Modeling and Methods of Signal Separations with Applications in Spectroscopic Sensing

Yuanchang Sun

Department of Mathematics and Statistics

Florida International University
Acknowledgements
Raman spectroscopy

In the summer of 2003, an unclaimed bag was reported on the New York subway. It contained a plastic bag containing about 5 lbs. of a white powder...

Figure: Identification of an unknown white powder on New York Subway by Raman Spectroscopy. Images are from www.perkinelmer.com
Nuclear Magnetic Resonance (NMR)

Figure: NMR spectra of metabolites in biofluids. Plot from High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. (2008)
Figure: Hyperspectral Remote Sensing. Image is from www.markelowitz.com.
Applications and challenges

- Homeland security
- Atmospheric sciences
- Medical sciences
- Speech processing
- and more...

- Multiple substances, absence of spectral references
- Measurement error, environment noise
- Spectral distortions (variability)
Linear Mixture Model

Consider the linear model

$$X = AS + N,$$

where

- $X \in \mathbb{R}^{m \times p}$: the mixed signals
- $A \in \mathbb{R}^{m \times n}$: the mixing coefficients
- $S \in \mathbb{R}^{n \times p}$: the target (source) signals

1. Blind signal separation (BSS)
2. Data matching
3. Partially blind signal separation (pBSS)
BSS METHODS

- Independent component analysis (ICA)
- Nonnegative Matrix Factorization (NMF)
- Methods based on convex geometry of the data
Figure: a NMR spectrum of a mixture of three chemical compounds.
Stand alone peaks (SAP), (Naanaa & Nuzillard,'05)
an example

\[ X_{3 \times p} = A_{3 \times 3} S_{3 \times p}, \text{ view each column of } X \text{ as a point in 3-space} \]

\[
\begin{bmatrix}
X(:, 1), X(:, 2), \ldots, X(:, p)
\end{bmatrix}
\]

\[ = [A(:, 1), A(:, 2), A(:, 3)] \times \begin{pmatrix}
\ast & \ldots & \ast & 1 & o & o & \ast & \ldots & \ast \\
\ast & \ldots & \ast & o & 1 & o & \ast & \ldots & \ast \\
\ast & \ldots & \ast & o & o & 1 & \ast & \ldots & \ast
\end{pmatrix},
\]

**Remark:** \(X(k,:)\) represents the \(k\)-th row, \(X(:,k)\) the \(k\)-th column.
**Convex cone: vertex representation**

**Figure:** A cloud of data points (columns of $X$), left plot, is rescaled to lie on a plane determined by the vertices.
Identification of $A$, noiseless case

To identify vertexes of the cone, the optimization is suggested for each $k$

$$c = \min_{j=1,j\neq k} \sum_{j=1}^{p} \lambda_j$$

s.t.

$$\sum_{j=1,j\neq k}^{p} X(:,j)\lambda_j = X(:, k), \lambda_j \geq 0.$$  

$X(:, k)$ is an edge of the convex cone if and only if the optimal objective function value $c^*$ is greater than 1.
Noise case

If there is noise, the following optimization is suggested for each $k$

$$\text{minimize } \text{score} = \| \sum_{j=1, j \neq k}^{p} X(:, j) \lambda_j - X(:, k) \|_2^2, \quad k = 1, \ldots, p$$

subject to $\lambda_j \geq 0$. 
Numerical example (NMR data by Shaka’s group at UCI)
Separation results

Top: Spectral references; Bottom: Recovered spectral structures
NMR Spectra of Biofluids

Figure: NMR spectra of serum and urine.
Case 1
source condition: SAPDi, (S, Xin, ’11)

Assumption

For each $i \in \{1, 2, \ldots, n - 1\}$ there exists an $j_i \in \{1, 2, \ldots, p\}$ such that $S(i, j_i) \gg S(k, j_i)$ ($k = 1, \ldots, i - 1, i + 1, \ldots, n$).

Moreover, $s(n, j) > 0$ for $j \in \{1, 2, \ldots, p\}$, and there is a set $I \subset \{1, 2, \ldots, p\}$ such that $S(n, k) \gg S(i, k)$ for $k \in I$ and $i \in \{1, 2, \ldots, n - 1\}$.

Or $S$ contain $n - 1$ stand-alone-peak (SAP) source signals, and one wide-peak source signal with dominant intervals (Di) over other sources. Or SAPDi condition.

$$[X(:, 1), X(:, 2), \ldots, X(:, p)] = [A(:, 1), A(:, 2), A(:, 3)] \cdot \begin{pmatrix} * & \cdots & 1 & o & \cdots & o & \cdots & o \\ * & \cdots & o & 1 & \cdots & o & \cdots & o \\ * & \cdots & * & * & \cdots & * & \cdots & * \end{pmatrix}.$$
Identifying $A(:, 3)$

The dominant intervals in $S(3, :)$ imply

$$X(:, k) = A(:, 3) \cdot \ast + A(:, 1) \cdot o + A(:, 2) \cdot o, \quad \ast \gg o,$$

or certain columns of $X$ cluster around $A(:, 3)$. 
Model reduction

The reduced mixture data

\[ Y = \begin{bmatrix}
X(1,:) - X(3,:) \frac{A(1,3)}{A(3,3)} \\
X(2,:) - X(3,:) \frac{A(2,3)}{A(3,3)}
\end{bmatrix}, \]

which contains the stand-alone-peak sources \( S(1,:) \), \( S(2,:) \). The convex cone methods apply.
Recovery of $S(3, :)$

\[ S(3, :) = S_p * \mathcal{L}, \]

where \( \mathcal{L}(x, w) = \frac{1}{\pi} \frac{\frac{1}{2} w}{x^2 + \left(\frac{1}{2} w\right)^2} \) is the Lorentzian kernel.

\[
\begin{align*}
\left( A(:, 1), A(:, 2), S_p \right) &= \arg \min \mu \| S_p \|_1 \\
&\quad + \| X - [A(:, 1), A(:, 2)] \cdot S(1:2,:) - S_p * \mathcal{L} \|_2^2.
\end{align*}
\]
Extension to multiple wide-peak signals: Recursive BSS, 
(S, Xin,’12)

Backwards

Begin \( X^n = X, A^n = A; \)

for \( i = n \rightarrow 2 \) do

\hspace{1em} \text{Identify } A^i(:,i) \text{ from } X^i;

\hspace{1em} \text{Set } Y = X^i

\hspace{1em} \text{Reduce mixtures } Y \leftarrow Y - Y \frac{A^i(1:i-1,i)}{A^i(i,i)}

\hspace{1em} \text{Set } X^i = Y.

end for

Forwards

for \( i = 2 \rightarrow n \) do

\hspace{1em} \text{Solve } S(i,:) = S_p \ast L \text{ by minimizing}

\hspace{1em} S_p = \arg \min \mu \| S_p \|_1

\hspace{1em} + \| X^i - A \cdot S(1:i-1,:) - S_p \ast L \|_2^2.

end for
Numerical example
Results

Figure: The recovered source signals and the ground truth (right).
uBSS (S, Xin, ’11)

Let \((m, n)\) = number of (mixtures, sources). If \(m < n\), we have under-determined blind signal separation or uBSS.

**Figure:** Non-uniqueness. Left \((m, n) = (2, 3)\), right \((m, n) = (3, 5)\)
Figure: Non-uniqueness. Left \((m, n) = (3, 4)\), right \((m, n) = (3, 5)\)
Consider $m \geq 2$ mixtures and $n > m$ sources. We propose to strengthen SPA by the $(m - 1)$-tuplewise maximum overlap condition (MOC-SPA) on the source signals;

**Assumption**

*For each column of the source matrix $S$, there are at most $m - 1$ nonzero entries. Furthermore, for each $i \in \{1, 2, \ldots, n\}$ there exists an $j_i \in \{1, 2, \ldots, p\}$ such that $s_{i,j_i} > 0$ and $s_{k,j_i} = 0 (k = 1, \ldots, i - 1, i + 1, \ldots, n)$.*

We propose a degenerate mixing system: mixing matrix $A$ has one of its columns is a positive linear combination of other linearly independent columns. We shall call this assumption one column degenerate condition (OCDC).
Theorem

*Up to scaling and permutation, the uBSS problem (m > 3 sources, m + 1 mixtures) attains a unique factorization $A, S$, provided that $A$ satisfies OCDC, and $S$ satisfies MOC-SPA.*
summary

- study BSS conditions that differ from SPA. (Facet component analysis, Yin, Sun, Xin,'13)
- develop convex optimization methods for the data
- test and evaluate the algorithms on realistic data.
Given that a noisy mixed signal \( x = a_1 s_1 + a_2 s_2 + a_3 s_3 + \eta \), find their proportions \( a_1, a_2, a_3 \). The least squares solution is
\[
A = x S^T (S S^T)^{-1}
\]
which solves
\[
\min_A \| x - AS \|_2^2,
\]
where \( A = [a_1, a_2, a_3] \), \( S = [s_1; s_2; s_3] \).
Complications

Raman Spectroscopy there exist random spectral shifts. In hyperspectral data, the spectral signatures of the endmembers (source signals, or basis functions) can vary spectrally as well spatially from one image to another.

![Raman spectral reference of Methanol (red) and 20 instances under spectral variability (blue lines).](image1)

![Raman spectral reference of Acetonitrile (red) and 20 instances under spectral variability (blue lines).](image2)
Figure: Hyperspectral data, reference spectra (red lines) and 20 corresponding instances under spectral variability (blue dotted lines).
A Perturbed Linear Model

\[ x_i = \sum_{j=1}^{n} a_{ij} \cdot (s_j + \delta_{ij}) + \eta_i, \text{ for } i = 1, \ldots, m, \]

where \( x_i \) denotes the \( i \)th mixed signal, \( s_j \) is the \( j \)th source signal, \( a_{ij} \) is the proportion of the \( j \)th source signal in the \( i \)th mixture, \( \delta_{ij} \) denotes the perturbation of the \( j \)th source in the \( i \)th mixture. Note that \( \delta_{ij} \) is a vector of the same size of \( s_j \).

In matrix form

\[ X = AS + (A \odot \Delta)1_{mp} + N \]

Here symbol \( \odot \) means the element-wise multiplication.
Problem formulation

We formulate the following optimization problem

$$C(\Delta, A) = \frac{1}{2}\|X - AS - (A \odot \Delta)1_{mp}\|_F^2 + \text{penalty terms}.$$ 

One may use the Alternating Direction Method of Multipliers.
Random shifts and squeezing

- Source signal
- Shifted signal
- Source signal
- Squeezed signal
Mathematical model of random shifts

Let us consider modeling shift effect by writing each row of $X$ as $x_i(\nu), i = 1, \ldots, m$. It gives the row entries when discretized. Let the shift on row $s_j = s_j(\nu)$ of $S$ be denoted by $\xi_{ij}$ which means the shift of source $s_j$ in mixture $x_i$. The nonlinear mixing model with shift adjustment is:

$$x_i = \sum_{j=1}^{n} a_{ij} s_j(\nu + \xi_{ij}) + N_i ,$$

and the related minimization problem is

$$(a_{ij}, \xi_{ij}) = \arg \min \| x_i - \sum_{j=1}^{n} a_{ij} s_j(\nu + \xi_{ij}) \|_2 .$$
Augmented Least Squares

**AgLS** : fit the reference spectra along with their derivatives to the mixed signals.

The Taylor expansion of \( s_j(\nu + \xi_{ij}) \) is,

\[
s_j(\nu + \xi_{ij}) = s_j(\nu) + \xi_{ij}s'_j(\nu) + \frac{1}{2}\xi_{ij}^2s''_j(\nu) + \cdots ,
\]

The augmented fitting basis is \( \tilde{S} = [s_1; s'_1; s_2; s'_2; \cdots; s_n; s'_n] \) where only the first derivatives are included. We will then solve the following augmented least squares problem for the linear mixture model \( X = \tilde{A}\tilde{S} + N \) whose solution is \( \tilde{A} = X\tilde{S}^T(\tilde{S}\tilde{S}^T)^{-1} \), here \( \tilde{A} \) contain weights of all the source signals and their derivatives.
Augmented Maximum Likelihood Estimator

Consider the following model including the source spectra and their first derivatives along with random shifts variables,

\[
\chi_{ij} = \sum_{k=1}^{n} a_{ik} s_{kj} + \sum_{k=1}^{n} a_{ik} \xi_{ik} s'_{kj} + n_{ij}.
\]

Let \( \Gamma_k \in \mathbb{R}^{m \times p} \) with \( \Gamma_{k,ij} = a_{ik} s'_{kj}, k = 1, \cdots, n \). Notice that \( \Gamma_k = \Gamma_k(A) \) depends on the unknown mixing matrix \( A \), then we have

\[
\chi_{ij} = \sum_{k=1}^{n} a_{ik} s_{kj} + \sum_{k=1}^{n} \Gamma_{k,ij} \xi_{ik} + N_{ij}
\]
Assumptions

Assume that $\xi_{ik}$ are i.i.d. Gaussian $\mathcal{N}(0, \sigma_k^2)$, $k = 1, \cdots, n$ and $N_{ij} \sim \mathcal{N}(0, \tau^2)$, and $\xi_{ik}$ and $N_{ij}$ are independent.

Let $\text{vec}(X) = [x_{11}, x_{12}, \cdots, x_{1p}, \cdots, x_{m1}, \cdots, x_{mp}]^T$. Then $\text{vec}(X) \sim \mathcal{N}_{mp}(\text{vec}(AS), V)$.

$$V = \tau^2 I_{mp \times mp} + \sum_{k=1}^{n} \Omega_k \sigma_k^2 = \begin{pmatrix} V_1 & \mathbf{0} \\ \mathbf{0} & \ddots \\ \mathbf{0} & \mathbf{0} & V_m \end{pmatrix}_{mp \times mp}$$

where $\Omega_k \in \mathcal{R}^{mp \times mp} = \text{diag}\left( (\Gamma_k)_1,:), (\Gamma_k)_1,:), \cdots, (\Gamma_k)_i,:), (\Gamma_k)_i,:), \cdots, (\Gamma_k)_m,:), (\Gamma_k)_m,:) \right)$, here $$(\Gamma_k)_i,:) \in \mathcal{R}^{1 \times p} = a_{ik}s'_k, :.$$ Hence $\Omega_k$ is a block diagonal matrix, and $V_i$ is $p \times p$ matrix.
The log-likelihood function is

\[ L(A, \sigma^2, \tau^2, X, S) = -\frac{1}{2} \ln \det(V) - \frac{1}{2} \left( \text{vec}(X - AS)^T V^{-1} (X - AS) \right) - \frac{1}{2} \ln(2\pi) . \]

Notice that \( V = V(\sigma^2, A) \) is determined by \( \sigma^2 \) and \( A \). If \( V, \sigma^2, \) and \( \tau^2 \) are known, the maximized likelihood estimator (AgMLE) of \( A \) is the solution of \( m \) generalized least squares, where \( \hat{A}_{i,:} \), the AgMLE of \( A_{i,:) \) for the \( i \)th mixture is

\[ \hat{A}_{i,:} = \arg \min \| V_i^{-1/2} (X_{i,:} - A_i S)^T \|_2^2 , \quad (1) \]
Iterative Algorithm

Iterative approach for AgMLE of \((A, \sigma^2, \tau^2)\),

1. Start with an initial \(A^{(0)}\) obtained by the ordinary least squares, \(\min \| X - AS \|^2_2 \). Then \(\Gamma_{k,ij}^{(0)} = A_{ik}^{(0)} S_{kj} \).

2. Obtain \(\xi = (\xi_{ik})\) by solving 
   \[ (\xi_{ik}) = \arg \min \| (X - A^{(0)} S)_{i,:} - \sum_k (\Gamma_k)_{i,:} \xi_{i,k} \|^2_2 \]. Then
   \[
   \sigma_k^{(0)} = \sqrt{\frac{\sum_{i=1}^m \xi_{ik}^2}{m}},
   \]
   \[
   \tau^0 = \frac{1}{\sqrt{mp}} \| (X - A^{(0)} S)_{i,:) - \sum_k (\Gamma_k)_{i,:} \xi_{i,k} \|_2;
   \]
   \[
   V_i^{(0)} = \sum_{k=1}^n (\Gamma_k)_{i,:)^T (\Gamma_k)_{i,:} (\sigma_k^{(0)})^2.
   \]

3. Update \(A^{(0)}\) by \(A_{i,:}^{(1)} = \arg \min \| V_i^{-1/2} (X_{i,:] - A_i^{(0)} S)^T \|_2^2 \).

4. Set \(A^{(0)} \leftarrow A^{(1)}\) and iterate step 2-3 until it converges.

Y. Sun  
Modeling and Methods of Signal Separations
**Figure:** The result of the augmented least squares (AgLS), and its comparison to the ordinary least squares. The standard deviations of the random shifts of source signals are $\sigma_1 = \sigma_2 = 1$. The standard deviation of the noise is 0.05, the ground truth of the weights of source 1 is one.
**Figure:** The result of the augmented maximum likelihood estimator (AgMLE), and its comparison to the ordinary least squares. The standard deviations of the random shifts of source signals are $\sigma_1 = 1$ and $\sigma_2 = 1$, whose estimations are $\hat{\sigma}_1 = 1.0722$, $\hat{\sigma}_2 = 0.7617$. The standard deviation of the noise is $\tau = 0.05$ and its estimation is $\hat{\tau} = 0.1806$, the ground truth of the weights of source 1 is one.
Background

BSS

Data Fitting

Figure: The spectral references of trace gases HONO and NO$_2$, and their noisy mixture.

Y. Sun

Modeling and Methods of Signal Separations
Figure: Left: The computed coefficients of HONO and NO$_2$ by augmented least squares (AgLS). Right: By the augmented maximum likelihood estimator (AgMLE).
Ongoing and Future work

- Serial correlation effects among the spectral variability from mixture to mixture.
- Validate the model with more real data, e.g. hyperspectral datasets.
- Signals with discontinuations and jumps (with Kai Huang)
Figure: signal with jumps and oscillations.

One idea is to solve optimization problem for the data matching with shift,

\[(A, P) = \arg \min_{A, P \geq 0} \|X - ASP\|_2 + \lambda \|P\|_1 + \text{other terms},\]

\(P\) is the shift matrix (sparse). We shall pursue an iterative way to solve the problem.
References

1. (with W. Wu, J. Xin) Computational modeling of spectral data fitting with non-linear distortions. under review
2. (with K. Huang, J. Xin) Blind source separation methods for nearly degenerate mixtures with applications in spectroscopy. under review
9. (with C. Ridge, F. del Rio, AJ. Shaka, J. Xin) Postprocessing and sparse blind source separation of positive and partially overlapped data, Sig. Process., 91(8),