Speech Enhancement using Polynomial Eigenvalue Decomposition

by Vincent W. Neo, Christine Evers, Patrick A. Naylor

21 October 2019
1. Introduction
   Motivation for PEVD

2. Background
   Multichannel Signal Model
   Polynomial Matrices
   PEVD

3. Application Examples
   Broadband Example
   Speech Enhancement Example

4. Conclusion
Introduction
Motivation

- Single-channel subspace speech enhancement [Ephraim 1995; Hu 2002]
  - Use an EVD to decorrelate spectrally
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 does so instantaneously
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 does so 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

- PEVD for speech enhancement
  - Simultaneously captures correlation across space, time and frequency using a 3D tensor
  - Impose spatial decorrelation over a range of time shifts
  - No phase discontinuity
Motivation for PEVD

- PEVD for speech enhancement
  - 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]
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 for time-shift $\tau$ is

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

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

$Z$-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 outside $\pm W$, the function becomes negligibly small and $\mathcal{R}_{xx}(z) = \mathcal{R}_{xx}^p(z) = \mathcal{R}_{xx}^H(z^{-1})$. 

Speech Enhancement using PEVD
The PEVD of $R_{xx}(z)$ [McWhirter2007] is defined as

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

where $\Lambda(z), U(z)$ are the eigenvalue and eigenvector polynomial matrices.
The PEVD of $R_{xx}(z)$ [McWhirter2007] is defined as
\[
R_{xx}(z) \approx U^P(z) \Lambda(z) U(z) \Leftrightarrow \Lambda(z) \approx U(z) R_{xx}(z) U^P(z),
\]
where $\Lambda(z), U(z)$ are the eigenvalue and eigenvector polynomial matrices.

$U(z)$ can be interpreted as a filterbank so that
\[
y(z) = U(z)x(z) \Rightarrow R_{yy}(z) \approx \Lambda(z),
\]
indicating that the outputs, $y(z)$ 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

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

\[
\begin{pmatrix}
0 & 0 & .5 \\
.8 & 0 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}, \quad
\begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
-4 & 0 & 0 \\
.7 & 0 & 0 \\
\end{pmatrix}, \quad
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1 \\
\end{pmatrix}, \quad
\begin{pmatrix}
0 & -.4 & .7 \\
0 & 0 & 0 \\
.8 & 0 & 0 \\
0 & 0 & .5 \\
\end{pmatrix}
\]
Example of a Polynomial Matrix

After diagonalization using PEVD, $\Lambda(z)$:
Alternate Representation of Example

Equivalently, shown as:

Original $R_{xx}(z)$.

Diagonalized $Λ(z)$. 

Speech Enhancement using PEVD - 13 / 34
Application Examples
A rectangular pulse source signal arriving at the 3 sensors, corrupted by i.i.d. sensor noise: $\mathcal{N}(0, 0.1^2)$. 

![Signal, x(n) graphs](image-url)
Broadband Example: ST-Covariance

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

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

- **Signal, $y(n)$**
- **Amplitude**
  - $y_1(n)$
  - $y_2(n)$
  - $y_3(n)$
  - Time sample, $n$

- **Coefficients of $z$**
- **Powers of $z$**

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

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

Diagonalization using PEVD with $\delta = 0.0077$ gives:

Iter. count=0, Max. off-diagonal, $|g|=0.899$
Broadband Example: PEVD

Using $\mathcal{U}(z)$ from PEVD using $\delta = 0.004$ gives:

Weighted output, $y(n)$.

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

Speech Enhancement using PEVD - 19 / 34
If \( s(n) \) is a speech signal, uncorrelated with noise

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

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

The output

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

has the first element, \( y_1(z) \), as the denoised speech signal with space-time covariance matrix

\[
\mathbf{R}_{y_1y_1} = \begin{bmatrix}
\mathbf{U}_S^P(z) & 0
\end{bmatrix}
\begin{bmatrix}
\Lambda_S(z) & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\mathbf{U}_S(z) \\
0
\end{bmatrix}.
\]
Application: Speech Enhancement

Speech Enhancement using PEVD

\[ x(n) \rightarrow \text{Form ST-Cov} \rightarrow R_{xx}(z) \rightarrow \text{PEVD} \rightarrow \Lambda(z) \rightarrow U(z) \rightarrow \text{Eigenvector filterbank} \rightarrow y(n) \rightarrow \text{Enhanced Output} \]
Experiment Setup

Anechoic

diffuse babble
5 dB SNR

TIMIT speech
Evaluation

Comparative algorithms:
1. Log-Minimum Mean Square Error (Log-MMSE)
2. Multichannel Wiener Filter (MWF) - Relative Transfer Function (RTF) and noise estimator
3. Oracle-MWF (O-MWF) - given clean speech

Evaluation measures:
- SegSNR, fwSegSNR, STOI, PESQ
Clean Spectrogram

Clean speech signal, s(n)

Clean | Noisy | Log-MMSE | PEVD
Noisy Spectrogram (5 dB diffuse babble)

Speech Enhancement using PEVD
Log-MMSE-Enhanced Spectrogram

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

Time (s)
-0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8

Frequency (kHz)
-1 0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5

Clean | Noisy | Log-MMSE | PEVD

Speech Enhancement using PEVD - 26 / 34
PEVD-Enhanced Spectrogram

PEVD enhanced signal, $y_1(n)$

Time (s)

Frequency (kHz)

Clean | Noisy | Log-MMSE | PEVD

Speech Enhancement using PEVD - 27 / 34
## Comparison of Enhancement Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>(\Delta)SegSNR</th>
<th>(\Delta)fwSegSNR</th>
<th>(\Delta)STOI</th>
<th>(\Delta)PESQ</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](clean.png) ![Noisy](noisy.png) ![log-MMSE](log-MMSE.png) ![MWF](MWF.png) ![O-MWF](O-MWF.png) ![PEVD](PEVD.png)
Conclusion
Conclusion

- Polynomial covariance matrices and PEVD as a tool for processing broadband signals
- Eigenvector polynomial matrix produced by PEVD can be interpreted as a filterbank
  - Sensor signals passing through the filterbank produce strongly decorrelated outputs
- Subspace decomposition provided by PEVD can be used for speech enhancement
  - Performance approaches oracle MWF
  - No noticeable artifacts
References


Thank you