung, Y. Lau, Z. Chen, J. Sun, Y. Zhang, J. Wright, and A. Pasupathy, “Beyond the fourier transform:
224 A nonconvex optimization approach to microscopy analysis,” *Submitted*, 2017.
- 225 [4] D. Briers, D. D. Duncan, E. R. Hirst, S. J. Kirkpatrick, M. Larsson, W. Steenbergen, T. Stromberg, and O. B.
226 Thompson, “Laser speckle contrast imaging: theoretical and practical limitations,” *Journal of biomedical
227 optics*, vol. 18, no. 6, p. 066018, 2013.
- 228 [5] A. H. Song, F. Flores, and D. Ba, “Spike sorting by convolutional dictionary learning,” *arXiv preprint
229 arXiv:1806.01979*, 2018.
- 230 [6] J. Sun, Q. Qu, and J. Wright, “When are nonconvex problems not scary?” *arXiv preprint arXiv:1510.06096*,
231 2015.
- 232 [7] P.-A. Absil, R. Mahoney, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton
233 University Press, 2009.
- 234 [8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding and evaluating blind deconvolution
235 algorithms,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE,
236 2009, pp. 1964–1971.
- 237 [9] A. Benichoux, E. Vincent, and R. Gribonval, “A fundamental pitfall in blind deconvolution with sparse
238 and shift-invariant priors,” in *ICASSP-38th International Conference on Acoustics, Speech, and Signal
239 Processing-2013*, 2013.
- 240 [10] B. Huang, M. Bates, and X. Zhuang, “Super-resolution fluorescence microscopy,” *Annual review of
241 biochemistry*, vol. 78, pp. 993–1016, 2009.
- 242 [11] E. Candes and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution,” *Communications
243 on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- 244 [12] P. Campisi and K. Egiazarian, *Blind image deconvolution: theory and applications*. CRC press, 2016.
- 245 [13] H. G. Rey, C. Pedreira, and R. Q. Quiroga, “Past, present and future of spike sorting techniques,” *Brain
246 research bulletin*, vol. 119, pp. 106–117, 2015.
- 247 [14] J. Friedrich, P. Zhou, and L. Paninski, “Fast online deconvolution of calcium imaging data,” *PLoS
248 computational biology*, vol. 13, no. 3, p. e1005423, 2017.
- 249 [15] S. Baker and T. Kanade, “Limits on super-resolution and how to break them,” *IEEE Transactions on
250 Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.
- 251 [16] G. Shtengel, J. A. Galbraith, C. G. Galbraith, J. Lippincott-Schwartz, J. M. Gillette, S. Manley, R. Sougrat,
252 C. M. Waterman, P. Kanchanawong, M. W. Davidson, *et al.*, “Interferometric fluorescent super-resolution
253 microscopy resolves 3d cellular ultrastructure,” *Proceedings of the National Academy of Sciences*, vol. 106,
254 no. 9, pp. 3125–3130, 2009.
- 255 [17] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE
256 transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- 257 [18] D. Wipf and H. Zhang, “Revisiting bayesian blind deconvolution,” *The Journal of Machine Learning
258 Research*, vol. 15, no. 1, pp. 3595–3634, 2014.
- 259 [19] S. Ling and T. Strohmer, “Blind deconvolution meets blind demixing: Algorithms and performance bounds,”
260 *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4497–4520, 2017.
- 261 [20] P. Walk, P. Jung, G. E. Pfander, and B. Hassibi, “Blind deconvolution with additional autocorrelations via
262 convex programs,” *arXiv preprint arXiv:1701.04890*, 2017.
- 263 [21] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*. Springer Science & Business
264 Media, 2013, vol. 87.
- 265 [22] B. T. Polyak, “Some methods of speeding up the convergence of iteration methods,” *USSR Computational
266 Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.
- 267 [23] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,”
268 *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- 269 [24] T. Pock and S. Sabach, “Inertial proximal alternating linearized minimization (ipalm) for nonconvex and
270 nonsmooth problems,” *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1756–1787, 2016.

- 271 [25] C. Jin, P. Netrapalli, and M. I. Jordan, "Accelerated gradient descent escapes saddle points faster than
272 gradient descent," *arXiv preprint arXiv:1711.10456*, 2017.
- 273 [26] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.
- 274 [27] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1 -minimization: Methodology and
275 convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.
- 276 [28] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, "Sparse reconstruction by separable approximation,"
277 *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.
- 278 [29] L. Xiao and T. Zhang, "A proximal-gradient homotopy method for the sparse least-squares problem," *SIAM
279 Journal on Optimization*, vol. 23, no. 2, pp. 1062–1091, 2013.
- 280 [30] A. Agarwal, S. Negahban, and M. J. Wainwright, "Fast global convergence rates of gradient methods for
281 high-dimensional statistical recovery," in *Advances in Neural Information Processing Systems*, 2010, pp.
282 37–45.
- 283 [31] C. Grienberger and A. Konnerth, "Imaging calcium in neurons," *Neuron*, vol. 73, no. 5, pp. 862–885, 2012.
- 284 [32] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction
285 microscopy (storm)," *Nature methods*, vol. 3, no. 10, p. 793, 2006.
- 286 [33] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic
287 optical reconstruction microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, 2008.
- 288 [34] B. Huang, H. Babcock, and X. Zhuang, "Breaking the diffraction barrier: super-resolution imaging of
289 cells," *Cell*, vol. 143, no. 7, pp. 1047–1058, 2010.
- 290 [35] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson,
291 J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution,"
292 *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.
- 293 [36] S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation
294 localization microscopy," *Biophysical journal*, vol. 91, no. 11, pp. 4258–4272, 2006.
- 295 [37] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal
296 of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
- 297 [38] E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield,
298 W. Yang, et al., "Simultaneous denoising, deconvolution, and demixing of calcium imaging data," *Neuron*,
299 vol. 89, no. 2, pp. 285–299, 2016.
- 300 [39] A. Giovannucci, J. Friedrich, P. Gunn, J. Kalfon, B. L. Brown, S. A. Koay, J. Taxidis, F. Najafi, J. L.
301 Gauthier, P. Zhou, et al., "Caiman an open source tool for scalable calcium imaging data analysis," *Elife*,
302 vol. 8, p. e38173, 2019.
- 303 [40] H. Bristow and S. Lucey, "Optimization methods for convolutional sparse coding," *arXiv preprint
304 arXiv:1406.2407*, 2014.
- 305 [41] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new
306 algorithms," *IEEE Transactions on Computational Imaging*, 2018.
- 307 [42] J. Sulam, V. Palyan, Y. Romano, and M. Elad, "Multilayer convolutional sparse modeling: Pursuit and
308 dictionary learning," *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4090–4104, 2018.
- 309 [43] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind deconvolution using a total
310 variation prior," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 12–26, 2008.
- 311 [44] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently,"
312 *arXiv preprint arXiv:1703.00887*, 2017.