Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization

Shuyang Ling

Department of Mathematics, UC Davis

Oct.18th, 2016

Acknowledgements

Research in collaboration with:

- Prof.Xiaodong Li (UC Davis)
- Prof.Thomas Strohmer (UC Davis)
- Dr.Ke Wei (UC Davis)

This work is sponsored by NSF-DMS and DARPA.

Outline

- Applications in image deblurring and wireless communication
- Mathematical models and convex approach
- A nonconvex optimization approach towards blind deconvolution

What is blind deconvolution?

What is blind deconvolution?

Suppose we observe a function $\boldsymbol{y}$ which consists of the convolution of two unknown functions, the blurring function $\boldsymbol{f}$ and the signal of interest $\boldsymbol{g}$, plus noise $\boldsymbol{w}$. How to reconstruct $\boldsymbol{f}$ and $\boldsymbol{g}$ from $\boldsymbol{y}$?

$$\boldsymbol{y} = \boldsymbol{f} * \boldsymbol{g} + \boldsymbol{w}.$$

It is obviously a highly ill-posed bilinear inverse problem...

- Much more difficult than ordinary deconvolution...but has important applications in various fields.
- Solvability? What conditions on $\boldsymbol{f}$ and $\boldsymbol{g}$ make this problem solvable?
- How? What algorithms shall we use to recover $\boldsymbol{f}$ and $\boldsymbol{g}$?

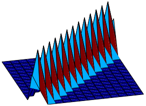
Why do we care about blind deconvolution?

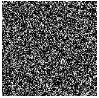
Image deblurring

Let $\boldsymbol{f}$ be the blurring kernel and $\boldsymbol{g}$ be the original image, then $\boldsymbol{y} = \boldsymbol{f} * \boldsymbol{g}$ is the blurred image.

Question: how to reconstruct $\boldsymbol{f}$ and $\boldsymbol{g}$ from $\boldsymbol{y}$?

$\boldsymbol{y}$ blurred image $=$ $\boldsymbol{f}$ blurring kernel $*$ $\boldsymbol{g}$ original image $+$ $\boldsymbol{w}$ noise


=

*

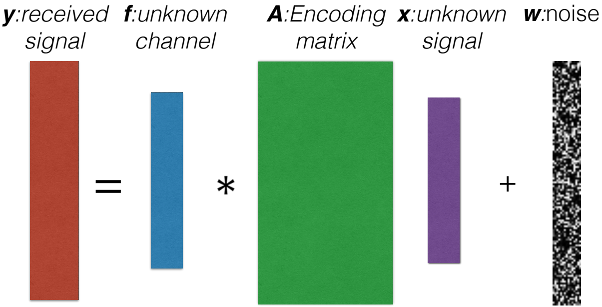
+


Why do we care about blind deconvolution?

Joint channel and signal estimation in wireless communication

Suppose that a signal $\boldsymbol{x}$, encoded by $\boldsymbol{A}$, is transmitted through an unknown channel $\boldsymbol{f}$. How to reconstruct $\boldsymbol{f}$ and $\boldsymbol{x}$ from $\boldsymbol{y}$?

$$\boldsymbol{y} = \boldsymbol{f} * \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}.$$



Subspace assumptions

We start from the original model

$$\boldsymbol{y} = \boldsymbol{f} * \boldsymbol{g} + \boldsymbol{w}.$$

As mentioned before, it is an ill-posed problem. Hence, this problem is unsolvable without further assumptions...

Subspace assumption

Both $\boldsymbol{f}$ and $\boldsymbol{g}$ belong to known subspaces: there exist known tall matrices $\boldsymbol{B} \in \mathbb{C}^{L \times K}$ and $\boldsymbol{A} \in \mathbb{C}^{L \times N}$ such that

$$\boldsymbol{f} = \boldsymbol{B}\boldsymbol{h}_0, \quad \boldsymbol{g} = \boldsymbol{A}\boldsymbol{x}_0,$$

for some unknown vectors $\boldsymbol{h}_0 \in \mathbb{C}^K$ and $\boldsymbol{x}_0 \in \mathbb{C}^N$.

Model under subspace assumption

In the frequency domain,

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{f}} \odot \hat{\boldsymbol{g}} + \boldsymbol{w} = \operatorname{diag}(\hat{\boldsymbol{f}})\hat{\boldsymbol{g}} + \boldsymbol{w},$$

where " $\odot$ " denotes entry-wise multiplication. We assume $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ are both of length L.

Subspace assumption

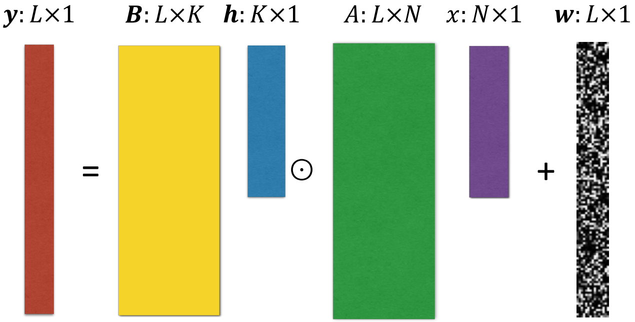
Both $\hat{\boldsymbol{f}}$ and $\hat{\boldsymbol{g}}$ belong to known subspaces: there exist known tall matrices $\hat{\boldsymbol{B}} \in \mathbb{C}^{L\times K}$ and $\hat{\boldsymbol{A}} \in \mathbb{C}^{L\times N}$ such that

$$\hat{\boldsymbol{f}} = \hat{\boldsymbol{B}}\boldsymbol{h}_0, \quad \hat{\boldsymbol{g}} = \hat{\boldsymbol{A}}\boldsymbol{x}_0,$$

for some unknown vectors $\boldsymbol{h}_0 \in \mathbb{C}^K$ and $\boldsymbol{x}_0 \in \mathbb{C}^N$. Here $\hat{\boldsymbol{B}} = \boldsymbol{F}\boldsymbol{B}$ and $\hat{\boldsymbol{A}} = \boldsymbol{F}\boldsymbol{A}$.

The degree of freedom for unknowns: $K + N$; number of constraint: L. To make the solution identifiable, we require $L \geq K + N$ at least.

Remarks on subspace assumption



Subspace assumption is flexible and useful in applications.

- In imaging deblurring, $\boldsymbol{B}$ can be the support of the blurring kernel; $\boldsymbol{A}$ is a wavelet basis.
- In wireless communication, $\boldsymbol{B}$ corresponds to time-limitation of the channel and $\boldsymbol{A}$ is an encoding matrix.

Mathematical model

$$\boldsymbol{y} = \mathrm{diag}(\boldsymbol{B}\boldsymbol{h}_0)\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w},$$

where $\frac{\boldsymbol{w}}{d_0} \sim \frac{1}{\sqrt{2}}\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_L) + \frac{\mathrm{i}}{\sqrt{2}}\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_L)$ and $d_0 = \|\boldsymbol{h}_0\|\|\boldsymbol{x}_0\|$.

Mathematical model

$$\boldsymbol{y} = \operatorname{diag}(\boldsymbol{B}\boldsymbol{h}_0)\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w},$$

where $\frac{\boldsymbol{w}}{d_0} \sim \frac{1}{\sqrt{2}}\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_L) + \frac{\mathrm{i}}{\sqrt{2}}\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_L)$ and $d_0 = \|\boldsymbol{h}_0\|\|\boldsymbol{x}_0\|$.
One might want to solve the following nonlinear least squares problem,

$$\min F(\boldsymbol{h}, \boldsymbol{x}) := \|\operatorname{diag}(\boldsymbol{B}\boldsymbol{h})\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2.$$

Difficulties:

1. **Nonconvexity:** F is a nonconvex function; algorithms (such as gradient descent) are likely to get trapped at local minima.
2. **No performance guarantees.**

Convex approach and lifting

Two-step convex approach

(a) Lifting: convert bilinear to linear constraints

(b) Solving a SDP relaxation to recover $\boldsymbol{hx}^*$.

Convex approach and lifting

Two-step convex approach

(a) Lifting: convert bilinear to linear constraints

(b) Solving a SDP relaxation to recover $\boldsymbol{hx}^*$.

Step 1: lifting

Let $\boldsymbol{a}_i$ be the i-th column of $\boldsymbol{A}^*$ and $\boldsymbol{b}_i$ be the i-th column of $\boldsymbol{B}^*$.

$$y_i = (\boldsymbol{B}\boldsymbol{h}_0)_i \boldsymbol{x}_0^* \boldsymbol{a}_i + w_i = \boldsymbol{b}_i^* \boldsymbol{h}_0 \boldsymbol{x}_0^* \boldsymbol{a}_i + w_i,$$

Let $\boldsymbol{X}_0 := \boldsymbol{h}_0 \boldsymbol{x}_0^*$ and define the linear operator $\mathcal{A} : \mathbb{C}^{K \times N} \to \mathbb{C}^L$ as,

$$\mathcal{A}(\boldsymbol{Z}) := \{\boldsymbol{b}_i^* \boldsymbol{Z} \boldsymbol{a}_i\}_{i=1}^L = \{\langle \boldsymbol{Z}, \boldsymbol{b}_i \boldsymbol{a}_i^* \rangle\}_{i=1}^L.$$

Then, there holds

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{w}.$$

In this way, $\mathcal{A}^*(\boldsymbol{z}) = \sum_{i=1}^L z_i \boldsymbol{b}_i \boldsymbol{a}_i^* : \mathbb{C}^L \to \mathbb{C}^{K \times N}$.

Convex relaxation and state of the art

Step 2: nuclear norm minimization

Consider the convex envelop of rank($\boldsymbol{Z}$): nuclear norm $\|\boldsymbol{Z}\|_* = \sum \sigma_i(\boldsymbol{Z})$.

$$\min \|\boldsymbol{Z}\|_* \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{Z}) = \mathcal{A}(\boldsymbol{X}_0).$$

Convex optimization can be solved within polynomial time.

Convex relaxation and state of the art

Step 2: nuclear norm minimization

Consider the convex envelop of rank($\boldsymbol{Z}$): nuclear norm $\|\boldsymbol{Z}\|_* = \sum \sigma_i(\boldsymbol{Z})$.

$$\min \|\boldsymbol{Z}\|_* \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{Z}) = \mathcal{A}(\boldsymbol{X}_0).$$

Convex optimization can be solved within polynomial time.

Theorem [Ahmed-Recht-Romberg 11]

Assume $\boldsymbol{y} = \operatorname{diag}(\boldsymbol{B}\boldsymbol{h}_0)\boldsymbol{A}\boldsymbol{x}_0$, $\boldsymbol{A} : L \times N$ is a complex Gaussian random matrix,

$$\boldsymbol{B}^*\boldsymbol{B} = \boldsymbol{I}_K, \quad \|\boldsymbol{b}_i\|^2 \leq \frac{\mu_{\max}^2 K}{L}, \quad L\|\boldsymbol{B}\boldsymbol{h}_0\|_\infty^2 \leq \mu_h^2,$$

the above convex relaxation recovers $\boldsymbol{X} = \boldsymbol{h}_0\boldsymbol{x}_0^*$ exactly with high probability if

$$C_0 \max(\mu_{\max}^2 K, \mu_h^2 N) \leq \frac{L}{\log^3 L}.$$

Pros and Cons of Convex Approach

Pros and Cons

- Pros: Simple and comes with theoretic guarantees
- Cons: Computationally too expensive to solve SDP

Our Goal: **rapid, robust, reliable nonconvex approach**

- Rapid: linear convergence
- Robust: stable to noise
- Reliable: provable and comes with theoretical guarantees; number of measurements close to information-theoretic limits.

A nonconvex optimization approach?

An increasing list of nonconvex approach to various problems:

- Phase retrieval: by Candés, Li, Soltanolkotabi, Chen, Wright, etc...
- Matrix completion: by Sun, Luo, Montanari, etc...
- Various problems: by Recht, Wainwright, Constantine, etc...

Two-step philosophy for provable nonconvex optimization

(a) Use spectral initialization to construct a starting point inside *"the* ***basin of attraction***";

(b) Simple gradient descent method.

The key is to build up "the basin of attraction".

Building *"the basin of attraction"*

The basin of attraction relies on the following three observations.

Observation 1: Unboundedness of solution

- If the pair $(\boldsymbol{h}_0, \boldsymbol{x}_0)$ is a solution to $\boldsymbol{y} = \operatorname{diag}(\boldsymbol{B}\boldsymbol{h}_0)\boldsymbol{A}\boldsymbol{x}_0$, then so is the pair $(\alpha\boldsymbol{h}_0, \alpha^{-1}\boldsymbol{x}_0)$ for any $\alpha \neq 0$.
- Thus the blind deconvolution problem always has infinitely many solutions of this type. We can recover $(\boldsymbol{h}_0, \boldsymbol{x}_0)$ only up to a scalar.
- It is possible that $\|\boldsymbol{h}\| \gg \|\boldsymbol{x}\|$ (vice versa) while $\|\boldsymbol{h}\| \cdot \|\boldsymbol{x}\| = d_0$. Hence we define $\mathcal{N}_{d_0}$ to balance $\|\boldsymbol{h}\|$ and $\|\boldsymbol{x}\|$:

$$\mathcal{N}_{d_0} := \{(\boldsymbol{h}, \boldsymbol{x}) : \|\boldsymbol{h}\| \leq 2\sqrt{d_0}, \|\boldsymbol{x}\| \leq 2\sqrt{d_0}\}.$$

Building *"the basin of attraction"*

Observation 2: Incoherence

Our numerical experiments have shown that the algorithm's performance depends on how much $\boldsymbol{b}_l$ and $\boldsymbol{h}_0$ are correlated.

$$\mu_h^2 := \frac{L\|\boldsymbol{B}\boldsymbol{h}_0\|_\infty^2}{\|\boldsymbol{h}_0\|^2} = L\frac{\max_i |\boldsymbol{b}_i^*\boldsymbol{h}_0|^2}{\|\boldsymbol{h}_0\|^2}, \quad \textbf{the smaller } \mu_h\textbf{, the better.}$$

Therefore, we introduce the $\mathcal{N}_\mu$ to control the incoherence:

$$\mathcal{N}_\mu := \{\boldsymbol{h} : \sqrt{L}\|\boldsymbol{B}\boldsymbol{h}\|_\infty \leq 4\mu\sqrt{d_0}\}.$$

"Incoherence" is not a new idea. In matrix completion, we also require the left and right singular vectors of the ground truth cannot be too "aligned" with those of measurement matrices $\{\boldsymbol{b}_i\boldsymbol{a}_i^*\}_{1\leq i\leq L}$. The same philosophy applies here.

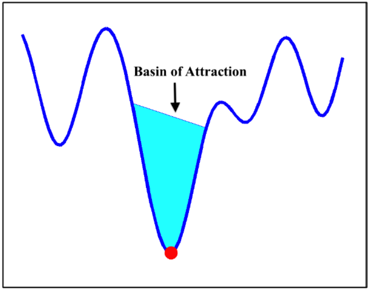
Building *"the basin of attraction"*

Observation 3: "Close" to the ground truth

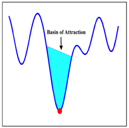
We define $\mathcal{N}_\varepsilon$ to quantify closeness of $(\boldsymbol{h}, \boldsymbol{x})$ to true solution, i.e.,

$$\mathcal{N}_\varepsilon := \{(\boldsymbol{h}, \boldsymbol{x}) : \|\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F \leq \varepsilon d_0\}.$$

We want to find an initial guess close to $(\boldsymbol{h}_0, \boldsymbol{x}_0)$.



Based on the three observations above, we define the three neighborhoods (denoting $d_0 = \|h_0\|\|x_0\|$ and $0 < \varepsilon \leq \frac{1}{15}$):



$$
\begin{aligned}
\mathcal{N}_{d_0} &:= \{(\boldsymbol{h}, \boldsymbol{x}) : \|\boldsymbol{h}\| \leq 2\sqrt{d_0}, \|\boldsymbol{x}\| \leq 2\sqrt{d_0}\} \\
\mathcal{N}_{\mu} &:= \{\boldsymbol{h} : \sqrt{L}\|\boldsymbol{B}\boldsymbol{h}\|_{\infty} \leq 4\mu\sqrt{d_0}\} \\
\mathcal{N}_{\varepsilon} &:= \{(\boldsymbol{h}, \boldsymbol{x}) : \|\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F \leq \varepsilon d_0\}.
\end{aligned}
$$

We first obtain a good initial guess $(\boldsymbol{u}_0, \boldsymbol{v}_0) \in \mathcal{N}_{d_0} \cap \mathcal{N}_{\mu} \cap \mathcal{N}_{\varepsilon}$, which is followed by regularized gradient descent.

Objective function: a variant of projected gradient descent

The objective function $\widetilde{F}$ consists of two parts: F and G:

$$\min_{(\boldsymbol{h},\boldsymbol{x})} \quad \widetilde{F}(\boldsymbol{h},\boldsymbol{x}) := \underbrace{F(\boldsymbol{h},\boldsymbol{x})}_{\text{least squares term}} + \underbrace{G(\boldsymbol{h},\boldsymbol{x})}_{\text{regularization term}}$$

where $F(\boldsymbol{h},\boldsymbol{x}) := \|\mathcal{A}(\boldsymbol{h}\boldsymbol{x}^*) - \boldsymbol{y}\|^2 = \|\operatorname{diag}(\boldsymbol{B}\boldsymbol{h})\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2$ and

$$G(\boldsymbol{h},\boldsymbol{x}) := \rho\Big[\underbrace{G_0\left(\frac{\|\boldsymbol{h}\|^2}{2d}\right) + G_0\left(\frac{\|\boldsymbol{x}\|^2}{2d}\right)}_{\mathcal{N}_{d_0}:\ \text{balance } \|\boldsymbol{h}\| \text{ and } \|\boldsymbol{x}\|} + \underbrace{\sum_{l=1}^{L} G_0\left(\frac{L|\boldsymbol{b}_l^*\boldsymbol{h}|^2}{8d\mu^2}\right)}_{\mathcal{N}_{\mu}:\ \text{impose incoherence}}\Big].$$

Here $G_0(z) = \max\{z-1, 0\}^2$, $\rho \approx d^2$, $d \approx d_0$ and $\mu \geq \mu_h$.
Regularization forces iterates $(\boldsymbol{u}_t, \boldsymbol{v}_t)$ inside $\mathcal{N}_{d_0} \cap \mathcal{N}_{\mu} \cap \mathcal{N}_{\varepsilon}$.

Algorithm: Wirtinger Gradient Descent

Step 1: Initialization via spectral method and projection:

1: Compute $\mathcal{A}^*(\boldsymbol{y})$, (since $\mathbb{E}(\mathcal{A}^*(\boldsymbol{y})) = \boldsymbol{h}_0\boldsymbol{x}_0^*$);

2: Find the leading singular value, left and right singular vectors of $\mathcal{A}^*(\boldsymbol{y})$, denoted by $(d, \hat{\boldsymbol{h}}_0, \hat{\boldsymbol{x}}_0)$ respectively;

3: $\boldsymbol{u}_0 := \mathcal{P}_{\mathcal{N}_\mu}(\sqrt{d}\hat{\boldsymbol{h}}_0)$ and $\boldsymbol{v}_0 := \sqrt{d}\hat{\boldsymbol{x}}_0$;

4: Output: $(\boldsymbol{u}_0, \boldsymbol{v}_0)$.

Step 2: Gradient descent with constant stepsize η:

1: **Initialization:** obtain $(\boldsymbol{u}_0, \boldsymbol{v}_0)$ via Algorithm 1.

2: **for** $t = 1, 2, \dots,$ **do**

3: $\quad \boldsymbol{u}_t = \boldsymbol{u}_{t-1} - \eta\nabla\widetilde{F}_{\boldsymbol{h}}(\boldsymbol{u}_{t-1}, \boldsymbol{v}_{t-1})$

4: $\quad \boldsymbol{v}_t = \boldsymbol{v}_{t-1} - \eta\nabla\widetilde{F}_{\boldsymbol{x}}(\boldsymbol{u}_{t-1}, \boldsymbol{v}_{t-1})$

5: **end for**

Main theorem

Theorem: [Li-Ling-Strohmer-Wei, 2016]

Let $\boldsymbol{B}$ be a tall partial DFT matrix and $\boldsymbol{A}$ be a complex Gaussian random matrix. If the number of measurements satisfies

$$L \geq C(\mu_h^2 + \sigma^2)(K + N)\log^2(L)/\varepsilon^2,$$

(i) then the initialization $(\boldsymbol{u}_0, \boldsymbol{v}_0) \in \frac{1}{\sqrt{3}}\mathcal{N}_{d_0} \bigcap \frac{1}{\sqrt{3}}\mathcal{N}_{\mu} \bigcap \mathcal{N}_{\frac{2}{5}\varepsilon}$;
(ii) the regularized gradient descent algorithm creates a sequence $(\boldsymbol{u}_t, \boldsymbol{v}_t)$ in $\mathcal{N}_{d_0} \cap \mathcal{N}_{\mu} \cap \mathcal{N}_{\varepsilon}$ satisfying

$$\|\boldsymbol{u}_t\boldsymbol{v}_t^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F \leq (1-\alpha)^t \varepsilon d_0 + c_0\|\mathcal{A}^*(\boldsymbol{w})\|$$

with high probability where $\alpha = \mathcal{O}\left(\frac{1}{(1+\sigma^2)(K+N)\log^2 L}\right)$

Remarks

(a) If $\boldsymbol{w} = \boldsymbol{0}$, $(\boldsymbol{u}_t, \boldsymbol{v}_t)$ converges to $(\boldsymbol{h}_0, \boldsymbol{x}_0)$ linearly.

$$\|\boldsymbol{u}_t \boldsymbol{v}_t^* - \boldsymbol{h}_0 \boldsymbol{x}_0^*\|_F \leq (1-\alpha)^t \varepsilon d_0 \to 0, \text{ as } t \to \infty$$

(b) If $\boldsymbol{w} \neq \boldsymbol{0}$, $(\boldsymbol{u}_t, \boldsymbol{v}_t)$ converges to a small neighborhood of $(\boldsymbol{h}_0, \boldsymbol{x}_0)$ linearly.

$$\|\boldsymbol{u}_t \boldsymbol{v}_t^* - \boldsymbol{h}_0 \boldsymbol{x}_0^*\|_F \to c_0 \|\mathcal{A}^*(\boldsymbol{w})\|, \text{ as } t \to \infty$$

where

$$\|\mathcal{A}^*(\boldsymbol{w})\| = \mathcal{O}\left(\sigma d_0 \sqrt{\frac{(K+N)\log L}{L}}\right) \to 0, \text{ if } L \to \infty.$$

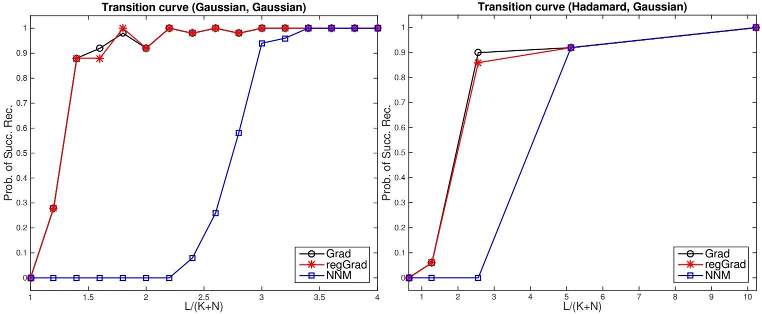
As L is becoming larger and larger, the effect of noise diminishes. (Recall linear least squares.)

Numerical experiments

Nonconvex approach v.s. convex approach:

$$\min_{(\boldsymbol{h},\boldsymbol{x})} \widetilde{F}(\boldsymbol{h},\boldsymbol{x}) \quad \text{v.s.} \quad \min \|\boldsymbol{Z}\|_* \quad s.t. \|\mathcal{A}(\boldsymbol{Z}) - \boldsymbol{y}\| \le \eta.$$

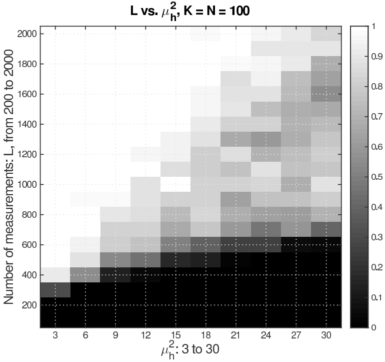
Nonconvex method requires fewer measurements to achieve exact recovery than convex method. Moreover, if $\boldsymbol{A}$ is a partial Hadamard matrix, our algorithm still gives satisfactory performance.

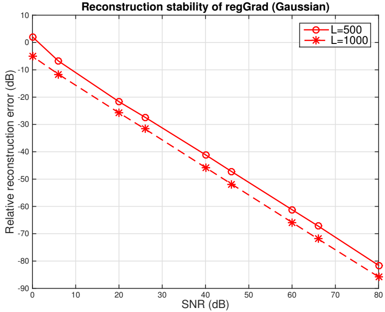


$K = N = 50$, $\boldsymbol{B}$ is a low-frequency DFT matrix.

L v.s. Incoherence μ_h^2 and stability

- The number of measurements L does depend linearly on μ_h^2.
- Our algorithm yields stable recovery if the observation is noisy.



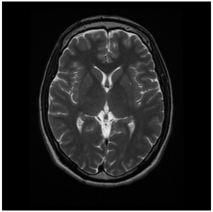


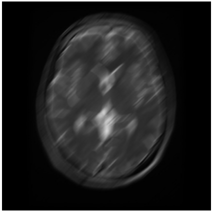
Here $K = N = 100$.

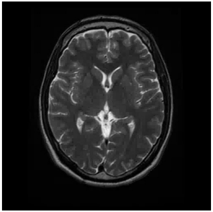
MRI Image deblurring:

Here $\boldsymbol{B}$ is a partial DFT matrix and $\boldsymbol{A}$ is a partial wavelet matrix.

When the subspace $\boldsymbol{B}, (K = 65)$ or support of blurring kernel is known:
$\boldsymbol{g} \approx \boldsymbol{Ax}$: image of 512×512; $\boldsymbol{A}$: wavelet subspace corresponding to the $N = 20000$ largest Haar wavelet coefficients of $\boldsymbol{g}$.

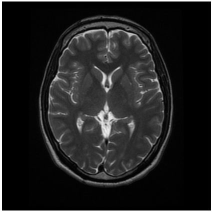


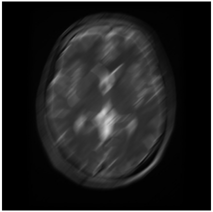


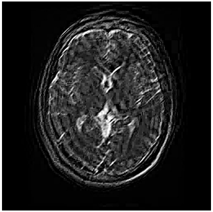


MRI Imaging deblurring:

When the subspace $\boldsymbol{B}$ or support of blurring kernel is unknown:
we assume the support of blurring kernel is contained in a small box;
$N = 35000$.







Important ingredients of proof

The first three conditions hold over "the basin of attraction" $\mathcal{N}_{d_0} \cap \mathcal{N}_{\mu} \cap \mathcal{N}_{\varepsilon}$.

Condition 1: Local Regularity Condition

Guarantee sufficient decrease in each iterate and linear convergence of $\widetilde{F}$:

$$\|\nabla \widetilde{F}(\boldsymbol{h}, \boldsymbol{x})\|^2 \geq \omega \widetilde{F}(\boldsymbol{h}, \boldsymbol{x})$$

where $\omega > 0$ and $(\boldsymbol{h}, \boldsymbol{x}) \in \mathcal{N}_{d_0} \cap \mathcal{N}_{\mu} \cap \mathcal{N}_{\varepsilon}$.

Condition 2: Local Smoothness Condition

Governs rate of convergence. Let $\boldsymbol{z} = (\boldsymbol{h}, \boldsymbol{x})$. There exists a constant C_L (Lipschitz constant of gradient) such that

$$\|\nabla \widetilde{F}(\boldsymbol{z} + t\Delta \boldsymbol{z}) - \nabla \widetilde{F}(\boldsymbol{z})\| \leq C_L t \|\Delta \boldsymbol{z}\|, \quad \forall 0 \leq t \leq 1,$$

for all $\{(\boldsymbol{z}, \Delta \boldsymbol{z}) : \boldsymbol{z} + t\Delta \boldsymbol{z} \in \mathcal{N}_{d_0} \cap \mathcal{N}_{\mu} \cap \mathcal{N}_{\varepsilon}, \forall 0 \leq t \leq 1\}$.

Important ingredients of proof

Condition 3: Local Restricted Isometry Property

Transfer convergence of objective function to convergence of iterates.

$$\frac{3}{4}\|\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F^2 \leq \|\mathcal{A}(\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*)\|^2 \leq \frac{5}{4}\|\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F^2$$

holds uniformly for all $(\boldsymbol{h}, \boldsymbol{x}) \in \mathcal{N}_{d_0} \cap \mathcal{N}_\mu \cap \mathcal{N}_\varepsilon$.

Condition 4: Robustness Condition

Provide stability against noise.

$$\|\mathcal{A}^*(\boldsymbol{w})\| \leq \frac{\varepsilon d_0}{10\sqrt{2}}.$$

where $\mathcal{A}^*(\boldsymbol{w}) = \sum_{l=1}^L w_l \boldsymbol{b}_l \boldsymbol{a}_l^*$ is a sum of L rank-1 random matrices. It concentrates around $\mathbf{0}$.

Two-page proof

Condition 1 + 2 $\Longrightarrow$ Linear convergence of $\widetilde{F}$

Proof.

Let $\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \eta\nabla\widetilde{F}(\boldsymbol{z}_t)$ with $\eta \leq \frac{1}{C_L}$. By using modified descent lemma,

$$\begin{aligned}\widetilde{F}(\boldsymbol{z}_t + \eta\nabla\widetilde{F}(\boldsymbol{z}_t)) &\leq \widetilde{F}(\boldsymbol{z}_t) - (2\eta + C_L\eta^2)\|\nabla\widetilde{F}(\boldsymbol{z}_t)\|^2 \\ &\leq \widetilde{F}(\boldsymbol{z}_t) - \eta\omega\widetilde{F}(\boldsymbol{z}_t)\end{aligned}$$

which gives $\widetilde{F}(\boldsymbol{z}_{t+1}) \leq (1-\eta\omega)^t\widetilde{F}(\boldsymbol{z}_0)$. □

Two-page proof: continued

Condition 3 $\Longrightarrow$ Linear convergence of $\|\boldsymbol{u}_t\boldsymbol{v}_t^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F$.

It follows from $\widetilde{F}(\boldsymbol{z}_t) \geq F(\boldsymbol{z}_t) \geq \frac{3}{4}\|\boldsymbol{u}_t\boldsymbol{v}_t^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F^2$. Hence, linear convergence of objective function also implies linear convergence of iterates.

Condition 4 $\Longrightarrow$ Proof of stability theory

If L is sufficiently large, $\mathcal{A}^*(\boldsymbol{w})$ is small since $\|\mathcal{A}^*(\boldsymbol{w})\| \to 0$. There holds

$$\|\mathcal{A}(\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*) - \boldsymbol{w}\|^2 \approx \|\mathcal{A}(\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*)\|^2 + \sigma^2 d_0^2.$$

Hence, the objective function behaves "almost like" $\|\mathcal{A}(\boldsymbol{h}\boldsymbol{x}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*)\|^2$, the noiseless version of F if the sample size is sufficiently large.

Conclusion: The proposed algorithm is the first blind deconvolution algorithm that is numerically efficient, robust against noise and comes with rigorous recovery guarantees under subspace conditions.

Outlook and Conclusion

Conclusion: The proposed algorithm is the first blind deconvolution algorithm that is numerically efficient, robust against noise and comes with rigorous recovery guarantees under subspace conditions.

- Can we remove the regularizers $G(\boldsymbol{h}, \boldsymbol{x})$ in the blind deconvolution?
- Can we generalize it to blind-deconvolution-blind-demixing problem, i.e., $\boldsymbol{y} = \sum_{i=1}^{r} \operatorname{diag}(\boldsymbol{B}_i \boldsymbol{h}_i) \boldsymbol{A}_i \boldsymbol{x}_i$?
- Can we show if similar result holds for other types of $\boldsymbol{A}$?
- What if $\boldsymbol{x}$ or $\boldsymbol{h}$ is sparse/both of them are sparse?
- Better choice of $\boldsymbol{B}$ in image deblurring?
- **See details**: Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization, *arXiv:1606.04933*.