Multi-frame Full-rank Spatial Covariance Analysis for Underdetermined BSS in Reverberant Environment

Hiroshi Sawada, Rintaro Ikeshita, Keisuke Kinoshita, Tomohiro Nakatani

NTT Communication Science Laboratories, NTT Corporation
Background

- BSS: Blind Source Separation

\[ y_1 \approx s_1 \]
\[ y_2 \approx s_2 \]

<table>
<thead>
<tr>
<th></th>
<th>overdetermined</th>
<th>underdetermined</th>
</tr>
</thead>
<tbody>
<tr>
<td>low reverberant</td>
<td>ICA</td>
<td>FCA</td>
</tr>
<tr>
<td>high reverberant</td>
<td>WPE</td>
<td></td>
</tr>
</tbody>
</table>

- ICA: Independent Component Analysis
- FCA: Full-rank spatial Covariance Analysis
- WPE: Weighted Prediction Error

\{ \text{BSS methods} \} \quad \text{a blind dereverberation method}
FCA model

- **Observation vector**
  - sum of source components
  \[ x_t = \sum_{n=1}^{N} c_{nt} \]
  \[ c_{nt} = \begin{bmatrix} c_{1nt} \\ \vdots \\ c_{Mnt} \end{bmatrix} \in \mathbb{C}^M \]

- **Source component vector**
  - follows a zero-mean Gaussian distribution
  \[ p(c_{nt}) = \mathcal{N}(c_{nt} | 0, C_{nt}) \]
  \[ C_{nt} = s_{nt} A_n \]

- **Parameters**
  \[ \theta = \{ \{ s_{nt} \}^T_{t=1}, A_n \}^N_{n=1} \]
### FCA and its Extensions

<table>
<thead>
<tr>
<th>Original</th>
<th>Conventional</th>
<th>Our proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FCA</strong></td>
<td><strong>FCAd</strong></td>
<td><strong>mfFCA</strong></td>
</tr>
<tr>
<td>[s_{n1} \quad s_{n2} \quad s_{n3} \quad s_{n4} \quad s_{n5}]</td>
<td>[s_{n1} \quad s_{n2} \quad s_{n3} \quad s_{n4} \quad s_{n5}]</td>
<td>[s_{n1} \quad s_{n2} \quad s_{n3} \quad s_{n4} \quad s_{n5}]</td>
</tr>
<tr>
<td>[\begin{pmatrix} x_1 \ x_2 \ x_3 \ x_4 \ x_5 \end{pmatrix}]</td>
<td>[\begin{pmatrix} x_1 \ x_2 \ x_3 \ x_4 \ x_5 \end{pmatrix}]</td>
<td>[\begin{pmatrix} x_1 \ x_2 \ x_3 \ x_4 \ x_5 \end{pmatrix}]</td>
</tr>
<tr>
<td>[s_{n3}]</td>
<td>[c_{n3}]</td>
<td>[c_{n3}]</td>
</tr>
<tr>
<td>[\begin{pmatrix} x_1 \ x_2 \ x_3 \ x_4 \ x_5 \end{pmatrix}]</td>
<td>[\begin{pmatrix} x_1 \ x_2 \ x_3 \ x_4 \ x_5 \end{pmatrix}]</td>
<td>[\begin{pmatrix} x_1 \ x_2 \ x_3 \ x_4 \ x_5 \end{pmatrix}]</td>
</tr>
</tbody>
</table>

- **FCA** components include delayed source components with time lags $\mathcal{L} = \{1, 2\}$.
- **FCAd** introduces multi-frame analysis.
- **mfFCA** extends the multi-frame approach for enhanced source separation.

### Formulas

- Original: $A_n$
- Conventional: $A_n^{(1)}$, $A_n^{(2)}$
- Our proposal: $A_n$, $A_n^{(0,1)}$, $A_n^{(0,2)}$, $A_n^{(1,0)}$, $A_n^{(1,2)}$, $A_n^{(2,0)}$, $A_n^{(2,1)}$, $A_n^{(2)}$
Proposed mfFCA model

- Multi-frame vectors
  \[ \bar{x}_t = \begin{bmatrix} x_t \\ x_{t+1} \\ x_{t+2} \end{bmatrix} \]
  \[ \text{observation} \]

- Source component vector \( \bar{c}_{nt} \)
  - zero-mean Gaussian distribution
  \[ p(\bar{c}_{nt}) = \mathcal{N}(\bar{c}_{nt} \mid 0, \bar{C}_{nt}) \]
  - covariance matrix has larger dimensionality
  \[ \bar{c}_{nt} = s_{nt} \bar{A}_n \]

- Observation vector \( \bar{x}_t \)
  - zero-mean Gaussian distribution
  \[ p(\bar{x}_t \mid \theta) = \mathcal{N}(\bar{x}_t \mid 0, \bar{X}_t) \]
  - covariance matrix has additional terms specific to mfFCA
  \[ \bar{X}_t = \begin{bmatrix} X_t \\ \cdots \\ X_{t+l_L} \end{bmatrix} + \sum_{n=1}^{N} \text{BoffDiag} \bar{c}_{nt} \]
mfFCA: EM algorithm

- For optimizing parameters
  \[ \theta = \{ \{ s_{nt} \}_{t=1}^T, \tilde{A}_n \}_{n=1}^N \]

- E-step
  - conditional distribution
    \[ p(\tilde{c}_{nt} | \bar{x}_t, \theta) = \mathcal{N}(\tilde{c}_{nt} | \mu_{nt}^{(\tilde{c})}, \Sigma_{nt}^{(\tilde{c})}) \]
    - mean vector
      \[ \mu_{nt}^{(\tilde{c})} = \tilde{C}_{nt} \tilde{X}_t^{-1} \bar{x}_t \]
    - covariance matrix
      \[ \Sigma_{nt}^{(\tilde{c})} = \tilde{C}_{nt} - \tilde{C}_{nt} \tilde{X}_t^{-1} \tilde{C}_{nt} \]

- M-step
  - optimize parameters
    \[
    \tilde{A}_n \leftarrow \frac{1}{T} \sum_{t=1}^T \frac{1}{s_{nt}} \tilde{C}_{nt} \\
    s_{nt} \leftarrow \frac{1}{M(L+1)} \text{tr} \left( \tilde{A}_n^{-1} \tilde{C}_{nt} \right)
    
    \text{with} \quad \tilde{C}_{nt} = \mu_{nt}^{(\tilde{c})} \mu_{nt}^{(\tilde{c})H} + \Sigma_{nt}^{(\tilde{c})}
    \]
Experiments

Conditions
- \( M = 3 \) microphones
- \( N = 4 \) sources
- 6-second speeches
- reverberation time: 130 ms to 450 ms

Separation performances
- measured in signal-to-distortion ratios (SDRs)
- did not aim for dereverberation (we used source images with reverberations at microphones as reference signals)
Overall results

5 methods

- **FCA**
  - original, baseline
- **FCAd**
  - conventional
    - $\mathcal{L} = \{2\}$
  - $\mathcal{L} = \{2, 4\}$
- **mfFCA**
  - our proposal
    - $\mathcal{L} = \{2\}$
    - $\mathcal{L} = \{2, 4\}$

mfFCA outperformed the original FCA by around 2 dB
Convergence behavior

Low reverberant case

High reverberant case

- FCA parameters were initialized by the procedure shown in [29].
- The first 5 iterations were by the original FCA model and updates.
Conclusion

- **A new FCA model**  
  - source components span multiple time frames  
  - modeled with covariance matrix of larger dimensionality

\[ p(\tilde{c}_{nt}) = \mathcal{N}(\tilde{c}_{nt} | 0, \tilde{C}_{nt}) \]

\[ \tilde{c}_{nt} = s_{nt}\tilde{A}_{n} \]

- **Developed**  
  - the whole probabilistic models and EM algorithm

- **Experimental results**  
  - show that the proposed method considerably improved the separation performance for underdetermined reverberant convolutive mixtures

- **Future work**  
  - evaluating the dereverberation capability of mfFCA  
  - reducing the computational complexity further (we have already accelerated the algorithm computation by a GPU)