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?

Suppose we observe a function \( y \) which is the convolution of two unknown functions, the blurring function \( f \) and the signal of interest \( g \), plus noise \( w \). How to reconstruct \( f \) and \( g \) from \( y \)?

\[
y = f * g + w.
\]

It is obviously a highly ill-posed bilinear inverse problem... but important in signal processing.
Suppose that a signal $x$, encoded by $A$, is transmitted through an unknown channel $f$. How to reconstruct $f$ and $x$ from $y$?

$$y = f \ast Ax + w.$$
Blind deconvolution meets demixing?

\[
y = \sum_{i=1}^{r} f_i \ast g_i + w
\]

User 1
- \( f_1 \): channel
- \( g_1 \): signal

User \( i \)
- \( f_i \): channel
- \( g_i \): signal

User \( r \)
- \( f_r \): channel
- \( g_r \): signal

Decoder

Estimate \((f_1, g_1)\)
Estimate \((f_i, g_i)\)
Estimate \((f_r, g_r)\)
Blind deconvolution and blind demixing

We start from the original model

\[ y = \sum_{i=1}^{r} f_i * g_i + w. \]

This is even more difficult than blind deconvolution, since this is a “mixture” of blind deconvolution problem.

More assumptions

- Each impulse responses \( f_i \) has maximum delay spread \( K \).
  \[ f_i(n) = 0, \quad \text{for } n > K. \]
- \( g_i := \tilde{A}_i x_i \) is the signal \( x_i \in \mathbb{C}^N \) encoded by matrix \( \tilde{A}_i \in \mathbb{C}^{L \times N} \) with \( L > N \).
Model under subspace assumption

In the frequency domain,

\[
\hat{y} = \sum_{i=1}^{r} \hat{f}_i \odot \hat{g}_i + w = \sum_{i=1}^{r} \text{diag}(\hat{f}_i)\hat{g}_i + w,
\]

where “\(\odot\)” denotes entry-wise multiplication. We assume \(y\) and \(\hat{y}\) are both of length \(L\).

Subspace assumption

Denote \(F\) as the \(L \times L\) DFT matrix.

- Let \(h_i \in \mathbb{C}^K\) be the first \(K\) nonzero entries of \(f_i\) and \(B_i\) be a low-frequency DFT matrix. There holds,
  \[
  \hat{f}_i = Ff_i = B_i h_i.
  \]

- \(\hat{g}_i := A_i x_i\) where \(A_i := F \tilde{A}_i\) and \(x_i \in \mathbb{C}^N\).
Mathematical model

Finally, we end up with the following model,

Model with subspace constraint

\[ y = \sum_{i=1}^{r} \text{diag}(B_i h_i) A_i x_i + w, \]

**Goal:** We want to recover \((h_i, x_i)_{i=1}^{r}\) from \((y, B_i, A_i)_{i=1}^{r}\).

**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.

\[ y = \text{diag}(B h) A x + w. \]
**Nonconvex optimization?**

**Naive approach?**

We may want to try nonlinear least squares approach:

\[
\min_{(u_i, v_i)} \left\| \sum_{i=1}^{r} \text{diag}(B_i u_i) A_i v_i - y \right\|^2.
\]

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:

\[
\min_{(u_i, v_i)} \left\| \sum_{i=1}^{r} \text{diag}(B_i u_i) A_i v_i - y \right\|^2.
\]

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 \( \{h_i x_i^*\}_{i=1}^{r} \)
Convex approach of demixing problem

**Step 1: lifting**

Let $a_{i,l}$ be the $l$-th column of $A_i^*$ and $b_{i,l}$ be the $l$-th column of $B_i^*$.

$$ y_l = \sum_{i=1}^{r} (B_i h_i) x_i^* a_{i,l} + w_l = \sum_{i=1}^{r} b_{i,l}^* h_i x_i^* a_{i,l} + w_l, $$

Let $X_i := h_i x_i^*$ and define the linear operator $A_i : \mathbb{C}^{K \times N} \rightarrow \mathbb{C}^L$ as,

$$ A_i(Z) := \{b_{i,l}^* Z a_{i,l}\}_{l=1}^{L} = \{\langle Z, b_{i,l} a_{i,l}^* \rangle\}_{l=1}^{L}. $$

Then, there holds

$$ y = \sum_{i=1}^{r} A_i(X_i) + w. $$

- **Advantage:** linear constraints (convex constraints)
- **Disadvantage:** dimension increases
Recast as rank-\(r\) matrix recovery

We rewrite \(y = \sum_{i=1}^{r} \text{diag}(B_i h_i) A_i x_i\) as

\[
y_l = \left\langle \begin{bmatrix} h_1 x_1^* & 0 & \cdots & 0 \\ 0 & h_2 x_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_r x_r^* \end{bmatrix}, \begin{bmatrix} b_{1,l} a_{1,l}^* & 0 & \cdots & 0 \\ 0 & b_{2,l} a_{2,l}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{r,l} a_{r,l}^* \end{bmatrix} \right\rangle
\]

- 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 \( \{Z_i\}_{i=1}^{r} \) from

\[
\min \sum_{i=1}^{r} \|Z_i\|_* \quad \text{subject to} \quad \sum_{i=1}^{r} A_i(Z_i) = y.
\]

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

**Key question:** does the solution to the SDP above really give \( \{h_i x_i^*\}_{i=1}^{r} \)? What conditions are needed?
State-of-the-art: blind deconvolution ($r = 1$)

**Nuclear norm minimization**

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

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

Convex optimization can be solved within polynomial time.

**Theorem [Ahmed-Recht-Romberg 14]**

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

$$B^*B = I_K, \quad L\|Bh\|_\infty^2 \leq \mu_h^2,$$

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

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

Theorem [Ling-Strohmer 15]

Each $B_i \in \mathbb{C}^{L \times K}$ partial DFT matrix with $B_i^* B_i = I_K$ and each $A_i$ is a Gaussian random matrix, i.e., each entry in $A_i$ i.i.d $\sim \mathcal{N}(0, 1)$. Let $\mu_h^2$ be as defined in $\mu_h^2 = L \max_{1 \leq i \leq r} \left\| \frac{B_i h_i}{h_i} \right\|^2$. If

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

then the solution to convex relaxation satisfies

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

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

- Our result is a generalization of Ahmed-Romberg-Recht’s result to $r > 1$.
- $B_i$ have other choices other than DFT matrix.
- Incoherence $\mu_h^2$ does affect the result.
Remark

- Our result is a generalization of Ahmed-Romberg-Recht’s result to $r > 1$.
- $B_i$ have other choices other than DFT matrix.
- Incoherence $\mu_h^2$ does affect the result.
- $r^2$ is not optimal. One group has claimed to reduce the sampling complexity from $r^2$ to $r$.
- Empirically, $A_i$ can be other matrices besides Gaussian. However, no theories exist so far.
**Numerics: does $L$ really scales linearly with $r$?**

$A_i$ 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 $A_i = D_i H$ where $H$ is a partial Hadamard matrix ($\pm 1$ orthogonal matrix). $D_i$ is a random $\pm 1$ 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 $L \approx O(r(K + N))$ gives exact recovery. Here $r = 2$ and $L \approx 1.5r(K + N)$. 

![Fixed L = 128, A: a Gaussian random matrix](image1)

![Fixed L = 128, A: a partial Hadamard matrix](image2)
We observe strong linear correlation between minimal required $L$ and $\mu_h^2$:

$$L \propto \max_{1 \leq i \leq r} \frac{\|B_i h_i\|_\infty^2}{\|h_i\|_2^2}.$$
Stability theorem

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

\[
\min \sum_{i=1}^{r} \| Z_i \|_* \quad \text{subject to} \quad \| \sum_{i=1}^{r} A_i(Z_i) - \hat{y} \| \leq \eta. \quad (1)
\]

Theorem

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

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

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

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

\[ \text{SNR (dB)} \]
\[ \text{Average Relative Error of 10 Samples (dB)} \]

L = 256, r = 15, A: Partial Hadamard matrix

L = 256, r = 3, A: Gaussian
Sketch of proof

Let’s consider the noiseless version,

\[
\min \sum_{i=1}^{r} \|Z_i\|_* \quad \text{subject to} \quad \sum_{i=1}^{r} A_i(Z_i) = y.
\]

Difficulties:
- \(X_i\) 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** $T_i$

\[
\max_{1 \leq i \leq r} \| \mathcal{P}_{T_i} A_i^* A_i \mathcal{P}_{T_i} - \mathcal{P}_{T_i} \| \leq \frac{1}{4}
\]

where $T_i = \{ h_i h_i^* Z + (I - h_i h_i^*) Z x_i x_i^* \}$.

2. **Local incoherence property**

\[
\max_{i \neq j} \| \mathcal{P}_{T_i} A_i^* A_j \mathcal{P}_{T_j} \| \leq \frac{1}{4r}
\]

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

\[
\| \mathcal{P}_{T_i} (A_i^* \lambda) - h_i x_i^* \|_F \leq \frac{1}{5r \gamma}, \quad \| \mathcal{P}_{T_i}^\perp (A_i^* \lambda) \| \leq \frac{1}{2}
\]

where $\gamma := \max \{ \| A_i \| \}$.
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 $A_i$?
- Fast algorithms: extend our nonconvex optimization framework to this blind-deconvolution-blind-demixing scenario.
- Can we develop a theoretical framework where the signals $x_i$ belong to some non-linear subspace, e.g. for sparse $x_i$?