T2: Applications of Polynomial Eigenvalue Decomposition to Multichannel Broadband Signal Processing
Part 4: Speech Processing Applications

Vincent W. Neo
Soydan Redif
Stephan Weiss
Patrick A. Naylor

Monday, 04 Sep 2023

Outline

Imperial College London

Introduction

PEVD for Speech Processing Applications

Imperial College London

Speech signals captured by multiple microphones are increasingly common

- Broadband nature of speech
- Temporal correlations especially when there is reverberation

Processing approaches include

- Splitting broadband into multiple narrowband signals via FFT [Cohen2002; Markovich2009]

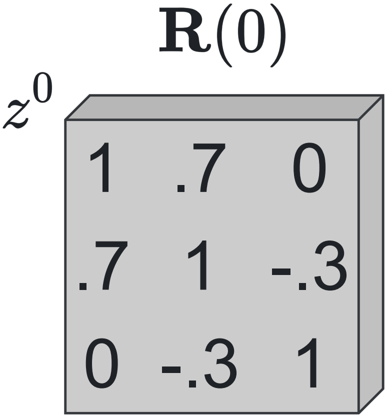
  ⇒ Ignores spectral coherence and correlations between frequency bands
- Classical subspace methods use the (instantaneous) spatial covariance matrix [Asano2000]

  ⇒ Inadequate for convolutively mixed and/or broadband signals

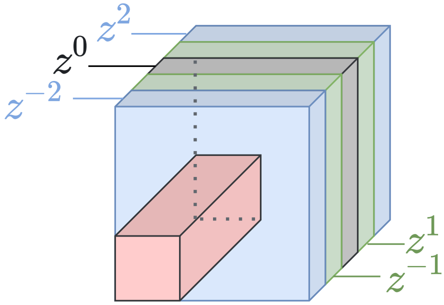
Promising Approach: Polynomial Matrix Eigenvalue Decomposition (PEVD)
⇒This Talk: Iterative Time-domain Algorithms e.g. SMD [Redif2015]

Recap: From EVD to PEVD

What Is A Polynomial Matrix?

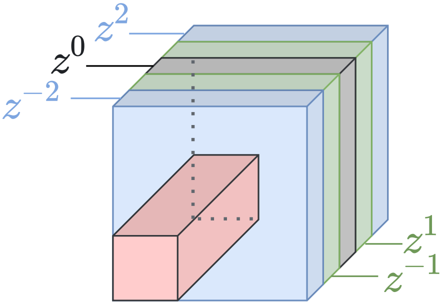


Matrix example.



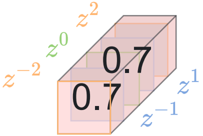
Polynomial matrix example.

What Is A Polynomial Matrix?



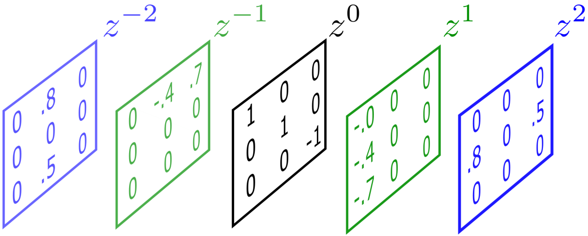
Polynomial with matrix coefficients.

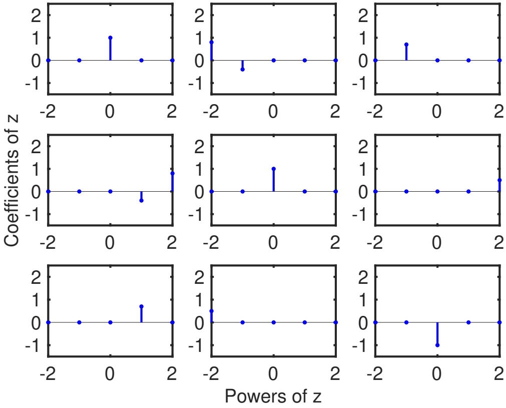
$$r_{31}(z) = 0.7z^{-1} + 0.7z^{1}$$



Matrix with polynomial elements.

What Is A Polynomial Matrix?



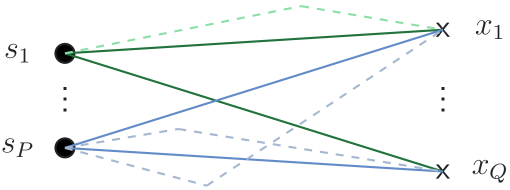


Polynomial with matrix coefficients.

Matrix with polynomial elements.

Multi-channel Signal Model

Consider the following MIMO system.



- Source signals: $s_1, \ldots, s_P$
- Received signals: $x_1, \ldots, x_Q$
- Multi-paths with different delays $\Rightarrow$ convolutive mixtures (polynomials in z)
- $h_{p,q}$: channel from pth source to qth sensor, J-th order FIR filter

Room Reverberation

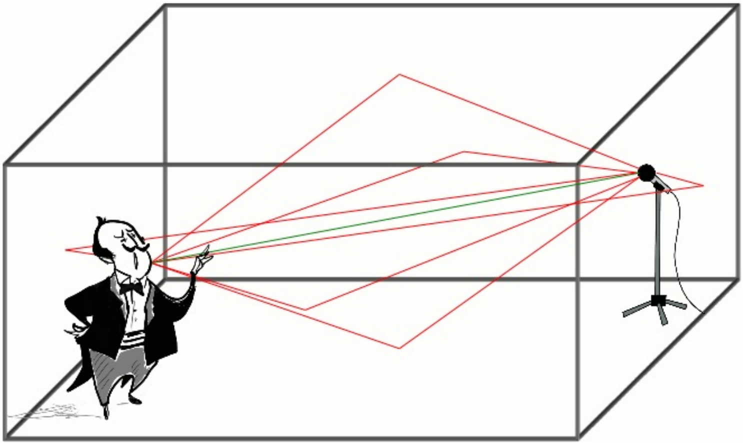
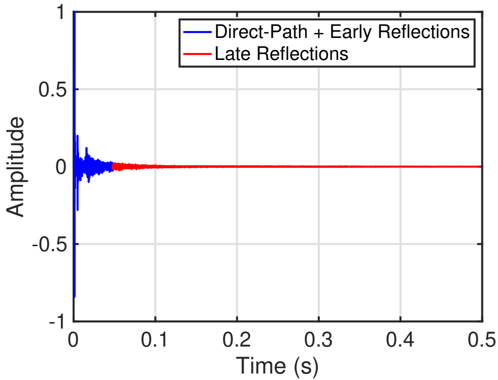


Figure is taken from the DREAMS project on the Imperial SAP website.

Reverberant Channel Model

The q-th channel modelled as a FIR filter: $\mathbf{h}_q = \mathbf{h}_{q,dp} + \mathbf{h}_{q,er} + \mathbf{h}_{q,lr}$



An example of a room impulse response.

Multichannel Reverberant Signal Model

For a single speaker, the received signal at the q-th sensor with time index n is

$$x_q(n) = \mathbf{h}_q^{\mathrm{T}} \mathbf{s}_0(n) + v_q(n) = \tilde{s}_q(n) + \tilde{v}_q(n)$$

where

- $\tilde{s}_q(n)$ is the speech associated with the direct-path and early reflections,
- $\tilde{v}_q(n)$ is the noise and late reflections,
- $\mathrm{s}_0(n)$ is the anechoic speech signal,
- $v_q(n)$ is the noise signal at the q-th sensor.

The data vector collected from Q sensors is

$$\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_Q(n)]^{\mathrm{T}} \in \mathbb{R}^Q.$$

Multi-channel Signal Model

Assuming stationarity, the space-time covariance matrix is

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau)]$$

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

Instantaneous spatial covariance matrix: $\mathbf{R}_{\mathbf{xx}}(0) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)]$ using $\tau = 0$.

The z-transform denoted by $\mathbf{R}_{\mathbf{xx}}(\tau) \circ\!\!-\!\!\bullet\ \mathcal{R}_{\mathbf{xx}}(z)$ is a para-Hermitian polynomial matrix

$$\mathcal{R}_{\mathbf{xx}}(z) = \sum_{\tau=-W}^{W} \mathbf{R}_{\mathbf{xx}}(\tau) z^{-\tau},$$

where $\mathbf{R}_{\mathbf{xx}}(\tau) \approx 0$ for $|\tau| > W$, calligraphy $\mathcal{R}$ for tensor and regular $\mathbf{R}$ for matrix.

Decorrelation Using PEVD vs EVD

Imperial College London

PEVD of $\mathcal{R}_{\mathbf{xx}}(z)$ is defined as [McWhirter2007; Redif2015; Neo2019b; Weiss2018]

$$\mathcal{R}_{\mathbf{xx}}(z) \approx \boldsymbol{\mathcal{U}}(z)\boldsymbol{\Lambda}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) \Leftrightarrow \boldsymbol{\Lambda}(z) \approx \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\mathcal{R}_{\mathbf{xx}}(z)\boldsymbol{\mathcal{U}}(z),$$

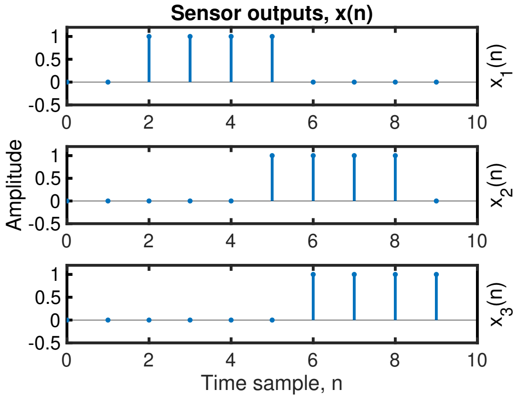
where $\boldsymbol{\Lambda}(z), \boldsymbol{\mathcal{U}}(z) \bullet\!-\!\circ\ \mathrm{U}(n)$ are the eigenvalue, eigenvector polynomial matrices respectively, $\mathcal{R}^{\mathrm{P}}_{\mathbf{xx}}(z) = \mathcal{R}^{\mathrm{H}}_{\mathbf{xx}}(1/z^*)$, and signals $\mathbf{y}(n) = \sum_{\nu} \mathrm{U}^{\mathrm{H}}(-\nu)\mathbf{x}(n-\nu)$ are strongly decorrelated.

EVD of $\mathbf{R}_{\mathbf{xx}}(0)$, the matrix coefficient of z^0, is:

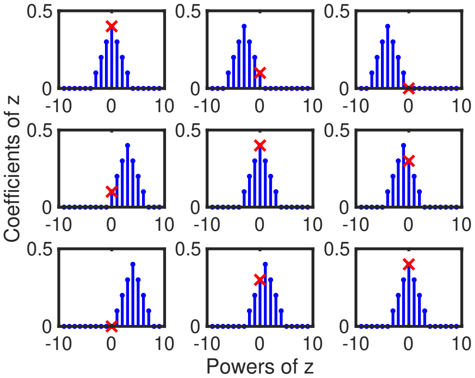
$$\mathbf{R}_{\mathbf{xx}}(0) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} \Leftrightarrow \boldsymbol{\Lambda} = \mathbf{U}^{\mathrm{H}}\mathbf{R}_{\mathbf{xx}}(0)\mathbf{U} = \mathbb{E}\{\mathbf{y}(n)\mathbf{y}^{\mathrm{H}}(n)\},$$

where signals $\mathbf{y}(n) = \mathbf{U}^{\mathrm{H}}\mathbf{x}(n)$ are instantaneously decorrelated.

Example: rectangular source arriving at the 3 sensors



Received signals $\mathbf{x}(n)$.



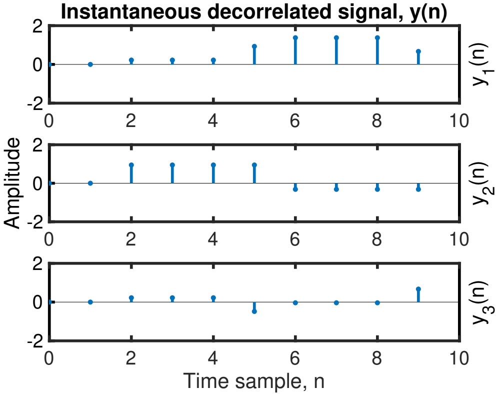
ST-covariance, $\mathcal{R}_{\mathbf{xx}}(z)$.

- instantaneous covariance marked by red crosses

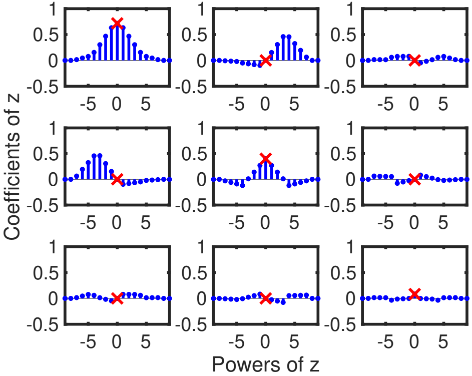
Rectangular Example: EVD

Imperial College London

Using $\mathbf{U}$ from EVD gives $\mathbf{y}(n)$ and its $\mathcal{R}_{\mathbf{yy}}(z)$:



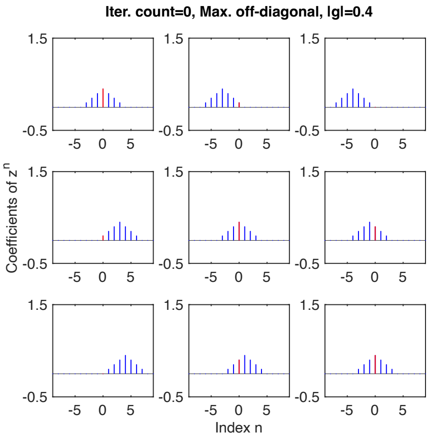
Weighted output, $\mathbf{y}(n)$.



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

Rectangular Example: PEVD Algorithm

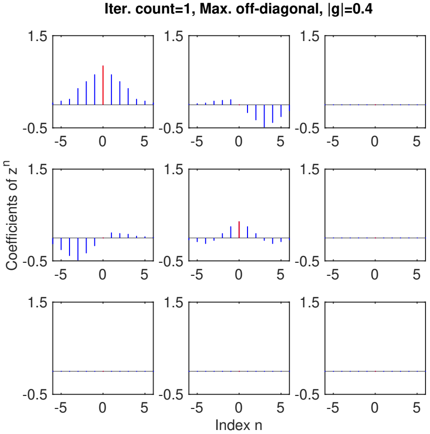
Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



Rectangular Example: PEVD Algorithm

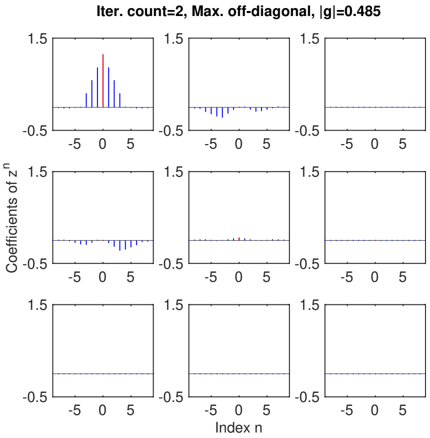
Imperial College London

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



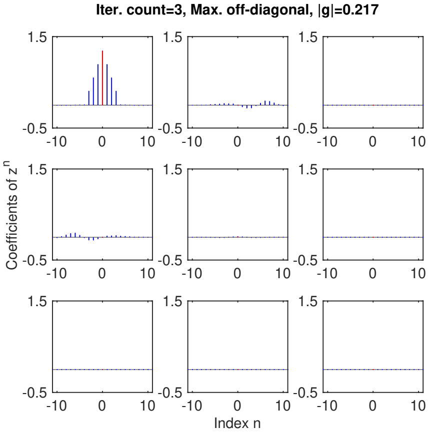
Rectangular Example: PEVD Algorithm

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



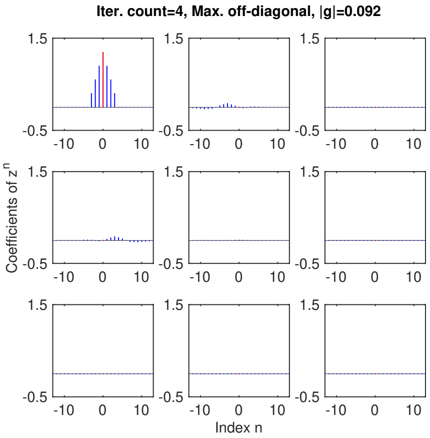
Rectangular Example: PEVD Algorithm

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



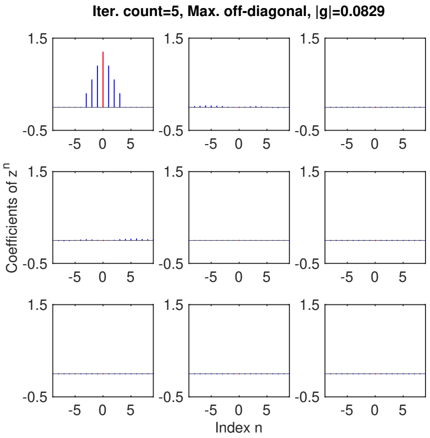
Rectangular Example: PEVD Algorithm

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



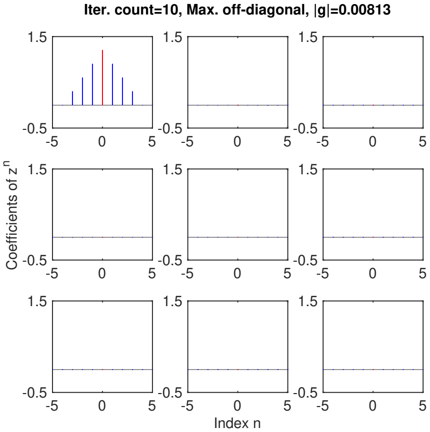
Rectangular Example: PEVD Algorithm

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



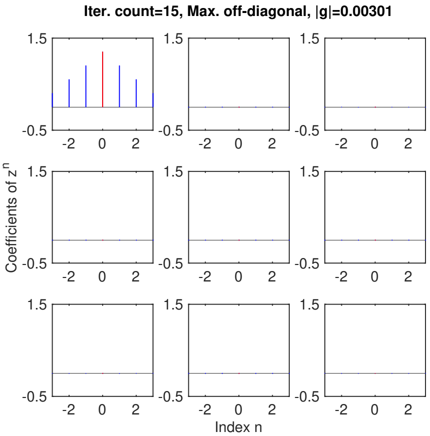
Rectangular Example: PEVD Algorithm

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives



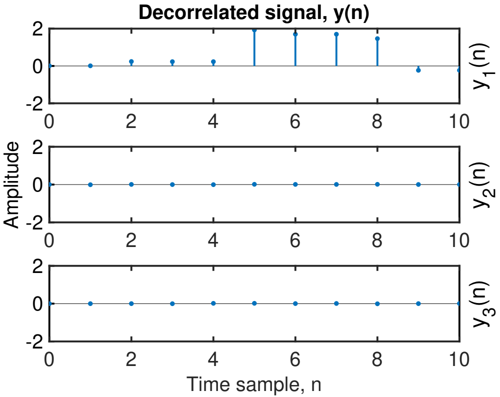
Rectangular Example: PEVD Algorithm

Performing PEVD using SMD: $\delta = 0.010, \mu = 10^{-4}, L = 500$ gives

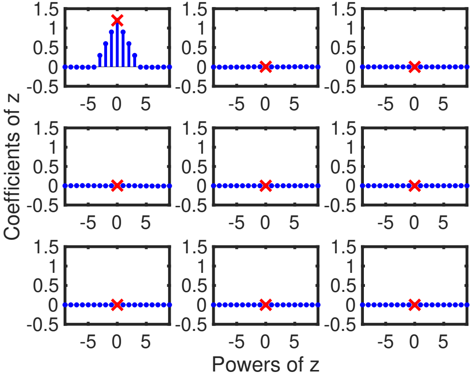


Rectangular Example: PEVD

Using $\mathcal{U}(z)$ from PEVD gives $\mathbf{y}(n)$ and its $\mathcal{R}_{\mathbf{yy}}(z)$:



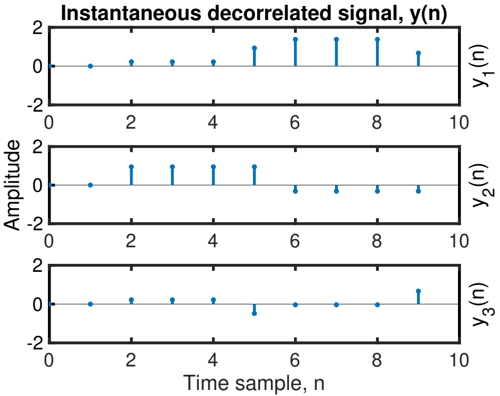
Weighted output, $\mathbf{y}(n)$.



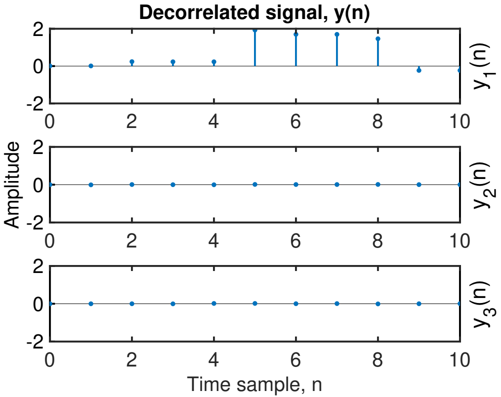
ST-covariance, $\mathcal{R}_{\mathbf{yy}}(z)$.

Rectangular Example: PEVD

Output signals are:



EVD outputs.

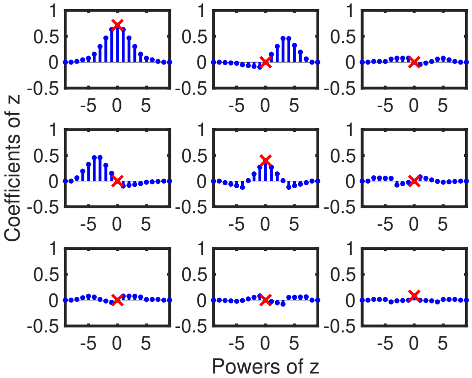


PEVD outputs.

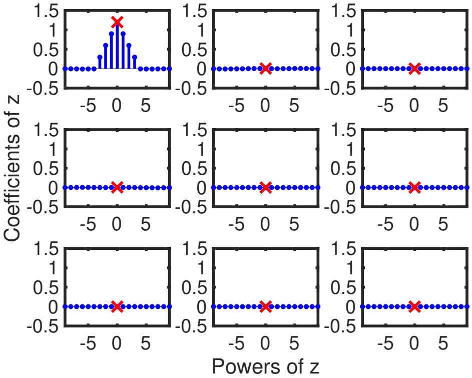
Rectangular Example: PEVD

Imperial College London

Space-time covariance matrix of the outputs:



Space-time covariance of EVD output.



Space-time covariance of PEVD output.

Summary of EVD vs PEVD

For the rectangular signal example,

- EVD: only decorrelate instantaneously $\rightarrow$ cannot recover rectangular
- PEVD: impose (strong) decorrelation over time shifts $\rightarrow$ recover rectangular

Interpretation

- EVD: Any function $=$ linear combination of the bases
- PEVD: Any function $=$ linear combination of the bases and its shifted versions

$\Rightarrow$ PEVD uses fewer number of basis for the same representation as EVD

$\Rightarrow$ PEVD achieves greater compression with fewer components

Key Ingredients: (i) Space-time Covariance Matrix; (ii) Strong decorrelation property of PEVD; (iii) Broadband steering vector

Speech Processing Applications Using PEVD

Imperial College London

I. Enhancement of Noisy Reverberant Speech
   V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition", in: *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 29, pp.3255-3266, Oct 2021.

II. Signal Compaction for Spherical Microphone Array Processing
   V. W. Neo, C. Evers, S. Weiss, and P. A. Naylor, "Signal compaction using polynomial EVD for spherical array processing with applications", in: *IEEE/ACM Trans. Audio, Speech, Lang. Process*, to appear.

III. Fixed Beamformer Design Using PEVD
   V. W. Neo, E. d'Olne, A. H. Moore, and P. A. Naylor, "Fixed beamformer design using polynomial eigenvalue decomposition", in: *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, pp.1-5, Sep 2022.

Application I. Enhancement of Noisy Reverberant Speech

Enhancement of Noisy Reverberant Speech

The subspace decomposed by PEVD gives

$$\mathcal{R}_{\mathbf{xx}}(z) = \left[\begin{array}{c|c} \boldsymbol{\mathcal{U}}_s(z) & \boldsymbol{\mathcal{U}}_v(z) \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_s(z) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Lambda}_v(z) \end{array}\right] \left[\begin{array}{c} \boldsymbol{\mathcal{U}}_s^{\mathrm{P}}(z) \\ \hline \boldsymbol{\mathcal{U}}_v^{\mathrm{P}}(z) \end{array}\right],$$

with orthogonal signal, $\{\cdot\}_s$ and noise subspaces, $\{\cdot\}_v$.

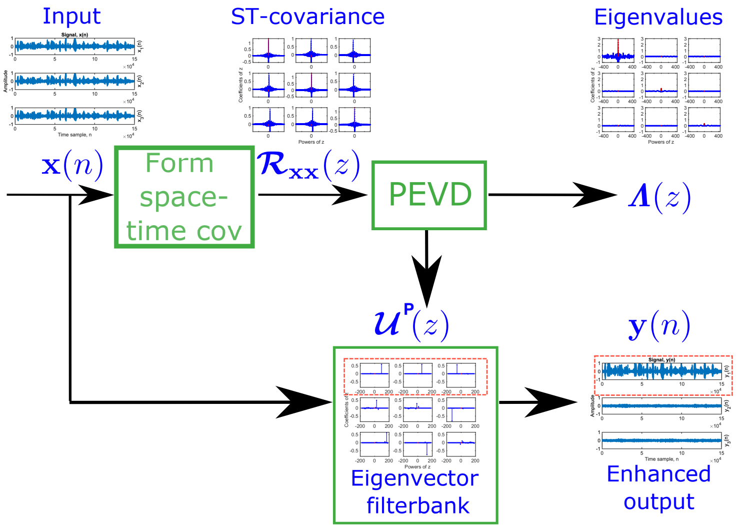
The strongly decorrelated output generated using $\boldsymbol{\mathcal{U}}(z) \bullet\!\!-\!\!\circ \mathbf{U}(n)$

$$\mathbf{y}(n) = \sum_{\nu} \mathbf{U}^{\mathrm{H}}(-\nu)\mathbf{x}(n-\nu),$$

has the first element, $y_1(n)$, as the denoised and enhanced speech signal with the space-time covariance matrix

$$\mathcal{R}_{\mathbf{yy}} = \left[\begin{array}{c|c} \boldsymbol{\mathcal{U}}_s(z) & \mathbf{0} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_s(z) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\mathcal{U}}_s^{\mathrm{P}}(z) \\ \hline \mathbf{0} \end{array}\right].$$

Speech Enhancement Algorithm [Neo2021]



⇒ Achieves significant

- noise reduction
- dereverberation
- speech enhancement

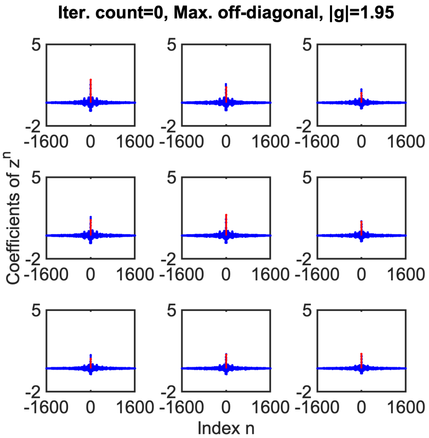
⇒ Arbitrary arrays

⇒ No noise estimation

⇒ No noticeable processing artefacts

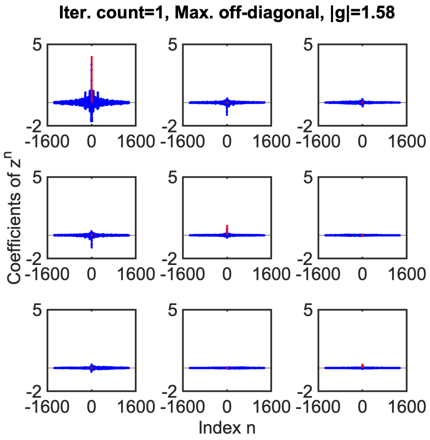
Example: PEVD Algorithm

Algorithm converges when $|g| < 1.68 \times 10^{-2}$



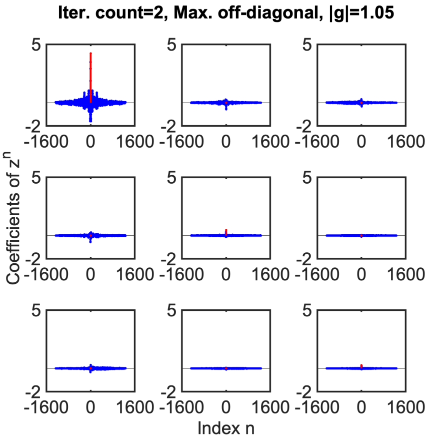
Example: PEVD Algorithm

Algorithm converges when $|g| < 1.68 \times 10^{-2}$



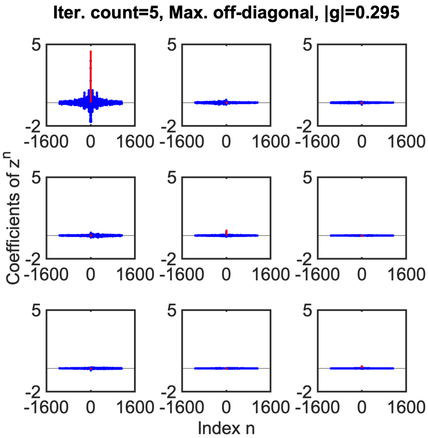
Example: PEVD Algorithm

Algorithm converges when $|g| < 1.68 \times 10^{-2}$



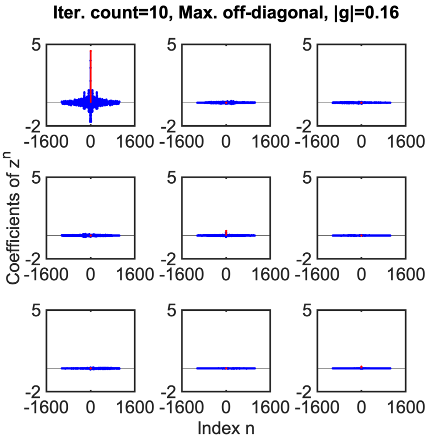
Example: PEVD Algorithm

Algorithm converges when $|g| < 1.68 \times 10^{-2}$



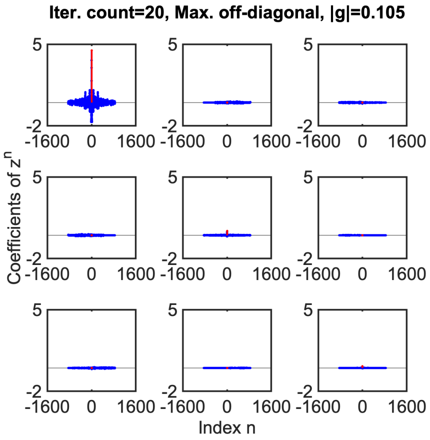
Example: PEVD Algorithm

Algorithm converges when $|g| < 1.68 \times 10^{-2}$



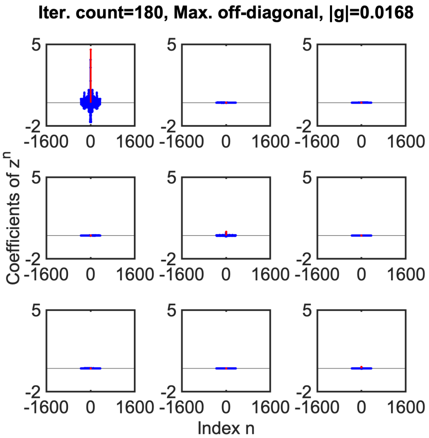
Example: PEVD Algorithm

Algorithm converges when $|g| < 1.68 \times 10^{-2}$



Example: PEVD Algorithm

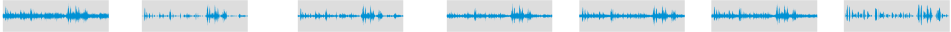
Algorithm converges when $|g| < 1.68 \times 10^{-2}$



Comparison of Speech Enhancement Algorithms

Algorithm	ΔSegSNR	ΔFwSegSNR	ΔSTOI	ΔPESQ
COLSUB	**8.35 dB**	**7.67 dB**	-0.018	-0.20
log-MMSE	3.67 dB	3.05 dB	-0.058	-0.12
MCSUB	-1.52 dB	-1.04 dB	-0.010	-0.03
MWF	1.06 dB	0.78 dB	-0.005	0.02
OMWF	0.57 dB	-0.44 dB	**0.084**	**0.17**
PEVD	2.96 dB	2.88 dB	0.078	0.11

Noisy COLSUB log-MMSE MWF OMWF PEVD Clean



Listening Examples: https://vwn09.github.io/pevd-enhance/

Application II. Signal Compaction for Spherical Microphone Arrays

PEVD for Spherical Microphone Arrays [Neo2023a]

- PEVD-based speech enhancement [Neo2019a; Neo2020; Neo2021]
  - Use PEVD to impose spatial decorrelation over a range of time shifts
  - Effective for noise reduction and dereverberation
  - Robust for linear and arbitrary array geometries
  - Works well for distributed arrays [d'Olne2022]

$\Rightarrow$ Limitation: Complexity at best $\propto$ $(\#\text{ of signals for processing})^3$

Can we exploit known information such as array geometry?

Focus: Spherical Microphone Array due to its relevance in hearing aids, sound field decomposition and reproduction, robot audition

Spherical Harmonics Transform (SHT)

The ℓ-th order, m-th degree eigenbeam signal, associated with the real SH basis function $\Upsilon_\ell^m(\mathbf{r}_q)$ and quadrature sampling weight α_q, is

$$\chi_\ell^m(n) \approx \sum_{q=1}^{Q} \alpha_q x(n, \mathbf{r}_q) \Upsilon_\ell^m(\mathbf{r}_q).$$

Recovery of the microphone signal at spherical coordinate $\mathbf{r}_q$ uses a weighted sum of the SH

$$x(n, \mathbf{r}_q) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \chi_\ell^m(n) \Upsilon_\ell^m(\mathbf{r}_q)$$

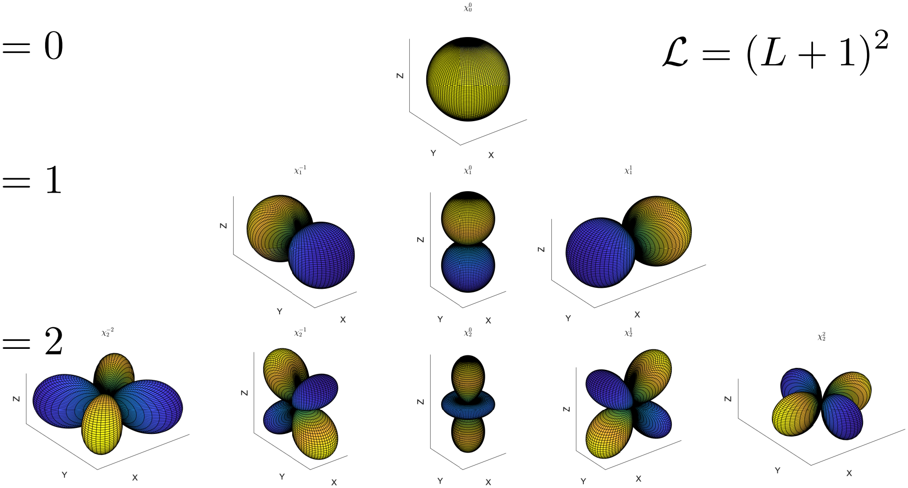
and alias-free spatial reconstruction requires $Q \geq (L+1)^2$ where L is the maximum SH order of the sound field and $\mathscr{L} \triangleq (L+1)^2$ eigenbeams.

$\ell = 0$

$$\mathcal{L} = (L+1)^2$$

$\ell = 1$

$\ell = 2$



$m = -2$ $\quad m = -1$ $\quad m = 0$ $\quad m = 1$ $\quad m = 2$

Complexity Analysis of a Single Example

ACE Lecture Room 2 with 10 dB Babble Noise

SH Order, L	0	1	2	3	4
# Eigenbeams $\mathcal{L}$	1	4	9	16	25
Approx. Error, $\varepsilon(\%)$	31.7	13.3	8.4	5.4	2.3
Complexity Factor, β	-	0.002	0.022	0.125	0.477

*$\beta = (\frac{\mathcal{L}}{Q})^3$, where $Q = 32$ microphones.

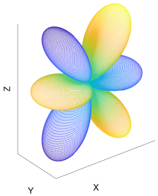
$\Longrightarrow$ A small number of eigenbeams (of a sufficiently large SH order) can sufficiently represent the microphone signals. This compact signal representation can reduce the number of signals used for PEVD processing.

Modal/Eigen-Beamformer Outputs

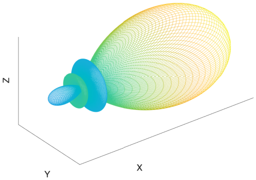
Given target source DoA (θ_p, ϕ_p), eigenbeam signals $\chi_\ell^m(n)$ are steered or used to form P beamformer outputs $\boldsymbol{\psi}(n) = [\psi_1(n), \ldots, \psi_P(n)]^{\mathrm{T}}$.

If w_ℓ^m represents modal beamformer weight or $\{0, 1\}$ for eigenbeam selection

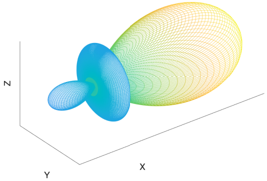
$$\psi_p(n) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} w_\ell^m \chi_\ell^m(n)$$



Steered eigenbeam.



Maximum directivity index (MaxDir) beamformer.



Modified hyper-cardioid (MHCARD) beamformer.

Space-time Covariance of Beamformer Outputs

Assuming stationarity, the space-time covariance matrix of P modal outputs is

$$\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau) = \mathbb{E}[\boldsymbol{\psi}(n)\boldsymbol{\psi}^{\mathrm{T}}(n-\tau)],$$

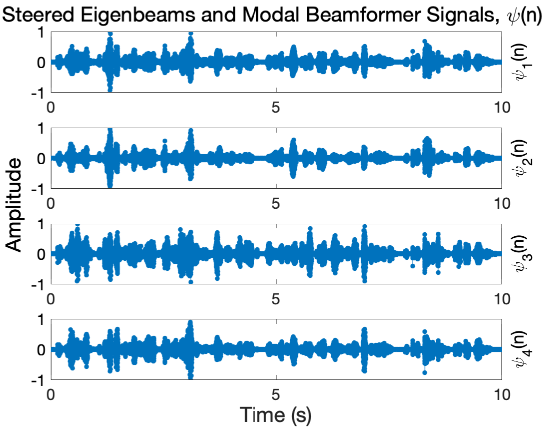
where $(i, j)^{\mathrm{th}}$ element is the correlation function $r_{ij}(\tau) = \mathbb{E}[\psi_i(n)\psi_j(n-\tau)]$ and τ is the time-shift.

Z-transform of $\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau)$ is a para-Hermitian polynomial matrix

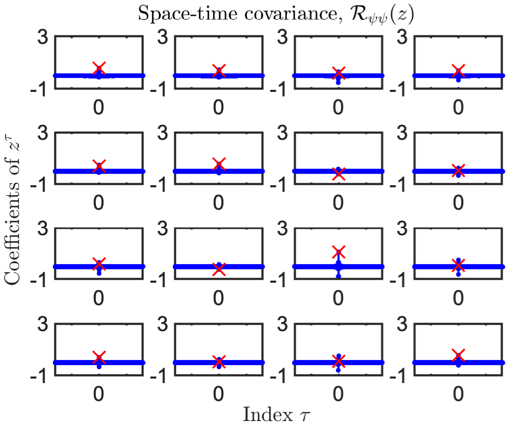
$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z) = \sum_{\tau=-W}^{W} \mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau) z^{-\tau},$$

where $\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau) \approx 0$ for $|\tau| > W$, calligraphic $\boldsymbol{\mathcal{R}}$ for polynomial matrices and regular $\mathbf{R}$ for matrices.

Example: Polynomial Matrix from ST-Covariance



Modal signals $\psi(n)$.



Polynomial matrix, $\mathcal{R}_{\psi\psi}(z)$.

Polynomial Matrix Eigenvalue Decomposition

The PEVD of $\mathcal{R}_{\psi\psi}(z)$ is [McWhirter2007]

$$\mathcal{R}_{\psi\psi}(z) \approx \mathcal{U}^{\mathrm{P}}(z)\boldsymbol{\Lambda}(z)\mathcal{U}(z),$$

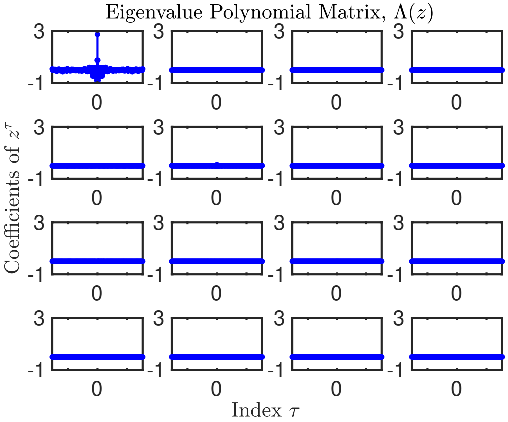
where $\boldsymbol{\Lambda}(z), \mathcal{U}(z)$ contain the eigenvalues and eigenvectors and $\mathcal{R}^{\mathrm{P}}_{\psi\psi}(z) = \mathcal{R}^{\mathrm{H}}_{\psi\psi}(1/z^*)$.

Subspace decomposition by the PEVD generates strongly decorrelated outputs:

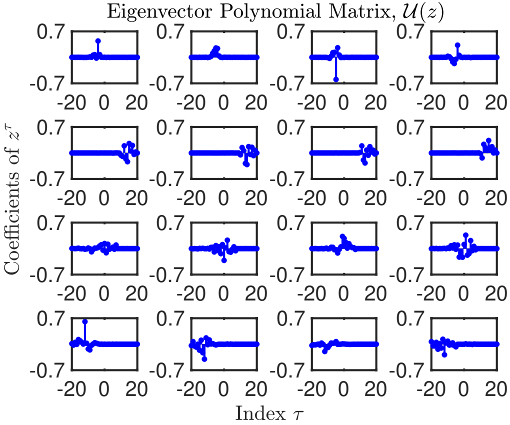
$$\mathcal{R}_{\psi\psi}(z) = \left[\; \mathcal{U}^{\mathrm{P}}_{s}(z) \;\middle|\; \mathcal{U}^{\mathrm{P}}_{s\perp}(z) \;\right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_s(z) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Lambda}_{s\perp}(z) \end{array}\right] \left[\begin{array}{c} \mathcal{U}_s(z) \\ \hline \mathcal{U}_{s\perp}(z) \end{array}\right],$$

associated with orthogonal target source, $\{\cdot\}_s$ and interferer, $\{\cdot\}_{s\perp}$ subspaces.

Example: PEVD Algorithm Outputs

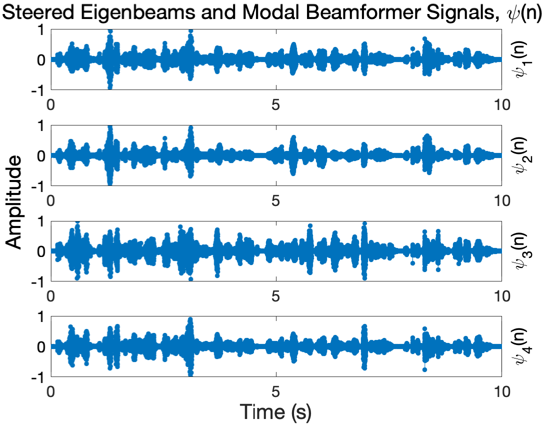


Eigenvalue polynomial matrix, $\boldsymbol{\Lambda}(z)$.

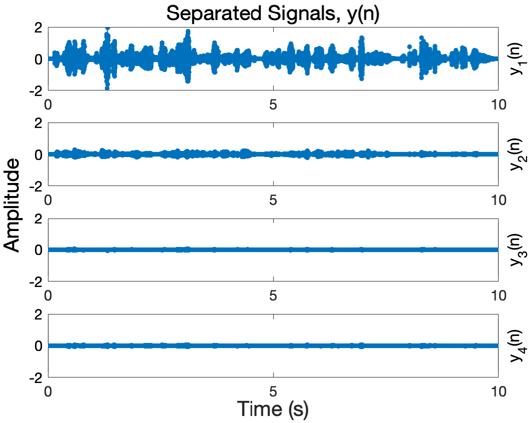


Eigenvector polynomial matrix, $\boldsymbol{\mathcal{U}}(z)$.

Example: Filterbank Output

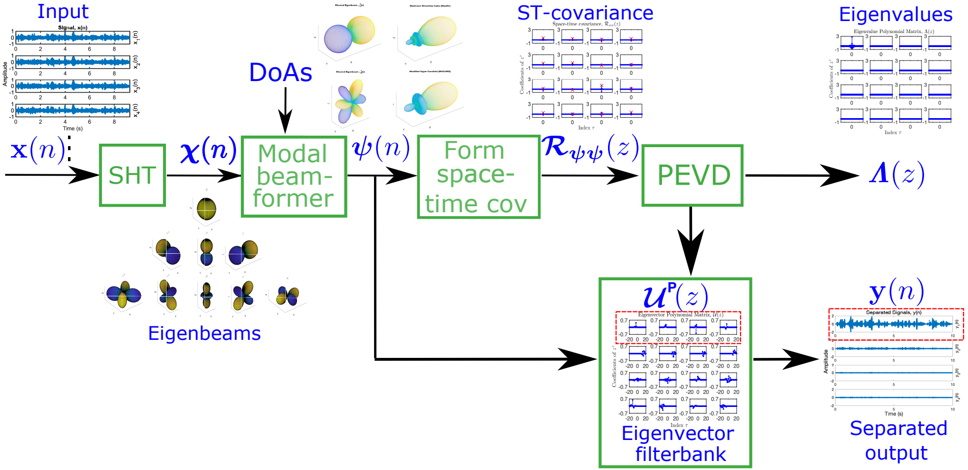


Modal signals $\psi(n)$.



Extracted source signals, $y(n)$.

Signal Compaction Framework [Neo2023b]

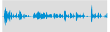


$\Longrightarrow$ SHT+PEVD: close to optimum signal compaction at reduced complexity while achieving good informed source separation and blind speech enhancement performance.

Source Separation Performance - Anechoic

Algorithm	ΔSDR	ΔSIR	ΔSAR	ΔSTOI	ΔPESQ
AuxIVA	17.7 dB	25.3 dB	11.4 dB	0.21	1.05
FastMNMF	20.6 dB	**35.2 dB**	13.8 dB	0.21	1.28
ILRMA	19.5 dB	31.3 dB	12.8 dB	0.21	1.21
MaxDir	3.9 dB	3.4 dB	4.7 dB	0.07	0.22
PEVD	**21.8 dB**	25.3 dB	**16.4 dB**	**0.24**	**1.39**

Received



MaxDir

AuxIVA

FastMNMF

ILRMA

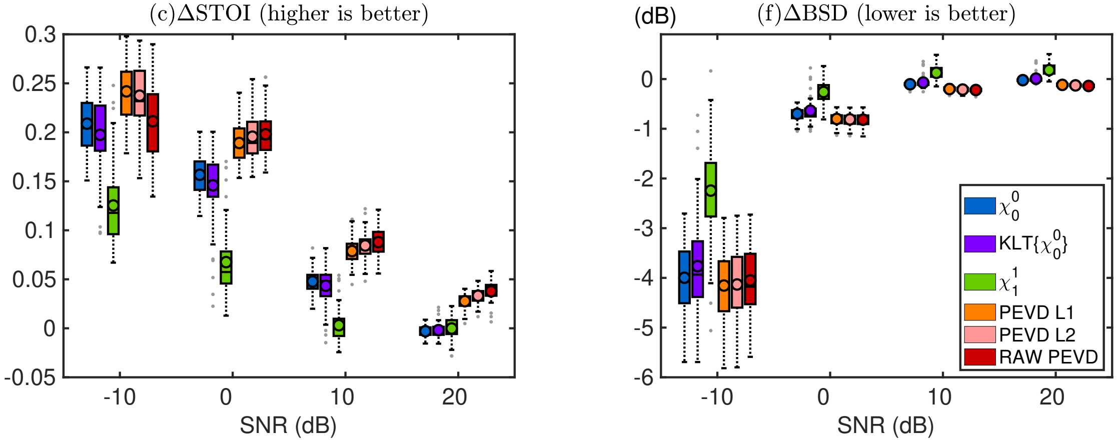
PEVD

Clean

Listening Examples: https://vwn09.github.io/pevd-smap/

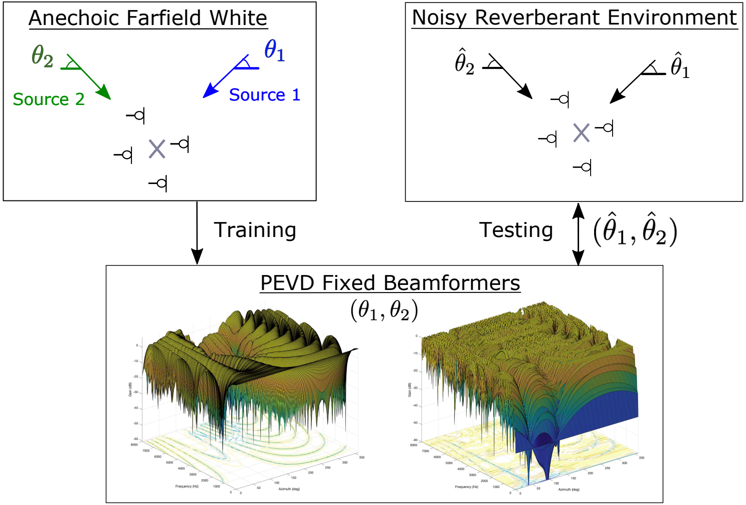
Speech Enhancement Using SHT+PEVD

Condition: ACE Lecture Room 2, Babble Noise



Application III. Fixed Beamformer Design Using PEVD

Fixed Beamformer Design Using PEVD [Neo2022a]



$\Rightarrow$ PEVD eigenvectors interpreted as beamformers

$\Rightarrow$ Arbitrary arrays

$\Rightarrow$ Avg. filter length: 114

$\Rightarrow$ Source separation performance approach data-dependent MVDR and LCMV beamformers

PEVD Beamformer Training

Rewriting as $\mathbf{H}(n) \circ\!\!-\!\!\bullet\, \boldsymbol{\mathcal{H}}(z) \in \mathbb{C}^{P\times Q}$, where each element is $h_{p,q}(n)$:

$$\boldsymbol{\mathcal{R}}_{\mathbf{x}}(z) = \boldsymbol{\mathcal{H}}^{\mathrm{P}}(z)\,\boldsymbol{\mathcal{R}}_{\mathbf{s}}(z)\,\boldsymbol{\mathcal{H}}(z) + \sigma_v^2\,\mathbf{I}\ ,$$

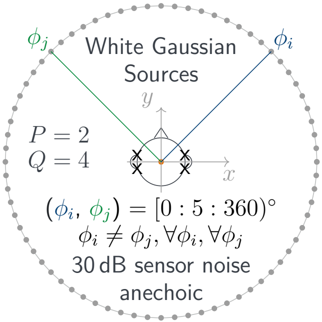
with spatially and temporally white noise $\mathbf{v}(n)$ of equal power σ_v^2.

With i.i.d. source signals and each drawn from $N(0,1)$, $\boldsymbol{\mathcal{R}}_{\mathbf{s}}(z) = \mathbf{I} \in \mathbb{C}^{P\times P}$.

Applying PEVD and rearranging:

$$\boldsymbol{\Lambda}(z) - \sigma_v^2\,\mathbf{I} = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\,\boldsymbol{\mathcal{H}}^{\mathrm{P}}(z)\,\boldsymbol{\mathcal{H}}(z)\,\boldsymbol{\mathcal{U}}(z)\ .$$

Training Using Simulated Impulse Responses



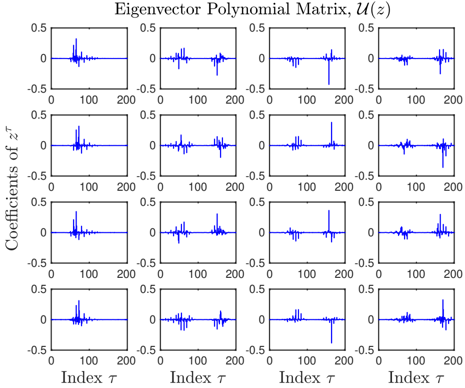
PEVD Beamformer Training

Diagonalization $\implies \boldsymbol{\mathcal{U}}(z)$ spatially decorrelate the acoustic channels $\boldsymbol{\mathcal{H}}(z)$.
By evaluating on the unit circle, beampattern response at frequency Ω:

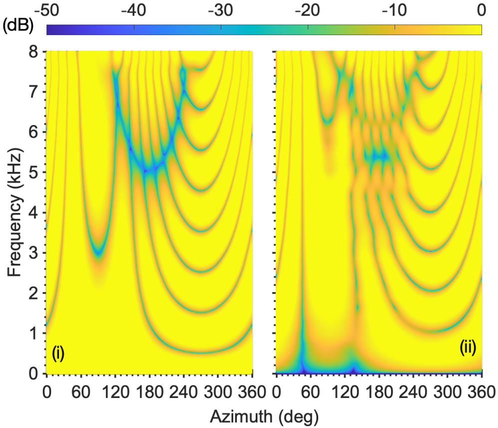
$$\mathbf{B}(\phi, \Omega) = [\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{a}_\phi(z)]\Big|_{z=e^{j\Omega}} ,$$

where $\mathbf{a}_\phi(n) \circ\!\!-\!\!\bullet\; \boldsymbol{a}_\phi(z) \in \mathbb{C}^Q$ is the broadband steering vector using array geometry and the qth element is $a_q(n) = \mathrm{sinc}(nT_s - \Delta\tau_q)$ with sampling period T_s and relative time delay $\Delta\tau_q$.

Eigenvector Filterbank Interpretation



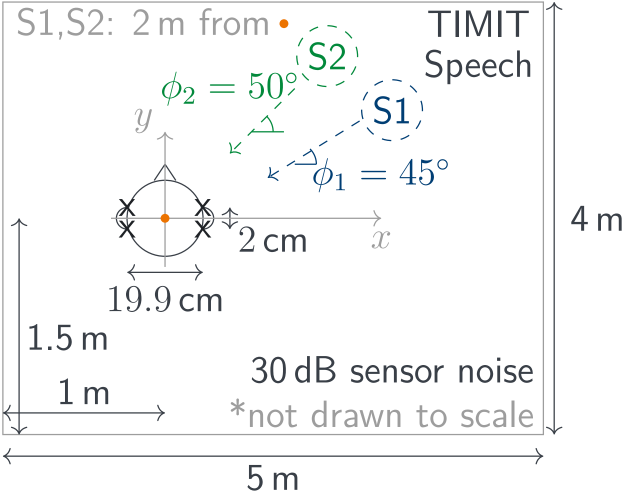
PEVD filterbank for $\{45°, 50°\}$



PEVD beamformer for $\{45°, 50°\}$

$\Rightarrow$ Since each element is a FIR filter, each PEVD eigenvector can be interpreted as a filterbank.

Testing Setup For Separation of Two Speakers



Results for Separation of Two Speakers

Algorithm	S1 ($\phi_1 = 45°$)		S2 ($\phi_2 = 50°$)	
	ΔSTOI	ΔSIR (dB)	ΔSTOI	SIR (dB)
PEVD $\{45°, 50°\}$	0.002	-0.034	**0.204**	15.752
PEVD $\{50°, 45°\}$	**0.123**	16.703	0.004	0.247
MVDR	0.113	13.487	0.186	12.435
LCMV	0.047	**19.986**	0.156	**23.522**

⇒ First PEVD output is similar to delay-and-sum that maximizes correlation in the look direction

⇒ Orthogonality across range of time lags as enforced by PEVD places a deep null in the first look direction in the Second PEVD output

Listening Examples:

https://vwn09.github.io/research/pevd-beamformer-iwaenc

Other Speech Processing Applications

PEVD has also been used for other speech processing applications

- Distributed array processing [d'Olne2022]
- Voice activity detection [Neo2022c; Neo2022b]
- Sound source localization [Hogg2021]

Conclusions

Conclusions

- Polynomial matrices are suited for modelling multichannel broadband signals e.g. speech
- PEVD achieves diagonalization of the space-time covariance over a range of time lags vs instantaneous spatial covariance for a single lag using EVD
- PEVD provides a more compact signal representation using fewer components compared to EVD
- Strong decorrelation and orthogonalizing properties of PEVD are useful for a variety of applications including speech enhancement, signal compaction, source separation and beamformer design

References

Asano, F., S. Hayamizu, T. Yamada, and S. Nakamura (2000). "Speech enhancement based on the subspace method". In: *IEEE Trans. Speech Audio Process.* 8.5, pp. 497–507.

Cohen, I. and B. Berdugo (Jan. 2002). "Noise estimation by minima controlled recursive averaging for robust speech enhancement". In: *IEEE Signal Process. Lett.* 9.1, pp. 12–15.

d'Olne, E., V. W. Neo, and P. A. Naylor (Aug. 2022). "Speech enhancement in distributed microphone arrays using polynomial eigenvalue decomposition". In: *Proc. European Signal Process. Conf. (EUSIPCO)*, pp. 55–59.

Hogg, Aidan O. T., Christine Evers, and Patrick A. Naylor (June 2021). "Multichannel overlapping speaker segmentation using multiple hypothesis tracking of acoustic and spatial features". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*.

Markovich, Shmulik, Sharon Gannot, and Israel Cohen (Aug. 2009). "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals". In: *IEEE Trans. Audio, Speech, Lang. Process.* 17.6, pp. 1071–1086.

McWhirter, J. G., P. D. Baxter, T. Cooper, S. Redif, and J. Foster (May 2007). "An EVD algorithm for para-Hermitian polynomial matrices". In: *IEEE Trans. Signal Process.* 55.5, pp. 2158–2169.

Neo, V. W., E. d'Olne, A. H. Moore, and P. A. Naylor (Sept. 2022a). "Fixed beamformer design using polynomial eigenvalue decomposition". In: *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1–5.

Neo, V. W., C. Evers, and P. A. Naylor (Oct. 2019a). "Speech enhancement using polynomial eigenvalue decomposition". In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 125–129.

References

- Neo, V. W., C. Evers, and P. A. Naylor (2020). "PEVD-based speech enhancement in reverberant environments". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, pp. 186–190.
- Neo, V. W., C. Evers, and P. A. Naylor (Oct. 2021). "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 29, pp. 3255–3266.
- Neo, V. W., C. Evers, S. Weiss, and P. A. Naylor (2023a). "Signal compaction using polynomial EVD for spherical array processing with applications". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.*
- Neo, V. W. and P. A. Naylor (2019b). "Second order sequential best rotation algorithm with Householder transformation for polynomial matrix eigenvalue decomposition". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, pp. 8043–8047.
- Neo, V. W., S. Redif, J. G. McWhirter, J. Pestana, I. K. Proudler, S. Weiss, and P. A. Naylor (2023b). "Polynomial eigenvalue decomposition for multichannel broadband signal processing". In.
- Neo, V. W., S. Weiss, S. W. McKnight, A. O. T. Hogg, and P. A. Naylor (Sept. 2022b). "Polynomial eigenvalue decomposition-based target speaker voice activity detection in the presence of competing talkers". In: *Proc. Intl. Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1–5.
- Neo, V. W., S. Weiss, and P. A. Naylor (Sept. 2022c). "A polynomial subspace projection approach for the detection of weak voice activity". In: *Sensor Signal Process. for Defence Conf. (SSPD)*, pp. 81–85.

References

- Redif, S., S. Weiss, and J. G. McWhirter (Jan. 2015). "Sequential matrix diagonalisation algorithms for polynomial EVD of para-Hermitian matrices". In: *IEEE Trans. Signal Process.* 63.1, pp. 81–89.
- Weiss, S., J. Pestana, and I. K. Proudler (May 2018). "On the existence and uniqueness of the eigenvalue decomposition of a para-Hermitian matrix". In: *IEEE Trans. Signal Process.* 66.10, pp. 2659–2672.

Thank you

Enhancement

Spherical

Beamformer