Shift Invariant Data Decomposition

Morten Mørup
Informatics and Mathematical Modeling
Intelligent Signal Processing
Technical University of Denmark
Collaborators

Kristoffer Hougaard Madsen and Lars Kai Hansen and Mikkel N. Schmidt
Informatics and Mathematical Modeling
Intelligent Signal Processing
Technical University of Denmark

Sidse M. Arnfred, Dr. Med. PhD
Cognitive Research Unit
Hvidovre Hospital
University Hospital of Copenhagen
Outline

- Shifted Non-negative Matrix Factorization
- Shifted Independent Component analysis
- Generalization to tensors (i.e., the PARAFAC model)
- Shift Invariant Sparse Coding
Factor Analysis

Spearman ~1900

\[ V \approx WH \]

Independent Component Analysis (ICA)
rows of \( H \) statistically independent
(P. Common, Bell & Sejnowski ~1995)

Non-negative Matrix Factorization (NMF):
\[ V_{n,m}, W_{n,d}, H_{d,m} \geq 0 \]
Multiplicative updates

\[ \frac{\partial C(\beta)}{\partial \beta_i, m} = \frac{\partial C(\beta)^+}{\partial \beta_i, m} - \frac{\partial C(\beta)^-}{\partial \beta_i, m} \]

\[ \beta_{i,m} = \beta_{i,m} - \mu_{i,m} \frac{\partial C(\beta)}{\partial \beta_i, m}, \quad \mu_{i,m} = \frac{\beta_{i,m}}{\partial C(\beta)^+} \]

\[ \beta_{i,m} = \beta_{i,m} - \frac{\beta_{i,m}}{\partial C(\beta)^+} \left( \frac{\partial C(\beta)^+}{\partial \beta_i, m} - \frac{\partial C(\beta)^-}{\partial \beta_i, m} \right) = \beta_{i,m} \frac{\partial C(\beta)^-}{\partial \beta_i, m} \frac{\partial C(\beta)^+}{\partial \beta_i, m} \]

Step size parameter

\[ \beta_{i,m}^{t+1} \leftarrow \beta_{i,m}^t \left( \frac{\partial C(\beta)^-}{\partial \beta_i, m} \frac{\partial C(\beta)^+}{\partial \beta_i, m} \right)^\alpha \]

\[ \frac{\partial C(\beta)^+}{\partial \beta_i, m} < \frac{\partial C(\beta)^-}{\partial \beta_i, m} \quad \beta_{i,m}^{t+1} \rightarrow \beta_{i,m}^t \]

\[ \frac{\partial C(\beta)^+}{\partial \beta_i, m} > \frac{\partial C(\beta)^-}{\partial \beta_i, m} \quad \beta_{i,m}^t \rightarrow \beta_{i,m}^{t+1} \]
Non-negative matrix factorization

\[ V_{i,j} \geq 0 \quad , \quad W_{i,d} \geq 0 \quad \text{and} \quad H_{d,j} \geq 0 \]

\[ C_{LS} = \frac{1}{2} \| V - WH \|_F^2 = \frac{1}{2} \sum_{i,j} (V_{i,j} - (WH)_{i,j})^2 \]

\[ W_{i,d} \leftarrow W_{i,d} \frac{(VH^T)_{i,d}}{(WHH^T)_{i,d}} \]

\[ H_{d,j} \leftarrow H_{d,j} \frac{(W^TV)_{d,j}}{(W^TWH)_{d,j}} \]

\[ C_{KL} = \sum_{i,j} V_{i,j} \log \frac{V_{i,j}}{(WH)_{i,j}} - V_{i,j} + (WH)_{i,j} \]

\[ W_{i,d} \leftarrow W_{i,d} \frac{\sum_j (WH)_{i,j} H_{d,j}}{\sum_j H_{d,j}} \]

\[ H_{d,j} \leftarrow H_{d,j} \frac{\sum_i W_{i,d} V_{i,j}}{\sum_i W_{i,d}} \]

(Some other approaches: Active Set, projected gradient, barrier functions, exponentiation)

NMF gives Part based representation

(Lee & Seung – Nature 1999)
Maximum likelihood (ML) ICA approach

\[ [A', S', H'] = \text{SVD}(X) \]

\[ W' = A'S' \]

Notice decomposition then ambiguous since

\[ WH = (W'Q^{-1})(QH') = W'H' \]

Thus ICA forms objective for ambiguity \( Q \) that minimizes:

\[ p(H|Q) = \prod_{m} p(H_m|Q) = \prod_{m} |\text{det}(Q)|p(QH_m) \]

Equivalently we derive two step procedure

Shift Invariant Subspace Analysis (SISA)

Shifted Independent Component Analysis (SICA)

Assume Independence

Change of variable principle
The shift problem

Convolutive ICA/NMF (echo effects, Smaragdis 2003)

\[ V_{n,m} = \sum_{d,\tau} W_{n,d}^{\tau} H_{d,m-\tau} + E_{n,m} \]

Shifted ICA/NMF (One specific delay between each sensor and source)

\[ V_{n,m} = \sum_{d} W_{n,d} H_{d,m-\tau_{n,d}} + E_{n,m} \]
History of shift

- Bell & Sejnowski 1995 (Sketched how to handle time delays in networks)
- Torkkola 1996 (Further developed Bell and Sejnowski’s work)
- Emile & Comon 1998 (Delay in model based on equally mixed sources formed by moving averages)
- Hong and Harshman 2003 – shifted factor analysis (a procedure based on exhaustive search over integer shifts – model conjectured unique)
- Yeredor 2003 (Solved the ICA model with shifts by joint diagonalization (sources=sensors) of the source cross spectra based on the AC-DC algorithm with non-integer shifts for the 2x2 system)
- Yeredor 2005 (extension to complex signals)
Why shifted ICA/NMF

Causes of shifts for instance
Doppler effect, Time of arrival differences
- Magnetic resonance spectra (Du et al, 2005)
- Astronomical spectrometers (red shift) (Pauca et al. 2006)
- Fluorescence spectra (Gobinet et al. 2004)
- PET imaging (Kim et al. 2001, Lee et al., 2001, Bödvarsson et al. 2007)
- Sound recording (delays between source and sensor due to propagation delay)
Generative model
Notation and LS-objective

- $\mathbf{U}$ and $\tilde{\mathbf{U}}$ denotes same matrix in time and frequency domain respectively.

- $\tilde{\mathbf{U}}^H$ denotes the conjugate transpose of $\tilde{\mathbf{U}}$

- $\mathbf{U}_{d,m-\tau} \sim \tilde{\mathbf{U}}_{d,f} e^{-i2\pi \frac{f-1}{M} \tau}$

- $\mathbf{U} \bullet \mathbf{V}$ denotes the direct product, i.e. element-wise multiplication.

- $(e^{-i2\pi \frac{f-1}{M} \tau})_{n,d} = e^{-i2\pi \frac{f-1}{M} \tau_{n,d}}$

- $\tilde{\mathbf{U}}(f) = \mathbf{U} \bullet e^{-i2\pi \frac{f-1}{M} \tau}$

- $\mathbf{U}_d$ $d^{th}$ column, $\mathbf{U}_{n,:}n^{th}$ row and $\mathbf{U}_{n,d}$ a given element of $\mathbf{U}$.

\[
C_{LS}(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_{n,m} (\mathbf{V}_{n,m} - \sum_d \mathbf{W}_{n,d} \mathbf{H}_{d,m-\tau_{n,d}})^2
\]

\[
= \frac{1}{2M} \| \tilde{\mathbf{V}}_f - (\mathbf{W} \bullet e^{-i2\pi \frac{f-1}{M} \tau}) \tilde{\mathbf{H}}_f \|^2_F = \frac{1}{2M} \| \tilde{\mathbf{V}}_f - \mathbf{W}(f) \tilde{\mathbf{H}}_f \|^2_F
\]

Follows from Parsevals identity with the above notation
**W update**

Let \( \tilde{H}^{(n)}_{d,f} = \tilde{H}_{d,f} e^{-i2\pi \frac{f-1}{M} \tau_{n,d}} \) denote the delayed version of the source signal \( \tilde{H}_{d,f} \) to the \( n \)th channel. The shift ICA/NMF model can then be stated as

\[
V_{n,:} = W_{n,:} H^{(n)} + E_{n,:}.
\]

This is the regular ICA/NMF problem which can be solved by the least squares ICA/NMF-update

\[
\text{SISA : } W_{n,:) = V_{n,:) / (H^{(n)} H^{(n)T}),}
\]

\[
\text{NMF : } W_{n,d} = W_{n,:} \frac{V_{n,:} H^{(n)T}_{d,:}}{W_{n,:} H^{(n)} H^{(n)T}_{d,:}}.
\]
**H update**

\[
C_{LS} = \frac{1}{2M} \| \tilde{V}_f - \tilde{W}^{(f)} \tilde{H}_f \|_F^2
\]

\[
G_f = \frac{\partial C_{LS}}{\partial \tilde{H}_f} = -\frac{1}{M} \tilde{W}^{(f)H} (\tilde{X}_f - \tilde{W}^{(f)} \tilde{H}_f)
\]

**ShiftNMF**

\[
\tilde{G}_f^+ = \frac{1}{M} \tilde{W}^{(f)H} \tilde{W}^{(f)} \tilde{H}_f
\]

\[
\tilde{G}_f^- = \frac{1}{M} \tilde{W}^{(f)H} \tilde{X}_f
\]

**SISA**

\[
G_f = (\tilde{W}^{(f)T} \tilde{W}^{(f)}) \backslash \tilde{X}_f
\]

\[
H_{d,n} = H_{d,n} \left( \frac{G_{d,n}^-}{G_{d,n}^+} \right) ^\alpha
\]
Update of the shifts ($\tau$)

$$C_{LS} = \frac{1}{2M} \sum_{f} (\tilde{V}_f - \langle \tilde{V} \rangle_f) (W \cdot e^{-i2\pi \frac{f-1}{M} \tau}) \tilde{H}_f$$

Warning!
Prone to local minima
Update of shifts ($\tau$) based on Cross-correlation

$$R_{n,m} = V_{n,m} - \sum_{d \neq d'} W_{n,d} H_{d,m-\tau_{n,d}}$$

$$C_{LS} = \frac{1}{2} \sum_{n,m} (V_{n,m} - \sum_{d} W_{n,d} H_{d,m-\tau_{n,d}})^2$$

$$= \frac{1}{2} \sum_{n,m} (R_{n,m} - W_{n,d'} H_{d',m-\tau_{n,d'}})^2$$

$$= \frac{1}{2} \|R\|^2 - \sum_{n} W_{n,d'} \sum_{m} R_{n,m} H_{d',m-\tau_{n,d'}} + \frac{1}{2} \|WH\|^2$$

Independent of $\tau$ Cross correlation $R$ and $H$ Independent of $\tau$

$$\tilde{c}_f = \tilde{R}_{n,f}^* \tilde{H}_{d',f}$$

$$t = \arg \max_m c_m, \quad \tau_{n,d'} = t - (M + 1).$$

The value of $W_{n,d'}$ corresponding to this delay is given by

$$W_{n,d'} = \frac{c_t}{H_{d'} : H_{d'}^T}.$$
Shift Invariant Subspace Analysis

Simulated Factors

True $A_1$

True $S_1$

True $T_1$

Estimated Factors (SISA)

Amb. Est. $A_1$

Amb. Est. $S_1$

Amb. Est. $T_1$

Amb. Est. $A_2$

Amb. Est. $S_2$

Amb. Est. $T_2$

Amb. Est. $A_3$

Amb. Est. $S_3$

Amb. Est. $T_3$
Shifted Independent Component Analysis

Define, $\tilde{U}_f = \tilde{Q}(f)\tilde{H}_f$, i.e. the sources at frequency $f$ when transformed according to the rotation and shift ambiguity described in the previous section. The ambiguity can be resolved by maximizing the log-likelihood assuming the (non-gaussian) Laplace distribution $p(\tilde{U}_f) \propto e^{-|\tilde{U}_{d,f}|}$, i.e.

$$p(\tilde{H}_f|Q, \tilde{\tau}) = \prod_f p(\tilde{H}_f|Q, \tilde{\tau}) = \prod_f |\text{det}(\tilde{Q}(f))|p(\tilde{Q}(f)\tilde{H}_f)$$ (1)

Such that the log-likelihood as a function of $Q$ and $\tilde{\tau}$ becomes

$$\mathcal{L}(Q, \tilde{\tau}) = \sum_f \ln |\text{det}(\tilde{Q}(f))| - \sum_d |\tilde{Q}(f)\tilde{H}_f|_d$$ (2)

By maximizing $\mathcal{L}(W, \tilde{\tau})$ $W$ and $\tilde{\tau}$ is estimated and a new unambiguous $H$ solution found by $\tilde{H}_f = \tilde{Q}(f)\tilde{H}_f$. The corresponding mixing and delays can be estimated alternating between the $W$ and $\tau$ update. We initialized $W$ as $W = WQ^{-1}$ and $\tau_{i,d}$ by the cross-correlation procedure.
Shifted Independent Component Analysis

Simulated Factors

Estimated Factors (SICA)
Algorithm assumptions

- **Sources** \( H \) and measured signal \( V \) have to be periodic
  If signals are not periodic a window function can be employed. However this is not trivial to implement in the \( \tau \)-update and slows down the algorithm significantly. Zero padding is simple and fast but introduces a bias towards small delays.

- **Noise** \( E \) assumed homoscedetic (normal) iid.
  If non-homoscedatic use weighted least squares. Algorithm works for Least squares due to Parseval’s identity. No such identity exists for other types than the least squares objective.
Extensions to tensors

Factor Analysis

\[ W_d \]

PARAFAC

\[ A^{(1)}_d \]

\[ \sum_d \]

\[ A^{(3)}_d \]

\[ V_{i_1 i_2} \approx \sum_{d=1}^{D} W_{i_1 d} H_{i_2 d} \]

\[ V_{i_1 i_2 i_3} \approx \sum_{d=1}^{D} A^{(1)}_{i_1 d} A^{(3)}_d \]

Not Unique

Unique
The Candecomp/PARAFAC (CP) model

\[ x_{i,j,k} = \sum_{d} a_{i,d} b_{j,d} c_{k,d} + \varepsilon_{i,j,k} \]

\[ x_{(1)} = a (c \odot b)^T + e_{(1)} \quad \Rightarrow \quad a \leftarrow x_{(1)} (c \odot b)^{T\dagger} \]

\[ x_{(2)} = b (c \odot a)^T + e_{(2)} \quad \Rightarrow \quad b \leftarrow x_{(2)} (c \odot a)^{T\dagger} \]

\[ x_{(3)} = c (b \odot a)^T + e_{(3)} \quad \Rightarrow \quad c \leftarrow x_{(3)} (b \odot a)^{T\dagger} \]

\[ (c \odot b) = (c_1 \otimes b_1 \, c_2 \otimes b_2 \, \ldots \, c_D \otimes b_D) \]

\[ x_{(1)} = x_{1 \times J K}, \quad x_{(2)} = x_{J \times I K}, \quad x_{(3)} = x_{K \times I J} \]

The CP model is unique if

\[ k_a + k_b + k_c \geq 2D + 2 \]

where \( k_A \) is the k-rank denoting the smallest subset of columns of \( A \) that is guaranteed to be linearly independent. Thus, \( k_A \leq \text{rank}(A) \).
Shifted CP model

\[ x_{i,j,k} = \sum_{d} A_{i,d} B_{j-\tau_{k,d}} C_{k,d} + \varepsilon_{i,j,k} \]

\[ X_{(1)} = AZ^T + E_{(1)} \quad \Rightarrow \quad A \leftarrow X_{(1)} Z^T \]

\[ \tilde{X}_{(2),f} = \tilde{B}_{f} : (\tilde{C}(f) \odot A)^T + \tilde{E}_{(2),f} \quad \Rightarrow \quad \tilde{B}_{f} : \leftarrow \tilde{X}_{(2),f} : (\tilde{C}(f) \odot A)^T \]

\[ X_{(3),k} = C_{k} : (B(k) \odot A)^T + E_{(3),k} \quad \Rightarrow \quad C_{k} : \leftarrow X_{(3),k} : (B(k) \odot A)^T \]

\[ r_{j}^{(k,d')} = \sum_{i} R_{i,j,k} A_{i,d'} \quad \tilde{c}_{k,d'}(f) = r_{f}^{(k,d')}^{*} \tilde{B}_{f,d'} . \]

\[ t_{k,d'} = \arg \max_{t} |c_{k,d'}(t)| \quad \tau_{k,d'} = t_{k,d'} - (J + 1) . \]

\[ C_{k,d'} = \frac{c_{k,d'}(t_{k,d'})}{B_{d'}^T B_{d'}} . \]
Synthetic EEG data

True components

Est. comp. PARAFAC

Est. comp. SPARAFAC

Informatics and Mathematical Modelling / Intelligent Signal Processing

Berkeley 2007
EEG data from Visual Paradigm

Shifted CP

CP
True Evoked Potential (EP)

Reconstructed EP component 18 and 20
fMRI data visual paradigm

\[ x_{i,j,k} = \sum_d A_{i,d} B_{j-d} \cdot C_{k,d} + \epsilon_{i,j,k} \]
Shift invariant sparse coding
(Solving the shift problem using sparse coding)

\[ X(x, y) \approx \sum_d \sum_{u, v} \alpha_d(u, v) \Phi_d(x - u, y - v) \]

Non-negative matrix 2D de-convolution introduced in:
Schmidt and Mørup, 2006d
Extended to Sparse coding
Mørup et al. 2007b
Informatics and Mathematical Modelling / Intelligent Signal Processing

Objective function:

\[
\frac{1}{2} \sum_{c,x,y} (X_c(x, y) - \sum_{c,d} s_{c,d} \sum_{u,v} \alpha_d(u, v) \Phi_d(x - u, y - v))^2 + \beta \sum_{d,x,y} \Phi_d(x, y)
\]

Solved by multiplicative updates (i.e., \(X, s, \alpha, \Phi \geq 0\))

\[
I_c(u, v) = s_{c,d} \alpha_d(u, v)
\]

\[
\Phi_d(x - u, y - v)
\]
Analysis of mono signal of mixed Organ and Piccolo

\[ X(x, y) \approx \sum_{d,u,v} \alpha_d(u, v) \Psi_d(x - u, y - v) \]

(Music data taken from Y.-G. Zhang, 2005)

Analysis of mono piano music

Organ

Piccolo

Organ and Piccolo

L-curve

Estimated Organ

Estimated Piccolo
**Conclusion:**
Modelling data using shift invariance seems highly relevant for a wide range of data types.

Two approaches presented: 1) estimation of specific shift in the frequency domain 2) Estimation of shifts using sparse coding.

Degeneracies encountered in the Candecomp/PARAFAC (CP) model appear to vanish when allowing for shifts. Thus degeneracy in CP often a result of sources of the data being shifted.

References

Carroll, J. D. and Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, Psychometrika 35 1970 283—319
Donoho, D. and Stodden, V. When does non-negative matrix factorization give a correct decomposition into parts? NIPS2003
Donoho D. For most large underdetermined systems of linear equations the minimal \( l_1 \)-norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics, 59(6):797829, 2006.
Egbert, J. and Kermer, E. Sparse coding and NMF. In Neural Networks volume 4, pages 2529-2533, 2004
Lathauwer, Lieven De and Moor, Bart De and Vandewalle, Joos MULTILINEAR SINGULAR VALUE DECOMPOSITION.SIAM J. MATRIX ANAL. APPL.2000 (21)1253–1278
Lee, D.D and Seung, H.S. Learning the parts of objects by non-negative matrix factorization, NATURE 1999
Morup, M., Hansen, L. K., Parnes, Josef, Hermann, C., Arnfred, S. M., Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG Neuroimage NeuroImage 29 938 – 947, 2006a
Norrgaard, L and Riddler, C.Rank unification factor analysis applied to flow injection analysis with photodiode-array detection Chemometrics and Intelligent Laboratory Systems 1994 (23) 107-114
Smardegis, P. Non-negative Matrix Factor deconvolution: Extraction of multiple sound sources from monophonic inputs. International Symposium on independent Component Analysis and Blind Source Separation (ICA)
Smilde, Age K. Smilde and Tauller, Roma and Saurina, Javier and Bro, Rasmus, Calibration methods for complex second-order data Analytica Chimica Acta 1999 237-251
Tamaras G. Kolda Multilinear operators for higher-order decompositions technical report Sandia national laboratory 2006 SAND2006-2081.
Tucker, L. R. Some mathematical notes on three-mode factor analysis Psychometrika 31 1966 279—311
References cont.


