Short-and-Sparse Deconvolution

A Geometric Approach

Qing Qu

Center for Data Science

New York University

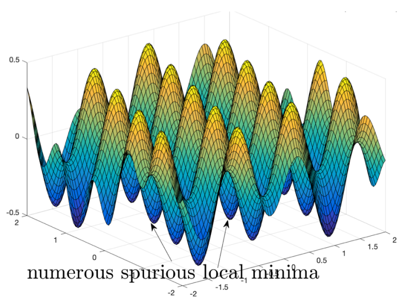
Joint with **Yenson Lau**, **Han-wen Kuo**, **Pengcheng Zhou**, and **John Wright** (Columbia U.)

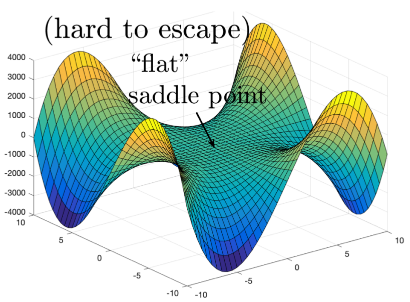
August 8, 2019

Geometric Analysis for Nonconvex Optimization

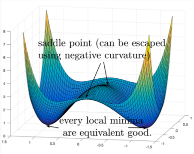
“bad” local minima

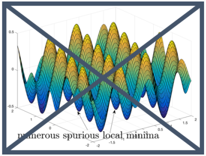
“flat” saddle point

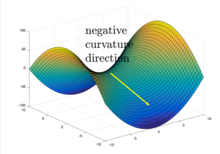


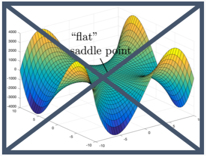


Geometric Analysis for Nonconvex Optimization



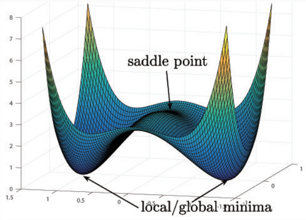






- Phase retrieval [Sun et al. ’18]
- Dictionary learning [Sun et al. ’16]
- Orthogonal tensor decomposition [Ge et al., ’15]
- Low-rank matrix recovery and completion

  [Ge et al., ’16, Bhojanapalli et al., ’16]
- Phase synchronization [Boumal ’17]
- Community detection [Bandeira et al., ’17]
- Shallow neural network [Song et al., ’18]



Short-and-Sparse (SaS) Deconvolution

Can we recover **short** kernel $\boldsymbol{a}_0$ & **sparse** signal $\boldsymbol{x}_0$ from

$$\boxed{\boxed{\boldsymbol{y} \quad = \quad \boldsymbol{a}_0 \quad \circledast \quad \boldsymbol{x}_0 \quad ?}}$$

We assume $m \gg n_0$, such that

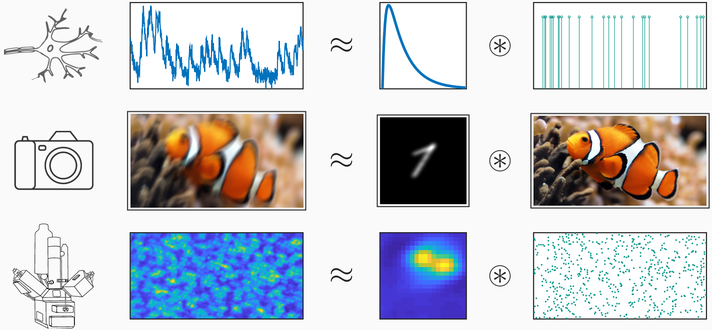
- $\boldsymbol{a}_0 \in \mathbb{R}_0^n$ is short kernel;
- $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{x}_0 \in \mathbb{R}^m$ are long sequence.

Short-and-Sparse (SaS) Deconvolution

Can we recover **short** kernel a_0 & **sparse** signal x_0 from

$$\boxed{y \quad = \quad a_0 \quad \circledast \quad x_0 \quad ?}$$

Numerous applications in *neuroscience*, *computer vision*, and *computational imaging*, etc:



$$y \quad \approx \quad a_0 \quad \circledast \quad x_0$$

Outline

Intrinsic Property

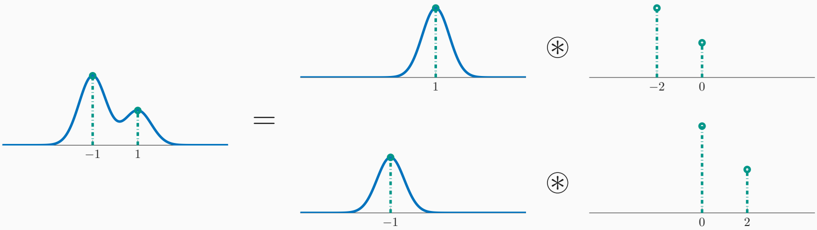
Problem Formulation and Geometry

Using Geometric Intuitions to Build Practical Algorithms

Applications

Intrinsic Property I: Symmetry

- **Shift Symmetry**: $\boldsymbol{a}_0 \circledast \boldsymbol{x}_0 = \mathrm{s}_\ell[\boldsymbol{a}_0] \circledast \mathrm{s}_{-\ell}[\boldsymbol{x}_0]$
- **Scaling Symmetry**: $\boldsymbol{a}_0 \circledast \boldsymbol{x}_0 = \alpha\boldsymbol{a}_0 \circledast \alpha^{-1}\boldsymbol{x}_0$

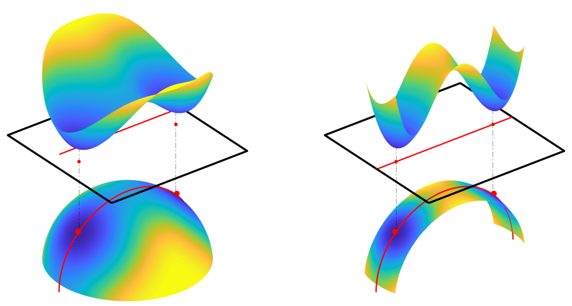


$$\boldsymbol{y} = \alpha\, s_\ell[\boldsymbol{a}_0] \circledast (1/\alpha)\, s_{-\ell}[\boldsymbol{x}_0]$$

Symmetry Leads to Nonconvex Problems

- Scaling is easy to handle, e.g., $\|\boldsymbol{a}\| = 1$;
- Shift symmetry creates equivalent solutions:

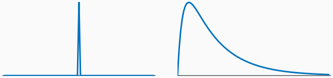
$$(\boldsymbol{a}_0, \boldsymbol{x}_0) \quad = \quad \left(\mathrm{s}_\ell\left[\boldsymbol{a}_0\right], \mathrm{s}_{-\ell}\left[\boldsymbol{x}_0\right]\right)$$



Intrinsic Property II: Sparsity & Coherence

Coherence of kernel $\boldsymbol{a}_0$:

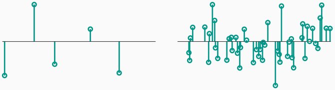
$$\mu \quad = \quad \max_{i \neq j} \left| \left\langle s_i[\boldsymbol{a}_0] \, , \, s_j[\boldsymbol{a}_0] \right\rangle \right|$$



$\mu \downarrow$ $\quad$ $\mu \uparrow$

Sparsity of signal $\boldsymbol{x}_0$:

sparsity level $\quad \theta \quad \in \quad (0, 1)$

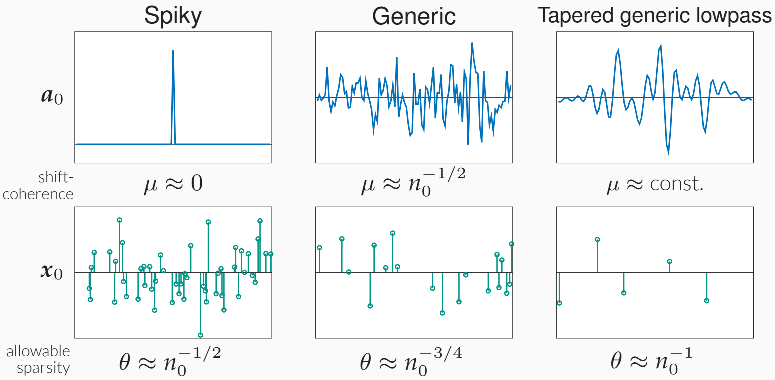


$\theta \downarrow$ $\quad$ $\theta \uparrow$

The problem becomes more difficult when

- Coherence $\mu \uparrow$ (solutions become closer)
- Sparsity $\theta \uparrow$ (more unknowns)

Intrinsic Property II: Sparsity - Coherence Tradeoff



Larger coherence μ of a_0 results in sparser x_0 we can solve.

Outline

Intrinsic Property

Problem Formulation and Geometry

Using Geometric Intuitions to Build Practical Algorithms

Applications

Nonconvex Formulation

Bilinear Lasso (BL): a natural nonconvex formulation

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

- Study nonconvex landscape over the sphere,

$$\varphi_{\mathrm{BL}}(\boldsymbol{a}) := \min_{\boldsymbol{x}}\left\{\Psi_{\mathrm{BL}}(\boldsymbol{a},\boldsymbol{x})\right\}, \quad \boldsymbol{a}\in\mathbb{S}^{n-1}.$$

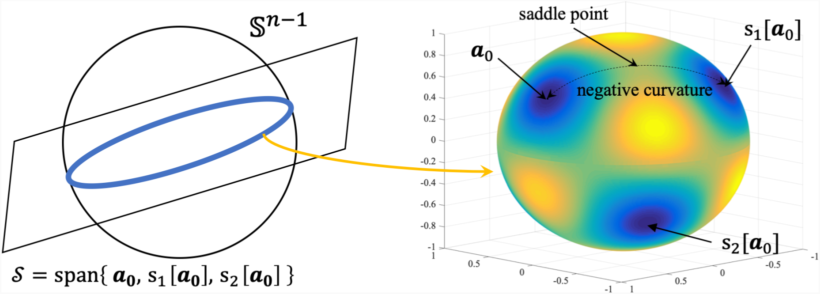
- When $\mu(\boldsymbol{a}_0)$ small, simplify the analysis through linearization [Kuo et al. ICML'19],

$$\varphi_{\mathrm{BL}}(\boldsymbol{a}) \approx \varphi_{\mathrm{ABL}}(\boldsymbol{a}).$$

Geometry I: Incoherent Case

Over subspace spanned by shifts of $\boldsymbol{a}_0$

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

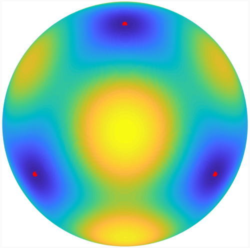


$$\varphi_{\mathrm{ABL}}(\boldsymbol{a}) = \min_{\boldsymbol{x}} \Psi_{\mathrm{ABL}}(\boldsymbol{a}, \boldsymbol{x}), \quad \boldsymbol{a} \in \mathbb{S}^{n-1},$$

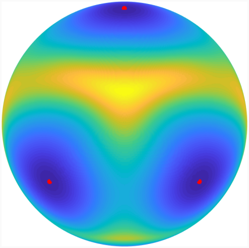
Geometry I: Incoherent Case

Over subspace spanned by shifts of $\boldsymbol{a}_0$

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



$\varphi_{\mathrm{ABL}}(\boldsymbol{a})$

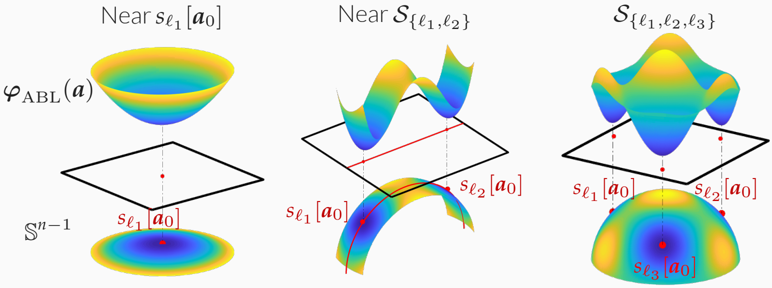


$\varphi_{\mathrm{BL}}(\boldsymbol{a})$

Geometry I: Incoherent Case

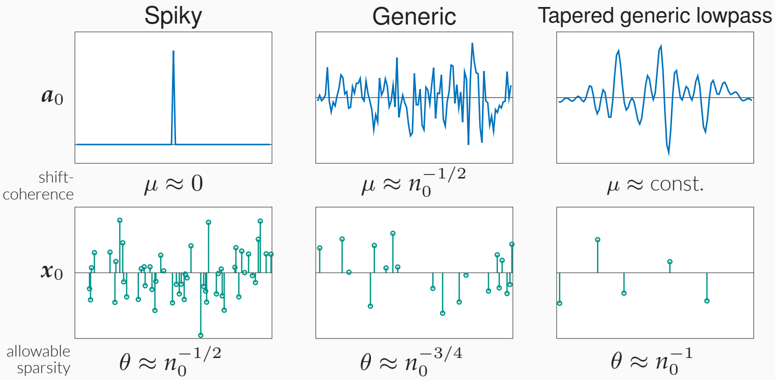
Over subspace spanned by shifts of $\boldsymbol{a}_0$

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



- Local minimizers are near shifts of $\boldsymbol{a}_0$;
- Negative Curvature breaks symmetry between shifts.

Geometry: Incoherent vs. Coherent Cases

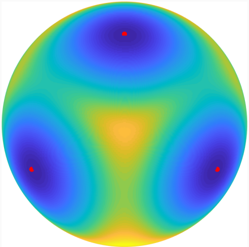


Larger coherence μ of $\boldsymbol{a}_0$ results in sparser $\boldsymbol{x}_0$ we can solve.

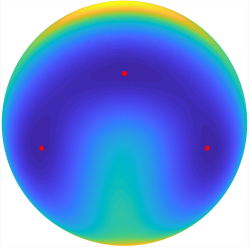
Geometry II: Coherent Case

Over subspace spanned by shifts of $\boldsymbol{a}_0$

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



$\varphi_{\mathrm{BL}}(\boldsymbol{a})$, small $\mu(\boldsymbol{a})$



$\varphi_{\mathrm{BL}}(\boldsymbol{a})$, large $\mu(\boldsymbol{a})$

Outline

Intrinsic Property

Problem Formulation and Geometry

Using Geometric Intuitions to Build Practical Algorithms

Applications

A Vanilla Algorithm

Alternating Descent Method (ADM):

- Fix $\boldsymbol{a}$ and take *proximal gradient* on $\boldsymbol{x}$:

$$\boldsymbol{x} \leftarrow \operatorname{prox}\left(\boldsymbol{x} - \tau \cdot \nabla_{\boldsymbol{x}} \varphi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x})\right)$$

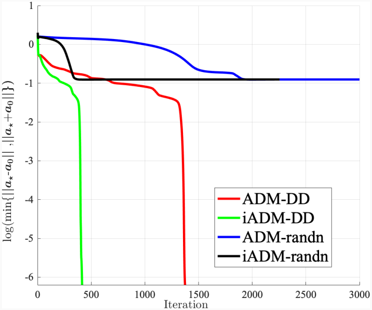
- Fix $\boldsymbol{x}$ and take a *Riemannian gradient* on $\boldsymbol{a}$:

$$\boldsymbol{a} \leftarrow \mathcal{P}_{\mathbb{S}^{n-1}}\left(\boldsymbol{a} - t \cdot \operatorname{grad}_{\boldsymbol{a}} \varphi_{\mathrm{BL}}(\boldsymbol{a}, \boldsymbol{x})\right)$$

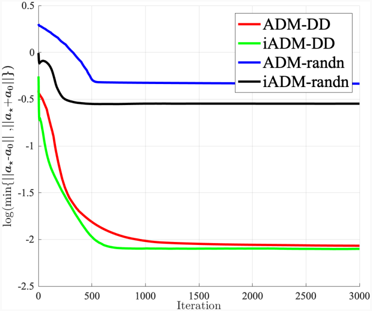
- Alternate until convergence.

From Geometry to Optimization I

Data-driven initialization: use part of the measurement y to avoid spurious local minimizer



small $\mu(a)$

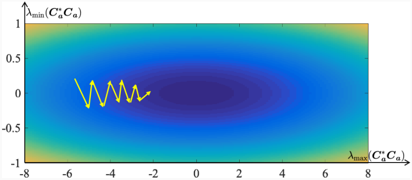


large $\mu(a)$

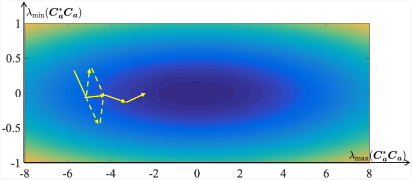
From Geometry to Optimization II

High coherence leads to **ill-conditioned** problem

without momentum



with momentum



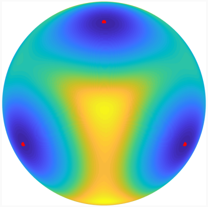
Idea: **momentum acceleration**

$$z^{(k+1)} \leftarrow z^{(k)} - \underbrace{\alpha \nabla f(z^{(k)})}_{\text{gradient direction}} + \underbrace{\beta \left(z^{(k)} - z^{(k-1)} \right)}_{\text{inertial term}},$$

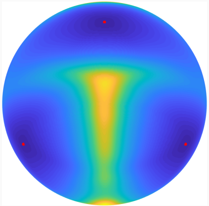
From Geometry to Optimization III

Homotopy continuation via adaptive updating penalty λ

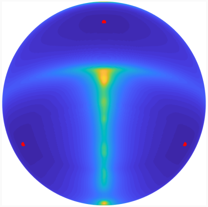
$$\boxed{\min_{\boldsymbol{a},\boldsymbol{x}} \frac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{a}\circledast\boldsymbol{x}\right\|^2+\lambda\left\|\boldsymbol{x}\right\|_1, \quad \text{s.t. } \left\|\boldsymbol{a}\right\|=1.}$$



$\lambda = 5 \times 10^{-1}$



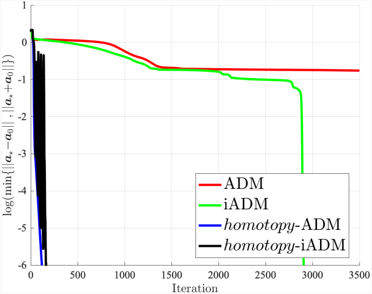
$\lambda = 5 \times 10^{-2}$



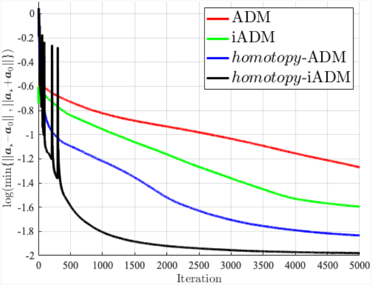
$\lambda = 5 \times 10^{-3}$

Comparison I - Iterate Convergence

Comparison of Iterate convergence on proposed methods



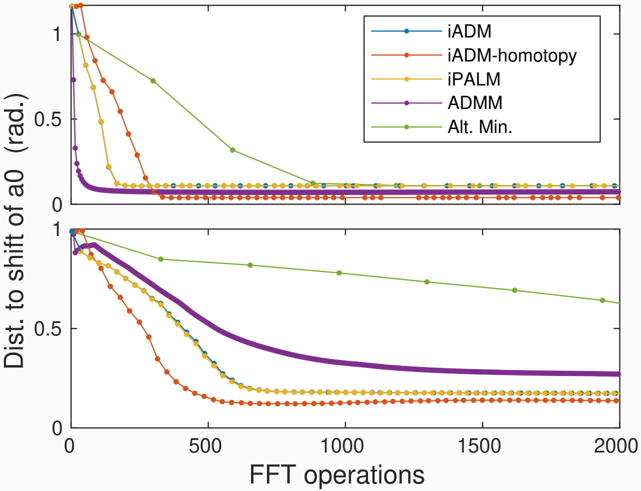
small $\mu(\boldsymbol{a})$



large $\mu(\boldsymbol{a})$

Comparison II - Computation Time

Compare with ADMM and Alternating Minimization



Outline

Intrinsic Property

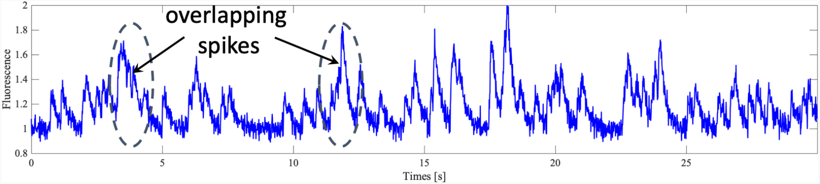
Problem Formulation and Geometry

Using Geometric Intuitions to Build Practical Algorithms

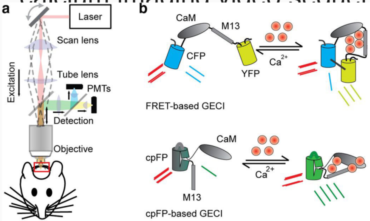
Applications

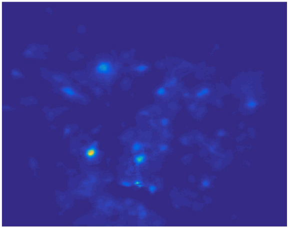
Application I: Calcium Imaging

calcium traces extracted from the video sequence

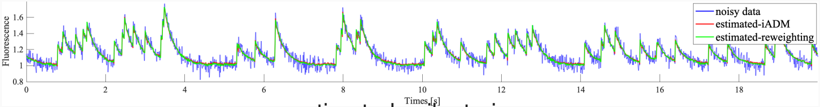


calcium imaging video sequence

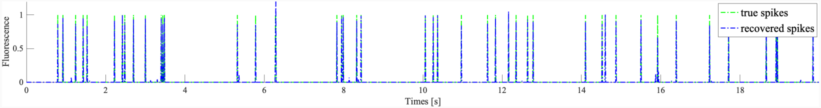


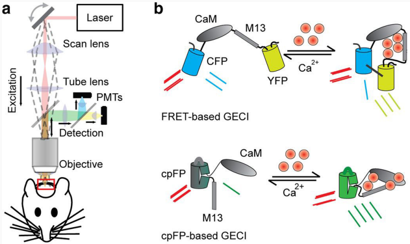


raw vs. estimated calcium sequence

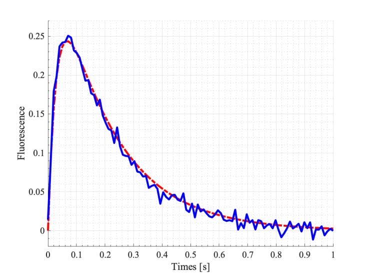


estimated spike train $\boldsymbol{x}$



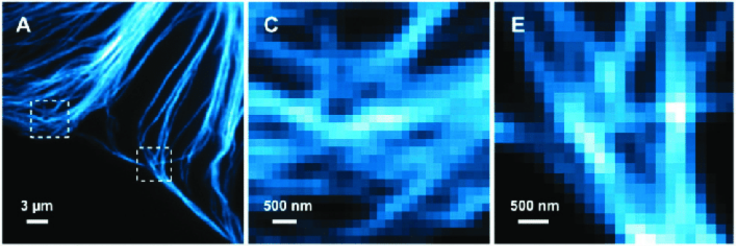


estimated kernel $\boldsymbol{a}$

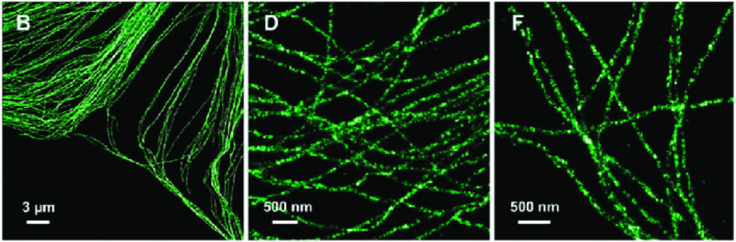


Application II: Super-resolution Microscopy

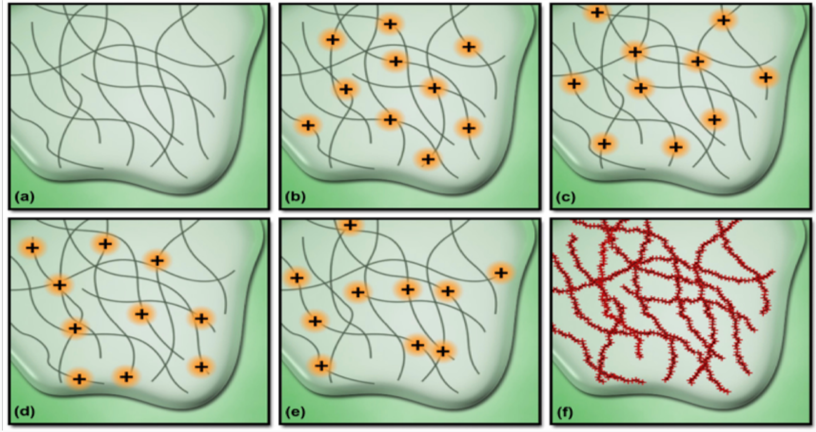
Conventional fluorescent optical microscopy



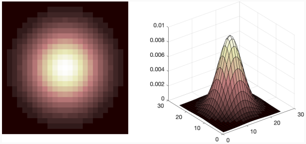
Stochastic Optical Reconstruction Microscopy (STORM)



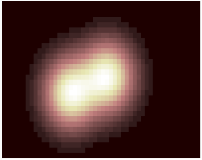
Stochastic and sparse activation of fluorophores



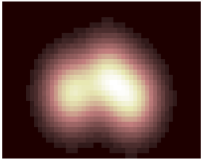
$$a(t_1, t_2) = C \exp\left(-\frac{t_1^2 + t_2^2}{2\sigma^2}\right)$$



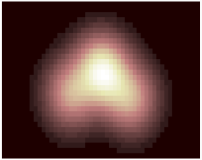
- overlapping PSFs due to high density



2 overlapping PSFs

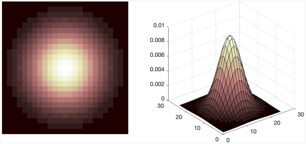


3 overlapping PSFs

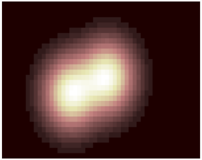


4 overlapping PSFs

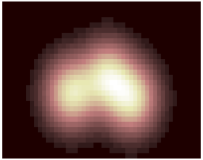
$$a(t_1, t_2) = C \exp\left(-\frac{t_1^2 + t_2^2}{2\sigma^2}\right)$$



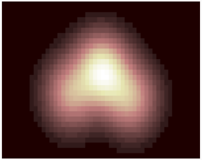
- overlapping PSFs due to high density



2 overlapping PSFs

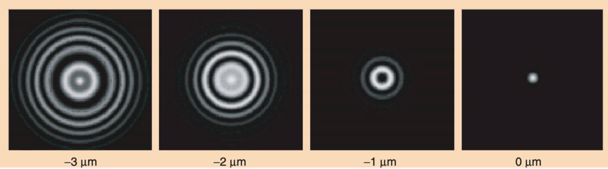


3 overlapping PSFs

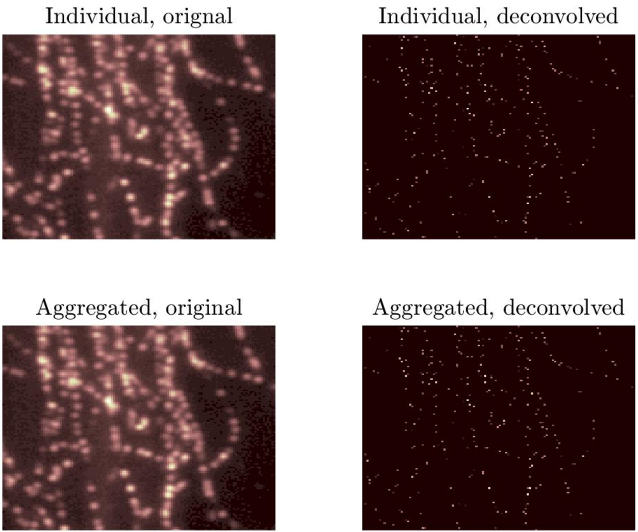


4 overlapping PSFs

- PSF distortion/aberration due to defocus

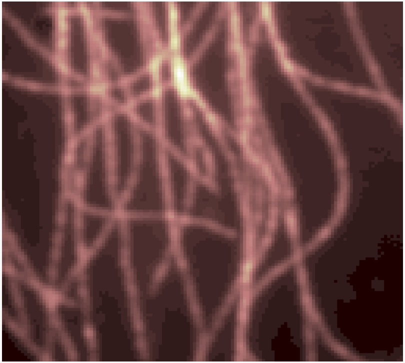


Application II: Super-resolution Microscopy

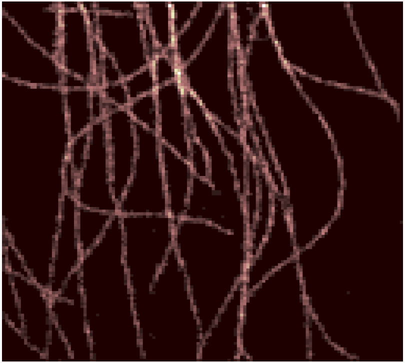


Application II: Super-resolution Microscopy

orignal image



deconvolved image



Extensions: Convolutional Dictionary Learning

- **Task:** recover multiple **short** kernels $\{\boldsymbol{a}_{0k}\}_{k=1}^{N}$ and **sparse** signals $\{\boldsymbol{x}_{0k}\}_{k=1}^{N}$ from

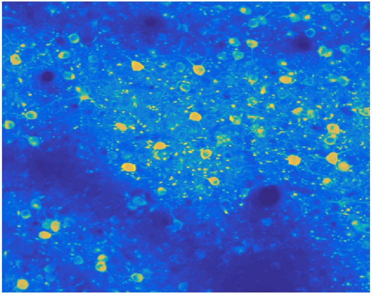
$$\boldsymbol{y} = \sum_{k=1}^{N} \boldsymbol{a}_{0k} \circledast \boldsymbol{x}_{0k}.$$

- **Formulation:** optimization over product of spheres

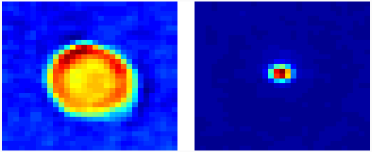
$$\min_{\boldsymbol{a}_k, \boldsymbol{x}_k} \frac{1}{2} \left\| \boldsymbol{y} - \sum_{k=1}^{N} \boldsymbol{a}_k \circledast \boldsymbol{x}_k \right\|^2 + \lambda \sum_{k=1}^{N} \|\boldsymbol{x}_k\|_1, \quad \text{s.t. } \|\boldsymbol{a}_k\| = 1.$$

Application I: Dendrites Classification in Calcium Imaging

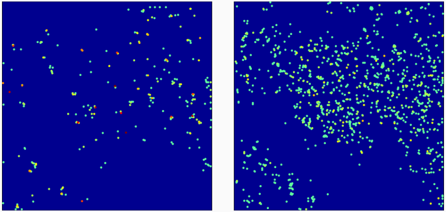
two-photon calcium image $\boldsymbol{Y}$



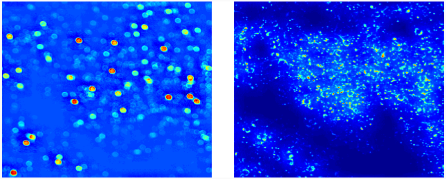
kernel $\boldsymbol{A}_k$ $(k = 1, 2)$



activation map $\boldsymbol{X}_k$ $(k = 1, 2)$

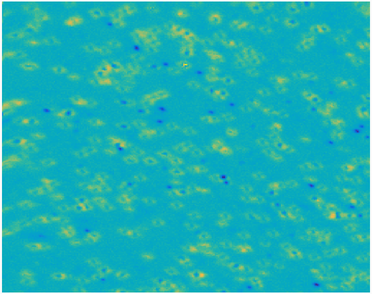


reconstruction $\boldsymbol{Y}_k = \boldsymbol{A}_k \circledast \boldsymbol{X}_k$ $(k = 1, 2)$

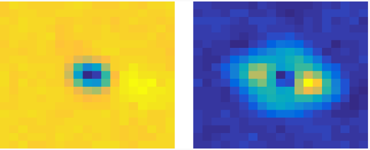


Application II: Defects Detection in Crystal Lattice

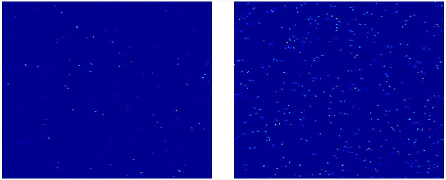
STM image $\boldsymbol{Y}$



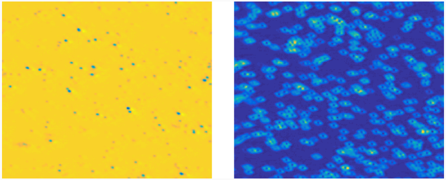
kernel $\boldsymbol{A}_k$ $(k = 1, 2)$



activation map $\boldsymbol{X}_k$ $(k = 1, 2)$



reconstruction $\boldsymbol{Y}_k = \boldsymbol{A}_k \circledast \boldsymbol{X}_k$ $(k = 1, 2)$



Conclusion

- Combining geometric intuition with practical heuristics to design practical nonconvex algorithms;
- Broad applications in computational imaging, computational neuroscience;
- More ideas are needed to study the landscape of $\varphi_{\mathrm{BL}}(\boldsymbol{a})$.

https://deconvlab.github.io/

https://github.com/qingqu06/sparse_deconvolution

THANK YOU!

…AND NYU MSDSE

MOORE - SLOAN
DATA SCIENCE ENVIRONMENTS