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)

August 7, 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 $y_i, a, x_i \in \mathbb{R}^n$.

♦ We write

\[
Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \end{bmatrix} \in \mathbb{R}^{n \times p}
\]
\[
X = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix} \in \mathbb{R}^{n \times p}
\]
Motivation I: Super-resolution Microscopy

Conventional fluorescent optical microscopy

Stochastic Optical Reconstruction Microscopy (STORM)

1. Image courtesy of Xiaowei Zhuang
Stochastic and sparse activation of fluorophores
Application II: Super-resolution Microscopy

Individual, original

Individual, deconvolved

Aggregated, original

Aggregated, deconvolved
Motivation II: Geophysics

Sparse 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 \textbf{sparse} $\{x_i\}_{i=1}^p$ simultaneously?

\begin{itemize}
    \item \textbf{Sparse} signal $x_i$
    \[
        x_i \sim_{i.i.d.} \text{Bernoulli - Gaussian}(\theta)
    \]
    \item We write
    \[
    Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \end{bmatrix} \in \mathbb{R}^{n \times p}
    \]
    \[
    X = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix} \in \mathbb{R}^{n \times p}
    \]
\end{itemize}
Symmetry Leads to Nonconvex Problems

♦ Shift Symmetry: \( y_i = a \odot x_i = s_\ell[a] \odot s_{-\ell}[x_i] \)

♦ Scaling Symmetry: \( y_i = a \ast x_i = \alpha a \ast \alpha^{-1} x_i \)
Symmetry Leads to Nonconvex Problems

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

♦ Shift symmetry creates equivalent solutions:

$$(a, \{x_i\}_{i=1}^p) = (s_\ell [a], \{s_{-\ell} [x_i]\}_{i=1}^p)$$
Strict saddle: benign optimization landscape\textsuperscript{2}

Noisy gradient with random init.\textsuperscript{3} solves sparse deconvolution\textsuperscript{4} (Li et al., NeurIPS’18)

This work: gradient descent + random init. $\rightarrow$ linear convergence to target solutions

2. Sun et al, When are nonconvex problems not scary? 2016
## Comparison with Literature

<table>
<thead>
<tr>
<th>Methods</th>
<th>Wang et al.(^5)</th>
<th>Li et al.(^6)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumptions</td>
<td>(a) spiky &amp; invertible, (x_i \sim_{i.i.d.} BG(\theta))</td>
<td>(a) invertible, (x_i \sim_{i.i.d.} BR(\theta))</td>
<td>(a) invertible, (x_i \sim_{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 \mathcal{O}(1/\sqrt{n}), \quad p \geq \tilde{\Omega}(n))</td>
<td>(\theta \in \mathcal{O}(1), \quad p \geq \tilde{\Omega}(\max {n, \kappa^8} \frac{n^8}{\varepsilon^8}))</td>
<td>(\theta \in \mathcal{O}(1), \quad p \geq \tilde{\Omega}(\max {n, \frac{\kappa^8}{\mu^2}} n^4))</td>
</tr>
<tr>
<td>Time Complexity</td>
<td>(\tilde{\mathcal{O}}(p^4n^5 \log(1/\varepsilon)))</td>
<td>(\tilde{\mathcal{O}}(pn^{13}/\varepsilon^8))</td>
<td>(\tilde{\mathcal{O}}(pn^5 + pn \log(1/\varepsilon)))</td>
</tr>
</tbody>
</table>

Outline

Nonconvex Problem Formulation

Geometric Analysis of Optimization Landscape

From Geometry to Efficient Optimization

Experimental Results
Assumptions

Task: recover both $a$ and $\{x_i\}_{i=1}^p$ from multiple

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

♦ Sparse signal $x_i$

\[x_i \sim_{i.i.d.} \text{Bernoulli} - \text{Gaussian}(\theta)\]

♦ Invertible kernel $a$

\[
C_a = F^* \text{ diag } (\hat{a}) F, \quad \text{or} \quad |\hat{a}| > 0.
\] (invertible)

Here, $\hat{a} = Fa$ and $F$ is unnormalized DFT matrix.
Problem Formulation

Let $h$ be the **inverse kernel** of $a$, $\hat{h} = a^{\circ -1}$ or $a \odot h = 1$,

\[
C_h \cdot Y = C_h \cdot C_a \cdot X = I \quad \text{sparse}
\]

Solve the problem

\[
\min_q \frac{1}{np} \left\| C_q Y \right\|_0 \quad \text{s.t.} \quad \left\{ \begin{array}{l}
q \neq 0 \\
\text{prevent trivial solution}
\end{array} \right.
\]

\[
\text{to recover } \hat{a} = s_{\ell} \left[ \alpha \hat{q}_{\star}^{\circ -1} \right] \text{ up to a shift-scaling symmetry.}
\]
Nonconvex Relaxation

♦ Original Problem:

$$\min_q \frac{1}{np} \sum_{i=1}^{p} \|C_{yi}q\|_0, \quad \text{s.t. } q \neq 0.$$ 

♦ Relaxed Problem:

$$\min_q \frac{1}{np} \sum_{i=1}^{p} \underbrace{H_\mu(C_{yi}Pq)}_{\text{smooth sparsity function}}, \quad \text{s.t. } q \in \mathbb{S}^{n-1}.$$ 

- $H_\mu(\cdot)$ is smooth Huber loss for promoting sparsity.
- $P$ is a preconditioning matrix.
Huber loss vs $\ell^1$-loss

$$\min_q \frac{1}{np} \sum_{i=1}^p H_\mu (C y_i P q), \quad \text{s.t. } q \in S^{n-1}.$$
Preconditioning I - Condition of $C_a$ Matters

\[
\min_q \frac{1}{np} \sum_{i=1}^{p} H_{\mu} \left( C_{y_i} P q \right), \quad \text{s.t.} \quad q \in S^{n-1}.
\]

♦ Preconditioning matrix

\[
P = \left( \frac{1}{\theta np} \sum_{i=1}^{p} C_{y_i}^T C_{y_i} \right)^{-1/2} \approx \left( C_a^T C_a \right)^{-1/2},
\]

♦ Orthogonalize the kernel $C_a$

\[
C_{y_i} P = C_{x_i} \underbrace{C_a P}_{R} \approx C_{x_i} C_a \left( C_a^T C_a \right)^{-1/2} \quad \text{orthogonal} \ Q
\]
Preconditioning II - Geometric Illustration

\( \ell^1 \)-loss, \( \times \)  
\( \ell^4 \)-loss, \( \times \)  
\( \ell^1 \)-loss, \( \checkmark \)  
Huber-loss, \( \checkmark \)  
Huber-loss, \( \checkmark \)  
\( \ell^4 \)-loss, \( \checkmark \)
Preconditioning III: Problem Reduction

Given

\[ C_{y} P q \approx C_{x} Q q, \]

suppose \( Q = I \), our problem reduces to

\[
\min_{q} f(q) = \frac{1}{np} \sum_{i=1}^{p} H_{\mu}(C_{x_i} q), \quad q \in \mathbb{S}^{n-1}.
\]

This implies that standard basis \( \{ \pm e_i \}_{i=1}^{n} \) are global solutions.
Outline

Nonconvex Problem Formulation

Geometric Analysis of Optimization Landscape

From Geometry to Efficient Optimization

Experimental Results
Symmetric Sets Excluding Saddle Points

Study optimization landscape for union of sets

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

For each set,

- It contains exactly one solution \( \pm e_i \);
- It excludes all saddle points;
- For some small \( \xi = \frac{1}{5 \log n} \), random initialization falls in \( S_{\xi}^{i\pm} \) with Prob. \( \geq 1/2 \).

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

- Light blue region denotes \( S_{\xi}^{i\pm} \) with \( \xi = 0 \).
- Red dots create \( S_{\xi}^{i\pm} \) with \( \xi > 0 \).
- Yellow region of radius \( \mu \) contains target solution.

– Image courtesy of Dar Gilboa.
Theorem

When \( p \geq \Omega \left( \text{poly}(n) \right) \), for each \( S^i_\xi \), w.h.p. we have

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

for all

\[
q \in S^i_\xi \cap \left\{ q \in \mathbb{S}^{n-1} \mid \sqrt{1 - q_i^2} \geq \mu \right\}.
\]

♦ Large \( \nabla f(q) \) even when \( q \to e_i \);
♦ Linear convergence of gradient descent.
Theorem

When \( p \geq \Omega(\text{poly}(n)) \), for each \( S_{\xi}^{i^+} \), w.h.p. we have

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

for all \( q \in S_{\xi}^{i^+} \) and any \( q_j \) such that \( j \neq i \) and \( q_j^2 \geq \frac{1}{3} q_i^2 \).

♦ Make sure GD iterates stay within \( S_{\xi}^{i^+} \);
Outline

Nonconvex Problem Formulation

Geometric Analysis of Optimization Landscape

From Geometry to Efficient Optimization

Experimental Results
Algorithmic Pipeline I - vanilla RGD

♦ Random initialization $q^{(0)} \in S_{\xi}^{i\pm}$ 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 $S_{\xi}^{i\pm}$, 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 \( r = q_\star \),

\[
\min_q \zeta(q) := \frac{1}{np} \sum_{i=1}^{p} \| C_{yi} P q \|_1, \quad \text{s.t. } \langle r, q \rangle = 1
\]

via projected subgradient descent

\[
q^{(k+1)} = q^{(k)} - \tau^{(k)} \cdot P_{r^\perp} g^{(k)},
\]

with \( \tau^{(k+1)} = \beta \tau^{(k)} \), it converges linearly

\[
\| q^{(k)} - q_{tgt} \| \leq \eta^k, \quad \eta \in (0, 1),
\]

thanks to local sharpness of \( \zeta(q) \).
### Comparison with Literature

<table>
<thead>
<tr>
<th>Methods</th>
<th>Wang et al.(^8)</th>
<th>Li et al.(^9)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumptions</td>
<td>( a \text{ spiky &amp; invertible, } x_i \sim_{\text{i.i.d.}} BG(\theta) )</td>
<td>( a \text{ invertible, } x_i \sim_{\text{i.i.d.}} BR(\theta) )</td>
<td>( a \text{ invertible, } x_i \sim_{\text{i.i.d.}} BG(\theta) )</td>
</tr>
<tr>
<td>Formulation</td>
<td>( \min_{|q|_\infty=1} |C_q Y|_1 )</td>
<td>( \max_{q \in S^{n-1}} |C_q PY|_4^4 )</td>
<td>( \min_{q \in S^{n-1}} H_\mu(C_q PY) )</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 \mathcal{O}(1/\sqrt{n}), p \geq \tilde{\Omega}(n) )</td>
<td>( \theta \in \mathcal{O}(1), p \geq \tilde{\Omega}(\max{n, \kappa^8} n^8) )</td>
<td>( \theta \in \mathcal{O}(1), p \geq \tilde{\Omega}(\max{n, \kappa^8 \mu^2} n^4) )</td>
</tr>
<tr>
<td>Time Complexity</td>
<td>( \tilde{\mathcal{O}}(p^4n^5 \log(1/\varepsilon)) )</td>
<td>( \tilde{\mathcal{O}}(pn^{13}/\varepsilon^8) )</td>
<td>( \tilde{\mathcal{O}}(pn^5 + pn \log(1/\varepsilon)) )</td>
</tr>
</tbody>
</table>

Outline

Nonconvex Problem Formulation

Geometric Analysis of Optimization Landscape

From Geometry to Efficient Optimization

Experimental Results
Experiment I: Convergence Comparison

Phase 1: RGD
Phase 2: LP Rounding

- $\ell^1$-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

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

Iteration Number

Values range from $-25$ to $0$ on the y-axis.
Experiment II: Phase Transition

\( \ell^1 \)-loss  \hspace{2cm} \text{Huber loss}  \hspace{2cm} \ell^4 \)-loss

We conjecture \( p \geq \Omega(\text{poly log } n) \) is sufficient.
Experiment III: Super-resolution Microscopy

Observation

Ground truth

Huber-loss

$\ell^4$-loss

Ground truth

Huber-loss

$\ell^4$-loss
Conclusion

♦ Vanilla RGD with random init. solves nonconvex problems to global optimizer efficiently;

♦ Extension to other problems, e.g., convolutional dictionary learning, blind gain calibration;

♦ Smooth vs nonsmooth optimization;

♦ Improve sample complexity from $\text{poly}(n)$ to $\text{poly log}(n)$. 
Conclusion

♦ Vanilla RGD with random init. solves nonconvex problems to global optimizer efficiently;

♦ Extension to other problems, e.g., convolutional dictionary learning, blind gain calibration;

♦ Smooth vs nonsmooth optimization;

♦ Improve sample complexity from $\text{poly}(n)$ to $\text{poly log}(n)$.

More Stories on Sparse Deconvolution, Room H3007, Thursday, 1:30pm.
THANK YOU!

...AND