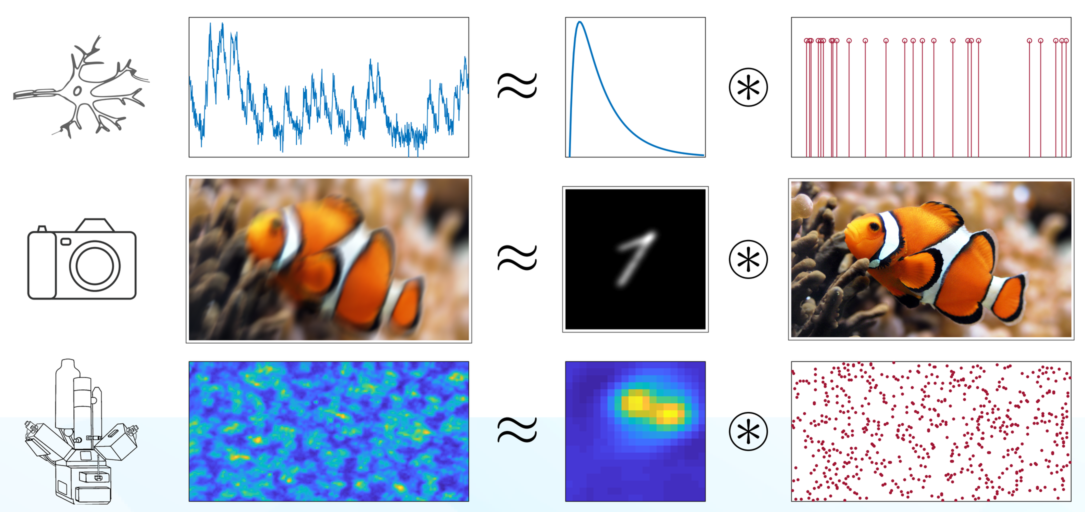
Short-and-Sparse Deconvolution – A Geometric Approach

Yenson Lau[†,⋆], Qing Qu[‡,⋆], Han-Wen (Henry) Kuo[†], Pengcheng Zhou[†], Yuqian Zhang[◇], and John Wright[†]

[†]Columbia University, [‡]New York University, [◇]Cornell University, [⋆]denote equal contribution

Short-and-Sparse (SaS) Model

- Model signals containing repeated (short) motifs:

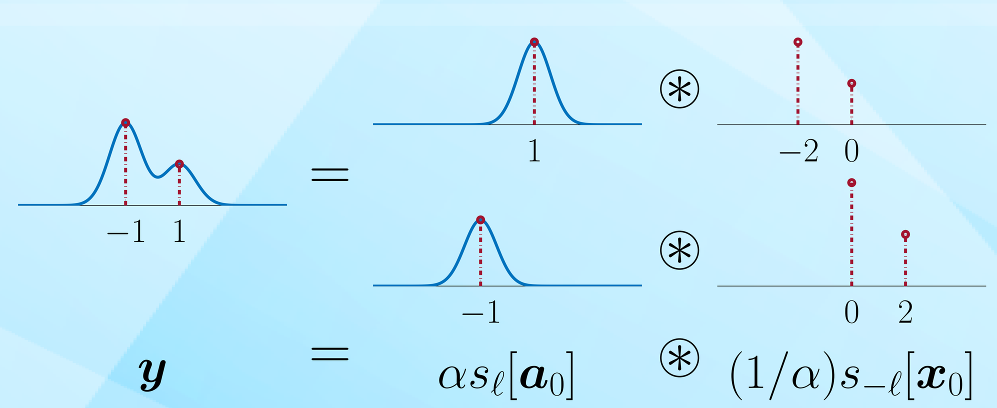


$$\boldsymbol{y} \approx \underbrace{\boldsymbol{a}_0}_{\text{short kernel}} \circledast \underbrace{\boldsymbol{x}_0}_{\text{sparse signal}}$$

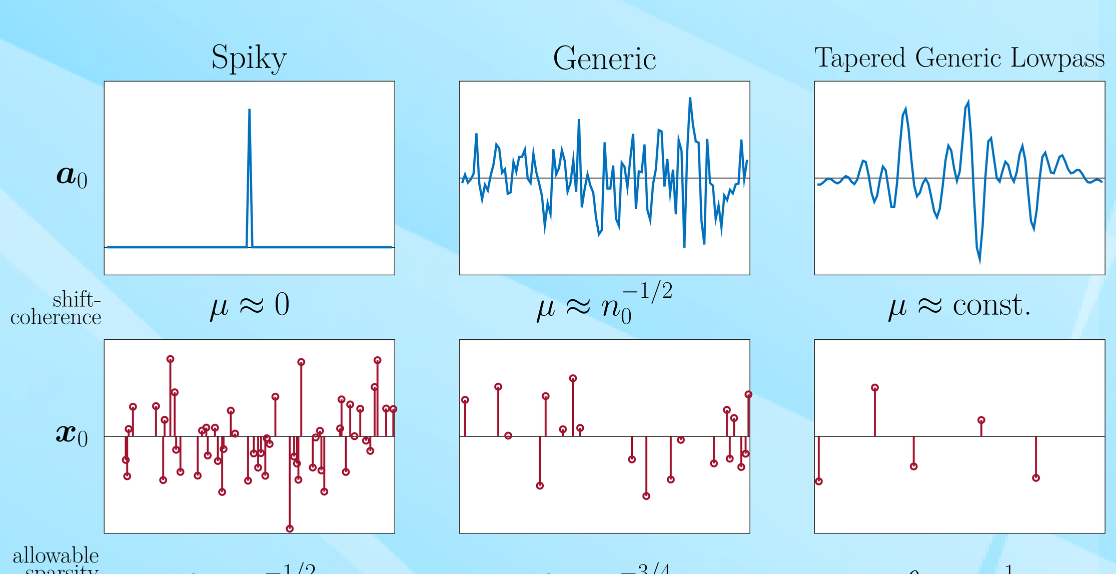
Problem: SaS Deconvolution (SaSD)

Given convolution $\boldsymbol{y} = \boldsymbol{a}_0 \circledast \boldsymbol{x}_0 \in \mathbb{R}^m$ between short $\boldsymbol{a}_0 \in \mathbb{R}^{n_0}$ $(n_0 \ll m)$, and sparse $\boldsymbol{x}_0 \in \mathbb{R}^m$, recover both $\boldsymbol{a}_0$ and $\boldsymbol{x}_0$.

Two Intrinsic Properties of SaSD



- **Symmetry:** All scale & shifts of $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ are *equivalent* solutions;



- Intrinsic shift symmetry leads to nonconvex problems.

- **Sparsity-coherence tradeoff.** For $\boldsymbol{a}_0 \in \mathbb{S}^{n_0-1}$, larger coherence of $\boldsymbol{a}_0$ (in other words, smoother)

$$\mu(\boldsymbol{a}_0) = \max_{\ell \neq 0} |\langle \boldsymbol{a}_0, s_\ell[\boldsymbol{a}_0] \rangle| \in (0, 1],$$

results in sparser $\boldsymbol{x}_0$ to be recovered, and vice versa.
  - In practice, kernels tend to be smooth with $\mu(\boldsymbol{a}_0) \approx 1$.

Problem Formulation

- **Bilinear Lasso (BL).** Natural nonconvex objective

$$\min_{\boldsymbol{a} \in \mathbb{S}^{n-1},\, \boldsymbol{x}} \Psi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x}) := \underbrace{\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{a} \circledast \boldsymbol{x}\|^2}_{\text{data fidelity}} + \lambda \cdot \underbrace{\|\boldsymbol{x}\|_1}_{\text{sparsity}}.$$

  - Optimize with $\boldsymbol{a} \in \mathbb{R}^n$ and $n = 3n_0 - 2$, to eliminate boundary effects (shift-truncations).
  - Break scale symmetry by spherical constraint $\boldsymbol{a} \in \mathbb{S}^{n-1}$;
- **Approximate Bilinear Lasso (ABL)** [2]

$$\Psi_{\mathrm{ABL}}(\boldsymbol{a}, \boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{x}\|^2 - \langle \boldsymbol{y}, \boldsymbol{a} \circledast \boldsymbol{x} \rangle + \frac{1}{2}\|\boldsymbol{y}\|^2 + \lambda \|\boldsymbol{x}\|_1,$$

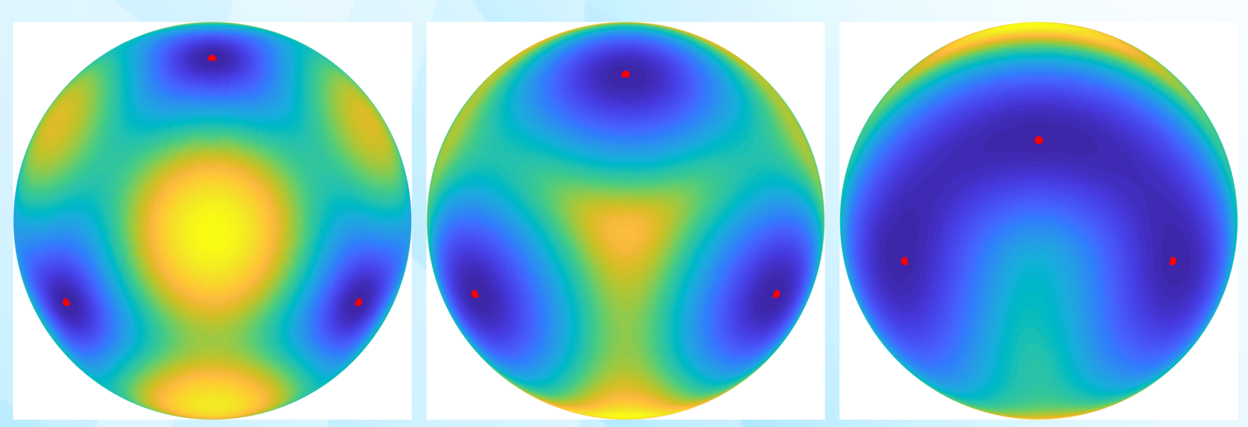
  - $\min_{\boldsymbol{x}} \Psi_{\mathrm{ABL}}(\boldsymbol{a}, \boldsymbol{x})$ has closed-form solution;
  - $\Psi_{\mathrm{ABL}}(\boldsymbol{a}, \boldsymbol{x}) \approx \Psi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x})$ when $\mu(\boldsymbol{a}_0) \searrow 0$. The approximation breaks when $\mu(\boldsymbol{a}_0)$ is large.

Nonconvex Optimization Landscape

- **Nonconvex landscape over the sphere.** Study

$$\Psi_{\mathrm{BL}}(\boldsymbol{a}) := \min_{\boldsymbol{x}} \Psi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x}), \quad \Psi_{\mathrm{ABL}}(\boldsymbol{a}) := \min_{\boldsymbol{x}} \Psi_{\mathrm{ABL}}(\boldsymbol{a}, \boldsymbol{x}).$$

  - The objectives are convex w.r.t. $\boldsymbol{x}$;
  - $\dim(\boldsymbol{a}) \ll \dim(\boldsymbol{x})$, the space of $\boldsymbol{a}$ is where measure concentrates.

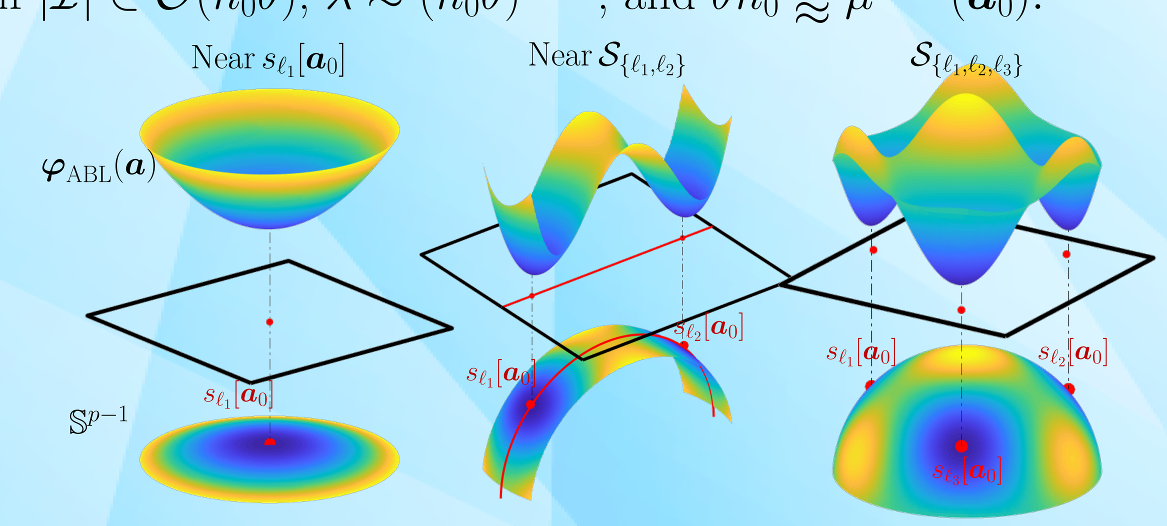


(a) $\Psi_{\mathrm{ABL}}(\boldsymbol{a})$, $\mu(\boldsymbol{a}_0) \approx 0$ (b) $\Psi_{\mathrm{BL}}(\boldsymbol{a})$, $\mu(\boldsymbol{a}_0) \approx 0$ (c) $\Psi_{\mathrm{BL}}(\boldsymbol{a})$, $\mu(\boldsymbol{a}_0) \approx 1$

- **Benign geometry over subspace of shifts** [2]

$$\mathcal{S}_{\mathcal{I}} \doteq \left\{ \textstyle\sum_{\ell \in \mathcal{I}} \alpha_\ell s_\ell[\boldsymbol{a}_0] \;:\; \alpha_\ell \in \mathbb{R} \right\} \cap \mathbb{S}^{n-1}$$

if $|\mathcal{I}| \in \mathcal{O}(n_0\theta)$, $\lambda \approx (n_0\theta)^{-1/2}$, and $\theta n_0 \lesssim \mu^{-1/2}(\boldsymbol{a}_0)$.



- Every local minimizer is a shift of the kernel $\boldsymbol{a}_0$;
- Saddle point exhibits negative curvature in symmetry breaking directions;
- Theoretical justified for $\Psi_{\mathrm{ABL}}(\boldsymbol{a})$ with $\mu(\boldsymbol{a}_0) \approx 0$ [2], less theoretically well-understood when $\Psi_{\mathrm{BL}}(\boldsymbol{a})$ or $\mu(\boldsymbol{a}_0) \approx 1$.

Convolutional Dictionary Learning

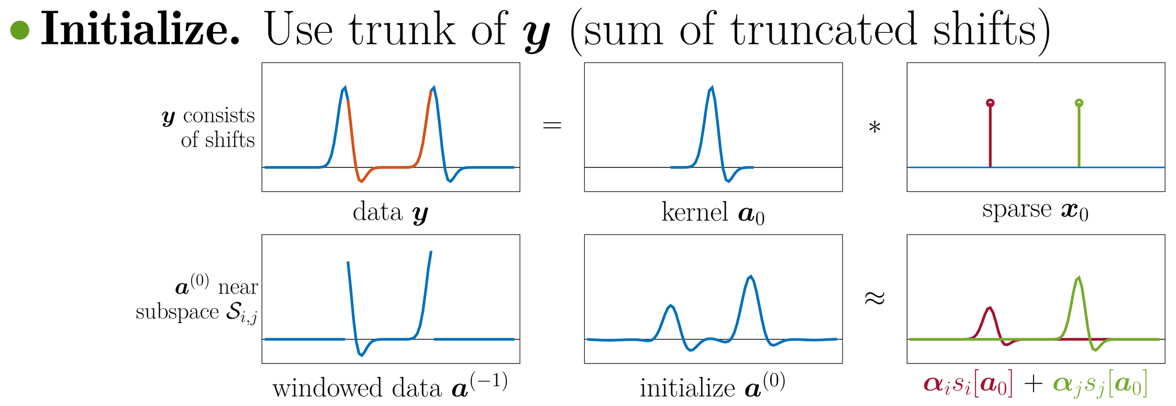
- To recover multiple unknown kernels, we optimize

$$\min_{\boldsymbol{a}_k, \boldsymbol{x}_k} \frac{1}{2}\left\| \boldsymbol{y} - \sum_{k=1}^N \boldsymbol{a}_k \circledast \boldsymbol{x}_k \right\|^2 + \lambda \sum_{k=1}^N \|\boldsymbol{x}_k\|_1, \quad \text{s.t. } \|\boldsymbol{a}_k\| = 1$$

by adapting similar ideas for recovering a single kernel.

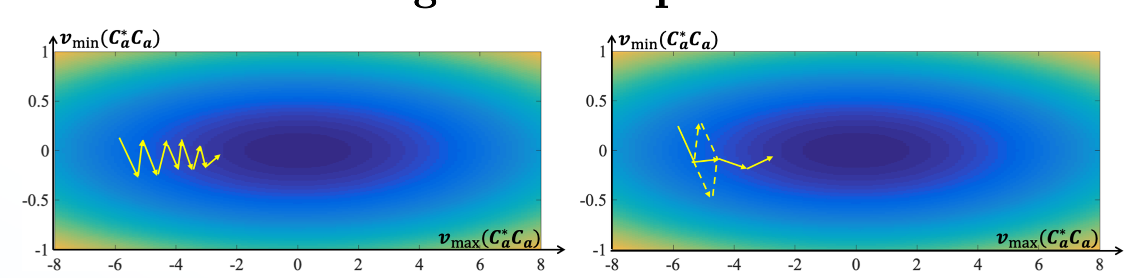
Optimization from Geometric Intuitions

- **Initialize.** Use trunk of $\boldsymbol{y}$ (sum of truncated shifts)



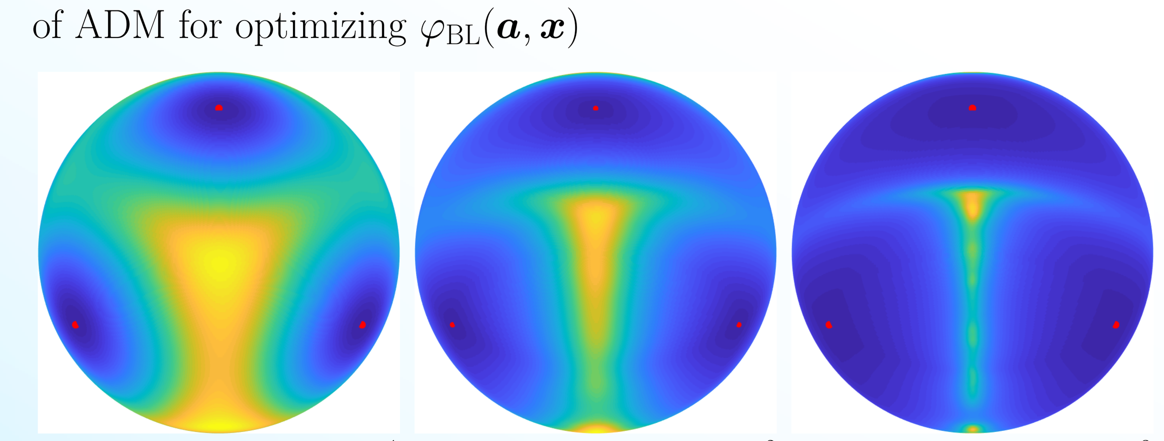
where $\boldsymbol{a}^{(0)} = -\mathcal{P}_{\mathbb{S}^{n-1}} \nabla \Psi_{\mathrm{BL}}\left(\mathcal{P}_{\mathbb{S}^{n-1}}(\boldsymbol{a}^{(-1)})\right)$.

- **Alternating descent method (ADM)**
  1. Fix $\boldsymbol{a}$ and take *proximal gradient* on $\boldsymbol{x}$:
  $$\boldsymbol{x} \leftarrow \operatorname{prox}\left(\boldsymbol{x} - \tau \cdot \nabla_{\boldsymbol{x}} \Psi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x})\right)$$
  2. Fix $\boldsymbol{x}$ and take a *Riemannian gradient* on $\boldsymbol{a}$:
  $$\boldsymbol{a} \leftarrow \mathcal{P}_{\mathbb{S}^{n-1}}\left(\boldsymbol{a} - t \cdot \operatorname{grad}_{\boldsymbol{a}} \Psi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x})\right)$$
  3. Alternate until convergence.
- **Heuristics for dealing coherent problems**



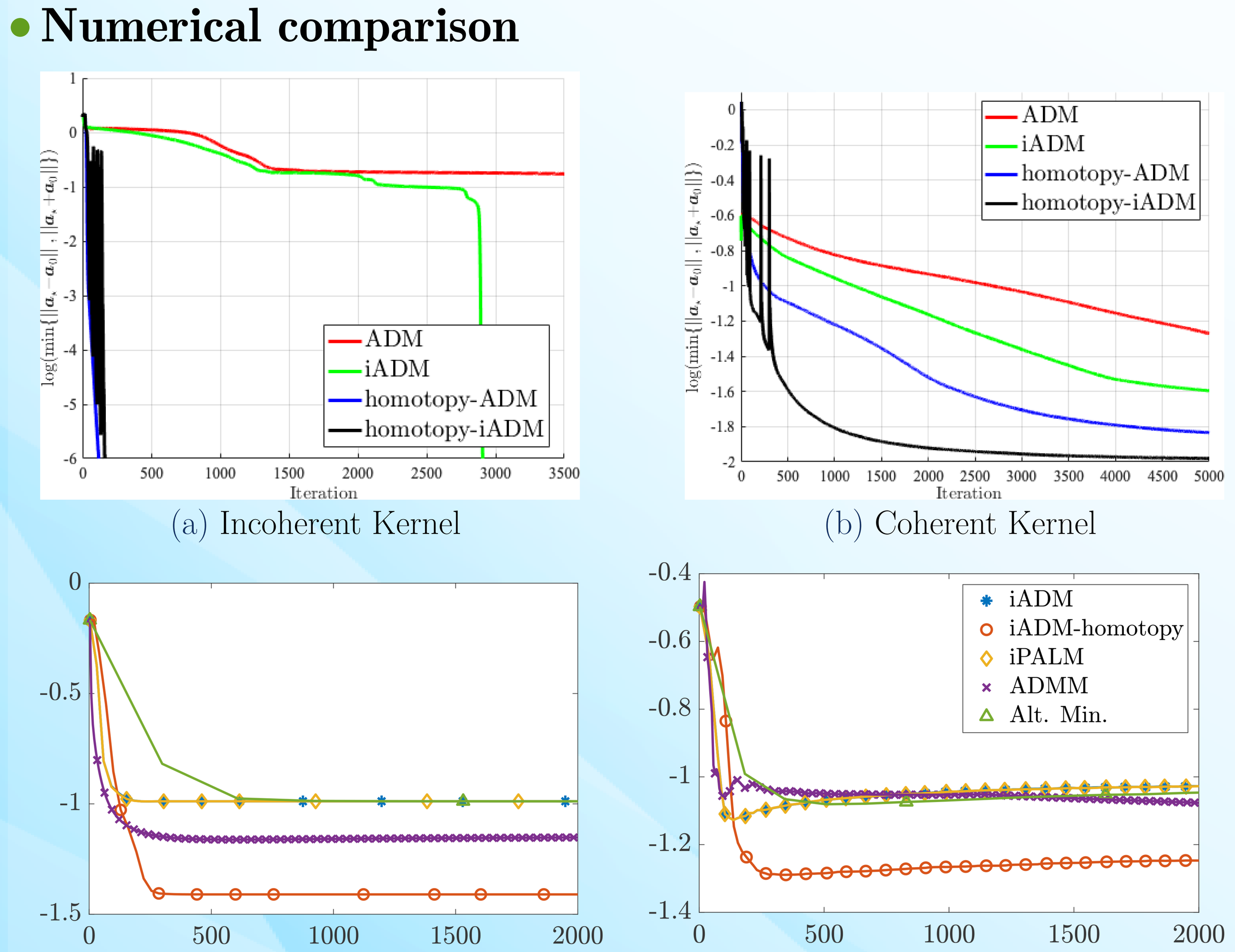
(a) Without momentum (b) With momentum

  - Momentum acceleration: deal with ill-conditioning by adding momentum term on descent steps for both $\boldsymbol{a}$ and $\boldsymbol{x}$.
  $$\boldsymbol{z}^{(k+1)} \leftarrow \boldsymbol{z}^{(k)} - \underbrace{\alpha \nabla f(\boldsymbol{z}^{(k)})}_{\text{gradient direction}} + \underbrace{\beta\left(\boldsymbol{z}^{(k)} - \boldsymbol{z}^{(k-1)}\right)}_{\text{inertial term}},$$
  - Homotopy method: adaptively shrink λ through the solution path $(\boldsymbol{a}, \boldsymbol{x})$ of ADM for optimizing $\varphi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x})$



(a) $\Psi_{\mathrm{BL}}(\boldsymbol{a})$, $\lambda = 5 \times 10^{-1}$ (b) $\Psi_{\mathrm{BL}}(\boldsymbol{a})$, $\lambda = 5 \times 10^{-2}$ (c) $\Psi_{\mathrm{BL}}(\boldsymbol{a})$, $\lambda = 5 \times 10^{-3}$

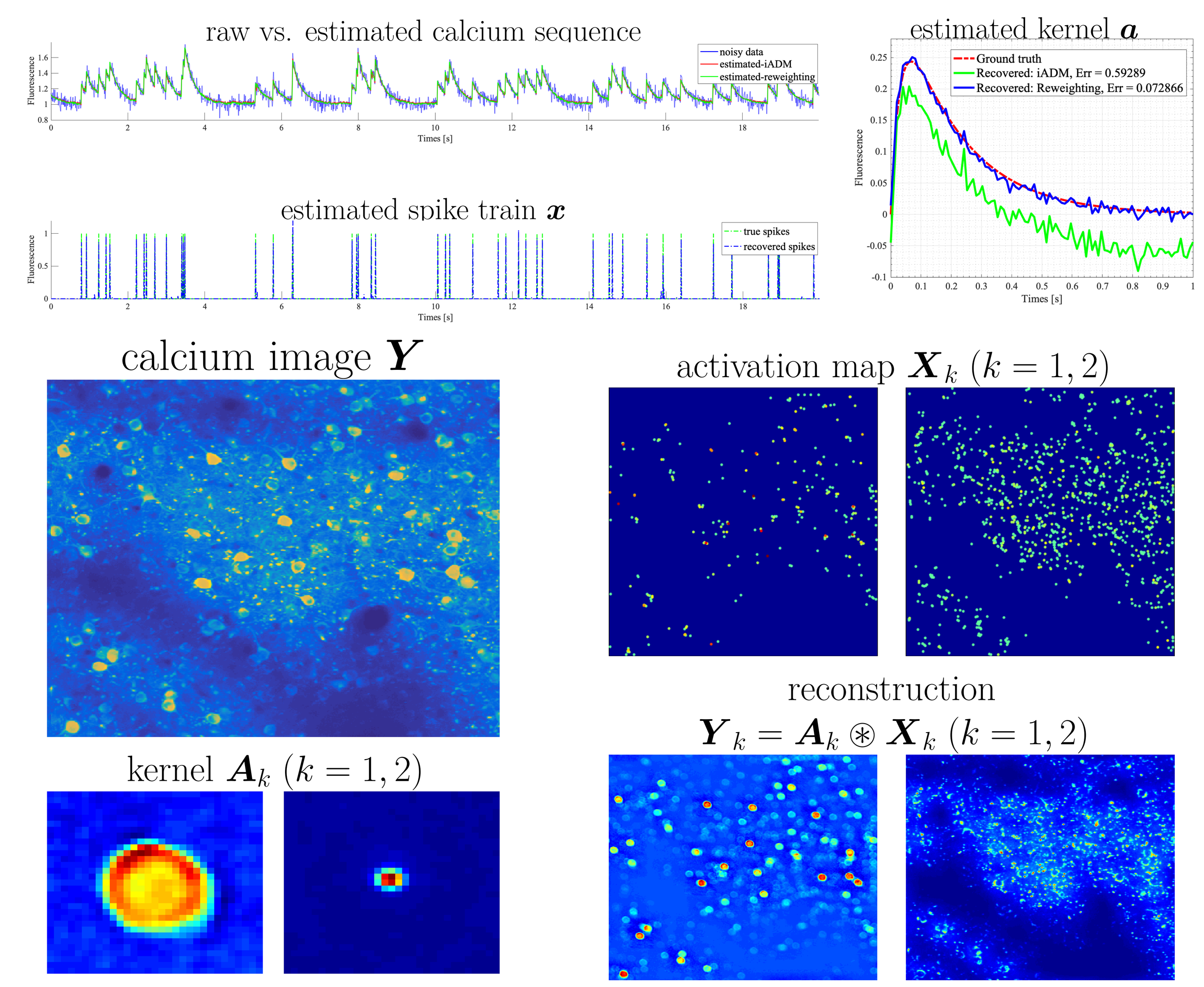
- **Numerical comparison**



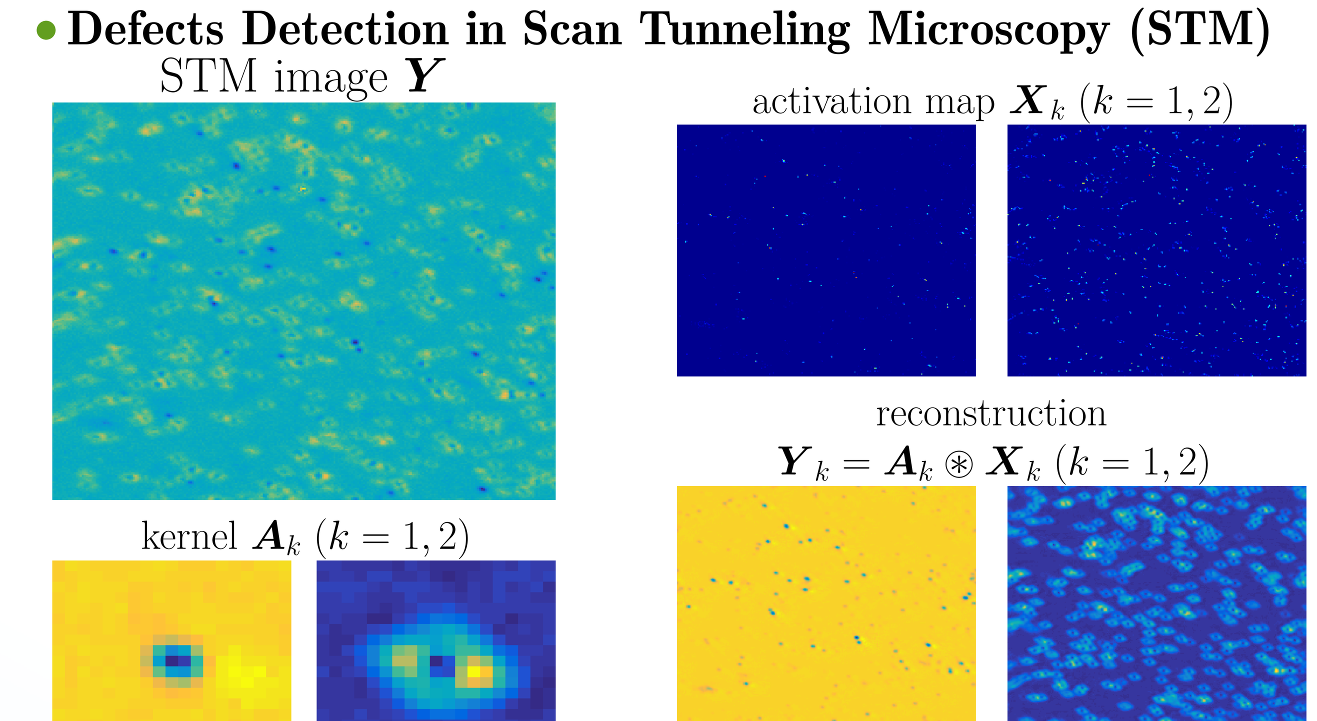
(a) Incoherent Kernel (b) Coherent Kernel (c) Incoherent Kernel (d) Coherent Kernel

Applications

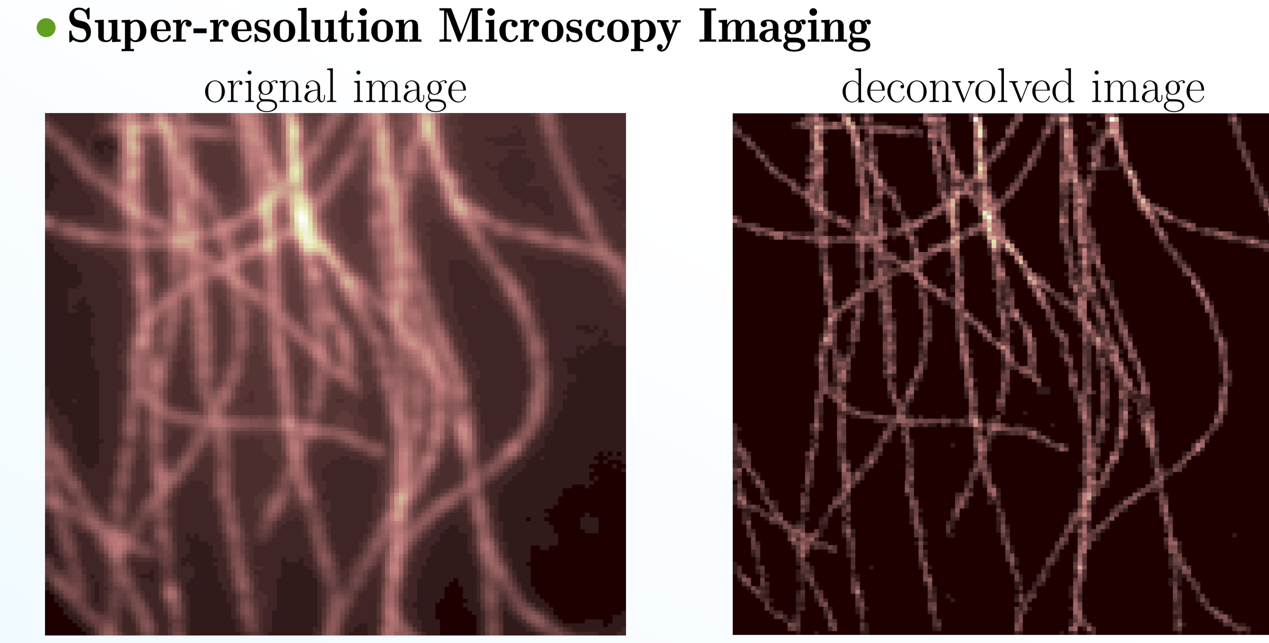
- **Calcium Imaging**



- **Defects Detection in Scan Tunneling Microscopy (STM)**



- **Super-resolution Microscopy Imaging**



- **Spike Sorting**

References

[1] Y. Lau, Q. Qu, and H. Kuo, P. Zhou, Y. Zhang, and J. Wright, "Short-and-sparse deconvolution – A geometric approach", *arXiv preprint arXiv:1908.10959, submitted to ICLR'20*, 2019.

[2] H. Kuo and Y. Lau, Y. Zhang, and J. Wright, "Geometry and Symmetry in Short-and-Sparse Deconvolution", *ICML*, 2019.

[3] Q. Qu and Y. Zhai, X. Li, Y. Zhang, and Z. Zhu, "Geometric analysis of nonconvex optimization landscapes for overcomplete learning". *submitted to ICLR'20*, 2019.

https://deconvlab.github.io/

https://github.com/qingqu06/sparse_deconvolution