Compressive Blind Source Separation

Yiyue Wu, Yuejie Chi, Robert Calderbank

Dept. of Electrical Engineering
Princeton University

Sept. 27, 2010
There are growing interests in applying sparse techniques to machine learning and image processing:

- SVM can be done in compressed domain [CJS09];
- Multi-label prediction via CS [HKLZ09];
- Bayesian inference for reconstruction [HC09, DWB08] [IMD06];
- Bayesian inference for image denoising, inpainting...
There are growing interests in applying sparse techniques to machine learning and image processing:

- SVM can be done in compressed domain [CJS09];
- Multi-label prediction via CS [HKLZ09];
- Bayesian inference for reconstruction [HC09, DWB08] [IMD06];
- Bayesian inference for image denoising, inpainting...

**This work and take-away message:** pick an interesting problem to showcase compressed measurements are as good as complete measurements as long as you have enough measurements.
Motivations

- Blind Source Separation (BSS) from conventional mixtures
  - Important in many areas: speech recognition, MIMO communications, etc.
**Motivations**

- **Blind Source Separation (BSS) from conventional mixtures**
  - Important in many areas: speech recognition, MIMO communications, etc.

- **In many cases, measurements are expensive:**
  - Body Area Networks (BAN): sensors are power hungry and need to last a few days to a few weeks.
  - Recent developments in Compressive Sensing (CS) provide an intriguing solution.
Motivations

- Blind Source Separation (BSS) from conventional mixtures
  - Important in many areas: speech recognition, MIMO communications, etc.
- In many cases, measurements are expensive:
  - Body Area Networks (BAN): sensors are power hungry and need to last a few days to a few weeks.
  - Recent developments in Compressive Sensing (CS) provide an intriguing solution.
- Our work: recover mixtures from small measurements:
  - Conventional methods like PCA and ICA may fail due to reduced dimensionality.
Background: Compressive Sensing

- Recovery of a sparse or compressible signal from a small number of linear measurements [Don06, CT05].

\[ y = \Phi x + n, \quad \Phi \in \mathbb{C}^{M \times N}, \quad M \ll N \]
Background: Compressive Sensing

- Recovery of a sparse or compressible signal from a small number of linear measurements [Don06, CT05].

\[ y = \Phi x + n, \quad \Phi \in \mathbb{C}^{M \times N}, \quad M \ll N \]

- Many classes of reconstruction algorithms available now:
  - "Old-fashioned" $\ell_1$ minimization.
  - Greedy algorithms: OMP, CoSaMP, GPSR...
  - Bayesian inference.
Background: Compressive Sensing

- Recovery of a sparse or compressible signal from a small number of linear measurements [Don06, CT05].

\[ y = \Phi x + n, \quad \Phi \in \mathbb{C}^{M \times N}, \quad M \ll N \]

- Many classes of reconstruction algorithms available now:
  - ”Old-fashioned” $\ell_1$ minimization.
  - Greedy algorithms: OMP, CoSaMP, GPSR...
  - Bayesian inference.

- Theoretical performance guarantee is usually given by Restricted Isometry Property (RIP) of measurement matrix satisfied by random matrices with high probability.
Background: Blind Source Separation

- Goal: to recover $T$ independent sources from $L$ observed mixture of sources, possibly corrupted by noise.

$$X = \Theta A + \epsilon$$

where

- $X \in \mathbb{C}^{N \times L}$ is the matrix of observations;
- $\Theta \in \mathbb{C}^{N \times T}$ is the matrix of sources;
- $\epsilon \in \mathbb{C}^{N \times L}$ is the noise;
- $A \in \mathbb{C}^{T \times L}$ is the mixing matrix.

- Without loss of generality, we let all representations lie in the wavelet domain.
  - particularly fit for image applications.
Approaches

- Separate procedures: Mixture Recovery + BSS

We address a **Bayesian answer** to this problem.

- Simplified procedures
- Better performance
Problem Formulation

- Compressed measurements of mixtures of the sources:

  \[ Y = \Phi \Theta A + N, \]

  so

  \[ p(Y_k|\Phi, \Theta, A, \alpha_k^N) \sim \mathcal{N}(\Phi \Theta A_k, (\alpha_k^N)^{-1}I), \]

  where \( Y_k \) and \( A_k \) are the \( k^{th} \) columns of \( Y \) and \( A \).

- To maximize the posterior distribution:

  \[
p(\Theta, A, N|Y, \Phi) 
  \propto p(Y|\Phi, \Theta, A, \alpha_N) \pi(N|\alpha_N) \pi(A|\alpha_A) \pi(\Theta|\alpha_\Theta)
  \]

  where \( \alpha = [\alpha_N, \alpha_A, \alpha_\Theta] \) is the set of the hyper parameters.
Hidden Markov Tree Model

- Model the statistical dependencies between wavelet-domain coefficients. [CNB98]
- Persistence:
  - Parent node large/small $\xrightarrow{h.p.}$ Child node large/small
  - A node large/small $\xrightarrow{h.p.}$ Adjacent nodes large/small
- Mixed Gaussian Model:

$$\theta^{s,i} \sim (1 - \pi^{s,i})\delta_0 + \pi^{s,i}\mathcal{N}(0, (\alpha^s)^{-1})$$

(Yuejie Chi, ICIP 2010: Compressive BSS)
Prior Distributions

- Prior distribution of noise variance:
  \[ \alpha_k^N \sim \text{Gamma}(a_0, b_0) \]

- Prior distribution of \( A = \{a_{ij}\} \):
  \[ a_{ij} \sim \mathcal{N}(\mu_{ij}, \alpha^{-1}_{ij}), \ 1 \leq i \leq T, 1 \leq j \leq L \]

- Prior distributions of \( \Theta \):
  \[ \theta^{s,i} \sim (1 - \pi^{s,i})\delta_0 + \pi^{s,i} \mathcal{N}(0, (\alpha^s)^{-1}) \]
  
  with \( \pi^{s,i} = \)
   \[
   \begin{cases}
   \pi^r \sim \text{Beta}(e^r_0, f^r_0), & \text{if } s = 1 \\
   \pi^{s0} \sim \text{Beta}(e^{s0}_0, f^{s0}_0), & \text{if } 2 \leq s \leq S, \theta^{p(s,i)} = 0 \\
   \pi^{s1} \sim \text{Beta}(e^{s1}_0, f^{s1}_0), & \text{if } 2 \leq s \leq S, \theta^{p(s,i)} \neq 0 
   \end{cases}
   \]

  \[ \alpha^s \sim \text{Gamma}(c_0, d_0) \]
The Gibbs sampler samples from the following conditional distributions at iteration $t$,

\[
\theta_{k}^{s,i}(t) \sim p(\theta_{k}^{s,i} | Y, \Phi, A(t-1), \alpha_{N}^{s}(t-1), \alpha_{k}^{s}(t-1), \pi_{k}^{s,i}(t-1)),
\]

\[
A(t) \sim p(A | Y, \Phi, A(t-1), \Theta(t-1), \alpha_{N}^{N}(t-1)),
\]

\[
\alpha_{k}^{N}(t) \sim p(\alpha_{k}^{N} | Y_{k}, \Phi, A_{k}(t-1), \Theta(t-1)),
\]

\[
\alpha_{k}^{s}(t) \sim p(\alpha_{k}^{s} | \theta_{k}^{s,i}(t-1)),
\]

\[
\pi_{k}^{s,i}(t) \sim p(\pi_{k}^{s,i} | \theta_{k}^{s,i}(t-1)),
\]

where $\{\theta_{k}^{s,i}(t)\}$ is the set of wavelet coefficients associated with the $k^{th}$ source in the $t^{th}$ iteration.

Note: all distributions belong to the exponential family!
Figure: Bayesian compressive blind separation of one-dimensional signals
One-dimensional BSS

Figure: Comparisons of recovered wavelet coefficients

- Original signals are sparse
- Our proposed method outperforms the separate procedure
Figure: Bayesian compressive blind separation of two images
Two-dimensional BSS

Figure: Comparisons of the first 256 recovered wavelet coefficients

- Original signals are not well sparse
Conclusions:

- We have addressed the blind source separation problem directly from the compressed mixtures obtained from compressive sensing measurements.
- Our approach outperforms the existing separate procedure.

Future work:

- Improving our proposed method in separation and recovery of nearly sparse signals.
- Incorporating dictionary learning in the inference procedure in 2-D case to obtain better performance.
Main References


THANK YOU!