

Short-and-Sparse Deconvolution

— A Geometric Approach

Yenson Lau^{*, †}, Qing Qu^{*, ‡}, Han-Wen Kuo[†], Pengcheng Zhou[§], Yuqian Zhang[¶] and John Wright^{†||}

Abstract—Short-and-sparse deconvolution (SaSD) is the problem of extracting localized, recurring motifs in signals with spatial and/or temporal structure. Variants of this problem arise in applications such as image deblurring, microscopy, neural spike sorting, and more. Short-and-sparse deconvolution is challenging in both theory and practice. Natural optimization formulations are nonconvex. Moreover, practical deconvolution problems involve smooth motifs (kernels) whose spectra decay rapidly, resulting in poor conditioning and numerical challenges. This paper is motivated by recent theoretical advances [ZLK⁺17], [KZLW19], which characterize the optimization landscape of a particular nonconvex formulation of SaS-BD. This is used to derive a *provable* algorithm which exactly solves certain non-practical instances of the SaSD problem. We leverage the key ideas from this theory (sphere constraints, data-driven initialization) to develop a *practical* algorithm, which performs well on data arising from a range of application areas. We highlight key additional challenges posed by the ill-conditioning of real SaSD problems, and suggest heuristics (acceleration, continuation, reweighting) to mitigate them. Experiments demonstrate both the performance and generality of the proposed method.

I. INTRODUCTION

Many signals arising in science and engineering can be modeled as superpositions of basic, recurring motifs, which encode critical information about a physical process of interest. Signals of this type can be modeled as the convolution¹ of a short kernel $\mathbf{a}_0 \in \mathbb{R}^{p_0}$ (the motif) with a longer sparse signal $\mathbf{x}_0 \in \mathbb{R}^m$ ($m \gg p_0$) which encodes the locations of the motifs in the sample:

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

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

This paper is motivated by recent theoretical advances in nonconvex optimization – and in particular, on the

^{*} These authors contributed equally to this work.

[†]Department of Electrical Engineering and Data Science Institute, Columbia University

[‡]Center for Data Science, New York University

[§]Department of Neuroscience, Columbia University

[¶]Department of Computer Science, Cornell University

^{||}Department of Applied Physics and Applied Mathematics, Columbia University

¹For simplicity, (I.1) uses cyclic convolution; algorithms are results also apply to linear convolution with minor modifications.

geometry of SaSD. [ZLK⁺17], [KZLW19] study particular optimization formulations for SaSD and show that the landscape is largely driven by the *problem symmetries* of SaSD. They derive provable methods for idealized problem instances, which exactly recover $(\mathbf{a}_0, \mathbf{x}_0)$ up to trivial ambiguities. While inspiring, these methods are *not practical* and perform poorly on real problem instances. In this paper, we leverage ideas from this theory, in combination with heuristics that address the conditioning of practical problems, to build a method that performs well on data arising in a range of application areas.

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

II. THE ROLE OF SYMMETRY IN SaSD

A. Symmetry and shift-coherence

A immediate, but 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

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

for any nonzero scalar α and cyclic shift $s_\ell[\cdot]$. Since recovery of \mathbf{a}_0 and \mathbf{x}_0 is equally acceptable up to some scaled shift, these *scale and shift symmetries* is largely drive the behavior of certain nonconvex optimization problems formulated for SaSD. Another important aspect of SaSD is the *shift-coherence* of its kernel,

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

Geometrically, SaSD is easier when $\mu(\mathbf{a}_0)$ is small, as the shifts of \mathbf{a}_0 are further apart on the sphere. A small $\mu(\mathbf{a}_0)$ also allows SaSD to be solved with denser \mathbf{x}_0 , as overlapping shifts are easier to distinguish (see Figure 1).

B. 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} ,

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

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

In the later subsections, we will see that the recovered kernel length p should be set slightly larger than p_0 .

The bilinear-lasso is a *nonconvex* optimization problem, as the shift symmetries of SaSD create discrete local minimizers in the objective landscape. The regularizing effect created by problem symmetries is a fairly general phenomenon [SQW15] and, as [KZLW19] shows, its influence extends far beyond local minimizers. The authors there analyze the *dropped quadratic* (DQ) objective

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

As we will see, this non-practical objective turns out to be a valid simplification for (BL) when the true kernel is itself incoherent, i.e. $\mu(\mathbf{a}_0) \simeq 0$. We are particularly interested in the objective of the *marginalized formulation*³

$$\min_{\mathbf{a} \in \mathcal{S}_p} \left[\varphi_{\text{DQ}}(\mathbf{a}) \doteq \min_{\mathbf{x} \in \mathbb{R}^m} \Phi_{\text{DQ}}(\mathbf{a}, \mathbf{x}) \right], \quad (\text{DQ})$$

since it removes \mathbf{x} via convex minimization and reduces the study of the objective landscape down to a manifold of lower dimension $p - 1$.

Regularity in the span of a few shifts: Under suitable conditions on \mathbf{a}_0 and \mathbf{x}_0 , φ_{DQ} enjoys a number of nice properties on the sphere. For instance, suppose $\mathbf{a} \simeq \alpha_1 s_{\ell_1}[\mathbf{a}_0] + \alpha_2 s_{\ell_2}[\mathbf{a}_0] \in \mathcal{S}_p$ is near the span of two shifts⁴ of \mathbf{a}_0 . If $\alpha_1 \simeq 1$ (or $\alpha_2 \simeq 0$), [KZLW19] asserts that \mathbf{a} is in a strongly convex region of φ_{DQ} with a single minimizer near $s_{\ell_1}[\mathbf{a}_0]$, and vice versa. Near the *balanced point* $\alpha_1 \simeq \alpha_2$, however, both $s_{\ell_1}[\mathbf{a}_0]$ and $s_{\ell_2}[\mathbf{a}_0]$ exert their influence on φ_{DQ} , creating a saddle-point characterized by large negative curvature along the direction of the two shifts, and positive curvature in orthogonal directions (Figure 2).

This characterization of φ_{DQ} – strong convexity near single shifts, and saddle-points near balanced points – extends to regions near superpositions of *several* shifts. In this case, regions between balanced points and individual shifts are characterized by a large negative gradient, pointing towards regions spanned by smaller subsets of shifts. It is this landscape geometry which makes efficient recovery of single shifts of \mathbf{a}_0 possible.

Optimization on the sphere: These nice properties of φ_{DQ} depend strongly on its restriction to the sphere. By uniformly increasing the curvature of φ_{DQ} [AMS09], the sphere prevents spurious local minimizers from being created as a result of any particular regions of the constraint surface with nonsmoothness or large positive curvature⁵.

C. Data-driven initialization

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

$$\mathbf{a}^{(0)} \doteq -\text{grad } \varphi_{\text{DQ}}([\mathbf{0}_{p-1} \ \tilde{\mathbf{y}}_i \ \mathbf{0}_{p-1}]). \quad (\text{II.2})$$

³ \mathbf{x}_0 can be recovered via convex optimization once \mathbf{a}_0 is found.

⁴Setting $p > p_0$ ensures that \mathcal{S}_p contains at least two shifts.

⁵Conversely, the popular ℓ_1 -norm constraint set tends to create trivial sparse minimizers w.r.t. \mathbf{a} [LWDF09], [BVG13], [ZLK⁺17].

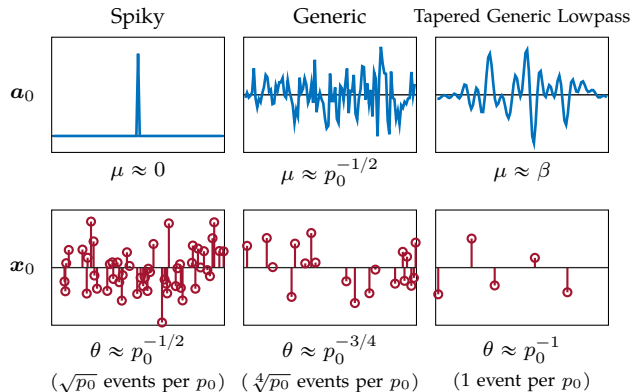


Fig. 1: Sparsity vs. coherence [KZLW19]: Smaller shift-coherence μ allows for higher sparsity rate θ (probability for any entry to be nonzero), and vice versa. With increasing difficulty: (a) A single spike δ , with $\mu(\delta) = 0$; (b) $\mathbf{a}_0 \sim \text{Unif}(\mathcal{S}_{p_0})$ has $\mu(\mathbf{a}_0) \approx p_0^{-1/2}$, allowing for $\theta \in \mathcal{O}(p_0^{-3/4})$; (c) sparsity cannot exceed $\theta \in \mathcal{O}(p_0^{-1})$ for a smooth low-pass filter, which has $\mu(\mathbf{a}_0) \simeq \text{constant}$.

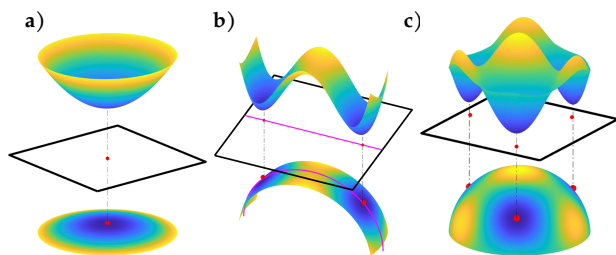


Fig. 2: Geometry of φ_{DQ} near superpositions of shifts [KZLW19]. Typical plot of (a) the strongly convex region of near a single shift; (b) the region between two shifts, containing a saddle-point with negative curvature pointing towards each shift, and positive curvature pointing away; (c) the span of three shifts of \mathbf{a}_0 .

Here, zero-padding with $\mathbf{0}_{p-1} \in \mathbb{R}^{p-1}$ and applying the Riemannian gradient helps to remove truncation effects from the window $\tilde{\mathbf{y}}_i$ [KZLW19]. However, the most important aspect of this initialization scheme is using the window $\tilde{\mathbf{y}}_i$ to put \mathbf{a} near the span of a few shifts of \mathbf{a}_0 .

D. A provable algorithm

With this initialization strategy, successful recovery of \mathbf{a}_0 and \mathbf{x}_0 now depends entirely on the regional landscape of φ_{DQ} . [KZLW19] shows that their characterization of the regional geometry of φ_{DQ} holds w.h.p. when the sparsity-coherence tradeoff $p_0\theta \lesssim (\mu(\mathbf{a}_0))^{-1/2}$ is satisfied. This implies that several popular descent methods, such as those from [GHJY15], [JNJ17], [AMS09], can be expected to locate some signed shift of \mathbf{a}_0 with proper initialization⁶.

III. DESIGNING A PRACTICAL SASD ALGORITHM

The dropped-quadratic problem (DQ) shows us an example of a regional landscape regularized by the sym-

⁶[KZLW19] also specifically proves algorithmic convergence for an algorithm based on the *curvilinear-search* method [Gol80].

metries of SaSD, and is nice enough that recovery up to a signed shift can be expected by popular first and second-order descent techniques when $\mu(\mathbf{a}_0) \simeq 0$.

As SaS applications are often motivated by sharpening or resolution tasks [HBZ09], [CFG14], [CE16], a practical algorithm must be able to handle cases where motifs are smooth and shift-coherent (i.e., $\mu(\mathbf{a}_0) \approx 1$). The (DQ) problem, on the other hand, is a valid approximation to the bilinear-lasso in only the most incoherent cases and fails otherwise, as exhibited by its poor sparsity-tradeoff rate. The correct starting place for solving SaSD in practical problem instances, therefore, is to rely on the (DQ) formulation as a “metaphor” for the bilinear-lasso under idealized settings (see Figure 5), and introduce additional computational heuristics, which we will describe below, to remedy the negative influences of large coherence.

A. Optimization via accelerated alternating descent

When \mathbf{a}_0 is shift-coherent, the Hessian of Ψ_λ becomes ill-conditioned as \mathbf{a} converges to single shifts, implying that the objective is akin to a “narrow valley”. This is known to cause excessive oscillations and slow convergence for first-order methods [Nes13] (Equation (III.1)). The momentum method [Pol64], [BT09] dampens such oscillations using an additional inertial term. For instance, consider augmenting gradient descent on some smooth $f(\mathbf{z})$ with the term \mathbf{w} ,

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

$$\mathbf{z}^{(k+1)} \leftarrow \mathbf{w}^{(k)} - \tau \cdot \nabla f(\mathbf{w}^{(k)}), \quad (\text{III.2})$$

where τ is the stepsize and α the inertial parameter⁷. As illustrated in Fig. 3, this additional term substantially reduces oscillations and improves convergence for ill-conditioned problems. For strongly convex problems, momentum improves iteration complexity⁸ from $\mathcal{O}(\kappa \log(\frac{1}{\epsilon}))$ to $\mathcal{O}(\sqrt{\kappa} \log(\frac{1}{\epsilon}))$ whilst retaining a similar computational complexity [BT09]. For nonconvex and nonsmooth problems momentum also improves escape from saddle-points [JNJ17] and convergence to local minimizers [PS16].

In Algorithm 1, we provide an inertial alternating descent method (iADM) for finding local minimizers of Ψ_λ . It modifies iPALM [PS16] to perform updates on \mathbf{a} via retraction on the sphere [AMS09]⁹. Computational performance on experiments is demonstrated by Figure 8.

B. Homotopy continuation

It is also possible to improve optimization by modifying the objective Ψ_λ directly through the sparsity penalty λ . Variations of this idea appear in both [ZLK⁺17] and [KZLW19], and can also help to mitigate the effects of large shift-coherence in practical problems.

When solving (BL) in the noise-free case, it is clear that larger choices of λ encourage sparser solutions for \mathbf{x} . Conversely, smaller choices of λ place local minimizers

⁷Setting $\alpha = 0$ reverts to vanilla gradient descent.

⁸ κ denotes the condition number of the Hessian, and ϵ the solution precision.

⁹The stepsizes t_k and τ_k are obtained by backtracking [NW06], [PS16] to ensure sufficient decrease for $\Psi_\lambda(\mathbf{a}^{(k)}, \mathbf{w}^{(k)}) - \Psi_\lambda(\mathbf{a}^{(k)}, \mathbf{x}^{(k+1)})$ and $\Psi_\lambda(\mathbf{z}^{(k)}, \mathbf{x}^{(k+1)}) - \Psi_\lambda(\mathbf{a}^{(k+1)}, \mathbf{w}^{(k+1)})$.

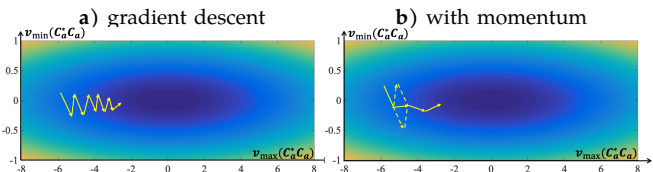


Fig. 3: Momentum acceleration. a) Iterates of gradient descent oscillate on ill-conditioned functions. b) Momentum dampens oscillation and speeds up convergence.

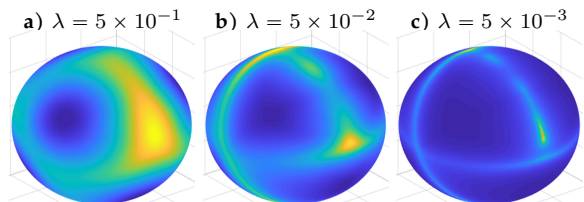


Fig. 4: Bilinear-lasso objective φ_λ on the sphere S_p , for $p = 3$ and varying λ . The function landscape of φ_λ flattens as sparse penalty λ decreases from left to right.

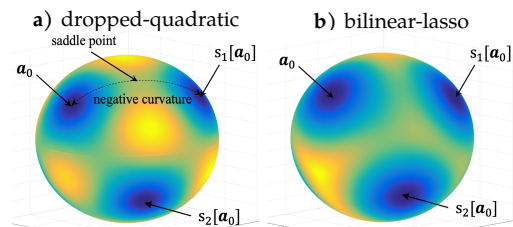


Fig. 5: The dropped-quadratic and bilinear-lasso loss in the span of three shifts. They exhibit qualitatively similar geometric properties, such as convexity near single-shifts, and negative curvature at balanced points.

of the marginal objective $\varphi_\lambda(\mathbf{a}) \doteq \min_{\mathbf{x}} \Psi_\lambda(\mathbf{a}, \mathbf{x})$ closer to signed-shifts of \mathbf{a}_0 by emphasizing reconstruction quality.

When $\mu(\mathbf{a}_0)$ is large, however, φ_λ becomes ill-conditioned as $\lambda \rightarrow 0$, due to the poor spectral conditioning of \mathbf{a}_0 . This leads to severe flatness near local minimizers, as illustrated by Figure 4, and can also lead to the creation of many spurious local minimizers when noise is present. At the expense of precision, larger values of λ force \mathbf{x} to be sparse, making Ψ_λ (restricted) strongly convex when \mathbf{a} is fixed. It is therefore important both for fast convergence and accurate recovery for λ to be chosen appropriately.

When problem parameters – such as the severity of noise, or p_0 and θ – are not known a priori, a homotopy continuation method [HYZ08], [WNF09], [XZ13] can be used to obtain a range of solutions for SaSD. Using initialization (II.2), Algorithm 2 first obtains a rough estimate $(\hat{\mathbf{a}}^{(1)}, \hat{\mathbf{x}}^{(1)})$ by solving (BL) with iADM using a large choice for $\lambda^{(1)}$; this estimate is refined by gradually decreasing $\lambda^{(n)}$ to produce the solution path $\{(\hat{\mathbf{a}}^{(n)}, \hat{\mathbf{x}}^{(n)}; \lambda^{(n)})\}$. Homotopy also ensures that \mathbf{x} remains sparse along the solution path, effectively providing the objective Ψ_λ with (restricted) strong convexity w.r.t. both \mathbf{a} and \mathbf{x} throughout optimization [ANW10]. As a result, homotopy achieves fast linear convergence for SaSD where sublinear convergence is expected otherwise (Figure 8).

Algorithm 1 Inertial Alternating Descent Method (iADM)

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

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

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:

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

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

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

$$\begin{aligned} \mathbf{z}^{(k)} &\leftarrow \mathcal{P}_{\mathcal{S}_p} \left(\mathbf{a}^{(k)} + \frac{\alpha}{\langle \mathbf{a}^{(k)}, \mathbf{a}^{(k-1)} \rangle} \cdot \mathcal{P}_{\mathbf{a}^{(k-1)}}(\mathbf{a}^{(k)}) \right) \\ \mathbf{a}^{(k+1)} &\leftarrow \mathcal{P}_{\mathcal{S}_p} \left(\mathbf{z}^{(k)} - \tau_k \cdot \text{grad}_{\mathbf{a}} \psi_\lambda(\mathbf{z}^{(k)}, \mathbf{x}^{(k+1)}) \right). \end{aligned}$$

end for

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 (II.2), $\hat{\mathbf{x}}^{(0)} = \mathbf{0} \in \mathbb{R}^m$.

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

C. Reweighted sparse penalization

When \mathbf{a}_0 is shift-coherent, minimization of the objective Ψ_λ with respect to \mathbf{x} becomes sensitive to perturbations, creating ‘‘smudging’’ effects on the recovered map \mathbf{x} . These resolution issues can be remedied with stronger *concave* regularizers. One can effectively minimize a $\log(\mathbf{x} + \epsilon)$ penalty, for example, by increasingly shrinking smaller entries of \mathbf{x} with a simple reweighting scheme [CWB08]. Figure 6, as well as calcium imaging experiments in Section IV-B and figs. 9 and 10, demonstrates improved estimation quality as a result of this method.

IV. EXPERIMENTS

A. Synthetic experiments on coherent motifs

We begin by simulations for which SaSD is solved by minimizing the bilinear-lasso (BL). We experiment with coherent kernels by discretizing the Gaussian window, $\mathbf{g}_{p,\sigma} \doteq \mathcal{P}_{\mathcal{S}_p} \left(\left[\exp \left(-\frac{(2i-p-1)^2}{\sigma^2(p-1)^2} \right) \right]_{i=1}^p \right)$, and setting the true kernel to $\mathbf{a}_0 = \mathbf{g}_{n_0,2}$. For incoherent problem settings, we sample $\mathbf{a}_0 \sim \text{Unif}(\mathcal{S}_{n_0})$ uniformly on the sphere.

1) *Recovery performance:* We test recovery probability for varying kernel lengths n_0 and sparsity rates θ . To ensure the problem size is sufficiently large, we set $m = 100n_0$.

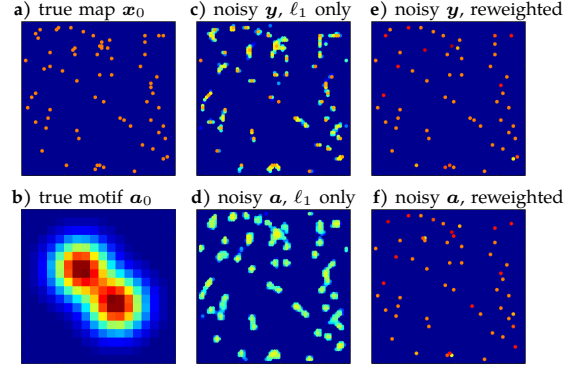


Fig. 6: Recovery of \mathbf{x}_0 with ℓ_1 -reweighting. (Left column) Truth signals. (c) Solving $\min_{\mathbf{x}} \Psi_\lambda(\mathbf{a}, \mathbf{x})$ with noisy data and coherent \mathbf{a}_0 leads to low-quality estimates of \mathbf{x} ; (d) performance suffers further when \mathbf{a} is a noisy estimate of \mathbf{a}_0 . (Right column) Reweighted ℓ_1 minimization alleviates this issue significantly.

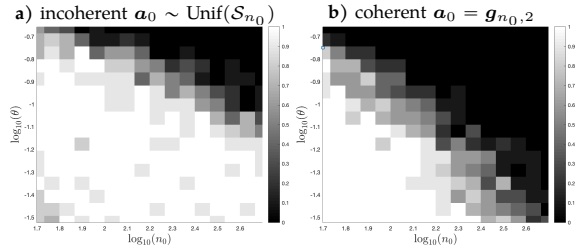


Fig. 7: Recovery performance for bilinear-lasso. For $\mathbf{x}_0 \sim \text{i.i.d. } \mathcal{BR}(\theta)$, success probability of SaS-BD by solving (BL), shown by increasing brightness, is large when the sparsity rate θ is sufficiently small compared to the length of \mathbf{a}_0 , and vice versa. Success with a fixed sparsity rate is more likely when \mathbf{a}_0 is incoherent.

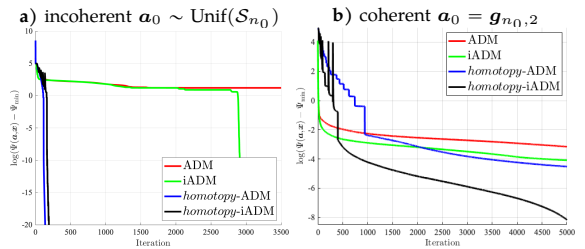


Fig. 8: Algorithmic convergence for bilinear-lasso. Convergence of function value Ψ_λ for iADM with $\alpha_k = (k-1)/(k+1)$ vs. $\alpha_k = 0$ (ADM); with and without homotopy. The final penalty is set to $\lambda = 0.3/\sqrt{n_0}\theta$. Homotopy significantly improves convergence rate. When \mathbf{a}_0 is coherent (b), convergence can be improved in the tail with momentum, but is slower overall compared to the incoherent case (a).

For each n_0 and θ , we randomly generate¹⁰ $\mathbf{x} \sim \text{i.i.d. } \mathcal{BR}(\theta)$ for both coherent and incoherent \mathbf{a}_0 . We solve ten trials of (BL) on clean observation data $\mathbf{a}_0 \otimes \mathbf{x}_0$ using iADM with $\lambda = \frac{10^{-2}}{\sqrt{n_0}\lambda}$. The probability of recovering a signed shift of \mathbf{a}_0 is shown in Figure 7. Recovery is likely when sparsity

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

is low compared to the kernel length. The smaller success region for the coherent problem setting shows its relative difficulty compared to the incoherent setting.

2) *Reweighting and homotopy*: We demonstrate the effects of momentum acceleration and homotopy on convergence of the objective Ψ_λ . We deconvolve clean observations with $n_0 = 10^2$, $m = 10^4$, and $\theta = n_0^{-3/4}$ for both coherent and incoherent \mathbf{a}_0 . **Algorithm 1** with data initialization is used to solve (BL) with $\lambda = \frac{0.3}{\sqrt{n_0\lambda}}$, with and without momentum ($\alpha = 0$) and homotopy. For iADM with momentum, we use an iteration dependent $\alpha_k = \frac{k-1}{k+2}$ [PS16]. With homotopy, we apply **Algorithm 2** with the initial estimate $\lambda^{(1)} = \max_\ell \langle s_\ell[\mathbf{a}^{(0)}], \mathbf{y} \rangle$ [XZ13], $\lambda^* = \frac{0.3}{\sqrt{n_0\lambda}}$, and $\eta = 0.8$, as well as the precision factor $\delta = 0.1$. The final solve of (BL), regardless of method, uses a precision of $\varepsilon^* = 10^{-6}$. The results show the effectiveness of momentum and homotopy on coherent problem settings, see **Figure 8**.

B. Calcium imaging applications

Based on **Section II**, solving the bilinear-lasso (BL) on the sphere provides a strong general starting point for SaS applications. Here, we show experimental results for data from calcium fluorescence imaging, a popular modality of observing the spiking activity of large neuronal populations *in vivo* [GK12]. Strong performance has also been demonstrated for microscopy and image deblurring, see [ZLK⁺17], [CSL⁺18].

1) *Sparse deconvolution of calcium signals*: Neural *spike trains* are temporal signals created by action potentials, each of which induce a *transient response* on the amount of calcium present in the surrounding environment. By viewing the transient and spike train as \mathbf{a}_0 and \mathbf{x}_0 respectively, calcium signals can be mathematically modeled as the SaS signal created by $\mathbf{a}_0 \otimes \mathbf{x}_0$. Although neither \mathbf{a}_0 or \mathbf{x}_0 are usually known in advance, deconvolution methods for recovering \mathbf{x}_0 have been proposed [FZP17] in the past by roughly modeling the transient $h(t)$ as the AR2 process $h_{\tau_1, \tau_2}(t) \simeq \exp(-t/\tau_1) - \exp(-t/\tau_2)$ for $t \geq 0$.

As kernels created by sampling $h(t)$ are shift-coherent, this is a challenging problem setting for SaSD. We show deconvolution results using iADM, with and without reweighting, for two cases: In **Figure 9**, calcium signals are simulated with $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{Bernoulli}(n_0^{-4/5}) \in \mathbb{R}^{10^4}$, and kernel $\mathbf{a}_0 \in \mathcal{S}_{100}$ by sampling the transient $h_{0.2, 0.03}(t)$ for $t \in [0, 1]$ s. We observe $\mathbf{y} = \mathbf{a}_0 \otimes \mathbf{x}_0 + \mathbf{n}$, with noise $\mathbf{n} \sim_{\text{i.i.d.}} \mathcal{N}(0, 5 \cdot 10^{-2})$. **Figure 10** demonstrates results using physical data¹¹. Although decent performance is provided by iADM alone, reweighting significantly suppresses noise and improves estimation quality for both data settings.

2) *Neuronal localization of calcium images*: The bilinear lasso easily be extended to handle multiple SaS signals. In calcium imaging, this can potentially be used to track the neurons in video sequences for calcium imaging, a challenging task due to (non)rigid motion, overlapping sources, and irregular background noise [PSG⁺16], [GFC⁺19].

We consider frames from calcium microscopy video obtained via the two-photon calcium microscopy dataset

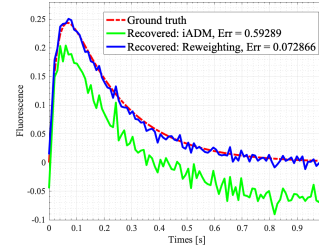


Fig. 9: Transient recovery from simulated calcium signals. (Green) Transient recovered using **Algorithm 2** without reweighting. (Blue) Reweighting significantly improves estimation quality.

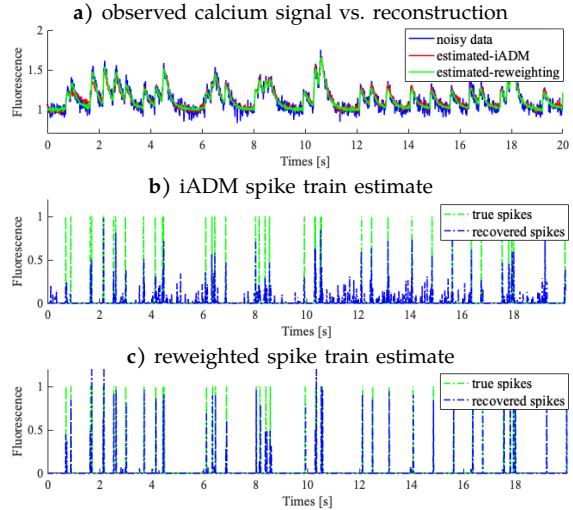


Fig. 10: Estimation of spike train from AR2 calcium signal. (a) Estimates of real data using **Algorithm 2**, with and without reweighting. (b) Spike train estimate with iADM alone, and (c) improved estimate and noise reduction with reweighting.

from the Allen Institute for Brain Science¹², shown in **Figure 11**. Each frame contains the cross section of several neurons and dendrites, which have distinct sizes. Each frame can thus be viewed as the sum of two SaS signals composed via 2D convolution, i.e. $\mathbf{Y} = \mathbf{A}_1[*] \mathbf{X}_1 + \mathbf{A}_2[*] \mathbf{X}_2$, with each comprised by neurons or dendrites exclusively.

By extending SaSD, we can apply **Algorithm 2** to solve the problem of recovering each of the kernels \mathbf{A}_k and occurrence maps \mathbf{X}_k . Solving this *convolutional dictionary learning* (SaS-CDL) problem allows us to better extract neuronal signals by, for example, removing the dendritic component from this image. Alternatively, we can apply this method to estimate the locations of neuron firings in video. As a result, the application of SaS-CDL as a denoising or analysis tool for calcium imaging videos provides a very promising direction future research.

V. SUMMARY

Due to their ability to model time-invariant signals, SaS problems are ubiquitous across science and engineering. However, the same property implies such problems are

¹¹Obtained at <http://spikefinder.codeneuro.org>.

¹²Obtained at <http://observatory.brain-map.org/visualcoding/>.

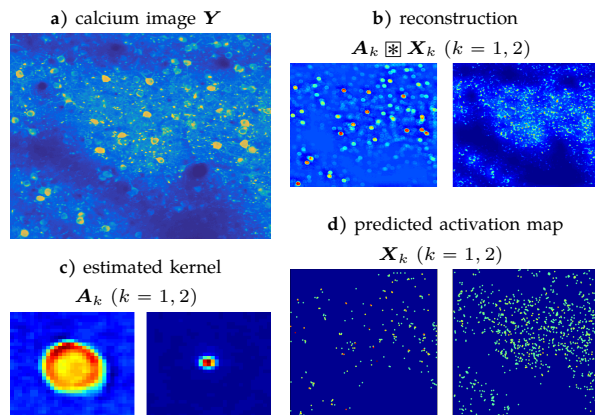


Fig. 11: Localization and classification for calcium images. (a) Original calcium image; (b) reconstructed images with the neuron (left) and dendrite (right) kernels specifically; (c) respective kernel estimates; (d) respective occurrence map estimates.

naturally formulated as nonconvex optimization problems. Although developing theory and intuition for such problems may seem intimidating at first glance, many nonconvex inverse problems – such as SaSD and SaS-CDL – are strongly regulated by their problem symmetries. Understanding this regularity and how they deviate in practical and non-ideal scenarios can provide a strong basis for developing algorithms involving SaS problems and other nonconvex inverse problems in practice.

REFERENCES

[AMS09] Pierre-Antoine Absil, Robert Mahoney, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[ANW10] Alekh Agarwal, Sahand Negahban, and Martin J Wainwright. Fast global convergence rates of gradient methods for high-dimensional statistical recovery. In *Advances in Neural Information Processing Systems*, pages 37–45, 2010.

[BT09] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[BVG13] Alexis Benichoux, Emmanuel Vincent, and Rémi Gribonval. A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors. In *ICASSP-38th International Conference on Acoustics, Speech, and Signal Processing-2013*, 2013.

[CE16] Patrizio Campisi and Karen Egiazarian. *Blind image deconvolution: theory and applications*. CRC press, 2016.

[CFG14] Emmanuel J. Candes and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[CSL⁺18] Sky Cheung, John Y. Shin, Yenson Lau, Zhengyu Chen, J. Sun, Yuqian Zhang, John Wright, and Abhay N Pasupathy. Dictionary learning in fourier transform scanning tunneling spectroscopy. 2018.

[CWB08] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

[FZP17] Johannes Friedrich, Pengcheng Zhou, and Liam Paninski. Fast online deconvolution of calcium imaging data. *PLoS computational biology*, 13(3):e1005423, 2017.

[GFG⁺19] Andrea Giovannucci, Johannes Friedrich, Pat Gunn, Jeremie Kalfon, Brandon L Brown, Sue Ann Koay, Jiannis Taxidis, Farzaneh Najafi, Jeffrey L Gauthier, Pengcheng Zhou, et al. Caiman an open source tool for scalable calcium imaging data analysis. *Elife*, 8:e38173, 2019.

[GHJY15] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Proceedings of The 28th Conference on Learning Theory*, pages 797–842, 2015.

[GK12] Christine Grienberger and Arthur Konnerth. Imaging calcium in neurons. *Neuron*, 73(5):862–885, 2012.

[Gol80] Donald Goldfarb. Curvilinear path steeplength algorithms for minimization which use directions of negative curvature. *Mathematical programming*, 18(1):31–40, 1980.

[HBZ09] Bo Huang, Mark Bates, and Xiaowei Zhuang. Super-resolution fluorescence microscopy. *Annual review of biochemistry*, 78:993–1016, 2009.

[HYZ08] Elaine T Hale, Wotao Yin, and Yin Zhang. Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

[JNJ17] Chi Jin, Praneeth Netrapalli, and Michael I Jordan. Accelerated gradient descent escapes saddle points faster than gradient descent. *arXiv preprint arXiv:1711.10456*, 2017.

[KZLW19] Han-Wen Kuo, Yuqian Zhang, Yenson Lau, and John Wright. Geometry and symmetry in short-and-sparse deconvolution. *arXiv preprint arXiv:1901.00256*, 63(7):4497–4520, 2019.

[LWDF09] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1964–1971. IEEE, 2009.

[Nes13] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[NW06] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[Pol64] Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[PS16] Thomas Pock and Shoham Sabach. Inertial proximal alternating linearized minimization (ipalm) for nonconvex and nonsmooth problems. *SIAM Journal on Imaging Sciences*, 9(4):1756–1787, 2016.

[PSG⁺16] Eftychios A Pnevmatikakis, Daniel Soudry, Yuanjun Gao, Timothy A Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 89(2):285–299, 2016.

[SQW15] Ju Sun, Qing Qu, and John Wright. When are nonconvex problems not scary? *arXiv preprint arXiv:1510.06096*, 2015.

[WNF09] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.

[XZ13] Lin Xiao and Tong Zhang. A proximal-gradient homotopy method for the sparse least-squares problem. *SIAM Journal on Optimization*, 23(2):1062–1091, 2013.

[ZLK⁺17] Yuqian Zhang, Yenson Lau, Han-Wen Kuo, Sky Cheung, Abhay Pasupathy, and John Wright. On the global geometry of sphere-constrained sparse blind deconvolution. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 4381–4389. IEEE, 2017.