Simultaneous Blind Deconvolution and Blind Demixing via Convex Programming

Shuyang Ling

Department of Mathematics, UC Davis

Nov.8th, 2016

Acknowledgements

Research in collaboration with:

Prof.Thomas Strohmer (UC Davis)

This work is sponsored by NSF-DMS and DARPA.

Outline

- Setup: blind deconvolution and demixing
- Convex relaxation and main result
- Numerics and idea of proof

What is blind deconvolution?

What is blind deconvolution?

Suppose we observe a function $\boldsymbol{y}$ which is 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 $\boldsymbol{y}$?

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

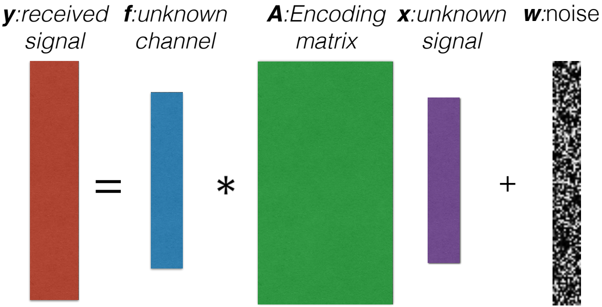
It is obviously a highly ill-posed bilinear inverse problem... but important in signal processing.

Blind deconvolution in wireless communication

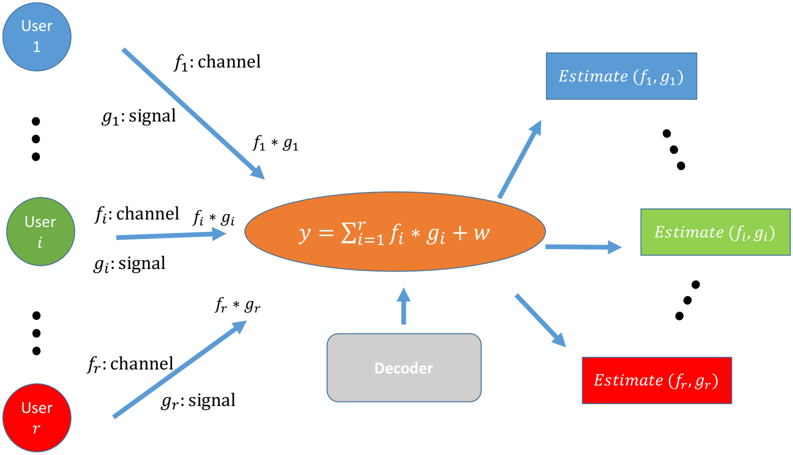
Joint channel and signal estimation in wireless communication

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 $\boldsymbol{y}$?

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



Blind deconvolution meets demixing?



Blind deconvolution and blind demixing

We start from the original model

$$\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{f}_i * \boldsymbol{g}_i + \boldsymbol{w}.$$

This is even more difficult than blind deconvolution, since this is a "mixture" of blind deconvolution problem

More assumptions

- Each impulse responses $\boldsymbol{f}_i$ has maximum delay spread K.

$$\boldsymbol{f}_i(n) = 0, \quad \text{for } n > K.$$

- $\boldsymbol{g}_i := \widetilde{\boldsymbol{A}}_i \boldsymbol{x}_i$ is the signal $\boldsymbol{x}_i \in \mathbb{C}^N$ encoded by matrix $\widetilde{\boldsymbol{A}}_i \in \mathbb{C}^{L \times N}$ with $L > N$.

Model under subspace assumption

In the frequency domain,

$$\hat{\boldsymbol{y}} = \sum_{i=1}^{r} \hat{\boldsymbol{f}}_i \odot \hat{\boldsymbol{g}}_i + \boldsymbol{w} = \sum_{i=1}^{r} \operatorname{diag}(\hat{\boldsymbol{f}}_i)\hat{\boldsymbol{g}}_i + \boldsymbol{w},$$

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

Subspace assumption

Denote $\boldsymbol{F}$ as the $L \times L$ DFT matrix.

- Let $\boldsymbol{h}_i \in \mathbb{C}^K$ be the first K nonzero entries of $\boldsymbol{f}_i$ and $\boldsymbol{B}_i$ be a low-frequency DFT matrix. There holds,

$$\hat{\boldsymbol{f}}_i = \boldsymbol{F}\boldsymbol{f}_i = \boldsymbol{B}_i \boldsymbol{h}_i.$$

- $\hat{\boldsymbol{g}}_i := \boldsymbol{A}_i \boldsymbol{x}_i$ where $\boldsymbol{A}_i := \boldsymbol{F}\widetilde{\boldsymbol{A}}_i$ and $\boldsymbol{x}_i \in \mathbb{C}^N$.

Mathematical model

Finally, we end up with the following model,

Model with subspace constraint

$$\boldsymbol{y} = \sum_{i=1}^{r} \mathrm{diag}(\boldsymbol{B}_i \boldsymbol{h}_i)\boldsymbol{A}_i \boldsymbol{x}_i + \boldsymbol{w},$$

Goal: We want to recover $(\boldsymbol{h}_i, \boldsymbol{x}_i)_{i=1}^r$ from $(\boldsymbol{y}, \boldsymbol{B}_i, \boldsymbol{A}_i)_{i=1}^r$.
Remark: The degree of freedom for unknowns: $r(K+N)$; number of constraint: L. To make the solution identifiable, we require $L \geq r(K+N)$ at least.

Special case if $r = 1$

In particular, if $r = 1$, it is a blind deconvolution problem.

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

Nonconvex optimization?

Naive approach?

We may want to try nonlinear least squares approach:

$$\min_{(\boldsymbol{u}_i,\boldsymbol{v}_i)} \left\| \sum_{i=1}^{r} \operatorname{diag}(\boldsymbol{B}_i \boldsymbol{u}_i)\boldsymbol{A}_i \boldsymbol{v}_i - \boldsymbol{y} \right\|^2 .$$

This gives a nonconvex objective function.

- May get stuck at local minima and no guarantees for recoverability.
- For $r = 1$, we have recovery guarantees by adding regularizers but not for $r > 1$.

Nonconvex optimization?

Naive approach?

We may want to try nonlinear least squares approach:

$$\min_{(\boldsymbol{u}_i,\boldsymbol{v}_i)} \left\| \sum_{i=1}^{r} \operatorname{diag}(\boldsymbol{B}_i \boldsymbol{u}_i)\boldsymbol{A}_i \boldsymbol{v}_i - \boldsymbol{y} \right\|^2.$$

This gives a nonconvex objective function.

- May get stuck at local minima and no guarantees for recoverability.
- For $r = 1$, we have recovery guarantees by adding regularizers but not for $r > 1$.

Two-step convex approach

(a) Lifting: convert nonconvex constraints to linear

(b) Solving a SDP relaxation and hope to recover $\{\boldsymbol{h}_i \boldsymbol{x}_i^*\}_{i=1}^r$

Convex approach of demixing problem

Step 1: lifting

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

$$y_l = \sum_{i=1}^{r} (\boldsymbol{B}_i \boldsymbol{h}_i)_l \boldsymbol{x}_i^* \boldsymbol{a}_{i,l} + w_l = \sum_{i=1}^{r} \boldsymbol{b}_{i,l}^* \boldsymbol{h}_i \boldsymbol{x}_i^* \boldsymbol{a}_{i,l} + w_l,$$

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

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

Then, there holds

$$\boldsymbol{y} = \sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{X}_i) + \boldsymbol{w}.$$

- Advantage: linear constraints (convex constraints)
- Disadvantage: dimension increases

Rank-r matrix recovery

Recast as rank-r matrix recovery

We rewrite $\boldsymbol{y} = \sum_{i=1}^{r} \text{diag}(\boldsymbol{B}_i \boldsymbol{h}_i)\boldsymbol{A}_i \boldsymbol{x}_i$ as

$$y_l = \left\langle \begin{bmatrix} \boldsymbol{h}_1\boldsymbol{x}_1^* & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{h}_2\boldsymbol{x}_2^* & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{h}_r\boldsymbol{x}_r^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{b}_{1,l}\boldsymbol{a}_{1,l}^* & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{b}_{2,l}\boldsymbol{a}_{2,l}^* & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{b}_{r,l}\boldsymbol{a}_{r,l}^* \end{bmatrix} \right\rangle$$

- Find a rank-r block diagonal matrix satisfying the linear constraints above.
- Finding such a rank-r matrix is also an NP-hard problem.

Convex relaxation

Nuclear norm minimization

Since this system is highly underdetermined, we hope to recover all $\{\boldsymbol{Z}_i\}_{i=1}^r$ from

$$\min \sum_{i=1}^{r} \|\boldsymbol{Z}_i\|_* \quad \text{subject to} \quad \sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{Z}_i) = \boldsymbol{y}.$$

Once we obtain $\{\hat{\boldsymbol{Z}}_i\}_{i=1}^r$, we can easily extract the leading left and right singular vectors from $\boldsymbol{Z}_i$ as the estimation of $(\boldsymbol{h}_i, \boldsymbol{x}_i)$.

Key question: does the solution to the SDP above really give $\{\boldsymbol{h}_i \boldsymbol{x}_i^*\}_{i=1}^r$? What conditions are needed?

State-of-the-art: blind deconvolution ($r = 1$)

Nuclear norm minimization

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

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

Convex optimization can be solved within polynomial time.

Theorem [Ahmed-Recht-Romberg 14]

Assume $\boldsymbol{y} = \text{diag}(\boldsymbol{Bh})\boldsymbol{Ax}$, $\boldsymbol{A} : L \times N$ is a Gaussian random matrix, and $\boldsymbol{B} \in \mathbb{C}^{L\times K}$ is a partial DFT matrix,

$$\boldsymbol{B}^*\boldsymbol{B} = \boldsymbol{I}_K, \quad L\|\boldsymbol{Bh}\|_\infty^2 \leq \mu_h^2,$$

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

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

Main results

Theorem [Ling-Strohmer 15]

Each $\boldsymbol{B}_i \in \mathbb{C}^{L\times K}$ partial DFT matrix with $\boldsymbol{B}_i^*\boldsymbol{B}_i = \boldsymbol{I}_K$ and each $\boldsymbol{A}_i$ is a Gaussian random matrix, i.e., each entry in $\boldsymbol{A}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$. Let μ_h^2 be as defined in $\mu_h^2 = L\max_{1\leq i\leq r} \frac{\|\boldsymbol{B}_i\boldsymbol{h}_i\|_\infty^2}{\|\boldsymbol{h}_i\|^2}$. If

$$L \geq C_{\alpha+\log r} r^2 \max\{K, \mu_h^2 N\} \log^3 L,$$

then the solution to convex relaxation satisfies

$$\hat{\boldsymbol{X}}_i = \boldsymbol{X}_i, \quad \text{for all } i = 1,\dots,r,$$

with probability at least $1 - \mathcal{O}(L^{-\alpha+1})$.

Remark

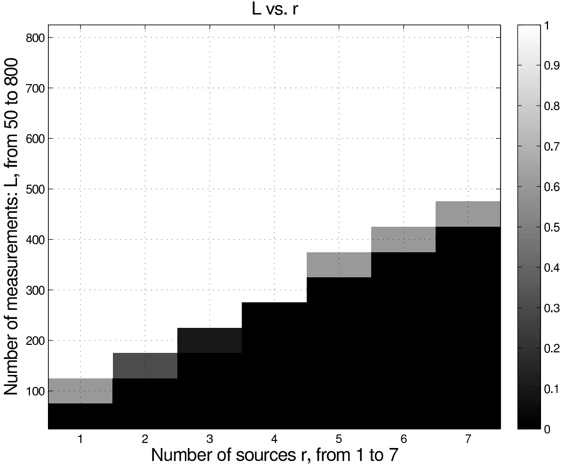
- Our result is a generalization of Ahmed-Romberg-Recht's result to $r > 1$.
- $\boldsymbol{B}_i$ have other choices other than DFT matrix.
- Incoherence μ_h^2 does affect the result.

Remark

- Our result is a generalization of Ahmed-Romberg-Recht's result to $r > 1$.
- $\boldsymbol{B}_i$ have other choices other than DFT matrix.
- Incoherence μ_h^2 does affect the result.
- r^2 is not optimal. One group has claimed to reduce the sampling complexity from r^2 to r.
- Empirically, $\boldsymbol{A}_i$ can be other matrices besides Gaussian. However, no theories exist so far.

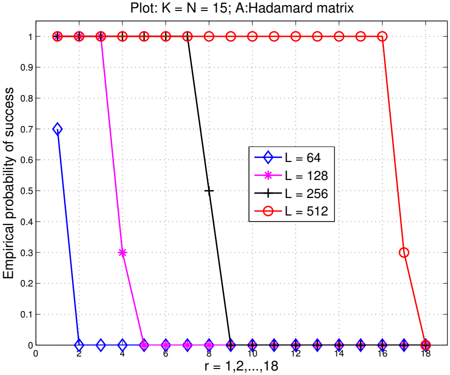
Numerics: does L really scales linearly with r?

$\boldsymbol{A}_i$ is chosen as Gaussian matrix. Here $K = 30$ and $N = 25$ are fixed.
Black: failure; White: success.



Numerics: does L really scales linearly with r?

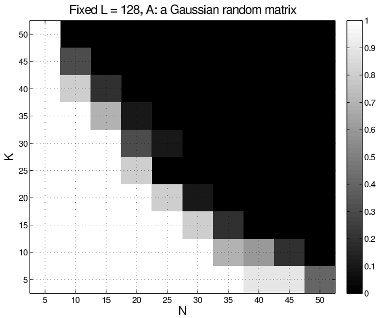
Choose $\boldsymbol{A}_i = \boldsymbol{D}_i \boldsymbol{H}$ where $\boldsymbol{H}$ is a partial Hadamard matrix (± 1 orthogonal matrix). $\boldsymbol{D}_i$ is a random ± 1 diagonal matrix.

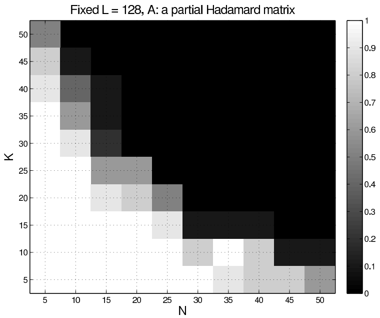


Number of unknowns $r(K+N) = 16 \cdot 30 = 480$ is slightly smaller than the number of constraints 512.

Numerics: L scales linearly with $K + N$.

Numerical simulations verify our theory that $L \approx \mathcal{O}(r(K+N))$ gives exact recovery. Here $r = 2$ and $L \approx 1.5r(K+N)$.

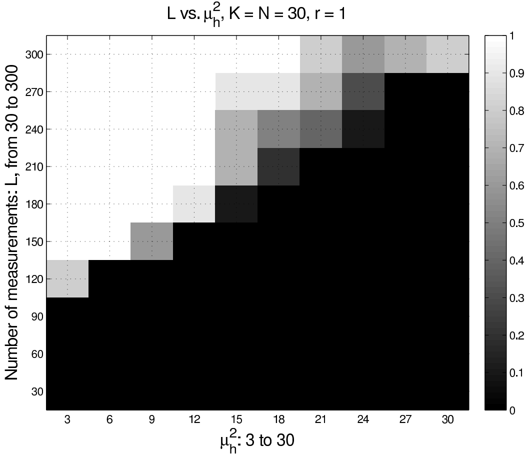




Numerics: L vs. μ_h^2

We observe strong linear correlation between minimal required L and μ_h^2:

$$L \propto \max_{1 \leq i \leq r} \frac{\|\boldsymbol{B}_i \boldsymbol{h}_i\|_\infty^2}{\|\boldsymbol{h}_i\|^2}.$$



Stability theorem

In reality measurements are noisy. Hence, suppose that $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{w}$ where $\boldsymbol{w}$ is noise with $\|\boldsymbol{w}\| \leq \eta$.

$$\min \sum_{i=1}^{r} \|\boldsymbol{Z}_i\|_* \quad \text{subject to} \quad \|\sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{Z}_i) - \hat{\boldsymbol{y}}\| \leq \eta. \tag{1}$$

Theorem

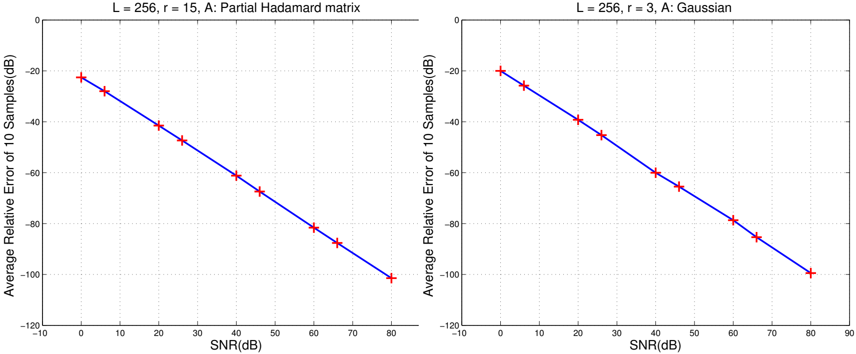
Assume we observe $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{w} = \sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{X}_i) + \boldsymbol{w}$ *with* $\|\boldsymbol{w}\| \leq \eta$. *Then, the minimizer* $\{\hat{\boldsymbol{X}}_i\}_{i=1}^{r}$ *satisfies*

$$\sqrt{\sum_{i=1}^{r} \|\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i\|_F^2} \leq Cr\sqrt{\max\{K, N\}}\eta.$$

with probability at least $1 - \mathcal{O}(L^{-\alpha+1})$.

Stability

We see that the relative error is linearly correlated with the noise in dB. Approximately, 10 units of increase in SNR leads to the *same* amount of decrease in relative error (in dB).



Sketch of proof

Let's consider the noiseless version,

$$\min \sum_{i=1}^{r} \|\boldsymbol{Z}_i\|_*, \quad \text{subject to} \sum_{i=1}^{r} \mathcal{A}_i(\boldsymbol{Z}_i) = \boldsymbol{y}.$$

Difficulties:

- $\boldsymbol{X}_i$ is asymmetric.
- How to deal with block diagonal structure?

Two-step proof

- Find a sufficient condition for exact recovery
- Construct an approximate dual certificate via golfing scheme

Sufficient condition

Three key ingredients to achieve exact recovery

1. Local isometry property on T_i

$$\max_{1\le i\le r} \|\mathcal{P}_{T_i}\mathcal{A}_i^*\mathcal{A}_i\mathcal{P}_{T_i} - \mathcal{P}_{T_i}\| \le \frac{1}{4}$$

   where $T_i = \{\boldsymbol{h}_i\boldsymbol{h}_i^*\boldsymbol{Z} + (\boldsymbol{I} - \boldsymbol{h}_i\boldsymbol{h}_i^*)\boldsymbol{Z}\boldsymbol{x}_i\boldsymbol{x}_i^*\}$.

2. Local incoherence property

$$\max_{i\ne j} \|\mathcal{P}_{T_i}\mathcal{A}_i^*\mathcal{A}_j\mathcal{P}_{T_j}\| \le \frac{1}{4r}$$

3. Existence of an approximate dual certificate, which is achieved via the celebrated golfing scheme). Find a $\boldsymbol{\lambda} \in \mathbb{C}^L$ such that for all $1 \le i \le r$,

$$\|\mathcal{P}_{T_i}(\mathcal{A}_i^*\boldsymbol{\lambda}) - \boldsymbol{h}_i\boldsymbol{x}_i^*\|_F \le \frac{1}{5r\gamma}, \quad \|\mathcal{P}_{T_i^\perp}(\mathcal{A}_i^*\boldsymbol{\lambda})\| \le \frac{1}{2}$$

   where $\gamma := \max\{\|\mathcal{A}_i\|\}$.

Conclusion and future work

- Can we derive a theoretical bound that scales linearly in r, rather than quadratic in r as our current theory? (It may have been solved!)
- Is it possible to develop satisfactory theoretical bounds for deterministic matrices $\boldsymbol{A}_i$?
- Fast algorithms: extend our nonconvex optimization framework to this blind-deconvolution-blind-demixing scenario.
- Can we develop a theoretical framework where the signals $\boldsymbol{x}_i$ belong to some non-linear subspace, e.g. for sparse $\boldsymbol{x}_i$?
- **See details in our paper**: *Blind Deconvolution Meets Blind Demixing: Algorithms and Performance Bounds. arXiv:1512.07730*.