Speech Enhancement using
Polynomial Eigenvalue Decomposition

Vincent W. Neo, Christine Evers, Patrick A. Naylor
21 October 2019
Introduction
Motivation

- Single-channel subspace speech enhancement [Ephraim1995; Hu2002]
  - Use an EVD to decorrelate spectrally
- Multi-channel subspace speech enhancement [Asano2000]
  - Use an EVD to decorrelate spatially

⇒ Limitation: Only decorrelates instantaneously

- Other methods typically use STFT to process [Cohen2002; Ephraim1984; Gannot2001; Markovich2009]
  - Use DFT to divide broadband into multiple narrowband signals
  - Require a 4D tensor to model the space, time, spectral correlations

⇒ Limitations: Lacks phase coherence across bands
  : Ignores correlation between bands
Motivation for PEVD

- Polynomial Eigenvalue Decomposition (PEVD)
  - Simultaneously captures correlation across space, time and frequency using a 3D tensor
  - Impose spatial decorrelation over a range of time shifts
  - No phase discontinuity

- PEVD-based broadband applications:
  - blind source separation [Redif2017]
  - adaptive beamforming [Weiss2015]
  - source identification [Weiss2017]

This Talk: PEVD for Speech Enhancement
Background
The received signal at the $q$-th sensor with time index $n$ is

$$x_q(n) = \sum_{j=0}^{J} h_q(n - j)s(j) + v_q(n),$$

where

- $s(n)$ is the source signal,
- $h_q(n)$ is the channel modelled as an order $J$ FIR filter,
- $v_q(n)$ is the noise signal at the $q$-th sensor.

The data vector collected from $Q$ sensors is

$$\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_Q(n)]^T.$$
Assuming stationarity, space-time covariance matrix is

\[ R_{xx}(\tau) = \mathbb{E}[x(n)x^H(n - \tau)], \]

where \((i, j)^{th}\) element is the correlation function \(r_{ij}(\tau) = \mathbb{E}[x_i(n)x_j^*(n - \tau)]\) and \(\tau\) is the time-shift.

Z-transform of \(R_{xx}(\tau)\) is a para-Hermitian polynomial matrix

\[ \mathcal{R}_{xx}(z) = \sum_{\tau=-W}^{W} R_{xx}(\tau) z^{-\tau}, \]

where \(R_{xx}(\tau) \approx 0\) for \(|\tau| > W\), calligraphy \(\mathcal{R}\) for tensor and regular \(R\) for matrix.
The PEVD of $R_{xx}(z)$ is defined as [McWhirter 2007]

$$R_{xx}(z) \approx U^P(z) \Lambda(z) U(z) \iff \Lambda(z) \approx U(z) R_{xx}(z) U^P(z),$$

where $\Lambda(z), U(z)$ are the eigenvalue and eigenvector polynomial matrices and $R_{xx}(z) = R_{xx}^P(z) = R_{xx}^H(z^{-1})$.

$U(z)$ is a filterbank for $x(z) \in \mathbb{C}^{Q \times 1 \times T}$ so that the outputs in

$$y(z) = U(z)x(z) \implies R_{yy}(z) \approx \Lambda(z),$$

are strongly decorrelated.
PEVD algorithms include:

- Second-order Sequential Best Rotation (SBR2) [McWhirter2007]
- Sequential Matrix Diagonalization (SMD) [Redif2015]
- Householder-like PEVD [Redif2011]
- Tridiagonal PEVD [Neo2019]
- Multiple-shift SBR2/SMD [Wang2015; Corr2014]
Example of a Polynomial Matrix

Typically compute $\mathbf{R}_{xx}(0) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^H(n)]:$

$\mathbf{R}_{xx}(0)$: instantaneous spatial covariance matrix / coefficient of $z^0$. 

\[
\begin{bmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & -1 \\
0 & 0 & -1 \\
\end{bmatrix}
\]
Example of a Polynomial Matrix

Before diagonalization, $\mathbf{R}_{xx}(z)$:

In this example, $z^0$ plane is diagonal but not at other planes.
Example of a Polynomial Matrix

After diagonalization using PEVD, $\Lambda(z)$:

\[
\begin{align*}
\text{\tiny \(z^{-2}\)} & : & \begin{bmatrix}
-0.03 & 0 & 0 \\
0 & 0.06 & 0 \\
0 & 0 & -0.02
\end{bmatrix} \\
\text{\tiny \(z^{-1}\)} & : & \begin{bmatrix}
13 & 0 & 0 \\
0 & 0.06 & 0 \\
0 & 0 & 0.07
\end{bmatrix} \\
\text{\tiny \(z^0\)} & : & \begin{bmatrix}
2.01 & 0 & 0 \\
0 & 0.36 & 0 \\
0 & 0 & -1.36
\end{bmatrix} \\
\text{\tiny \(z^1\)} & : & \begin{bmatrix}
-0.13 & 0 & 0 \\
0 & 0.06 & 0 \\
0 & 0 & 0.07
\end{bmatrix} \\
\text{\tiny \(z^2\)} & : & \begin{bmatrix}
-0.03 & 0 & 0 \\
0 & 0.06 & 0 \\
0 & 0 & -0.02
\end{bmatrix}
\end{align*}
\]
Alternate Representation of Example

Equivalently, expressed as:

Polynomial with matrix coefficients. Matrix with polynomial elements.
The same example can be represented as:

Original $\mathcal{R}_{xx}(z)$.

Diagonalized $\Lambda(z)$. 

Speech Enhancement using PEVD - 15 / 38
Application Examples
A rectangular pulse source signal arriving at the 3 sensors, corrupted by i.i.d. sensor noise: $\mathcal{N}(0, 0.1^2)$.

Source signal, $s(n)$.

Received signals, $x(n)$.
Broadband Example: ST-Covariance

Corresponding space-time covariance matrix, $\mathcal{R}_{xx}(z)$

- instantaneous covariance, $\mathbb{E}[x(n)x^H(n)]$, marked in red.
Using $\mathbf{U}$ from EVD gives:

**Signal, $y(n)$**

- $y_1(n)$
- $y_2(n)$
- $y_3(n)$

**Amplitude**

- Time sample, $n$

**Weighted output, $y(n)$**

**ST-covariance, $\mathcal{R}_{yy}(z)$**

**Coefficients of $z$**

- Powers of $z$
Diagonalization using PEVD with $\delta = 0.0077$ gives:

Iter. count=0, Max. off-diagonal, $|g|=0.899$
Using $\mathcal{U}(z)$ from PEVD using $\delta = 0.0077$ gives:

Weighted output, $y(n)$, with arbitrary delays compensated.

ST-covariance, $\mathcal{R}_{yy}(z)$. 
Proposed Methodology
If \( s(n) \) is a speech signal, uncorrelated with noise

\[
\mathbf{R}_{xx}(z) = \begin{bmatrix}
\mathbf{U}_S^P(z) & \mathbf{U}_V^P(z)
\end{bmatrix}
\begin{bmatrix}
\mathbf{Λ}_S(z) & 0 \\
0 & \mathbf{Λ}_V(z)
\end{bmatrix}
\begin{bmatrix}
\mathbf{U}_S(z) \\
\mathbf{U}_V(z)
\end{bmatrix}
\]

with orthogonal signal, \( \{\cdot\}_S \) and noise subspaces, \( \{\cdot\}_V \).

The output

\[ y(z) = \mathbf{U}(z)x(z), \]

has the first element, \( y_1(z) \in \mathbb{R}^{1 \times 1 \times T} \), as the denoised speech signal with space-time covariance matrix

\[
\mathbf{R}_{y_1y_1} = \begin{bmatrix}
\mathbf{U}_S^P(z) & 0
\end{bmatrix}
\begin{bmatrix}
\mathbf{Λ}_S(z) & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\mathbf{U}_S(z) \\
0
\end{bmatrix}
\]
Speech Enhancement Using PEVD

\[ \mathbf{x}(n) \rightarrow \text{Form space-time cov} \rightarrow \mathbf{R}_{xx}(z) \rightarrow \text{PEVD} \rightarrow \mathbf{U}(z) \rightarrow \mathbf{y}(n) \]

Input

ST-covariance

Eigenvalues

Enhanced output
Experimental Results
Speech in Noise (Anechoic)

diffuse babble
5 dB SNR

TIMIT speech
Comparative algorithms:

1. Log-Minimum Mean Square Error (Log-MMSE) [Ephraim 1984]
2. Multichannel Wiener Filter (MWF) - Relative Transfer Function (RTF) and noise estimator [Kuklasiński 2016]
3. Oracle-MWF (O-MWF) - Given clean speech [Doclo 2002]

Evaluation measures:

- Segmental SNR (SegSNR)
- Frequency weighted SegSNR (fwSegSNR) [Hu 2008]
- STOI [Taal 2011]
- PESQ [ITU-T P.862]
Clean Spectrogram

Clean speech signal, \( s(n) \)

Clean | Noisy | Log-MMSE | PEVD
Log-MMSE-Enhanced Spectrogram

Log-MMSE enhanced signal, $y_1(n)$

-clean-
-noisy-
-log-mmse-
-pevd-

Speech Enhancement using PEVD - 30 / 38
PEVD-Enhanced Spectrogram

PEVD enhanced signal, $y_1(n)$

Clean | Noisy | Log-MMSE | PEVD
## Comparison of Enhancement Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>ΔSegSNR</th>
<th>ΔfwSegSNR</th>
<th>ΔSTOI</th>
<th>ΔPESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-MMSE</td>
<td>3.69 dB</td>
<td>2.46 dB</td>
<td>-0.007</td>
<td>0.08</td>
</tr>
<tr>
<td>MWF</td>
<td>1.07 dB</td>
<td>1.54 dB</td>
<td>0.002</td>
<td>0.15</td>
</tr>
<tr>
<td>O-MWF</td>
<td>4.67 dB</td>
<td>4.04 dB</td>
<td>0.084</td>
<td>0.31</td>
</tr>
<tr>
<td>PEVD</td>
<td>4.30 dB</td>
<td>4.00 dB</td>
<td>0.080</td>
<td>0.29</td>
</tr>
</tbody>
</table>

### Clean, Noisy, log-MMSE, MWF, O-MWF, PEVD
Conclusion
Conclusion

- Polynomial covariance matrices and PEVD as a tool for processing broadband multichannel signals
  - Polynomial matrices can simultaneously capture the correlation across space, time and frequency
  - PEVD can impose stronger decorrelation than the EVD

- Proposed a speech enhancement algorithm using PEVD
  - Performance approaches oracle MWF when \( \text{SNR} \geq 5 \text{ dB} \)
  - No noticeable artifacts


Thank you
Figure Mean of the results for babble noise involving 150 trials.