Nonconvex Optimization for Multichannel Sparse Blind Deconvolution

Qing Qu

Center for Data Science

New York University

Joint with Xiao Li (CUHK) and Zhihui Zhu (JHU)

November 27, 2019
Multichannel Blind Deconvolution

Given multiple measurement \( y_i \) of circulant convolution

\[
y_i = a \ast x_i, \quad (1 \leq i \leq p),
\]

can we recover both \( a \) and \( \{x_i\}_{i=1}^{p} \) simultaneously?

- We assume \( y_i, a, x_i \in \mathbb{R}^n \).
- **Invertible** kernel \( a \).
- **Sparse** signal \( x_i \)

\[
x_i \sim_{i.i.d.} \text{Bernoulli} - \text{Gaussian}(\theta)
\]
Motivation: Super-resolution Microscopy

Conventional fluorescent optical microscopy

Stochastic Optical Reconstruction Microscopy\(^1\) (STORM)

1. Image courtesy of Xiaowei Zhuang
Symmetry Leads to Nonconvex Problems

♦ Scaling is easy to handle, e.g., $\|a\| = 1$;

♦ Shift symmetry creates equivalent solutions:

$$\left( a, \{x_i \}_{i=1}^p \right) = \left( s_{\ell} [a], \{s_{-\ell} [x_i] \}_{i=1}^p \right)$$
Nonconvex Formulation

\[
\min_{\mathbf{q}} \frac{1}{n_p} \sum_{i=1}^{p} H_\mu (\mathbf{C}_{y_i} \mathbf{P} \mathbf{q} ), \quad \text{s.t.} \quad \mathbf{q} \in S^{n-1}.
\]

♦ Preconditioning matrix

\[
\mathbf{P} = \left( \frac{1}{\theta n_p} \sum_{i=1}^{p} \mathbf{C}_{y_i}^\top \mathbf{C}_{y_i} \right)^{-1/2} \approx \left( \mathbf{C}_a^\top \mathbf{C}_a \right)^{-1/2},
\]

♦ Orthogonalize the kernel \( \mathbf{C}_a \)

\[
\mathbf{C}_{y_i} \mathbf{P} = \mathbf{C}_{x_i} \underbrace{\mathbf{C}_a \mathbf{P}}_{\text{R}} \approx \mathbf{C}_{x_i} \mathbf{C}_a \left( \mathbf{C}_a^\top \mathbf{C}_a \right)^{-1/2}.
\]
Benign Symmetry - I

Study optimization landscape for within each symmetric set

$$S_{\xi}^{i \pm} := \left\{ q \in S^{n-1} \mid \frac{|q_i|}{\|q_{-i}\|_{\infty}} \geq \sqrt{1 + \xi}, \ q_i \geq 0 \right\}, \quad \xi \in (0, +\infty),$$

Minimizer

\[ \xi = 0 \]

\[ \xi = \frac{1}{\sqrt{n \log(n)}} \]

\[ \xi \]

\[ -e_2 \]

\[ -e_3 \]

\[ e_1 \]

\[ e_2 \]

\[ e_3 \]
When \( p \geq \Omega({\text{poly}}(n)) \), for each set \( S_{\xi}^{i+} \) with \( i \in [n] \),

\[ \langle \nabla f(q), q_i q - e_i \rangle \geq \alpha(q) \cdot \| q - e_i \|, \]

for all \( q \in S_{\xi}^{i+} \cap \{ q \in \mathbb{S}^{n-1} | \sqrt{1 - q_i^2} \geq \mu \} \).

\[ \langle \nabla f(q), \frac{1}{q_j} e_j - \frac{1}{q_i} e_i \rangle \geq c \frac{\theta(1 - \theta)}{n} \frac{\xi}{1 + \xi}, \]

for all \( q \in S_{\xi}^{i+} \) and any \( q_j \) such that \( j \neq i \) and \( q_j^2 \geq \frac{1}{3} q_i^2 \).
Random initialization $q^{(0)} \in S^i_\xi$ with $P \geq 1/2$;

Phase I: Riemannian gradient descent (RGD)

$$q^{(k+1)} = P_{S^{n-1}} \left( q^{(k)} - \tau \cdot \text{grad} f(q^{(k)}) \right),$$

with constant $\tau$, stays in $S^i_\xi$, and produces a solution $q_*$ with

$$\left\| q_* - q_{tgt} \right\| \leq O(\mu)$$

in a linear rate, thanks to regularity condition.

Phase II: Solve LP rounding with $r = q_*$,

$$\min_q \zeta(q) := \frac{1}{np} \sum_{i=1}^p \left\| C_{y_i}Pq \right\|_1 \quad \text{s.t. } \langle r, q \rangle = 1$$

via projected subgradient descent with linear convergence.
## Comparison with Literature

<table>
<thead>
<tr>
<th>Methods</th>
<th>Wang et al.(^2)</th>
<th>Li et al.(^3)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumptions</td>
<td>(a) spiky &amp; invertible, (x_i \sim \text{i.i.d. } BG(\theta))</td>
<td>(a) invertible, (x_i \sim \text{i.i.d. } BR(\theta))</td>
<td>(a) invertible, (x_i \sim \text{i.i.d. } BG(\theta))</td>
</tr>
<tr>
<td>Formulation</td>
<td>(\min_{|q|_\infty = 1} |C_qY|_1)</td>
<td>(\max_{q \in S^{n-1}} |C_qPY|_4^4)</td>
<td>(\min_{q \in S^{n-1}} H_{\mu}(C_qPY))</td>
</tr>
<tr>
<td>Algorithm</td>
<td>interior point</td>
<td>noisy RGD</td>
<td>vanilla RGD</td>
</tr>
<tr>
<td>Recovery Condition</td>
<td>(\theta \in O(1/\sqrt{n}), \quad p \geq \tilde{\Omega}(n))</td>
<td>(\theta \in O(1), \quad p \geq \tilde{\Omega}(\max{n, \kappa^8}^n\sqrt{\varepsilon}))</td>
<td>(\theta \in O(1), \quad p \geq \tilde{\Omega}(\max{n, \kappa^8/\mu^2}^n4))</td>
</tr>
<tr>
<td>Time Complexity</td>
<td>(\tilde{O}(p^4n^5 \log(1/\varepsilon)))</td>
<td>(\tilde{O}(pn^{13}/\varepsilon^8))</td>
<td>(\tilde{O}(pn^5 + pn \log(1/\varepsilon)))</td>
</tr>
</tbody>
</table>

Experiment I: Convergence Comparison

Phase 1: RGD
Phase 2: LP Rounding

$\log_{10}(\min \{ \| a_x - a \|, \| a_x + a \| \})$

- $\ell^1$-loss
- Huber-loss, $\mu = 5 \times 10^{-1}$
- Huber-loss, $\mu = 5 \times 10^{-2}$
- Huber-loss, $\mu = 5 \times 10^{-3}$
- $\ell^4$-loss
Experiment II: Super-resolution Microscopy

Observation | Ground truth | Huber-loss | $\ell^4$-loss

Ground truth | Huber-loss | $\ell^4$-loss
Take home message

With random init., gradient descent solves sparse blind deconvolution in a linear rate.

Q. Qu, X. Li, Z. Zhu, A nonconvex approach for exact and efficient multichannel sparse blind deconvolution, NeurIPS’19 (spotlight)
THANK YOU!

...AND