THEORY OF POLYNOMIAL MATRIX EIGENVALUE DECOMPOSITION WITH APPLICATIONS TO SPEECH AND ACOUSTICS SIGNAL PROCESSING

Vincent Weisheng Neo

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy of Imperial College London

Speech and Audio Processing Lab
Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London, United Kingdom
October 2022

Copyright Declaration

Statement of Originality

I declare that this thesis and the research it contains are the product of my own work under the guidance of my thesis supervisors, Prof. Patrick A. Naylor and Dr. Christine Evers. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. The material of this thesis has not been submitted for any degree at any other academic or professional institution.

"I have no special talent. I am only passionately curious."

- Albert Einstein

Abstract

The eigenvalue decomposition (EVD) of covariance matrices is used in signal processing applications such as data compression, noise reduction, direction-of-arrival estimation, source separation and adaptive beamforming. These multi-channel covariance matrices are usually computed using the instantaneous data vector under the assumption of narrowband models. However, when broadband sources are involved, correlations across different sensors and temporal lags need to be considered. Hence, an EVD that removes correlations at a single time lag becomes inadequate for completely decorrelating the signals.

To simultaneously capture the correlations in space, time and frequency, space-time covariance polynomial matrices have been proposed as an appropriate model for multi-channel broadband signals. The processing of polynomial matrices has motivated the development of a number of polynomial matrix eigenvalue decomposition (PEVD) algorithms.

This thesis first presents an algorithmic enhancement to the well-known second-order sequential best rotation (SBR2) algorithm using Householder transformations motivated by ideas for the symmetric eigenvalue problem in numerical linear algebra. The thesis then introduces and extends polynomial matrices and PEVD to the field of speech and acoustics signal processing. A novel PEVD-based speech enhancement algorithm is proposed for noise reduction and dereverberation. Despite being a blind and unsupervised algorithm, the approach achieves good speech enhancement, and experiments show that it does not introduce any noticeable processing artefacts into the enhanced signal.

While the method performs well for arbitrary microphone arrays, the computational complexity scales at best cubically with the number of signals used for processing. When the array geometry information is known and exploited, certain pre-processing reduces the dimensions for PEVD processing. The thesis compares different signal representations using KLT, spherical harmonic transform (SHT) and PEVD for a spherical microphone array geometry. PEVD is shown to provide the most compact spatio-temporal representation

of the microphone signals at the cost of high complexity. In contrast, spherical harmonics provide a compact spatial representation of the sound field using eigenbeam signals at a relatively low cost. The thesis provides both theoretical and experimental evidence showing that the PEVD of a smaller number of eigenbeams achieves good speech enhancement and source separation without noticeable audible distortions.

One of the goals of this thesis is to provide a bridge between the theoretical developments of PEVD and the applications for signal enhancement, compact signal representation and source separation. The work in this thesis also points in the way for further development in PEVD and its applications to other speech and acoustics signal processing tasks.

Acknowledgements

I must first thank Defence Science and Technology Agency (DSTA), Singapore, for the generous support and sponsorship of my undergraduate and postgraduate studies. Without their support, this thesis would have remained imaginary. In fact, my interest in speech processing started during an undergraduate internship project hosted by DSTA.

Determined to learn everything about speech processing, I was fortunate to participate in an UROP project while doing an internship with Nuance Communications. I must thank my supervisor, Prof. Patrick Naylor, for his excellent guidance and wealth of experience that have profoundly influenced my technical interests. I am very grateful for his patience, encouragement and openness to ideas which have sparked many exciting ideas and discussions, especially the work on PEVD. I want to express my gratitude to Dr Christine Evers. She has acted like my co-supervisor, guiding and teaching me plenty about acoustics, research approaches, writing styles and conference etiquette.

As a new kid on the block, I must thank the hospitality and warmth offered by the PEVD research community. In particular, I am thankful to Prof. Stephan Weiss for his enthusiasm and willingness to share his expert knowledge and time so generously. I would also like to thank Dr Soydan Redif and Prof. Ian Proudler for inducting me into the community during the special session at my first ICASSP in 2019. I am looking forward to more opportunities to work together in the future.

I am very humbled and grateful to Prof. Danilo Mandic and Prof. Marc Moonen for being willing to spend their time understanding and examining my work. Both have directly and indirectly broadened and deepened my technical perspectives with their work, especially those related to adaptive signal processing, tensor decompositions, SVD and GEVD.

During my time at Imperial College London, I have been fortunate to have so many brilliant friends and colleagues in the Speech and Audio Processing (SAP) Lab and Communications and Signal Processing (CSP) Group. Many have endured my sales talk on

PEVD, made my time as a PhD student more fun, and I will always remember our experiences attending physical and virtual conferences together. In no order of preference, I would like to thank Mike Brookes, Peter Dawson, Emilie d'Olne, Eric Grinstein, Pierre Guiraud, Sina Hafezi, Daniel Jones, Anselm Lohmann, Simon McKnight, Alastair Moore, Francesco Nespoli, Vikas Tokala, Raquel Viciana-Abad, and Rebecca Vos from SAP Lab, Lina Brazinskaite and Vanessa Rodriguez-Gonzalez for all their help, and my friends outside the office including Jia Chen, Lejon Chua, Dominicus Guo, Aaron Hau, Chang Hong How, Patrick Kwai, Wan Yee Lam, Audrie Lim, Yuxiang Lim, Jonathan Ong, Pablo Peso, James Poh, Marc See, Matthew Tan, Yingying Yeo, Longfei Yin, and Ruoxi Yong for all my fond memories in the UK.

I need to thank Aidan Hogg for everything he has taught me, especially in British culture, DSP, education and life, and for being so dependable. More special mentions go to Yunhao Liu for all the extended time we spent eating, drinking, walking and working, and Edoardo David Santi for taking my physical fitness to new heights and for all the personal and mathematical talks we had. I hope this only marks the start of our deep friendships for many more years to come!

I must thank all the unwavering support and love I receive back home from my colleagues, friends, relatives and family. I am forever grateful to my parents, Richard and Susan, my Grandmother, my siblings and in-laws, Victor, Cindy, Sandy, Priscillia and Marcus. In particular, I dedicate Chapter 5 to my nieces and nephews, Hollyn, Kayden, Holton, Holyatt and Sierra, for this work has been carried out during lockdown in Singapore. The circumstance has constantly reminded me of the importance of speech enhancement and noise reduction.

Vincent W. Neo

London, UK

20 May 2022

Contents

Notations

Number Sets

$\mathbb{R}$	real space
$\mathbb{C}$	complex space
$\mathbf{a}$	vector (lower case, bold)
$\mathbf{A}$	matrix (upper case, bold)
$\boldsymbol{\mathcal{A}}(z)$	polynomial matrix (upper case, bold calligraphic, parameterized)

Matrices

σ	singular value
λ	eigenvalue
$\mathbf{0}$	zero matrix
$\mathbf{I}$	identity matrix
$\mathbf{\Sigma}$	singular value matrix
$\mathbf{\Lambda}$	eigenvalue matrix
$\mathbf{U}$	eigenvector matrix (columns)
$\mathbf{W}$	eigenvector matrix (rows)
$\mathbf{G}$	Givens rotation matrix
$\mathbf{H}$	Householder reflection matrix
$\mathbf{T}$	upper triangular matrix

Operators

j	complex number, $\sqrt{-1}$
$*$	convolution
$.^*$	complex conjugate
$.^{\mathrm{T}}$	transpose
$.^{\mathrm{H}}$	Hermitian
$.^{\mathrm{P}}$	para-Hermitian
$.^{-1}$	inverse
$\mathbb{E}$	expectation
$\Im$	imaginary part
$\Re$	real part
$\lvert.\rvert$	magnitude/modulus
$\angle$	angle/phase/argument
$\mathcal{F}$	Fourier transform
$\mathcal{Z}$	z-transform
$\dim$	dimension
$\sup$	supremum
$\lvert\lvert.\rvert\rvert_2$	Euclidean/L2-norm
$\lvert\lvert.\rvert\rvert_F$	Frobenius-norm

$||.||_\infty$ infinity-norm

$\mathcal{O}\{\cdot\}$ Big O notation

Polynomial Matrices

μ trim parameter

δ convergence target

$\iota/\ell = 1, \ldots, I/L$ iterations

$|g|$ maximum off-diagonal element

β maximum off-diagonal column

$\mathbf{\Lambda}(z)$ eigenvalue polynomial matrix

$\mathbf{\Sigma}(z)$ singular value polynomial matrix

$\mathcal{U}(z)$ eigenvector (columns)

$\mathcal{W}(z)$ eigenvector (rows)

$\mathcal{D}(z)$ delay polynomial matrix

$\mathbf{R}(\tau)$ space-time covariance matrix

$\mathbf{R}(0)$ instantaneous covariance matrix

$\mathcal{R}(z)$ z-transform of $\mathbf{R}(\tau)$

$\mathcal{T}(z)$ upper triangular polynomial matrix

Signal Processing

f frequency

Ω normalized angular frequency

F_s sampling frequency

T_s sampling period

τ time delay/temporal lag

ϕ phase

ϵ reconstruction error

J channel/filter order

W temporal window length

$k = 1, \ldots, K$ frequency bins

$t = 1, \ldots, T$ frames

$n = 1, \ldots, N$ time samples

$p = 1, \ldots, P$ sources

$q = 1, \ldots, Q$ sensors

$s_p(n)$ p^{th} source signal

$x_q(n)$ q^{th} sensor signal

$y_q(n)$ q^{th} processed signal

$v_q(n)$ q^{th} noise signal

$h_{p,q}$ $(p,q)^{\text{th}}$ channel impulse response

$\mathbf{s}$ source signals $\in \mathbb{R}^P$

$\mathbf{v}$ noise signals $\in \mathbb{R}^Q$

$\mathbf{x}$ received signals $\in \mathbb{R}^Q$

$\mathbf{y}$ processed signals $\in \mathbb{R}^Q$

$\{.\}_s$ signal (signal-plus-noise) subspace

$\{.\}_v$ noise (noise-only) subspace

Spherical Harmonic and Array Processing

r radius of the spherical array

$\mathcal{L}$ number of beamformer outputs

$P_\ell^{(m)}(\cdot)$ associated Legendre function

α quadrature weight

θ elevation angle downwards from the z-axis

ϕ azimuth angle from the x-axis towards the y-axis

$\mathbf{r}$ spherical coordinate, (r, θ, ϕ)

$\mathbf{w}, \mathbf{W}$ beamformer weights

$\ell = 0, \ldots, L$ spherical harmonic order

$m = -\ell, \ldots, \ell$ spherical harmonic degree

$Y_\ell^{(m)}$ ℓ-th order, m-th degree complex-valued spherical harmonic basis

$R_\ell^{(m)}$ ℓ-th order, m-th degree real-valued spherical harmonic basis function

$\chi_\ell^{(m)}$ ℓ-th order, m-th degree real-valued eigenbeam signal

ψ_q q^{th} modal beamformer output signal

$\boldsymbol{a}(z)$ broadband steering vector $\in \mathbb{C}^Q$

$\boldsymbol{u}_i(z)$ i^{th} polynomial eigenvector beamformer

$\mathbf{k}$ wavenumber

$\mathcal{B}(\mathbf{k}, z)$ broadband array response

Abbreviations

AuxIVA	auxiliary function-based independent vector analysis
BSD	Bark spectral distortion
DFT	discrete Fourier transform
DoA	direction of arrival
DRR	direct-to-reverberant ratio
DSP	digital signal processing
EVD	eigenvalue decomposition
FastMNMF	fast multi-channel non-negative matrix factorization
FIR	finite impulse response
F-norm	Frobenius-norm
FwSegSNR	frequency-weighted segmental signal-to-noise ratio
GEVD	generalized eigenvalue decomposition
GSC	generalized sidelobe canceller
GWPE	generalized weighted prediction error
ICA	independent component analysis
ILRMA	independent low-rank matrix analysis
ISCLP	integrated sidelobe canceller linear prediction Kalman filter
KLT	Karhunen-Loève transform
LCMV	linearly constrained minimum variance
LCMP	linearly constrained minimum power
LPC	linear prediction coding
MCSUB	multi-channel subspace
MIMO	multiple input multiple output
MINT	multi-channel inversion theorem
MNMF	multi-channel non-negative matrix factorization
MMSE	minimum mean square error

MPDR	minimum power distortionless response
MUSIC	multiple signal classification
MVDR	minimum variance distortionless response
MWF	multi-channel Wiener filter
NSRR	normalized signal-to-reverberant ratio
OMWF	oracle multi-channel Wiener filter
PCFB	principal component filter bank
PESQ	perceptual evaluation of speech quality
PEVD	polynomial matrix eigenvalue decomposition
PMD	polynomial matrix decomposition
PSD	power spectral density
RIR	room impulse response
RTF	relative transfer function
SAR	source-to-artefacts ratio
SBR2	second-order sequential best rotation
SegSNR	segmental signal-to-noise ratio
SDR	source-to-distortion ratio
SH	spherical harmonic
SHT	spherical harmonic transform
SIMO	single input multiple output
SIR	source-to-interference ratio
SMD	sequential matrix diagonalisation
SNR	signal-to-noise ratio
STFT	short-time Fourier transform
STOI	short-time objective intelligibility
SUT	strong uncorrelating transform
SVD	singular value decomposition
TDoA	time difference of arrival
WPD	weighted power minimum distortionless response
WPE	weighted prediction error

List of Figures

List of Tables

Introduction

Chapter 1

Introduction

1.1 Motivation

Multi-channel broadband signals exist in many essential commercial and military technologies such as speech processing, telecommunications, healthcare monitoring, seismic surveillance, radar, and sonar [1–4]. Facilitated by low manufacturing costs, there is widespread availability of multiple sensors like microphones on consumer products such as mobile phones, hearing aids, voice-controlled systems and robots. The spatial diversity present in the acquired signals can be exploited for source separation [5, 6], signal enhancement [7], channel coding [8], and source localization [9], thus improving application performance.

A common approach to solving these problems is to divide the broadband signal into multiple narrowband signals, for each of which a rich set of optimal linear algebraic tools can be exploited [10–12]. However, splitting the broadband signal into independent frequency bins neglects spectral coherence and ignores correlations between different discrete Fourier transform (DFT) bins [13, 14]. The explicit embedding of the cross-spectral terms has been proposed to incorporate these inter-bands correlations [15]. Another approach uses tapped delay line (TDL) processing [16], but the performance depends on the length of the TDL, which is challenging to determine in practice. These approaches highlight the lack of generic tools to solve broadband problems directly.

In classical subspace approaches such as the multiple signal classification (MUSIC) algorithm [17], the instantaneous spatial covariance matrix is computed using the complex outer product of the instantaneous data vector collected from the sensors. The eigenvalue decomposition (EVD) of these instantaneous Hermitian matrices is widely used in many important signal processing applications such as data compression [18], noise reduction [19], spectral estimation [20], blind source separation [21] and adaptive beamforming [22, 23]. This computation, however, assumes that the sources are narrowband, uncorrelated and instantaneously mixed.

When source signals are broadband or convolutively mixed, e.g. due to reverberation [24], the outputs are generally correlated in time across different sensors. Therefore, the time delays for broadband signals cannot be represented by only phase shifts or complex gain factors captured by the instantaneous spatial covariance matrix. Instead, the correlations across different sensors and temporal lags needs to be modelled such as by using a matrix of finite impulse response (FIR) filters in the form of a polynomial matrix.

The processing of polynomial matrices has motivated the development of a family of polynomial matrix eigenvalue decomposition (PEVD) methods [25–29], based on the pioneering second-order sequential best rotation (SBR2) algorithm [30]. Since then, PEVD has been applied to many multi-channel wideband signal processing applications such as adaptive beamforming [31], multi-channel Wiener filtering [32], blind source separation [33], source identification [34] and channel coding [35].

1.2 Contributions of Thesis

1.2.1 Research Objectives

The aims of this thesis are to explain and develop a deep theoretical understanding of the PEVD, propose enhancements to existing PEVD algorithms or new approaches, as well as use PEVD for speech processing applications such as signal enhancement and source separation.

1.2.2 Original Contributions

To the best of the author's knowledge, the novel contributions of this thesis are:

1. **Insights and analysis obtained from the comparison between the subspaces generated by the EVD and PEVD.** (Chapter 3)
   - The bases, generated by the PEVD eigenvector, and their translates span a shift-invariant subspace compared to the vector subspace spanned by the outputs generated by the eigenvectors computed from the well-known EVD [10, 11].
   - Signal reconstruction can be achieved using a linear combination of the strongly decorrelated outputs and their shifted versions. In contrast, for the EVD, signal reconstruction only uses linear combination of the computed outputs.
   - The rank of the PEVD subspace is smaller or equal to the rank of the EVD subspace.
2. **An extension to the SBR2 algorithm by using Householder transformation to reduce polynomial matrices to tri-diagonal form.** (Chapter 3, [29])
   - Using Householder reduction to first reduce the principal plane of the polynomial matrix to tri-diagonal form decreases the number of iterations and speeds up convergence.
3. **The use of a polynomial matrix as a signal model and PEVD-based algorithms for speech enhancement.** (Chapter 4, [36–39])
   - The use of polynomial matrix as a broadband, multi-channel signal model for processing speech signals.
   - A PEVD-based noise reduction algorithm for noisy speech in anechoic environments that does not introduce any audible artifacts while suppressing background noise.
   - A PEVD-based speech dereverberation algorithm for suppressing reverberation without introducing any audible artifacts.
   - A PEVD-based speech enhancement algorithm for noisy and reverberant speech in diverse acoustic environments that improves speech quality and intelligibility.
   - A blind and unsupervised PEVD-based speech enhancement algorithm that does not rely on any noise estimators and is suitable for arbitrary array geometries.

4. **Signal compaction using PEVD for spherical arrays with applications.** (Chapter 5, [40–42])
   - A comparison of the signal representation afforded by the SHT and PEVD for spherical arrays.
   - A proposed framework that combines SHT and PEVD for more efficient processing because of a more compact representation than microphone signals.
   - Application of the proposed framework for blind speech enhancement.
   - Application of the proposed framework for informed source separation using direction of arrival (DoA) information.
5. **Fixed beamformer design using PEVD.** (Chapter 6, [43])
   - A visual interpretation of the polynomial eigenvectors generated by the PEVD using a beamformer or spatial filter.
   - An analysis of the fixed beamformer designed by the PEVD.
   - A comparison of the PEVD fixed beamformers with classical data-dependent minimum variance distortionless response (MVDR) and linearly constrained minimum variance (LCMV) beamformers.

1.3 Publications

During the course of this work, the following have been published or are being considered for publication:

Journal Articles

1. V. W. Neo, C. Evers, S. Weiss, and P. A. Naylor, "Signal compaction using polynomial EVD for spherical array processing with applications," *submitted to IEEE/ACM Trans. Audio, Speech and Language Proc. (TASLP).*

2. V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition," *IEEE/ACM Trans. Audio, Speech and Language Proc. (TASLP)*, Oct. 2021.

Conference and Workshop Papers

1. S. W. McKnight, A. O. T. Hogg, V. W. Neo, and P. A. Naylor, "Studying human-based speaker diarization and comparing to state-of-the-art systems," *Proc. Asia Pacific Signal and Inform. Process. Assoc. Annu. Summit and Conf. (APSIPA ASC)*, Nov. 2022.

2. V. W. Neo, S. Weiss, and P. A. Naylor, "A polynomial subspace projection approach for the detection of weak voice activity," *in Proc. Sensor Signal Proc. for Defence Conf. (SSPD)*, Sep. 2022.

3. V. W. Neo, E. d'Olne, A. H. Moore, and P. A. Naylor, "Fixed beamformer design using polynomial eigenvalue decomposition," *in Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022.

4. V. W. Neo, S. Weiss, S. W. McKnight, A. O. T. Hogg, and P. A. Naylor, "Polynomial eigenvalue decomposition-based target speaker voice activity detection in the presence of competing talkers," *in Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022.

5. E. d'Olne, V. W. Neo, and P. A. Naylor, "Frame-based space-time covariance matrix estimation for polynomial eigenvalue decomposition-based speech enhancement," *in Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022.

6. E. d'Olne, V. W. Neo, and P. A. Naylor, "Speech enhancement in distributed microphone arrays using polynomial eigenvalue decomposition," *in Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2022.

7. S. W. McKnight, A. O. T. Hogg, V. W. Neo, and P. A. Naylor, "A study of salient modulation domain features for speaker identification," in *Proc. Asia Pacific Signal*

and Inform. Process. Assoc. Annu. Summit and Conf. (APSIPA ASC), Dec. 2021.

8. V. W. Neo, C. Evers, and P. A. Naylor, ”Polynomial matrix eigenvalue decomposition-based source separation using informed spherical microphone arrays,” in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2021.

9. A. O. T. Hogg, V. W. Neo, S. Weiss, C. Evers, and P. A. Naylor, ”A polynomial eigenvalue decomposition MUSIC approach for broadband sound source localization,” in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2021.

10. V. W. Neo, C. Evers, and P. A. Naylor, ”Polynomial matrix eigenvalue decomposition of spherical harmonics for speech enhancement,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2021.

11. V. W. Neo, C. Evers, and P. A. Naylor, ”Speech dereverberation performance of a polynomial-EVD subspace approach,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2020.

12. V. W. Neo, C. Evers, and P. A. Naylor, ”PEVD-based speech enhancement in reverberant environments,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2020.

13. V. W. Neo, C. Evers, and P. A. Naylor, ”Speech enhancement using polynomial eigenvalue decomposition,” in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2019.

14. V. W. Neo and P. A. Naylor, ”Second order sequential best rotation algorithm with Householder transformation for polynomial matrix eigenvalue decomposition,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2019.

1.4 Thesis Outline

The remaining chapters in this thesis are organized as follows:

Chapter 2 Polynomial Matrix Decomposition describes the multi-channel broadband signal model used in this thesis. The space-time covariance matrix, a generalized form of the spatial covariance matrix, is computed using the sensor data. The space-time covariance polynomial matrices are then processed using PEVD, which is the main focus of this thesis and will be reviewed along with other polynomial matrix decomposition techniques.

Chapter 3 New Insights, Theoretical & Algorithmic Development for PEVD compares EVD and PEVD and explains why PEVD is more suitable for multi-channel broadband signal processing applications by providing a compact signal representation, a concept further explored in Chapter 5. An interpretation of the polynomial eigenvectors is also pursued in Chapter 6. An algorithmic enhancement to the classical SBR2 is also proposed to speed up convergence while managing computational complexity.

Chapter 4 PEVD-based Speech Enhancement in Noisy Reverberant Environments first reviews the literature on speech enhancement and noise reduction algorithms. A PEVD-based speech enhancement algorithm is proposed. Applying the PEVD algorithm to the space-time covariance polynomial matrix computed using the microphone signals produces polynomial eigenvalues and eigenvectors. Speech enhancement is then achieved via a subspace reweighing. The proposed approach is compared against benchmark approaches under diverse acoustic conditions.

Chapter 5 Signal Compaction Using PEVD for Spherical Arrays with Applications picks up Chapter 3 and focuses on the compact representation of speech signals acquired by spherical microphone arrays. A comparison of the signal representation afforded by SHT and PEVD is provided, where it is shown to be advantageous to combine both approaches. Processing with the proposed framework improves representation using fewer computations, leading to effective speech enhancement and source separation.

Chapter 6 Fixed Beamformer Design Using PEVD proposes fixed beamformer designs using PEVD algorithms, which are lightweight and suitable for arbitrary arrays. This chapter also provides an insightful interpretation of the eigenvector filterbanks in the form of spatio-temporal filters. The fixed beamformers are then shown to perform comparably to classical data-dependent MVDR and LCMV beamformers for the separation of sources closely spaced by $5°$ in noisy reverberant environments.

Chapter 7 Conclusions summarizes the list of contributions in this thesis, presents the research findings, and points in the way for promising directions for PEVD developments and applications.

Technical Background

Chapter 2

Polynomial Matrix Decomposition

2.1 Introduction

This chapter first introduces a historical perspective on polynomial matrices and provides an overview of polynomial matrix eigenvalue decomposition (PEVD) algorithms. This will lead to multi-channel broadband signal processing where polynomial matrices naturally arise in the form of space-time covariance matrices. The need to process polynomial matrices has led to several factorization techniques. As the thesis will mainly focus on PEVD, this chapter will review the well-known second-order sequential best rotation (SBR2) and sequential matrix diagonalisation (SMD) classes of z-domain algorithms and briefly introduce frequency-domain approaches, related methods in conventional matrices and other polynomial matrix decomposition (PMD) techniques.

2.2 Overview of Polynomial Matrix Decomposition

2.2.1 Historical Perspectives and Background

Polynomial matrices are widely used in control theory and signal processing. In control theory, these matrices can describe multi-variable transfer functions for multiple input multiple output (MIMO) systems [44]. More recently, in digital signal processing (DSP), polynomial matrices are used to describe filterbank systems for polyphase decomposition in multi-rate processing [45].

Control systems are usually designed for continuous-time systems and hence, are analyzed in the Laplace domain. There, factorization of matrices in the Laplace variable s, such as the Smith and Smith-McMillan decomposition [45, 46], target unimodularity, which is critical in the control context to minimize time delay [30, 44]. While Smith and Smith-McMillan decompositions can also be used for discrete-time systems, they do not offer important properties for signal processing, such as losslessness or para-unitarity. Therefore, this thesis is devoted to the development and applications of polynomial matrix techniques that extend the utility of EVD in linear algebra [10–12] from the narrowband to the broadband case.

Classical subspace approaches such as the MUSIC algorithm apply an EVD to the instantaneous spatial covariance matrix. The instantaneous spatial covariance matrix can capture the phase shifts, or more generally complex gain factors, between narrowband signals arriving at different sensors. However, in multichannel broadband arrays or convolutively mixed signals, the outputs are generally correlated in time across different sensors. Therefore, the time delays for broadband signals cannot be represented by only phase shifts and needs to be explicitly modelled using a matrix of FIR filters. The relative time shifts are captured using the polynomial space-time covariance matrix, where decorrelation over a range of time shifts can be achieved using a PEVD.

While the initial work on PEVD was a numerical algorithm [30], the decomposition of an analytic, positive semi-definite para-Hermitian matrix like the space-time covariance matrix, assumed to contain Laurent polynomial elements, has subsequently been proven to exist [47–49], based on an analogous problem in the continuous-time domain [50]. In most cases, unique para-Hermitian eigenvalues and para-unitary eigenvectors exist but are of infinite order. However, being analytic, they permit good approximations using finite-length factors that are still helpful for many signal processing applications.

2.2.2 State of the Art Polynomial Matrix Eigenvalue Decomposition (PEVD) Algorithms

Polynomial matrix factorization techniques such as the polynomial QR and singular value decomposition (SVD) [51,52] exist and will be briefly discussed in Section 2.6. Instead, this thesis mainly focuses on the polynomial EVD of para-Hermitian polynomial matrices like the space-time covariance matrix, which began from McWhirter's pioneering work on the SBR2 algorithm [30]. This has led to a family of z-domain algorithms based on the SBR2 [30, 53, 54] and SMD [28, 29, 55]. These algorithms tend to generate spectrally majorized eigenvalues, which may not be analytic, as well as para-unitary eigenvectors.

PEVD algorithms generating fixed polynomial order eigenvectors have also been proposed [56, 57], although their convergence has not been proven. Another class of algorithms operate in the DFT-domain [58–60], where the lost associations between frequency bins need to be established, and maximally smooth analytic solutions can be obtained.

Since their inception, PEVD algorithms have been used in many multi-channel broadband signal processing applications such as beamformers [31], MIMO communications [61–63] and channel coding [35, 64], signal enhancement [32, 39], source separation [33, 41, 65, 66], source identification [34] and source localization [67–69].

2.3 Broadband Signal Model and Polynomial Matrices

2.3.1 Multi-channel Signal Model

The system comprising multiple sensors can either be a single input multiple output (SIMO) or multiple input multiple output (MIMO) system. For simplicity, the SIMO system with 1 source and Q receivers, as shown in Fig. 2.1, is first considered [70].

The noisy signal, $x_q(n)$, at the q-th sensor for sample index, $n = 0, 1, \ldots, N-1$, is

$$x_q(n) = \mathbf{h}_q^{\mathrm{T}} \mathbf{s}_0(n) + v_q(n) = s_q(n) + v_q(n) \ , \tag{2.1}$$

where $\mathbf{s}_0(n) = [s_0(n), s_0(n-1), \ldots, s_0(n-J)]^{\mathrm{T}}$ is the source signal, $\mathbf{h}_q = [h_{q,0}, h_{q,1}, \ldots, h_{q,J}]^{\mathrm{T}}$ represents the q-th channel modelled as a J-th order finite impulse response (FIR) filter, $s_q(n)$ is the source signal component, $v_q(n)$ is the additive noise signal received at the q-th sensor and $[\cdot]^{\mathrm{T}}$ denotes the transpose operator. The noise signals are assumed to be zero-mean with power σ_v^2, and uncorrelated at different sensors and with the source signal. Furthermore, the channel is also assumed to be stationary as a special case of Fig. 2.1. The received data vector for Q sensors at time sample n becomes

$$\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_Q(n)]^{\mathrm{T}} \in \mathbb{C}^Q , \tag{2.2}$$

with $\mathbf{s}(n)$ and $\mathbf{v}(n)$ similarly defined. Assuming stationarity, a typical subspace approach uses the spatial covariance matrix, $\mathbf{R_{xx}}$, which is Hermitian symmetric and defined as

$$\mathbf{R_{xx}} = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\} \in \mathbb{C}^{Q\times Q} , \tag{2.3}$$

where $[\cdot]^{\mathrm{H}}$ and $\mathbb{E}\{\cdot\}$ denote the Hermitian transpose and expectation operators, respectively. This matrix captures spatial information arising from the geometry of the array and position of the source. Performing an EVD on the spatial covariance matrix gives

$$\mathbf{R_{xx}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} , \tag{2.4}$$

where the diagonal matrix, $\boldsymbol{\Lambda}$, contains as its elements the eigenvalues with the corresponding orthonormal eigenvectors lying on the columns of $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_Q]$. In order to enforce uniqueness and avoid permutation ambiguity, eigenvalues are ordered decreasingly, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_Q \geq 0$. Ambiguities in eigenvectors can arise from arbitrary phase rotations or non-distinct eigenvalues [11, 12]. In the former, each eigenvector can be multiplied by any arbitrary phase shift ϕ_q without affecting orthonormality between eigenvectors, i.e., $\tilde{\mathbf{u}}_q = \mathbf{u}_q e^{j\phi_q}$ lies in the columns of $\tilde{\mathbf{U}}$ such that $\tilde{\mathbf{U}}^{\mathrm{H}}\tilde{\mathbf{U}} = \mathbf{I}$. If an eigenvalue is repeated M times, i.e. has an algebraic multiplicity of M, the corresponding M eigenvectors are ambiguous up to any arbitrary unitary matrix $\boldsymbol{\Phi} \in \mathbf{C}^{M\times M}$ such that $[\tilde{\mathbf{u}}_q, \ldots, \tilde{\mathbf{u}}_{q+M-1}] = [\mathbf{u}_q, \ldots, \mathbf{u}_{q+M-1}]\boldsymbol{\Phi}$.

Assuming that source and noise signals are uncorrelated, the covariance matrix can be decomposed into disjoint signal and noise subspaces such that (2.3) can be rewritten

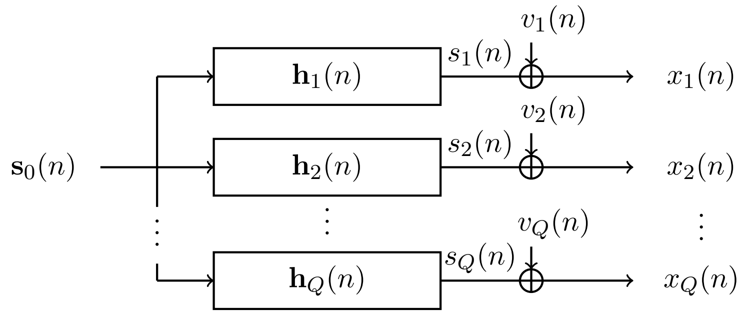


Figure 2.1: SIMO System with Q-outputs.

as

$$\mathbf{R}_{\mathbf{xx}} = (\sigma_s^2 + \sigma_w^2)\mathbf{u}_1\mathbf{u}_1^{\mathrm{H}} + \sum_{i=2}^{Q} \sigma_w^2 \mathbf{u}_i\mathbf{u}_i^{\mathrm{H}} = \mathbf{R}_{\mathbf{ss}} + \mathbf{R}_{\mathbf{vv}} \ , \tag{2.5}$$

where $\mathbf{R}_{\mathbf{ss}}$ and $\mathbf{R}_{\mathbf{vv}}$ are the signal and noise covariance matrices. For a single narrowband source, $\dim(\mathbf{R}_{\mathbf{ss}}) = 1$ and $\dim(\mathbf{R}_{\mathbf{vv}}) = Q$, with $\dim(\cdot)$ indicating the number of linearly independent basis vectors present in its argument. Exploiting the orthogonality between the assumed signal-plus-noise (or simply signal) and noise-only (or noise) subspace, the eigenvectors associated with the noise subspace, $\mathbf{U}_v = [\mathbf{u}_2, \ldots, \mathbf{u}_Q]$, lie in the nullspace of the signal covariance matrix such that $\mathbf{R}_{\mathbf{ss}}\mathbf{U}_v = \mathbf{0}$. This approach is used extensively for signal parameter estimation based on the classical MUSIC algorithm [17].

The decomposition in (2.4) only accounts for the instantaneous correlation because

$$\mathbf{\Lambda} = \mathbf{U}^{\mathrm{H}}\mathbf{R}_{\mathbf{xx}}\mathbf{U}, \tag{2.6}$$

indicates that the diagonalization is achieved for the spatial covariance matrix, which is computed based on the instantaneous outer product of the data vector, $\mathbf{x}(n)$. For a narrowband source signal, the instantaneous correlation sufficiently captures all necessary information, such as phase, for processing.

2.3.2 Space-time Covariance Matrix and Polynomial Matrices

In the applications described in this thesis, the broadband source signals received by the sensors exhibit spatial, spectral and temporal correlations. The space-time covariance matrix, that is a generalization of the (instantaneous) spatial covariance matrix, is defined as [30]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau)\} \in \mathbb{C}^{Q\times Q} \ , \tag{2.7}$$

where the $(p,q)^{\mathrm{th}}$ element, $r_{p,q}(\tau) = \mathbb{E}\{x_p(n)x_q^*(n-\tau)\}, \tau \in \mathbb{Z}$, is the correlation sequence between the data vector received at the p-th and q-th sensor, and $[\cdot]^*$ denotes the complex conjugate operation. The auto-correlation and cross-correlation sequences are obtained when $p = q$ and $p \neq q$ for $r_{p,q}(\tau)$, respectively. Note that the space-time covariance matrix changes with τ and the instantaneous spatial covariance matrix is the special case when $\tau = 0$ as in (2.3).

Consider the case when $\mathbf{s}_0(n)$ is a narrowband signal at frequency f_0 and only direct-path propagation. The signal at the q-th sensor, $x_q(n)$, is then related to the signal at the reference sensor, $x_1(n)$, by a constant phase shift, $\phi_q = 2\pi f_0\tau_q$, where τ_q is the fixed time delay for the signal to propagate from the reference to the q-th sensor. Consequently, (2.2) becomes

$$\mathbf{x}(n) = [x_1(n), x_1(n)\mathrm{e}^{-\mathrm{j}\phi_2}, \ldots, x_1(n)\mathrm{e}^{-\mathrm{j}\phi_Q}]^{\mathrm{T}} \ , \tag{2.8}$$

where $\mathrm{j} = \sqrt{-1}$ is the imaginary number, while the $(p,q)^{\mathrm{th}}$ element in (2.7) becomes

$$\begin{aligned} r_{p,q}(\tau) &= \mathbb{E}\{x_p(n)x_q^*(n-\tau)\} = \mathbb{E}\{[x_1(n)\mathrm{e}^{-\mathrm{j}\phi_p}][x_1(n-\tau)\mathrm{e}^{-\mathrm{j}\phi_q}]^*\} && (2.9)\\ &= \mathbb{E}\{x_1(n)x_1^*(n)\}\mathrm{e}^{-\mathrm{j}(\phi_p-\phi_q-2\pi f_0\tau)} \\ &= \mathbb{E}\{x_1(n)x_1^*(n)\}\mathrm{e}^{-\mathrm{j}\phi_{p,q}}e^{2\pi f_0\tau} \ , && (2.10)\end{aligned}$$

while noting that the expectation operation is performed over all time samples n and the terms inside the expectation are independent of τ. The phase difference between the $(p,q)^{\mathrm{th}}$ sensor pair, $\phi_{p,q} = \phi_p - \phi_q$ depends only on the array geometry and is associated with the time difference of arrival (TDoA). Since the array geometry is fixed and known *a priori*, $\phi_{p,q}$ is constant and can be computed at any τ to perform the necessary phase correction. It is therefore, convenient to choose $\tau = 0$ so that the instantaneous covariance matrix in (2.3), i.e., $\mathbf{R}_{\mathbf{xx}}(0)$, is used.

However, when broadband signals such as speech are involved, different frequency components, f, are affected by different phase shifts at the same time shift, τ_q, according to $\phi_q = 2\pi f \tau_q$. This implies that phase corrections for different frequency components at different time lags need to be adjusted for, and therefore, correlations across different sensors and temporal lags as in (2.7) need to be considered.

The z-transform of (2.7) [30],

$$\mathcal{R}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau) z^{-\tau} \ , \tag{2.11}$$

also denoted by $\mathbf{R}(\tau) \circ\!\!-\!\!\bullet \mathcal{R}(z)$ in short, is a para-Hermitian polynomial matrix. Consider example 2 from [30] which is given by

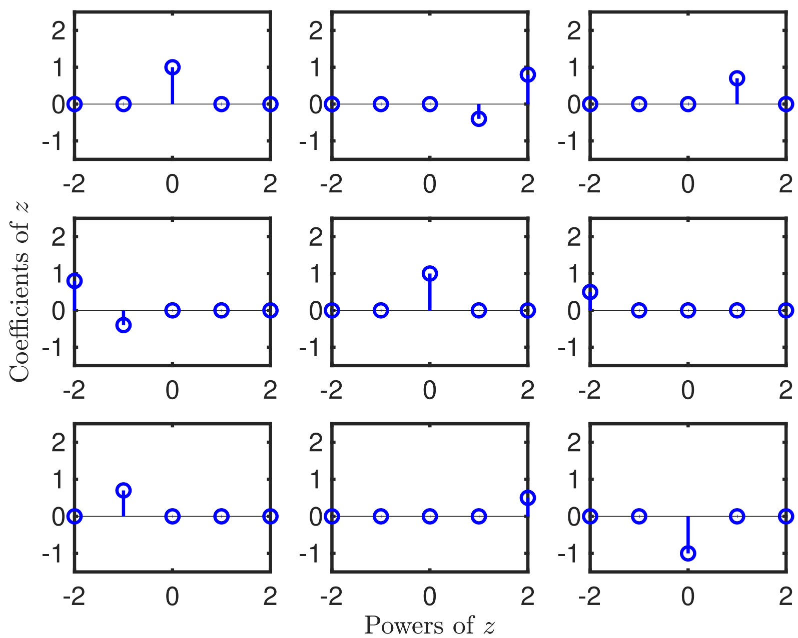
$$\mathcal{R}_2(z) = \begin{bmatrix} 1 & 0.8z^2 - 0.4z & 0.7z \\ 0.8z^{-2} - 0.4z^{-1} & 1 & 0.5z^{-2} \\ 0.7z^{-1} & 0.5z^2 & -1 \end{bmatrix} . \tag{2.12}$$

The same polynomial matrix can be interpreted as a matrix with polynomial elements as shown in Fig. 2.2(a) or a polynomial with matrix coefficients in Fig. 2.2(b). The former interpretation expresses each element as the temporal relation between that sensor pair, e.g., $r_{p,q}(z)$, while the latter shows the evolution of the spatial relation between all sensors with z, e.g., $\mathbf{R}_{\mathbf{xx}}(\tau)$ is the coefficient of $z^{-\tau}$. The instantaneous covariance matrix in (2.3) is the matrix coefficient of z^0.

2.4 PEVD Algorithms in the z-Domain

Several ways of factorizing polynomial matrices include the PEVD in this section, and the polynomial QR [71] and SVD [52] in Section 2.6. The pioneering work on PEVD began with the SBR2 algorithm [30], which will first be reviewed, and has inspired the SMD algorithm [28], another landmark work widely used in this thesis. The SBR2 and SMD classes of approaches have led to a family of PEVD algorithms in the z-domain.

(a) Matrix with polynomial elements.



(b) Polynomial with matrix coefficients.

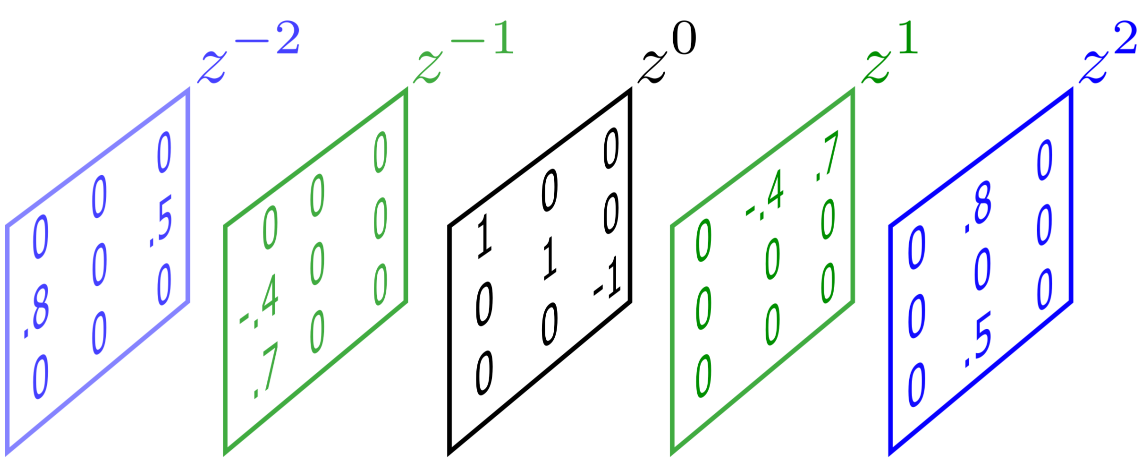


Figure 2.2: Different visualization and interpretation of (2.12).

2.4.1 Second-order Sequential Best Rotation for Symmetric PEVD

The PEVD of a para-Hermitian polynomial matrix is numerically computed using [30]

$$\boldsymbol{\mathcal{R}}(z) \approx \boldsymbol{\mathcal{U}}(z)\mathbf{\Lambda}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) = \boldsymbol{\mathcal{W}}^{\mathrm{P}}(z)\mathbf{\Lambda}(z)\boldsymbol{\mathcal{W}}(z) \ , \tag{2.13}$$

where the columns of the polynomial matrix, $\boldsymbol{\mathcal{U}}(z)$, correspond to the eigenvectors with their associated eigenvalues on the diagonal polynomial matrix, $\mathbf{\Lambda}(z)$, and the rows of $\boldsymbol{\mathcal{W}}(z) = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)$ contains the eigenvectors. In this thesis, both forms of the eigenvectors are used — $\boldsymbol{\mathcal{W}}(z)$ is adopted in SBR2 [30] and will be used when describing algorithmic enhancement; $\boldsymbol{\mathcal{U}}(z)$ will be used in applications since this follows standard linear algebra and signal processing conventions [11, 72]. Furthermore, $\boldsymbol{\mathcal{R}}(z)$ is also para-Hermitian and this is defined as

$$\boldsymbol{\mathcal{R}}(z) = \boldsymbol{\mathcal{R}}^{\mathrm{P}}(z) = \boldsymbol{\mathcal{R}}^{\mathrm{H}}(1/z^*) \ , \tag{2.14}$$

where $[\cdot]^{\mathrm{P}}$ is the para-Hermitian operator, that is complex conjugate or Hermitian transpose, followed by a time-reversal. Akin to conventional matrices, similarity transformations do not change the eigenvalues and hence numerical diagonalization is based on

$$\mathbf{\Lambda}(z) \approx \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{R}}(z)\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{W}}(z)\boldsymbol{\mathcal{R}}(z)\boldsymbol{\mathcal{W}}^{\mathrm{P}}(z) \ , \tag{2.15}$$

where the construction of $\boldsymbol{\mathcal{U}}(z)$ and $\boldsymbol{\mathcal{W}}(z)$ satisfies the para-unitary or lossless condition [45], given by

$$\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{U}}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) = \boldsymbol{\mathcal{W}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{W}}(z) = \boldsymbol{\mathcal{W}}(z)\boldsymbol{\mathcal{W}}^{\mathrm{P}}(z) = \mathbf{I} \ . \tag{2.16}$$

A. Para-unitary Polynomial Matrices

One such unitary matrix takes on the form of a Given's rotation [12],

$$\mathbf{G}_{(p,q)}(\theta,\phi) = \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \cos\theta & \mathbf{0} & \sin\theta e^{\mathrm{j}\phi} & \vdots \\ \vdots & \mathbf{0} & \mathbf{I}_{q-p-1} & \mathbf{0} & \vdots \\ \vdots & -\sin\theta e^{-\mathrm{j}\phi} & \mathbf{0} & \cos\theta & \vdots \\ \mathbf{0} & \dots & \dots & \dots & \mathbf{I}_{Q-q} \end{bmatrix} \in \mathbb{C}^{Q\times Q}\,, \tag{2.17}$$

designed to zero the off-diagonal element pair such that

$$\mathbf{G}_{(p,q)}(\theta,\phi) \begin{bmatrix} \ddots & \vdots & & \vdots & \iddots \\ \dots & r_{p,p} & \dots & r_{p,q} & \dots \\ & \vdots & \ddots & \vdots & \\ \dots & r_{q,p} & \dots & r_{q,q} & \dots \\ \iddots & \vdots & & \vdots & \ddots \end{bmatrix} \mathbf{G}^{\mathrm{H}}_{(p,q)}(\theta,\phi) = \begin{bmatrix} \ddots & \vdots & & \vdots & \iddots \\ \dots & r'_{p,p} & \dots & 0 & \dots \\ & \vdots & \ddots & \vdots & \\ \dots & 0 & \dots & r'_{q,q} & \dots \\ \iddots & \vdots & & \vdots & \ddots \end{bmatrix}\,, \tag{2.18}$$

where $r_{p,q}$ is the p-th row, q-th column matrix element and $[.]'$ denotes the resulting sub-matrix after multiplications. The values in (2.17) satisfying (2.18) are $\phi = \arg(r_{p,q})$ and $\tan 2\theta = \frac{2|r_{p,q}|}{r_{p,p}-r_{q,q}}$, where $\arg(\cdot)$ is the complex argument. Furthermore, the diagonal elements, $r'_{p,p}$ and $r'_{q,q}$ can be ordered in decreasing magnitude if θ is computed using the four quadrant inverse tangent to avoid permutation ambiguity arising from arbitrary ordering. If $r_{p,p} = 0$, the permutation matrix can be chosen to reorder the elements. The Givens rotation in (2.17) can be viewed as a degenerate polynomial matrix, in which the polynomial matrix only has non-zero coefficients for z^0, and also satisfies (2.16).

To move elements from one lag plane to another (for example z^{-2} to z^0) and change the polynomial order as shown in Fig. 2.2(b), the τ-th delay polynomial matrix applied to the

p-th row (column) upon a pre(post)-multiplication while satisfying(2.16) is

$$\mathscr{D}_{(p,\tau)}(z) = \mathrm{diag}([1, \ldots, 1, z^{-\tau}, 1 \ldots, 1]) = \begin{bmatrix} \mathbf{I}_{p-1} & 0 & 0 \\ 0 & z^{-\tau} & 0 \\ 0 & 0 & \mathbf{I}_{Q-p} \end{bmatrix}. \tag{2.19}$$

The operator, diag(·), forms a diagonal matrix from a vector argument, which is not one only at the p-th element. Although the goal of a PEVD algorithm is to diagonalize the polynomial matrix, $\mathscr{R}(z)$, it may be tempting to assume that the algorithm performs multiple EVD to different lags, based on Fig. 2.2(b). This is, however, not the principle in which PEVD adopts, since the elements in each plane are coupled. One can clearly observe that a diagonal matrix multiplication applied to the left of the polynomial matrix will scale all polynomial elements in a row-wise manner, demonstrating the coupling effect in polynomials.

B. SBR2 Algorithm

The SBR2 algorithm [30], used for PEVD, extends Jacobi's method for EVD to PEVD by factorizing (2.15) using a series of para-unitary transformations satisfying (2.16). Each para-unitary transformation can be interpreted as the transfer of energy from one channel to another while preserving the total energy of the system in all channels.

At each iteration, SBR2 begins by searching for the largest off-diagonal element, $|g|$, across all powers of z. $|g|$ is indexed by the p-th row, q-th column and the z^{τ}-th plane. This is compared against a predefined threshold, which is a user-defined positive parameter for the algorithm, δ. If $|g|$ exceeds δ, the element is brought to the principal plane, the plane of z^0, using a delay polynomial matrix, $\mathscr{D}_{(q,\tau)}(z)$, in (2.19). A Givens rotation, $\mathbf{G}_{(p,q)}(\theta, \phi)$, computed based on the principal plane using (2.18), is then applied to the entire polynomial matrix. Consequently, $|g|$ on the principal plane is zeroed out. Due to symmetry, the search is confined to half the off-diagonals and similarity transformation results in operations acting on dominant pairs rather than just elements.

The algorithm continues until the maximum off-diagonal element is less than the user-defined threshold value δ or the number of iterations ℓ reaches the user-defined limit. Upon reaching the stopping criterion, the overall para-unitary matrix which diagonalizes the polynomial matrix approximately after L iterations is

$$\boldsymbol{\mathcal{W}}(z) = \boldsymbol{\mathcal{W}}_L(z)\boldsymbol{\mathcal{W}}_{L-1}(z)\cdots\boldsymbol{\mathcal{W}}_1(z)\ , \tag{2.20}$$

where $\boldsymbol{\mathcal{W}}_\ell(z) = \mathbf{G}^{(\ell)}\boldsymbol{\mathcal{D}}^{(\ell)}(z)$ is the product of rotation and delay matrices during the ℓ-th iteration and satisfies (2.16).

Naturally, the number of delay operations may increase as the algorithm progresses with more iterations. Consequently, the order of the polynomial matrix may grow and become unnecessarily large. This will increase the search and computation times as well as storage requirements. For practical applications, $\tilde{\boldsymbol{\mathcal{W}}}(z)$ and $\tilde{\boldsymbol{\Lambda}}(z)$, approximations of $\boldsymbol{\mathcal{W}}(z)$ and $\boldsymbol{\Lambda}(z)$ in (2.13), are usually sufficient. This has led to the use of a reconstruction error measure, $\epsilon = \sum_{\forall z} \|\tilde{\boldsymbol{\mathcal{R}}}(z) - \boldsymbol{\mathcal{R}}(z)\|_F$, where $\|.\|_F$ is the Frobenius-norm (F-norm), which indicates the accuracy of the approximation. The error ϵ can also be expressed as a percentage of $\sum_{\forall z} \|\boldsymbol{\mathcal{R}}(z)\|_F$.

Several methods for trimming polynomial terms with small coefficients to keep the order compact have been proposed [27, 54, 73] but the energy-based technique [74] was chosen in SBR2. For a given ratio of squared F-norm, or energy permitted to be lost, $\tilde{\boldsymbol{\mathcal{R}}}(z)$ is approximated by the smallest $\tau_{\max}$ satisfying the inequality expressed as

$$\text{trim}(\boldsymbol{\mathcal{R}}(z), \mu) = \sum_{\tau=-\tau_{\max}}^{\tau_{\max}} \|\mathbf{R}(\tau)z^{-\tau}\|_F^2 \geq (1-\mu) \sum_{\tau=-\infty}^{\infty} \|\mathbf{R}(\tau)z^{-\tau}\|_F^2\ , \tag{2.21}$$

where μ is the trim parameter and $\mathbf{R}(\tau)$ are the matrix coefficients. This technique is energy-based because the sum of the squared F-norms across all the z planes is akin to the total energy of the signals in all channels. Planes that lie within $\pm\tau_{\max}$ and retain at least $(1-\mu)$ of the total energy are kept while planes that lie outside $\pm\tau_{\max}$ and accumulate up to μ of the total energy are discarded.

C. Convergence of the SBR2 Algorithm

To outline the proof for the convergence of the algorithm, the following quantities are defined as: the trace-norm squared of the principal plane [30], expressed as

$$N_1 = \sum_{p=1}^{Q} \left| r_{p,p}(0) \right|^2 , \tag{2.22}$$

the F-norm squared of the principal plane, written as

$$N_2 = \left\| \mathbf{R}(0) \right\|_F^2 , \tag{2.23}$$

the F-norm squared of the off-diagonal elements on the principal plane, given by

$$N_3 = \sum_{p=1}^{Q} \sum_{\substack{q=1 \\ p \neq q}}^{Q} |r_{p,q}(0)|^2 = N_2 - N_1 . \tag{2.24}$$

As $\mathcal{U}(z)$ is constructed according to (2.16), the square of the total F-norm of a polynomial matrix defined as

$$N_4 = \sum_{\forall \tau} \left\| \mathbf{R}(\tau) \right\|_F^2 , \tag{2.25}$$

is invariant to any lossless transformation. After applying the delay polynomial matrices to bring the maximum off-diagonal pair onto the principal plane followed by a Givens rotation, N_1 will increase by $2|g|^2$ at every iteration. For the ℓ-th iteration,

$$N_1^{(\ell)} = N_1^{(\ell-1)} + 2|g_\ell|^2 , \tag{2.26}$$

where $|g_\ell|$ is the magnitude of the largest off-diagonal element at the ℓ-th iteration. Since N_1 is monotonically increasing and upper bounded by a constant N_4,

$$N_1^{(\ell)} \leq N_4^{(\ell)} \quad \forall \ell , \tag{2.27}$$

it must have a supremum, S, such that

$$S = \sup_{\ell} N_1^{(\ell)} \ . \tag{2.28}$$

Consequently, for any $\epsilon > 0$, there must be an iteration number L, for which

$$|S - N_1^{(L)}| < 2\epsilon \ . \tag{2.29}$$

Furthermore, at subsequent iterations, $(L+\iota)$, $\iota > 0$, the maximum off-diagonal pair must satisfy

$$2|g_{L+\iota}|^2 \leq |S - N_1^{(L)}| < 2\epsilon \ , \tag{2.30}$$

implying that the maximum off-diagonal element squared $|g_L|^2$ is bounded by ϵ. Even though the trace-norm of the principal plane, N_1, increases monotonically with more iterations, the Givens rotation applied to other z-planes may increase the magnitude of the off-diagonal elements, i.e., $|g|$ is not guaranteed to decrease with ℓ.

Upon convergence of the SBR2 algorithm after L iterations, the original $\mathcal{R}(z)$ is diagonalized and approximates the eigenvalue polynomial matrix, $\mathbf{\Lambda}(z)$. The accumulated para-unitary matrices used for numerical diagonalization via similarity transformations in (2.15) give the polynomial eigenvector $\mathcal{W}(z)$.

2.4.2 Sequential Matrix Diagonalization Algorithm

The SMD algorithm is motivated by faster diagonalization using fewer number of iterations than SBR2. To achieve this, it aims to transfer more off-diagonal components onto the principal plane and diagonalize the principal plane in one-step.

A. Search and Delay Approach

During the ℓ-th iteration of the SBR2 algorithm, $\mathcal{W}_\ell(z)$ takes the form of a delay $\mathcal{D}_{q,\tau}^{(\ell)}(z)$ and Givens rotation $\mathbf{G}^{(\ell)}(\theta, \phi)$. A similarity transformation applied to a polynomial matrix using delay matrices affects one row and one column according to

$$\mathcal{R}'^{(\ell)}(z) = \mathcal{D}^{(\ell)}_{q,\tau}(z)\mathcal{R}^{(\ell)}(z)\left[\mathcal{D}^{(\ell)}_{q,\tau}(z)\right]^{\mathrm{P}} . \tag{2.31}$$

This implies that the delay matrix in (2.19) brings multiple elements, including the maximum off-diagonal element $|g|$, from one plane to the principal plane. It is, therefore, more efficient to search for the dominant off-diagonal column (row) rather than only an element, $|g|$. The search used in SMD for the ℓ-th iteration is based on

$$\{q,\tau\} = \arg\max_{q,\tau} \left\| \bar{\mathbf{r}}^{(\ell)}_q(\tau) \right\|_2 , \tag{2.32}$$

where the L_2-norm for the q-th column without the diagonal element is

$$\left\| \bar{\mathbf{r}}^{(\ell)}_q(\tau) \right\|_2 = \sqrt{\sum_{\substack{p=1 \\ p\neq q}}^{Q} \left| r^{(\ell)}_{p,q}(\tau) \right|^2} . \tag{2.33}$$

B. One-Step Diagonalization

A Givens rotation used in SBR2 only targets a pair of off-diagonal elements on the principal plane. To achieve complete diagonalization of the zero-lag plane, the SMD uses the eigenvector matrix $\mathbf{W}^{(\ell)}$ computed by applying an EVD on the matrix coefficient of z^0, i.e., $\mathbf{R}'^{(\ell)}(0) = \left[\mathbf{W}^{(\ell)}\right]^{\mathrm{H}} \boldsymbol{\Lambda}'^{(\ell)}\mathbf{W}^{(\ell)}$. The eigenvector matrix, with eigenvectors lying in the rows of $\mathbf{W}^{(\ell)}$, is then applied to the matrix coefficients across all lags given by

$$\mathcal{R}^{(\ell+1)}(z) = \mathbf{W}^{(\ell)}\mathcal{R}'^{(\ell)}(z)\left[\mathbf{W}^{(\ell)}\right]^{\mathrm{H}} . \tag{2.34}$$

For the SMD algorithm, the overall para-unitary matrix in (2.20) at the ℓ-th iteration is $\mathcal{W}_\ell(z) = \mathbf{W}^{(\ell)}\,\mathcal{D}^{(\ell)}_{q,\tau}(z)$.

C. Convergence of the SMD Algorithm

To simplify the notation, the quantity in (2.33) is denoted as $\beta^{(\ell)} = \left\|\bar{\mathbf{r}}_q^{(\ell)}(\tau)\right\|_2^2$ for the ℓ-th iteration. Assuming that the principal plane has been diagonalized in the $(\ell-1)$-th iteration, applying the delay matrices shifts the dominant row and column to the principal plane such that

$$N_1^{(\ell)} = N_1^{(\ell-1)} + 2\beta^{(\ell)} \tag{2.35}$$

is obtained after diagonalization. This implies that $N_1^{(\ell)}$ increases monotonically with ℓ while being upper bounded by a constant N_4, similarly to the condition established for the SBR2 algorithm in (2.27). This defines the existence of a supremum S such that

$$S = \sup_{\ell} N_1^{(\ell)} \ . \tag{2.36}$$

Therefore, for any $\epsilon > 0$, there exists an iteration number L, for which

$$|S - N_1^{(L)}| < 2\epsilon \ . \tag{2.37}$$

At subsequent iterations, $(L+\iota)$, $\iota > 0$, the dominant off-diagonal column L_2-norm squared must also satisfy

$$2\beta^{(L+\iota)} \leq |S - N_1^{(L)}| < 2\epsilon \ , \tag{2.38}$$

implying that $\beta^{(L)}$ is bounded by ϵ. Although N_1 is guaranteed to increase with the number of iterations as shown in (2.35), the unitary matrix $\mathbf{W}^{(\ell)}$, which diagonalizes the zero-lag plane, may increase the L_2-norm of the columns (rows) in other planes and result in an increase in $\beta^{(\ell+1)}$.

Similar to SBR2, upon convergence, the numerically diagonalized polynomial matrix $\mathbf{\Lambda}(z)$ contains the eigenvalues, while the accumulated para-unitary matrices form the eigenvector polynomial matrix $\boldsymbol{\mathcal{W}}(z)$.

2.4.3 Family of z-Domain PEVD Algorithms

At each iteration, the four steps involved in SBR2 and SMD are search, delay, zero and trim. Different variations have resulted in a family of z-domain PEVD algorithms.

Steps 1 and 2: Search and Delay

In the original SBR2 algorithm, the element pair to be zeroed is selected by searching for the element with the largest magnitude across all z-planes. This is equivalent to a search strategy based on the L_∞-norm expressed as

$$\{p, q, \tau\} = \arg\max_{p,q,\tau} \left\| \mathbf{r}_{p,q}(\tau) \right\|_\infty . \tag{2.39}$$

Each Givens rotation, which only affects 2 rows and columns, is then used in SBR2 to zero the maximum off-diagonal pairs on the principal plane. This means that more Givens rotations based on the unaffected rows and columns can be computed in parallel to eliminate more off-diagonal pairs. This parallel processing idea motivated an alternative search strategies that involve identifying multiple elements using a reduced search method, hence the term multiple shift maximum element (MSME) in [26, 53]. Another search strategy which targets column-row pairs such as the SMD is the L_2-norm in (2.32) [28].

Step 3: Zero

The Givens rotation is used to zero the off-diagonal pair at each iteration in SBR2. Because zeroing occurs pairwise during each step and the rotation applied to other lags may undo previous diagonalization effort, SBR2 tends to take many iterations. To speed up convergence, one-step diagonalization using a Householder-like factorization has been proposed [25, 55]. However, the factorization has no closed form and a solution is obtained by solving an optimization problem. Another approach, the SMD, performs an EVD on the principal plane. The eigenvector unitary matrix diagonalizes the principal plane and is also applied to planes across all lags to achieve quicker convergence [28].

Step 4: Trim

The energy-based truncation trimming method, adopted in SBR2, is based on the retention of significant coefficients relative to the proportion of the F-norm of the polynomial matrix [74]. Alternative approaches include the fixed polynomial order truncation method in [71], row-correction truncation applied to individual eigenvectors in [73], and the fixed direction delay strategy in the order-controlled multiple-shift SBR2 [54].

2.5 PEVD in the Frequency-Domain

The PEVD of a para-Hermitian polynomial matrix, $\mathcal{R}(z) \bullet\!\!-\!\!\circ \mathbf{R}(\tau)$, is [48,49]

$$\mathcal{R}(z) = \mathcal{U}(z)\mathbf{\Lambda}(z)\mathcal{U}^{\mathrm{P}}(z) \in \mathbb{C}^{Q\times Q} , \tag{2.40}$$

where both the eigenvector and eigenvalue polynomial matrices, $\mathcal{U}(z)$ and $\mathbf{\Lambda}(z)$, respectively, are assumed to exist as absolutely convergent Laurent series. Assuming that the region of convergence includes the unit circle, (2.40) can also be computed in the frequency-domain by evaluating at K discrete points on the unit circle such that [58,75]

$$\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega_k}) = \mathcal{R}(z)\Big|_{z=\mathrm{e}^{\mathrm{j}\Omega_k}} = \sum_{\tau} \mathbf{R}(\tau)\mathrm{e}^{-\mathrm{j}\Omega_k\tau}, \quad k = 0, 1, \ldots, K-1 , \tag{2.41}$$

where $\Omega_k = \frac{2\pi k}{K}$. The K-point DFT of (2.40) becomes [48]

$$\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega_k}) = \mathbf{U}(\mathrm{e}^{\mathrm{j}\Omega_k})\mathbf{\Lambda}(\mathrm{e}^{\mathrm{j}\Omega_k})\mathbf{U}^{\mathrm{H}}(\mathrm{e}^{\mathrm{j}\Omega_k}), \quad k = 0, 1, \ldots, K-1 , \tag{2.42}$$

approximating a PEVD using K independent EVDs pointwise in frequency. Like $\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega_k})$, the computed $\mathbf{U}(\mathrm{e}^{\mathrm{j}\Omega_k})$ and $\mathbf{\Lambda}(\mathrm{e}^{\mathrm{j}\Omega_k})$ in (2.42) are first assumed to be 2π periodic. These computations performed independently result in permutation ambiguity between channels across frequencies for the eigenvalues as well as the loss of phase coherence of the eigenvectors between frequency bins [58,75]. This ambiguity can be addressed by enforcing certain permutation constraints described below, while the phase alignment of eigenvectors between neighbouring bins can be formulated as an optimization problem [58,76,77].

2.5.1 Existence and Uniqueness of a PEVD

The polynomial matrix $\boldsymbol{\mathcal{R}}(z)$ is a matrix-valued function containing elements that are functions of the complex variable z [48, 78–80]. The conditions for the existence and uniqueness of the polynomial eigenvalues and eigenvectors are established for an analytic $\boldsymbol{\mathcal{R}}(z)$ [48], excluding the subband coding case in [49]. This section summarizes the results in [48, 49], beginning with a brief description of analytic functions.

Analytic Functions. Consider an analytic function $f(x)$ of a real-valued variable x. Analyticity results in a continuous, maximally smooth and infinitely differentiable function which implies the existence of a convergent power series or a Taylor series expansion about a point x_0 expressed as [80–82]

$$f(x) = \varphi_0 + \varphi_1(x - x_0) + \varphi_2(x - x_0)^2 + \cdots = \sum_{\ell=0}^{\infty} \varphi_\ell (x - x_0)^\ell \; . \tag{2.43}$$

The ℓ-th coefficient of the expansion in (2.43) is

$$\varphi_\ell = \frac{1}{\ell!} \frac{d^{(\ell)} f(x)}{dx^{(\ell)}}\bigg|_{x=x_0} = \frac{1}{\ell!} f^{(\ell)}(x_0) \; , \tag{2.44}$$

where $f^{(\ell)}(x_0)$ denotes the ℓ-th order derivative evaluated at x_0, $\ell = 0, 1, \ldots$, demonstrating the infinitely differentiable property of the analytic $f(x)$ for the real-valued x.

Similarly, for a complex-valued variable $z = x + \mathrm{j}y$ with $\Re(z) = x, \Im(z) = y$, the analytic or holomorphic function on some defined neighbourhood of z_0 is continuous and infinitely differentiable [83, 84]. For a given $\epsilon > 0$, the neighbourhood of z_0 satisfies $|z - z_0| < \epsilon$. A complex function $f(z) = u(x, y) + \mathrm{j}v(x, y)$, expressed in real and imaginary parts and is continuous in a region $\mathcal{C} \in \mathbb{C}$, is also continuous at $z_0 \in \mathcal{C}$ when

$$\lim_{z \to z_0} f(z) = f(z_0) \; , \tag{2.45}$$

and its derivative at z_0 is the limit which converges to

$$f^{(1)}(z_0) = \lim_{\Delta z \to 0} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z} \ , \tag{2.46}$$

where $\Delta z = z - z_0, \ z \neq z_0$. By writing $z_0 = x_0 + \mathrm{j}y_0$ and computing (2.46) for the real and imaginary parts separately, a necessary condition for differentiability at z_0 is the satisfaction of the Cauchy-Riemann equations at z_0 [83], given by

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \ , \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \ . \tag{2.47}$$

Therefore, if the function $f(z)$ is analytic on $\mathcal{C}$, (2.47) holds for every point in $\mathcal{C}$.

This thesis considers analytic matrix functions $\boldsymbol{\mathcal{A}}(z) : \mathbb{C} \to \mathbb{C}^{Q \times Q}$ which contain as elements, Laurent polynomials $a(z)$, with the unit circle included in the region of convergence. The time-domain representation $a(n) \circ\!\!-\!\!\bullet \ a(z)$ can be obtained from the inverse z-transform [83, 85], that is

$$a(n) = \frac{1}{2\pi \mathrm{j}} \oint_{\mathcal{C}} a(z) z^{n-1} dz \ , \tag{2.48}$$

where $\mathcal{C}$ represents an anti-clockwise closed contour within the region of convergence of the z-transform. By selecting the unit circle as the closed contour, the inverse z-transform becomes the inverse Fourier transform [85–87]

$$a(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} a(\mathrm{e}^{\mathrm{j}\Omega}) \mathrm{e}^{\mathrm{j}\Omega n} d\Omega \ , \tag{2.49}$$

where $a(\mathrm{e}^{\mathrm{j}\Omega})$ is assumed to be 2π periodic with a finite number of singularities. Then, the Laurent or power series is convergent and absolutely summable, e.g., $\sum_n |a(n)| < \infty$, thus guaranteeing uniqueness in the unit circle [48], i.e., $\mathbf{A}(\mathrm{e}^{\mathrm{j}\Omega}) = \boldsymbol{\mathcal{A}}(z)\Big|_{z=\mathrm{e}^{\mathrm{j}\Omega}}$. For absolute summability, a sufficient condition in the frequency domain is the so-called Hölder continuity [88], expressed as

$$\sup_{\Omega_1, \Omega_2 \in \mathbb{R}} \left| a(\mathrm{e}^{\mathrm{j}\Omega_1}) - a(\mathrm{e}^{\mathrm{j}\Omega_2}) \right| \leq C \left| \mathrm{e}^{\mathrm{j}\Omega_1} - \mathrm{e}^{\mathrm{j}\Omega_2} \right|^{\alpha} \ , \tag{2.50}$$

where $\alpha > 0.5$ and $C \in \mathbb{R}$ is a constant that depends on α.

In other words, if an arbitrary function $a(\mathrm{e}^{\mathrm{j}\Omega})$ is sufficiently smooth or Hölder continuous in the frequency-domain, the absolute summability of $a(n)$ is guaranteed in the time-domain. Then, the Fourier series can be used, implying the existence of a convergent Laurent series $a(z)$. Additionally, if the unit circle can be selected as the closed contour for the analytic continuation of $a(\mathrm{e}^{\mathrm{j}\Omega})$ via $z = \mathrm{e}^{\mathrm{j}\Omega}$, $a(z)$ exists as a convergent power series.

For an analytic para-Hermitian $\boldsymbol{\mathcal{R}}(z)$ which converges on the unit circle, EVDs can be computed pointwise on the unit circle to generate $\mathbf{\Lambda}(\mathrm{e}^{\mathrm{j}\Omega_k})$ and $\mathbf{U}(\mathrm{e}^{\mathrm{j}\Omega_k})$. When $\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega})$ is perturbed by $\Delta\Omega$, changes in its eigenvalues are bounded by [89]

$$\sum_{q=1}^{Q} \left|\lambda_q(\mathrm{e}^{\mathrm{j}\Omega}) - \lambda_q(\mathrm{e}^{\mathrm{j}(\Omega+\Delta\Omega)})\right| \leq \left\|\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega}) - \mathbf{R}(\mathrm{e}^{\mathrm{j}(\Omega+\Delta\Omega)}\right\|_F . \tag{2.51}$$

Since $\boldsymbol{\mathcal{R}}(z)$ is infinitely smooth and continuous, the term on the right of (2.51) vanishes as $\Delta\Omega \to 0$. This also implies the continuity of eigenvalues, i.e., left hand side of (2.51) goes to zero as $\Delta\Omega \to 0$. Hence, $\lambda_q(\mathrm{e}^{\mathrm{j}\Omega})$ $\forall q$, can always be chosen to be infinitely differentiable and analytic. The existence and uniqueness of $\boldsymbol{\mathcal{U}}(z)$ and $\mathbf{\Lambda}(z)$ in (2.40) as convergent power series can be deduced by inspecting K discrete points, $\mathbf{U}(\mathrm{e}^{\mathrm{j}\Omega_k})$ and $\mathbf{\Lambda}(\mathrm{e}^{\mathrm{j}\Omega_k})$ in (2.42), and can even be recovered via analytic continuation $z = \mathrm{e}^{\mathrm{j}\Omega}$. The results discuss the two cases in [48, 49].

Case 1: Distinct Eigenvalues. The distinct and non-overlapping eigenvalues $\lambda_q(\mathrm{e}^{\mathrm{j}\Omega})$, $q = 1, \ldots, Q$, computed from the EVD on the unit circle of an analytic $\boldsymbol{\mathcal{R}}(z)$, could be chosen to be analytic and ordered decreasingly for uniqueness, i.e., spectrally majorized for all frequencies Ω with strict inequality. Consequently, the eigenvalues in $\mathbf{\Lambda}(z)$ exist as unique analytic Laurent series and match $\mathrm{diag}([\lambda_1(\mathrm{e}^{\mathrm{j}\Omega}), \ldots, \lambda_Q(\mathrm{e}^{\mathrm{j}\Omega})])$ on the unit circle.

When eigenvalues are distinct and ordered, each ordered eigenvector can have an arbitrary phase response due to phase ambiguity described in Section 2.3.1. Then, as long as the phase response is chosen for $\mathbf{u}_q(\mathrm{e}^{\mathrm{j}\Omega})$ $\forall q$ to satisfy Hölder continuity, an absolutely convergent Laurent series for $\boldsymbol{u}_q(z)$ exists and matches with $\mathbf{u}_q(\mathrm{e}^{\mathrm{j}\Omega})$ on the unit circle.

Case 2: Eigenvalues with Algebraic Multiplicities. Here, eigenvalues are constrained to be continuous and, therefore, piecewise analytic on the unit circle. Then, permutations at points with multiplicity greater than one, i.e. intersection points, can be selected to be analytic if eigenvalues satisfy Lipschitz continuity for $m, n = \{1, \ldots, Q\}$ [48]

$$\sup_{\Omega_1, \Omega_2 \in \mathbb{R}} \left| \lambda_m(\mathrm{e}^{\mathrm{j}\Omega_1}) - \lambda_n(\mathrm{e}^{\mathrm{j}\Omega_2}) \right| \leq L \left| \mathrm{e}^{\mathrm{j}\Omega_1} - \mathrm{e}^{\mathrm{j}\Omega_2} \right|^{\alpha} . \tag{2.52}$$

Lipschitz continuity is stricter than Hölder continuity and is equal to (2.50) when $\alpha = 1$ and the constant is equal to the Lipschitz constant, i.e., $C = L$, defined as

$$L = \max_{\forall q, \Omega \in \mathbb{R} \setminus \mathcal{m}} \left| \frac{d}{d\Omega} \lambda_q(\mathrm{e}^{\mathrm{j}\Omega}) \right| , \tag{2.53}$$

which is computed over the set of frequencies Ω excluding $\mathcal{m}$ containing points with multiplicity greater than one. Consequently, eigenvalues $\lambda_q(\mathrm{e}^{\mathrm{j}\Omega})$, $q = 1, \ldots, Q$, which are formed from permutations of piecewise analytic segments, are Lipschitz continuous, and $\boldsymbol{\Lambda}(z)$ can exist as absolutely convergent Laurent series. The uniqueness of $\boldsymbol{\Lambda}(z)$ can be enforced by imposing constraints such as maximum smoothness or spectral majorization which will directly affect the order of the Laurent polynomial approximation of $\boldsymbol{\Lambda}(z)$.

Omitting the case when eigenvalues are identical, i.e., $\lambda_m(\mathrm{e}^{\mathrm{j}\Omega}) = \lambda_n(\mathrm{e}^{\mathrm{j}\Omega})\ \forall\Omega,\ n \neq m$, $m, n = \{1, \ldots, Q\}$, analytic eigenvectors exist when eigenvalues are selected to be analytic [50]. Each eigenvector incurs an ambiguity in phase rotation, providing some freedom in the choice of the phase response. If the phase response is selected such that $\mathbf{u}_q(\mathrm{e}^{\mathrm{j}\Omega})\ \forall q$ is Hölder continuous or even analytic in Ω, a convergent Laurent series $\boldsymbol{\mathcal{U}}(z)$ exists.

If the phase response of $\mathbf{u}_q(\mathrm{e}^{\mathrm{j}\Omega})$ is piecewise continuous with a finite number of discontinuities, a convergent Laurent series does not exist. Discontinuities can also arise when enforcing spectrally majorized for eigenvalues with algebraic multiplicity greater than one. However, due to periodicity in Ω, the Laurent polynomial $\hat{\boldsymbol{\mathcal{U}}}(z)$ can be approximated by the Fourier series, which converges to $\mathbf{U}(\mathrm{e}^{\mathrm{j}\Omega})$ on the unit circle except at those discontinuous points, where they converge to the mean values before and after the jumps.

Non-existence of Analytic Eigenvalues. The K-point DFT computed in (2.41) implies that $\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega})$ is 2π periodic, which must be satisfied at every frequency Ω_k such that $\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega_k}) = \mathbf{R}(\mathrm{e}^{\mathrm{j}(\Omega_k+2\pi N)})\ \forall k,\ N \in \mathbb{N}$. The observed set of points for the eigenvalues $\{\lambda_q(\mathrm{e}^{\mathrm{j}\Omega_k}) \,|\, q \in \{1,\ldots,Q\}\}$ will be the same as $\{\lambda_\mu(\mathrm{e}^{\mathrm{j}(\Omega_k+2\pi N)}) \,|\, \mu \in \{1,\ldots,Q\}\}$. Because of possible permutations at every frequency to achieve analyticity, the direct matching of points, i.e., $\mu(q) = q$, does not necessarily arise and cannot be assumed. Consequently, the periodicity of analytic eigenvalues and eigenvectors is at least $2\pi N$ for some $N \in \mathbb{N}$, where (2.42) is a special case when $N = 1$. The analytic EVD on the unit circle is [48]

$$\mathbf{R}(\mathrm{e}^{\mathrm{j}\Omega}) = \mathbf{U}(\mathrm{e}^{\mathrm{j}\Omega/N})\mathbf{\Lambda}(\mathrm{e}^{\mathrm{j}\Omega/N})\mathbf{U}^{\mathrm{H}}(\mathrm{e}^{\mathrm{j}\Omega/N}) \ , \tag{2.54}$$

with its eigenvalues and eigenvectors being analytic in Ω and exhibiting the same periodicity. From the discrete frequency points on the unit circle, analytic continuation via $z = \mathrm{e}^{\mathrm{j}\Omega}$ can only be used to derive $\mathbf{\Lambda}(z)$ and $\boldsymbol{\mathcal{U}}(z)$ when $N = 1$. Hence, for $z^{1/N}, N > 1$, the analytic solutions do not exist but can be approximated by Laurent polynomials.

To obtain (2.54), $\boldsymbol{\mathcal{R}}(z)$ can first be under-sampled by a factor of N expressed as

$$\boldsymbol{\mathcal{R}}_\lambda(z^N) = \boldsymbol{\mathcal{U}}(z)\mathbf{\Lambda}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) \ , \tag{2.55}$$

followed by evaluating at N points on the unit circle. This will result in [49]

$$\mathbf{\Lambda}(z) = \mathrm{diag}([\lambda(z), \lambda(z\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}}), \lambda(z\mathrm{e}^{\mathrm{j}2\frac{2\pi}{N}}), \ldots, \lambda(z\mathrm{e}^{\mathrm{j}(N-1)\frac{2\pi}{N}})]) \ , \tag{2.56}$$

which contains, as elements, frequency-shifted versions of $\lambda(z)$. The para-unitary matrix can be constructed using $\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{D}}(z)\mathbf{F}$, where $\boldsymbol{\mathcal{D}}(z) = \mathrm{diag}([1, z^{-1}, \ldots, z^{-N+1}])$ is a delay matrix, and $\mathbf{F}$ is the N-point DFT matrix. In this case, $\boldsymbol{\mathcal{R}}_\lambda(z^N)$ is pseudo-circulant expressed as [45, 49, 90]

$$\mathcal{R}_\lambda(z) = \mathcal{D}(z)\mathbf{F}\boldsymbol{\Lambda}(z)\mathbf{F}^{\mathrm{H}}\mathcal{D}^{\mathrm{P}}(z) \tag{2.57}$$

$$= \begin{bmatrix} \phi_0(z) & \phi_1(z) & \dots & \phi_{N-1}(z) \\ z^{-1}\phi_{N-1}(z) & \phi_0(z) & \dots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ z^{-1}\phi_1(z) & \dots & z^{-1}\phi_{N-1}(z) & \phi_0(z) \end{bmatrix} , \tag{2.58}$$

where $\phi_n(z)$ is the n-th polyphase component. The matrix, $\mathbf{F}\boldsymbol{\Lambda}(z)\mathbf{F}^{\mathrm{H}}$, which the DFT matrix can diagonalize, is circulant and the elements are over-sampled N times but time-shifted [90]. The time shifts are then corrected by $\mathcal{D}(z)$. The pseudo-circulant matrix, typically encountered in subband coding problems [25, 90], has frequency-shifted eigenvalues in (2.56), of which the analytic solutions do not exist. Although analytic eigenvalues and, consequently, analytic eigenvectors do not exist, a spectrally majorized solution which is absolutely convergent on the unit circle can be computed [49].

2.5.2 Spectrally Majorized Solutions

The family of z-domain algorithms based on SBR2 and SMD approximates (2.40) using Laurent polynomial matrix factors, $\boldsymbol{\Lambda}(z)$ and $\boldsymbol{\mathcal{U}}(z)$ [48, 49, 60]. When evaluating the diagonal polynomial matrix $\boldsymbol{\Lambda}(z)$ on the unit circle as in (2.42), the eigenvalues are spectrally majorized for all normalized angular frequencies Ω such that [30, 90]

$$\lambda_{q-1}(\mathrm{e}^{\mathrm{j}\Omega}) \geq \lambda_q(\mathrm{e}^{\mathrm{j}\Omega}) \; , \quad q = 2, \dots, Q \; , \tag{2.59}$$

where $\lambda_q(\mathrm{e}^{\mathrm{j}\Omega})$ is the $(q, q)^{\mathrm{th}}$ element of $\boldsymbol{\Lambda}(z)|_{z=\mathrm{e}^{\mathrm{j}\Omega}}$, or the power spectral density (PSD) of the q-th signal which generated $\mathcal{R}(z)$. This ordering implies that the assigned first output channel is always associated with the largest eigenvalue for all frequencies and the next channel gets the next largest accordingly. Similar to EVD in linear algebra, the order of the eigenvectors and eigenvalues is essential in factorization [11]. For PEVD, this order is preserved by using the same channel-to-eigenvalue assignments for the eigenvectors. The polynomial factors generated by SBR2 and SMD tend to be of much higher polynomial order because they approximate non-differentiable or discontinuous functions [60].

2.5.3 Maximally Smooth Analytic Solutions

Assuming analytic eigenvalues and eigenvectors exist, there is a path of factorization for $\boldsymbol{\mathcal{R}}(z)$ generating maximally smooth $\boldsymbol{\mathcal{U}}(z)$ and $\boldsymbol{\Lambda}(z)$ [78]. If the underlying eigenvalues generating $\boldsymbol{\mathcal{R}}(z)$ are not spectrally majorized, the solution in Section 2.5.2 is a frequency-dependent permutation of the analytic solution and tend to be of much higher polynomial order [60]. The channel-to-eigenvalue associations are established to obtain analytic solutions from the discrete points evaluated on the unit circle to maximize a smoothness measure [59, 60]. Another approach uses the inner product between eigenvectors in adjacent frequencies to establish the associations [58, 59]. Regardless, the same associations are used to permute the eigenvalues and eigenvectors.

2.6 Other Decompositions and Polynomial Extensions

Apart from the EVD in linear algebra [11], other matrix factorizations such as the QR, SVD and generalized eigenvalue decomposition (GEVD) have been found to be useful for signal processing applications and will be summarized in this section [10, 72]. These techniques have also been extended to polynomial matrices and will be introduced.

2.6.1 Related Matrix Decompositions

A. QR Decomposition

Any general complex rectangular matrix $\mathbf{A} \in \mathbb{C}^{P\times Q}$ can be factorized into

$$\mathbf{A} = \mathbf{Q}\mathbf{T} \Longleftrightarrow \mathbf{Q}^{\mathrm{H}}\mathbf{A} = \mathbf{T}\,, \tag{2.60}$$

where $\mathbf{Q} \in \mathbb{C}^{P\times P}$ is a unitary matrix and $\mathbf{T} \in \mathbb{C}^{P\times Q}$ is an upper triangular matrix. The QR factorization can be computed using Gram-Schmidt orthogonalization, Householder reflections or Givens rotations. The latter two are often preferred for numerical stability

[11, 12]. QR factorization is often used for solving least-squares and some eigenvalue problems [10,12]. For example, consider the least squares problem, which seeks a solution $\hat{\mathbf{x}}$ minimizing $\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\|^2$. The minimization implies that the derivative w.r.t. $\mathbf{x}$ equals zero, which gives the normal equation expressed as

$$\mathbf{A}^{\mathrm{H}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^{\mathrm{H}}\mathbf{b} \implies \hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{b}$$

$$\hat{\mathbf{x}} = (\mathbf{T}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{Q}\mathbf{T})^{-1}(\mathbf{T}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}})\mathbf{b} \implies \hat{\mathbf{x}} = \mathbf{T}^{-1}(\mathbf{T}^{\mathrm{H}})^{-1}\mathbf{T}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{b} \tag{2.61}$$

$$\hat{\mathbf{x}} = \mathbf{T}^{-1}\mathbf{Q}^{\mathrm{H}}\mathbf{b} \iff \mathbf{T}\hat{\mathbf{x}} = \mathbf{Q}^{\mathrm{H}}\mathbf{b}\ , \tag{2.62}$$

and can therefore be solved by back substitution, which is fast [11].

B. Singular Value Decomposition

A general matrix $\mathbf{A} \in \mathbb{C}^{P\times Q}$ of rank M can be factorized by the SVD using [10,91]

$$\mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{V} = \mathbf{\Sigma} \iff \mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\ , \tag{2.63}$$

where diagonal matrix $\mathbf{\Sigma} \in \mathbb{C}^{P\times Q}$ contains M positive-valued singular values. The columns of $\mathbf{U} \in \mathbb{C}^{P\times P}$ contain the (left) singular vectors which lie in the column space of $\mathbf{A}$ and diagonalizes $\mathbf{A}\mathbf{A}^{\mathrm{H}}$ such that

$$\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{U}\mathbf{\Lambda}^{(1)}\mathbf{U}^{\mathrm{H}}, \tag{2.64}$$

where $\mathbf{\Lambda}^{(1)} \in \mathbb{C}^{P\times P}$ is a diagonal matrix of the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$. The elements of $\mathbf{\Lambda}^{(1)}$ and $\mathbf{\Sigma}$ are related by $\lambda_{i,i}^{(1)} = \sigma_{i,i}^2,\ i = 1, \ldots, M$. Similarly, the (right) singular vectors in the columns of $\mathbf{V} \in \mathbb{C}^{Q\times Q}$, lie in the row space of $\mathbf{A}$ and diagonalizes $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ such that

$$\mathbf{A}^{\mathrm{H}}\mathbf{A} = \mathbf{V}\mathbf{\Lambda}^{(2)}\mathbf{V}^{\mathrm{H}}, \tag{2.65}$$

where $\mathbf{\Lambda}^{(2)} \in \mathbb{C}^{Q\times Q}$ is a diagonal matrix containing the eigenvalues of $\mathbf{A}^{\mathrm{H}}\mathbf{A}$, and the elements of $\mathbf{\Lambda}^{(2)}$ and $\mathbf{\Sigma}$ are related by $\lambda_{i,i}^{(2)} = \sigma_{i,i}^2,\ i = 1, \ldots, M$.

The SVD is widely used in many signal processing applications including data compression and dimensionality reduction, channel identification, and signal enhancement and recovery for images, speech, biomedical data and communications [10, 72].

C. Generalized Eigenvalue Decomposition

The GEVD of two matrices $\mathbf{A} \in \mathbb{C}^{Q \times Q}$ and $\mathbf{B} \in \mathbb{C}^{Q \times Q}$ is defined as [12, 92]

$$\mathbf{AV} = \mathbf{BV\Lambda} \ , \tag{2.66}$$

where the generalized eigenvectors lie in the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_Q] \in \mathbb{C}^{Q \times Q}$ with the corresponding generalized eigenvalues found on the diagonal of $\mathbf{\Lambda} \in \mathbb{C}^{Q \times Q}$, i.e., $[\lambda_1, \ldots, \lambda_Q]^{\mathrm{T}}$. The expression (2.66) can also be written as

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{B}\mathbf{v}_i \Longleftrightarrow (\mathbf{A} - \lambda_i \mathbf{B})\mathbf{v}_i = \mathbf{0} \quad \forall i \in \{1, \ldots, Q\} \ . \tag{2.67}$$

In this case, the set of matrices $(\mathbf{A} - \lambda \mathbf{B})$, $\lambda \in \mathbb{C}$ is called the matrix pencil of $(\mathbf{A}, \mathbf{B})$. The generalized eigenvalues λ satisfy $\det(\mathbf{A} - \lambda \mathbf{B}) = 0$, where $\det(\cdot)$ computes the determinant of a matrix. The generalized eigenvectors are the non-zero vectors $\mathbf{v}$ satisfying (2.67) and cause $\mathbf{A}$ to behave like a scaled version of $\mathbf{B}$, where the scaling factor is given by the generalized eigenvalue. The non-trivial solution $(\mathbf{\Lambda}, \mathbf{V})$ is an eigenpair of $(\mathbf{A}, \mathbf{B})$.

Another interpretation of the GEVD is the joint diagonalization of matrices $\mathbf{A}$ and $\mathbf{B}$, which can be expressed as

$$\mathbf{V}^{\mathrm{H}}\mathbf{AV} = \mathbf{\Lambda} \ , \tag{2.68}$$

$$\mathbf{V}^{\mathrm{H}}\mathbf{BV} = \mathbf{I} \ . \tag{2.69}$$

By pre-multiplying $\mathbf{\Lambda}$ in (2.68) by $\mathbf{V}^{\mathrm{H}}\mathbf{BV}$ and both sides by $(\mathbf{V}^{\mathrm{H}})^{-1}$, (2.68) reduces to (2.66). This implies that the generalized eigenvectors can simultaneously diagonalize matrices $\mathbf{A}$ and $\mathbf{B}$ according to (2.68) and (2.69), respectively. If $\mathbf{B}$ is non-singular and invertible, (2.66) reduces to the standard EVD of $\mathbf{C}$ given by

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda} \implies \mathbf{C}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda} \Longleftrightarrow \mathbf{C} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} \;, \tag{2.70}$$

where $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}$, which may not be Hermitian symmetric. Consequently, the eigenvalues may be complex-valued and the eigenvectors may not be unitary and orthogonal. Note that the EVD is a particular case of GEVD when $\mathbf{B} = \mathbf{I}$ in (2.66).

One method to compute (2.66) is based on joint diagaonlization. A Cholesky factorization is first used to obtain [11, 12]

$$\mathbf{B} = \mathbf{T}^{\mathrm{H}}\mathbf{T} \;, \tag{2.71}$$

where $\mathbf{T}^{\mathrm{H}}$ and $\mathbf{T}$ are lower and upper triangular matrices, respectively. The second step is to form an intermediate Hermitian matrix $\mathbf{C}$ according to

$$\mathbf{C} = \left(\mathbf{T}^{\mathrm{H}}\right)^{-1}\mathbf{A}\mathbf{T}^{-1} \;. \tag{2.72}$$

In the final step, an EVD is performed on $\mathbf{C}$ such that

$$\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} \;. \tag{2.73}$$

The generalized eigenvalue matrix is given by $\boldsymbol{\Lambda}$ while the generalized eigenvectors lie in the columns of $\mathbf{V} = \mathbf{T}^{-1}\mathbf{U}$. Equating (2.72) and (2.73), pre-multiplying by $\mathbf{T}^{\mathrm{H}}$ and post-multiplying both sides by $\mathbf{U}$ gives

$$\left(\mathbf{T}^{\mathrm{H}}\right)^{-1}\mathbf{A}\mathbf{T}^{-1} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} \Longrightarrow \mathbf{T}^{\mathrm{H}}\left(\mathbf{T}^{\mathrm{H}}\right)^{-1}\mathbf{A}\mathbf{T}^{-1}\mathbf{U} = \mathbf{T}^{\mathrm{H}}\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}}\mathbf{U} \tag{2.74}$$

$$\mathbf{A}\mathbf{V} = \mathbf{T}^{\mathrm{H}}\mathbf{U}\boldsymbol{\Lambda} \Longrightarrow \mathbf{A}\mathbf{V} = \mathbf{B}\mathbf{V}\boldsymbol{\Lambda} \;, \tag{2.75}$$

since $\mathbf{B}\mathbf{V} = \left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right)\left(\mathbf{T}^{-1}\mathbf{U}\right) = \mathbf{T}^{\mathrm{H}}\mathbf{U}$.

D. Strong Uncorrelating Transforms

In Section 2.3, diagonalization of the instantaneous covariance matrix is achieved by an EVD. Consider a complex-valued random variable $x = u + \mathrm{j}v$ containing real-valued components u and v with zero mean i.e., $\mathbb{E}\{x\} = 0$. The second-order moment, conventionally computed using the complex outer product, is given by

$$C = \mathbb{E}\{xx^*\} = \mathbb{E}\{(u + \mathrm{j}v)(u - \mathrm{j}v)\} = \mathbb{E}\{uu^*\} + \mathbb{E}\{vv^*\} + \mathrm{j}0 = \sigma_u^2 + \sigma_v^2 \,, \tag{2.76}$$

where σ_u^2 and σ_v^2 are the variances of u and v, respectively. The quantity in (2.76) does not fully capture the correlation between the random variable x and its complex conjugate x^*. It requires the complementary variance Γ defined as [93, 94]

$$\Gamma = \mathbf{E}\{xx\} = \mathbf{E}\{(u + \mathrm{j}v)^2\} = \mathbb{E}\{u^2\} + \mathbb{E}\{(\mathrm{j}v)^2\} + \mathbb{E}\{2\mathrm{j}uv\} = \sigma_u^2 - \sigma_v^2 + \mathrm{j}2\rho_{uv} \,, \tag{2.77}$$

where $\rho_{uv} = \mathbb{E}\{uv\}$ describes the correlation between u and v. The complex random variable x is called proper or circular if Γ vanishes to zero and this only happens when $\sigma_u^2 = \sigma_v^2$ and $\rho_{uv} = 0$. If $\Gamma \neq 0$, x is improper or non-circular [84, 94–96].

These definitions completely capture the second-order moments and have also been extended to complex-valued random vectors $\mathbf{x} \in \mathbb{C}^Q$. To jointly diagonalize the conventional covariance, $\mathbf{C} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\}$, and the complementary or pseudo-covariance $\boldsymbol{\Gamma} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$, the strong uncorrelating transform (SUT) results in [93]

$$\boldsymbol{\Phi}\mathbf{C}\boldsymbol{\Phi}^{\mathrm{H}} = \mathbf{I}, \quad \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^{\mathrm{T}} = \boldsymbol{\Lambda} \,. \tag{2.78}$$

An alternate approach is to use an augmented vector $\boldsymbol{\chi} = [\mathbf{x}, \mathbf{x}^{\mathrm{H}}]^{\mathrm{T}} \in \mathbb{C}^{2Q}$ to form an augmented covariance matrix expressed as [95]

$$\mathbb{E}\{\boldsymbol{\chi}\boldsymbol{\chi}^{\mathrm{H}}\} = \begin{bmatrix} \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} & \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\} \\ \mathbb{E}\{\mathbf{x}^*\mathbf{x}^{\mathrm{H}}\} & \mathbb{E}\{\mathbf{x}^*\mathbf{x}^{\mathrm{T}}\} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}^* & \mathbf{C}^* \end{bmatrix} \in \mathbb{C}^{2Q \times 2Q} \,. \tag{2.79}$$

The augmented covariance captures both types of covariance matrices and can be jointly diagonalized by the Takagi factorization, which is a special type of SVD [95]. In practice, for univariate non-circular data, it is sufficient to only diagonalize the pseduo-covariance giving rise to the so-called approximate uncorrelating transform (AUT) [97].

The SUT has been used for speech analysis and voice activity detection [98], rank reduction and transform coding [93], where the processing usually involves complex numbers represented in the short-time Fourier transform (STFT) domain. However, in the applications presented in this thesis, the considered real-valued speech signals are processed in the time domain. Consequently, the complementary covariance vanishes and the second-order moments are completely described by the conventional covariance matrix. As such, this thesis will focus only on the covariance matrix.

2.6.2 Polynomial QR and Singular Value Decompositions

A. Polynomial QR Decomposition

Like QR factorization in linear algebra, the polynomial QR (PQR) decomposition of $\boldsymbol{\mathcal{A}}(z) \in \mathbb{C}^{P\times Q}$ aims to find a para-unitary $\boldsymbol{\mathcal{Q}}(z) \in \mathbb{C}^{P\times P}$ such that [52]

$$\boldsymbol{\mathcal{A}}(z) \approx \boldsymbol{\mathcal{Q}}(z)\boldsymbol{\mathcal{T}}(z) \Longleftrightarrow \boldsymbol{\mathcal{Q}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{A}}(z) \approx \boldsymbol{\mathcal{T}}(z) \;, \tag{2.80}$$

where $\boldsymbol{\mathcal{T}}(z) \in \mathbb{C}^{P\times Q}$ is an upper triangular polynomial matrix. In [52], $\boldsymbol{\mathcal{Q}}^{\mathrm{P}}(z)$ comprises a series of delays $\boldsymbol{\mathcal{D}}(z)$ and Givens rotations $\mathbf{G}$ designed to create zeros below the diagonals of $\boldsymbol{\mathcal{A}}(z)$ across all lags to give $\boldsymbol{\mathcal{T}}(z)$ upon convergence. Consider a 2×1 column polynomial vector example $\boldsymbol{a}(z) = [a_1(z), a_2(z)]^T$, where the dominant element on the lower triangle is the coefficient of $z^{-\tau}$, i.e., $a_2(\tau)$. The para-unitary $\boldsymbol{\mathcal{Q}}_\ell^{\mathrm{P}}(z)$ for the ℓ-th iteration is

$$\boldsymbol{\mathcal{Q}}_\ell^{\mathrm{P}}(z) = \begin{bmatrix} 1 & 0 \\ 0 & z^{-\tau} \end{bmatrix} \begin{bmatrix} \mathrm{e}^{\mathrm{j}\alpha}\cos\theta & \mathrm{e}^{\mathrm{j}\phi}\sin\theta \\ -\mathrm{e}^{-\mathrm{j}\phi}\sin\theta & \mathrm{e}^{-\mathrm{j}\alpha}\cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z^{\tau} \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\alpha}\cos\theta & \mathrm{e}^{\mathrm{j}\phi}\sin\theta\, z^{\tau} \\ -\mathrm{e}^{-\mathrm{j}\phi}\sin\theta\, z^{\tau} & \mathrm{e}^{-\mathrm{j}\alpha}\cos\theta \end{bmatrix} , \tag{2.81}$$

where the parameters θ, ϕ and α are computed based on the respective coefficients

$$\tan(\theta) = \frac{|a_2(\tau)|}{|a_1(0)|}, \quad \phi = -\arg(a_2(\tau)), \quad \alpha = -\arg(a_1(0)) \,. \tag{2.82}$$

The selection of these parameters is designed to zero the dominant element, $a_2(\tau)$. The leftmost delay factor in (2.81) restores the original coefficients of z^0 to their original positions following an initial delay that brings the dominant element to the principal plane and a Givens rotation, represented by the third and second polynomial matrices, respectively. When the algorithm converges after L iterations with the polynomial elements beneath the diagonal becoming sufficiently small in magnitude, the overall para-unitary matrix $\mathcal{Q}^{\mathrm{P}}(z) = \mathcal{Q}_L^{\mathrm{P}}(z) \ldots \mathcal{Q}_1^{\mathrm{P}}(z)$ results in

$$\mathcal{Q}^{\mathrm{P}}(z) \begin{bmatrix} a_1(z) \\ a_2(z) \end{bmatrix} \approx \begin{bmatrix} t_1(z) \\ 0 \end{bmatrix} \,, \tag{2.83}$$

where $t_1(z)$ is the element in the upper triangular polynomial matrix $\mathcal{T}(z)$. Note that $\boldsymbol{a}(z)$ can be extended to describe arbitrary vectors in $\mathbb{C}^Q$ and matrices $\mathcal{A}(z)$ in $\mathbb{C}^{P\times Q}$.

PQR decomposition, like its linear algebra counterpart, is related to the eigenvalue problem [12] and has been used for MIMO channel equalization and decoding [51, 52].

B. Polynomial Singular Value Decomposition

The para-unitary polynomial matrices constructed from delay and Givens rotation matrices can also be used for the polynomial SVD (PSVD) of $\mathcal{A}(z) \in \mathbb{C}^{P\times Q}$ by performing two polynomial QR decompositions so that [52]

$$\mathcal{U}^{\mathrm{P}}(z)\mathcal{A}(z) \approx \mathcal{T}_1(z) \,, \tag{2.84}$$

$$\mathcal{V}^{\mathrm{P}}(z)\mathcal{T}_1^{\mathrm{P}}(z) \approx \mathcal{T}_2(z) \,, \tag{2.85}$$

where $\mathcal{U}(z) \in \mathbb{C}^{P\times P}$ and $\mathcal{V}(z) \in \mathbb{C}^{Q\times Q}$ are para-unitary, $\mathcal{T}_1(z) \in \mathbb{C}^{P\times Q}$ and $\mathcal{T}_2(z) \in \mathbb{C}^{Q\times P}$ are upper triangular polynomial matrices. The process is repeated by substituting

$\boldsymbol{\mathcal{A}}(z)$ in (2.84) with $\boldsymbol{\mathcal{T}}_2(z)$ in (2.85). For a sufficiently large number of iterations achieving approximate diagonalization of $\boldsymbol{\mathcal{T}}_2(z)$, i.e. $\boldsymbol{\mathcal{T}}_2^{\mathrm{P}}(z) \approx \boldsymbol{\Sigma}(z)$, the PSVD of $\boldsymbol{\mathcal{A}}(z)$ is [52]

$$\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{A}}(z)\boldsymbol{\mathcal{V}}(z) \approx \boldsymbol{\Sigma}(z) \Longleftrightarrow \boldsymbol{\mathcal{A}}(z) \approx \boldsymbol{\mathcal{U}}(z)\boldsymbol{\Sigma}(z)\boldsymbol{\mathcal{V}}^{\mathrm{P}}(z) \; . \tag{2.86}$$

The elements of the diagonal polynomial matrix $\boldsymbol{\Sigma}(z)$ are the singular values of $\boldsymbol{\mathcal{A}}(z)$. Furthermore, if $\boldsymbol{\mathcal{A}}(z)$ has a rank of M, only the first M elements of $\boldsymbol{\Sigma}(z)$ are non-zero.

Instead of (2.86), the PSVD can also be calculated using the PEVD of two para-Hermitian polynomial matrices, $\boldsymbol{\mathcal{R}}_1(z) \in \mathbb{C}^{P\times P}$ and $\boldsymbol{\mathcal{R}}_2(z) \in \mathbb{C}^{Q\times Q}$, according to [30, 52]

$$\boldsymbol{\mathcal{R}}_1(z) = \boldsymbol{\mathcal{A}}(z)\boldsymbol{\mathcal{A}}^{\mathrm{P}}(z) \approx \boldsymbol{\mathcal{U}}(z)\boldsymbol{\Lambda}_1(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) \; , \tag{2.87}$$

$$\boldsymbol{\mathcal{R}}_2(z) = \boldsymbol{\mathcal{A}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{A}}(z) \approx \boldsymbol{\mathcal{V}}(z)\boldsymbol{\Lambda}_2(z)\boldsymbol{\mathcal{V}}^{\mathrm{P}}(z) \; , \tag{2.88}$$

where the elements in the eigenvalue polynomial matrices $\boldsymbol{\Lambda}_1(z) = \boldsymbol{\Sigma}(z)\boldsymbol{\Sigma}^{\mathrm{P}}(z) \in \mathbb{C}^{P\times P}$ and $\boldsymbol{\Lambda}_2(z) = \boldsymbol{\Sigma}^{\mathrm{P}}(z)\boldsymbol{\Sigma}(z) \in \mathbb{C}^{Q\times Q}$ are the squares of the polynomial singular values, i.e. $\lambda_i(z) = \sigma_i^2(z),\ i = 1,\ldots,M,$ and $\lambda_i(z) = 0,\ i > M+1$ for the i-th element. When applying a PEVD using an iterative approach such as SBR2 on (2.87), the algorithm converges when the magnitude of the maximum off-diagonal element in $\boldsymbol{\Lambda}_1(z)$ and $\boldsymbol{\Lambda}_2(z)$ are less than the user-defined parameter δ. Instead of targeting $\boldsymbol{\Lambda}_1(z)$ and $\boldsymbol{\Lambda}_2(z)$ in the PEVD approach, the PQR method operates directly on $\boldsymbol{\Sigma}(z)$ and avoids this lack of direct control on the diagonality.

2.6.3 Polynomial Matrix Generalized Eigenvalue Decomposition

The GEVD has also been extended to polynomial matrices in [32]. The algorithm aims to jointly diagonalize two polynomial matrices, $\boldsymbol{\mathcal{A}}(z)$ and $\boldsymbol{\mathcal{B}}(z)$, such that

$$\boldsymbol{\mathcal{V}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{A}}(z)\boldsymbol{\mathcal{V}}(z) = \boldsymbol{\Lambda}(z) \; , \tag{2.89}$$

$$\boldsymbol{\mathcal{V}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{B}}(z)\boldsymbol{\mathcal{V}}(z) = \mathbf{I} \; , \tag{2.90}$$

where $\mathbf{\Lambda}(z)$ contains the generalized polynomial eigenvalues and $\mathcal{V}(z)$, which is no longer guaranteed to be para-unitary, has the generalized polynomial eigenvectors. Instead of using a Cholesky decomposition, commonly used in a GEVD algorithm [12], as explained in Section 2.6.1, the computation of $\mathbf{\Lambda}(z)$ and $\mathcal{V}(z)$ relies on a multi-channel spectral factorization approach followed by a PEVD.

The first step is a Cholesky-style decomposition based on a spectral factorization technique using a PEVD such that

$$\mathcal{B}(z) = \mathcal{U}(z)\mathbf{\Gamma}(z)\mathcal{U}^{\mathrm{P}}(z) = \mathcal{U}(z)\mathbf{\Gamma}^{(+)}(z)\mathbf{\Gamma}^{(-)}(z)\mathcal{U}^{\mathrm{P}}(z) = \mathcal{T}^{\mathrm{P}}(z)\mathcal{T}(z) \; . \tag{2.91}$$

The terms $\mathbf{\Gamma}^{(+)}(z)$ and $\mathbf{\Gamma}^{(-)}(z) = \left(\mathbf{\Gamma}^{(+)}(z)\right)^{\mathrm{P}}$ are the minimum and maximum phase components from the factorization of elements in the diagonal $\mathbf{\Gamma}(z)$ using [27, 99]. The Cholesky-style factor is $\mathcal{T}^{\mathrm{P}}(z) = \mathcal{U}(z)\mathbf{\Gamma}^{(+)}(z)$, and its inverse is [99]

$$\left(\mathcal{T}^{\mathrm{P}}(z)\right)^{-1} = \left(\mathbf{\Gamma}^{(+)}(z)\right)^{-1}\mathcal{U}^{\mathrm{P}}(z) \; . \tag{2.92}$$

The inverse of $\mathbf{\Gamma}^{(+)}(z) \bullet\!\!-\!\!\circ \mathbf{\Gamma}^{(+)}(\tau)$ is computed in the frequency domain and reciprocated before returning to the time domain using the z-domain representation. Note that the inverse of a para-unitary matrix is its para-Hermitian since $\mathcal{U}(z)\mathcal{U}^{\mathrm{P}}(z) = \mathbf{I}$.

Similar to the second step of GEVD introduced in (2.72), the intermediate para-Hermitian polynomial matrix is computed as

$$\mathcal{C}(z) = \left(\mathcal{T}^{\mathrm{P}}(z)\right)^{-1}\mathcal{A}(z)\mathcal{T}^{-1}(z) \; . \tag{2.93}$$

The final step uses a PEVD to decompose the intermediate polynomial matrix so that

$$\mathcal{C}(z) = \mathcal{U}(z)\mathbf{\Lambda}(z)\mathcal{U}^{\mathrm{P}}(z) \; , \tag{2.94}$$

where the elements in the diagonal $\mathbf{\Lambda}(z)$ are the polynomial generalized eigenvalues with their associated polynomial generalized eigenvectors in the columns of $\mathcal{V}(z) = \mathcal{T}^{-1}(z)\mathcal{U}(z)$.

Equating (2.93) and (2.94), pre- and post-multiplying by $\mathcal{T}^{\mathrm{P}}(z)$ and $\mathcal{U}(z)$ give

$$\begin{aligned}\mathcal{T}^{\mathrm{P}}(z)\Big(\mathcal{T}^{P}(z)\Big)^{-1}\mathcal{A}(z)\mathcal{T}^{-1}(z)\mathcal{U}(z) &= \mathcal{T}^{\mathrm{P}}(z)\mathcal{U}(z)\boldsymbol{\Lambda}(z)\mathcal{U}^{\mathrm{P}}(z)\mathcal{U}(z)\\ \mathcal{A}(z)\mathcal{V}(z) &= \mathcal{B}(z)\mathcal{V}(z)\boldsymbol{\Lambda}(z)\ ,\end{aligned} \tag{2.95}$$

where $\mathcal{B}(z)\mathcal{V}(z) = \Big(\mathcal{T}^{\mathrm{P}}(z)\mathcal{T}(z)\Big)\Big(\mathcal{T}^{-1}(z)\mathcal{U}(z)\Big) = \mathcal{T}^{\mathrm{P}}(z)\mathcal{U}(z)$ is used.

The polynomial matrix GEVD algorithm has been used for broadband multi-channel Wiener filter (MWF) [32], blind signal separation [65, 66] and communications [62]. Further details of the algorithm can be found in [32].

2.7 Chapter Summary

In this chapter, historical and technical perspectives of polynomial matrices have been provided. The multi-channel broadband signal and space-time covariance matrices leading to polynomial matrices have been introduced. PEVD methods in the z- and frequency domains are reviewed, emphasizing the SBR2 and SMD algorithms since they are instrumental in this thesis. The conditions for the existence and uniqueness of analytic polynomial matrix factors are also explained. Related decompositions for conventional (non-polynomial) matrices as well as polynomial matrix extensions such as the polynomial QR, polynomial SVD and the polynomial matrix GEVD have also been briefly discussed.

Contributions

PEVD Theory &
Algorithmic Development

Chapter 3

New Insights, Theoretical & Algorithmic Development for PEVD

3.1 Introduction

This chapter explores the use of (polynomial) subspace approaches in signal processing applications. A comparison between EVD and PEVD is provided to highlight the differences in the processing. It is also shown and explained why PEVD is more suited to multi-channel broadband signal processing, and the EVD can be viewed as a degenerate case of the PEVD. The pioneering SBR2 algorithm, which inspired a family of PEVD algorithms in the z-domain, takes many iterations for convergence. An algorithmic enhancement has been proposed to reduce the matrix to a tri-diagonal form using reflections before applying rotations to speed up convergence while managing computational complexity. This algorithmic enhancement work has been presented in [29].

The eigenvalue decomposition (EVD) of Hermitian matrices is widely used in many signal processing applications such as subspace decomposition for data compression [18], noise reduction [19], spectral estimation [20], blind source separation [21] and adaptive beamforming [22,23]. The matrix used is usually the multi-channel spatial covariance matrix, computed using the (instantaneous) complex outer product of the data vector. This computation assumes that sources are narrowband, uncorrelated and zero-mean.

When wideband signals are involved, time delays between sensors cannot be simply modelled as phase shifts. Instead, correlation across different sensors and temporal lags need to be considered giving rise to space-time covariance matrices, usually modelled as polynomial matrices. This has motivated the development of a family of PEVD algorithms [25–28], based on the second-order sequential best rotation (SBR2) [30], for wideband signal processing applications like polynomial MUSIC [69, 100], adaptive beamforming [31], multi-channel Wiener filtering [32], blind source separation [33], channel identification [34, 101] and channel coding [35].

At every iteration, SBR2 searches for the dominant off-diagonal element and seeks to zero it out using elementary delay and rotation operations. It exploits the para-Hermitian symmetry and is akin to Jacobi's eigenvalue algorithm [11]. Even though the SBR2 has been proven to converge, it tends to take many iterations because zeroing occurs pairwise during each step and the rotation applied to other lags may undo previous diagonalization efforts.

While there were attempts to achieve one-step diagonalization for faster convergence using a Householder-like factorization, no closed form solution is available [25, 55]. Another approach, the SMD [28], performs an EVD on the zero lag plane before applying the same unitary matrices across all lags for faster convergence at the expense of complexity.

Motivated by the idea of reducing the matrix to condensed form in [12, 102, 103], in this chapter, we show how to enhance SBR2 by reducing the zero lag plane to tri-diagonal form using elementary reflection before applying rotation to speed up convergence.

3.2 A Comparison Between EVD and PEVD

The space-time covariance matrix has been introduced in Section 2.3.2 but will be reviewed here. It has been shown that the instantaneous spatial covariance matrix, commonly used in many subspace-based approaches, is the matrix coefficient of z^0, also known as the zero-lag plane, as shown in Fig. 2.2. An EVD is applied to the instantaneous covariance matrix for subspace decomposition. A PEVD algorithm [28], on the other hand, operates on the space-time covariance matrix, and its goal is to achieve diagonalization across all planes of z. While it might seem reasonable to assume that diagonalization can be achieved by applying multiple EVDs independently to multiple matrix coefficients, different planes are coupled together due to temporal correlations. Hence, the PEVD can only systematically decouple the planes using a sequence of delay and zeroing matrices.

3.2.1 The Rank Problem and Properties of the Subspaces

To understand the differences as well as their implications, EVD and PEVD are compared for two illustrative examples representing a sinusoidal and broadband source signal, respectively. The analytical proofs for these examples are provided in Appendix A. The simplified signal model used in the analysis will first be described.

A. Signal Model for Illustrative Examples

The source signal $s(n)$ arriving at the q-th sensor is

$$x_q(n) = h_q(n) * s(n) , \tag{3.1}$$

where n is the sample index, $h_q(n)$ is the channel impulse response modelled as a propagation delay, and $*$ denotes the linear convolution operator. Noise is not considered in these examples. The multi-channel output is $\mathbf{x}(n) = [x_1(n), \ldots, x_Q(n)]^{\mathrm{T}}$, where $[\cdot]^{\mathrm{T}}$ is the transpose operator.

The space-time covariance matrix can be written as [30]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau)\} , \tag{3.2}$$

where $\tau \in \mathbb{Z}$ is the temporal lag parameter, $\mathbb{E}\{\cdot\}$ and $[\cdot]^{\mathrm{H}}$ are the expectation and Hermitian transpose operators, respectively. Classical subspace-based methods such as the MUSIC algorithm [17] consider (3.2) at $\tau = 0$ so that the (instantaneous) spatial covariance matrix can be written as

$$\mathbf{R}_{\mathbf{xx}}(0) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\} . \tag{3.3}$$

An EVD is applied to (3.3) so that the spatial covariance matrix can be factorized as

$$\mathbf{R}_{\mathbf{xx}}(0) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} , \tag{3.4}$$

where $\boldsymbol{\Lambda}$ and $\mathbf{U}$ are the eigenvalue and eigenvector matrices, respectively [11]. Further, rearranging (3.4) gives

$$\boldsymbol{\Lambda} = \mathbf{U}^{\mathrm{H}}\mathbf{R}_{\mathbf{xx}}(0)\mathbf{U} = \mathbf{U}^{\mathrm{H}}\mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\}\mathbf{U} = \mathbb{E}\{\mathbf{y}(n)\mathbf{y}^{\mathrm{H}}(n)\} = \mathbf{R}_{\mathbf{yy}}(0) , \tag{3.5}$$

where the output $\mathbf{y}(n) = \mathbf{U}^{\mathrm{H}}\mathbf{x}(n)$ is computed using a linear combination of the sensor signals $\mathbf{x}(n)$. The linear combination provided by the eigenvector matrix $\mathbf{U}^{\mathrm{H}}$ causes the outputs or bases in $\mathbf{y}(n)$ to be (instantaneously) decorrelated. This is equivalent to diagonalizing $\mathbf{R}_{\mathbf{xx}}(0)$, representing a spatial decorrelation at the zero lag, e.g., $\tau = 0$.

Applying a PEVD to the z-transform of (3.2), $\mathbf{R}_{\mathbf{xx}}(\tau) \circ\!\!-\!\!\bullet\, \mathcal{R}_{\mathbf{xx}}(z)$, gives

$$\mathcal{R}_{\mathbf{xx}}(z) = \boldsymbol{\mathcal{U}}(z)\boldsymbol{\Lambda}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) , \tag{3.6}$$

where $\boldsymbol{\Lambda}(z)$ and $\boldsymbol{\mathcal{U}}(z)$ are the eigenvalue and eigenvector polynomial matrices, respectively [48,49]. To diagonalize $\mathcal{R}_{\mathbf{xx}}(z)$, the inputs $\mathbf{x}(n)$ are filtered based on

$$\mathbf{y}(n) = \sum_{j} \mathbf{U}^{\mathrm{H}}(-j)\mathbf{x}(n-j) , \tag{3.7}$$

where the eigenvectors in $\boldsymbol{\mathcal{U}}(z) \bullet\!\!-\!\!\circ\, \mathbf{U}(\tau)$ are modelled using J-th order FIR filters. The space-time covariance polynomial matrix of the bases $\mathbf{y}(n)$ is

$$\mathcal{R}_{\mathbf{yy}}(z) = \boldsymbol{\Lambda}(z) . \tag{3.8}$$

In the case of PEVD, diagonalization is achieved over the range of z. This is strong decorrelation, equivalent to spatial decorrelation imposed over a range of time lags τ [30]. The diagonal elements at the zero-lag plane of any space-time covariance and eigenvalue polynomial matrices represent the power of the corresponding signals generating them.

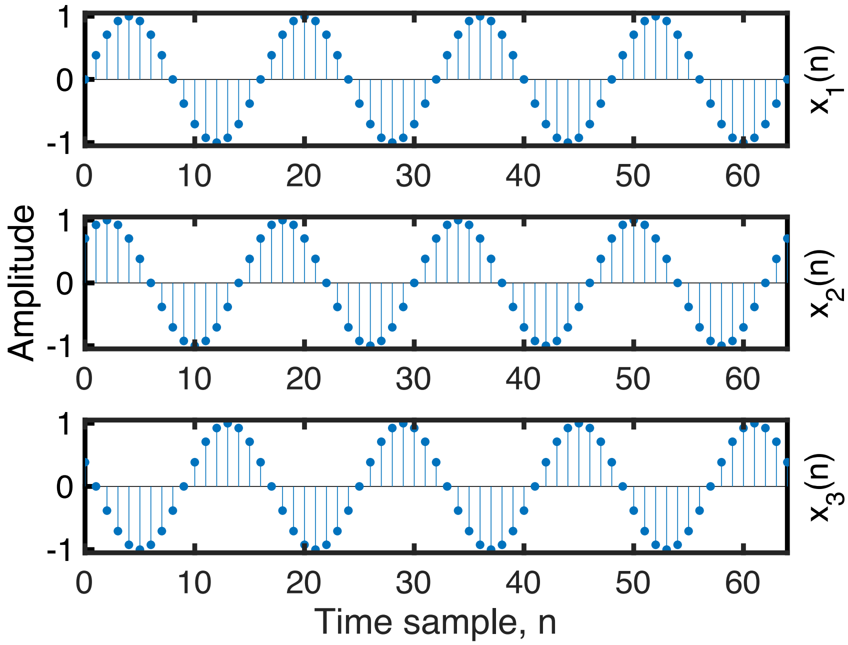


Figure 3.1: A sinusoidal source arriving at 3 sensors $\mathbf{x}(n)$.

B. Sinusoidal Signal Illustrative Example

A sinusoidal signal arriving at a 3-sensor array, as shown in Fig. 3.1, has the following space-time covariance matrix shown in Fig. 3.2. The instantaneous spatial covariance matrix is the coefficient of z^0 on the principal plane, marked by red crosses.

The EVD factorizes the spatial covariance matrix into eigenvalue and eigenvector matrices. The eigenvector is used to generate spatially decorrelated outputs using a linear combination of $\mathbf{x}(n)$ based on (3.5), as shown in Fig. 3.3. This is further supported by the space-time covariance of the output $\mathbf{y}(n)$ in Fig. 3.4, where the red crosses on the off-diagonals are zero-valued. The outputs in Fig. 3.3 show that $y_1(n)$ and $y_2(n)$ are orthogonal, i.e., the inner product between the two signals equals 0, as expected for sine and cosine functions at the same frequency. Consequently, the subspace rank, obtained by applying an EVD to the instantaneous covariance matrix, is 2.

The PEVD factorizes the space-time covariance matrix into eigenvalue and eigenvector

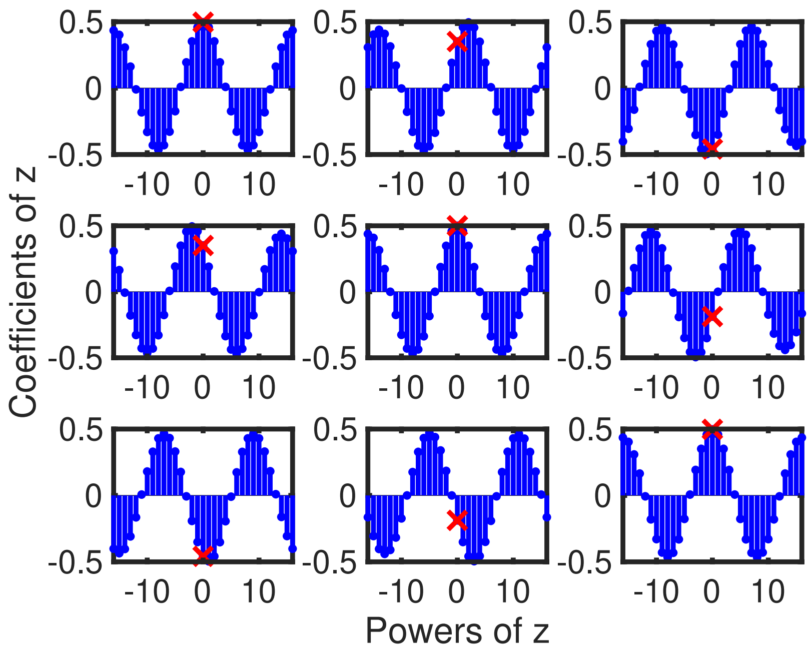


Figure 3.2: Space-time covariance matrix $\mathcal{R}_{\mathbf{xx}}(z)$, with the (p, q)-th subplot representing the correlation sequence between the p-th and q-th sensor. Red crosses represents the instantaneous correlation at $\tau = 0$.

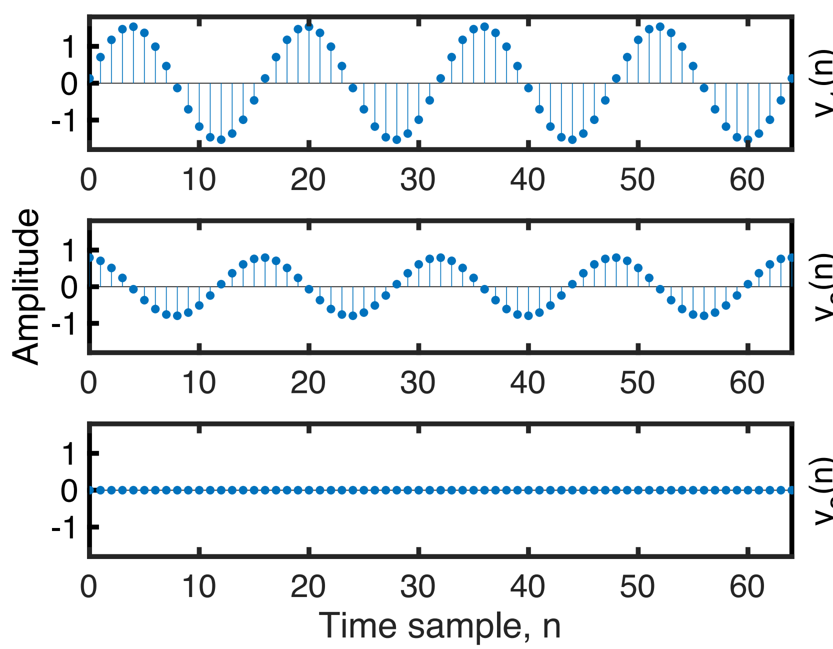


Figure 3.3: Instantaneously decorrelated output $\mathbf{y}(n)$ processed using eigenvector $\mathbf{U}$.

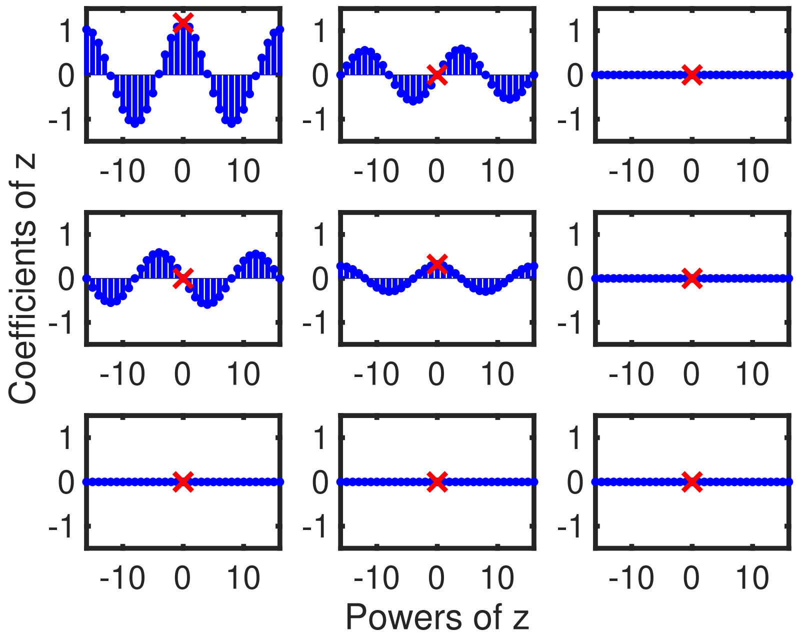


Figure 3.4: Space-time covariance matrix of instantaneously decorrelated output $\mathcal{R}_{\mathbf{yy}}(z)$, with the (p, q)-th subplot representing the correlation sequence between the p-th and q-th output in Fig. 3.3 and red crosses represent the instantaneous correlation at $\tau = 0$.

polynomial matrices. The sensor signals $\mathbf{x}(n)$ are filtered through the eigenvector polynomial matrix to generate strongly decorrelated outputs $\mathbf{y}(n)$, as shown in Fig. 3.5. The polynomial eigenvector, equivalent to a FIR filter bank, can exploit time delays to align the signals before combining them. Using a sequence of delay and zeroing matrices, the single sinusoidal output correctly reflects the signal subspace rank as 1, as shown in Fig. 3.6. In contrast to the set of bases generated by the EVD, the PEVD bases are strongly decorrelated, i.e., the inner product between any shifted basis pair equals 0 or exhibits orthogonality across a range of time lags.

C. Broadband Signal Illustrative Example

The second example considers a pulse broadband source signal arriving at the 3-channel array, as shown in Fig. 3.7. The envelopes of the correlation sequences are triangular as expected for a rectangular signal but are time-shifted according to the relative delays between channels, as shown in Fig. 3.8. Observe that two instantaneous spatial covariance

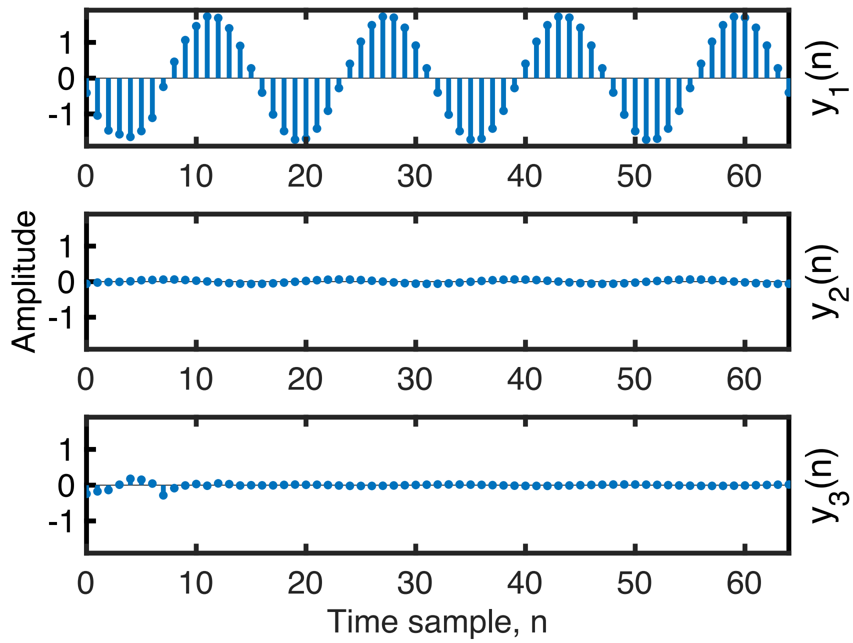


Figure 3.5: Strongly decorrelated output $\mathbf{y}(n)$ processed using eigenvector $\boldsymbol{\mathcal{U}}(z)$.

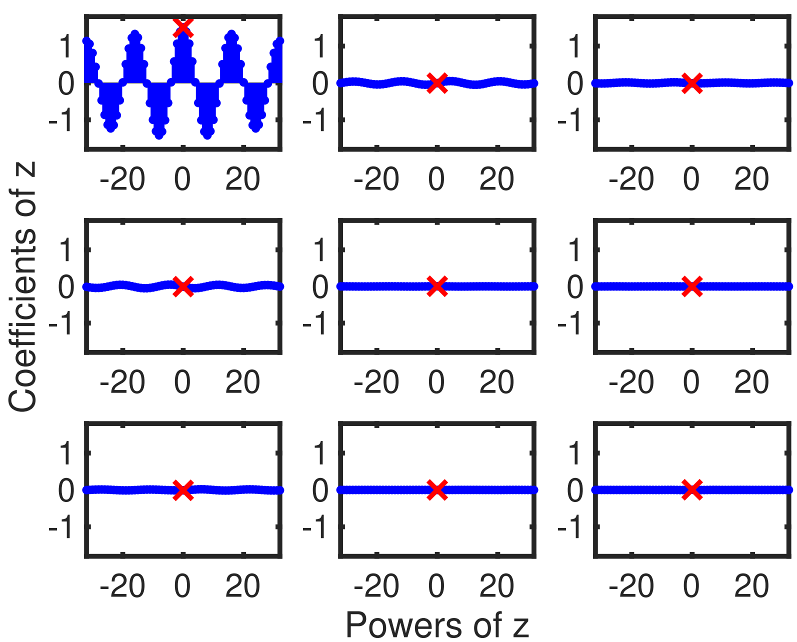


Figure 3.6: Space-time covariance matrix of strongly decorrelated output $\mathcal{R}_{\mathbf{yy}}(z)$. , with the (p, q)-th subplot representing the correlation sequence between the p-th and q-th output in Fig. 3.5. Red crosses represent the instantaneous correlation at $\tau = 0$.

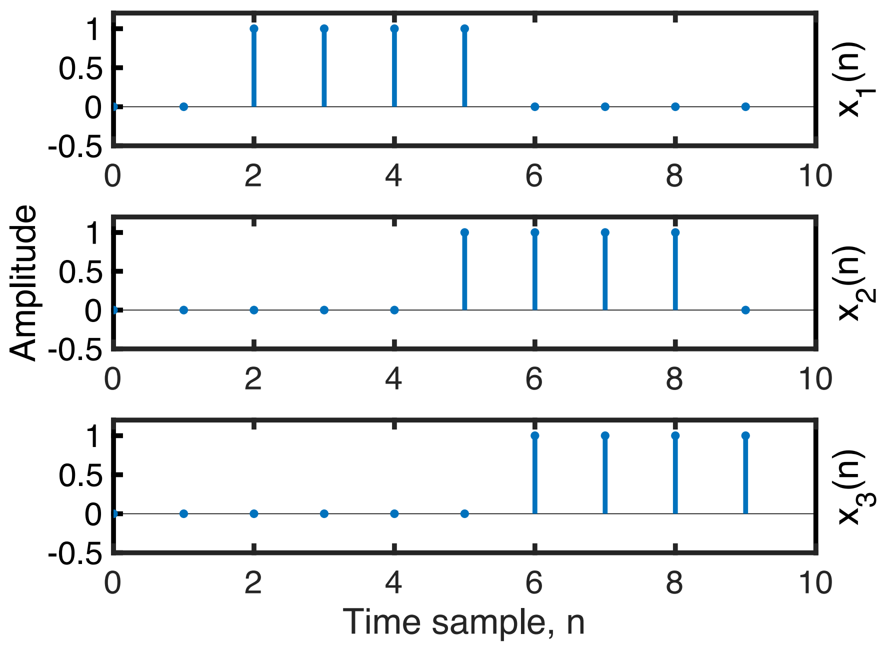


Figure 3.7: A broadband pulse source signal arriving at 3 sensors $\mathbf{x}(n)$.

matrix elements, $r_{1,3}(0)$ and $r_{3,1}(0)$, are zeros because $x_1(n)$ and $x_3(n)$ are orthogonal, i.e., their inner product equals zero.

The eigenvector matrix computed from the EVD of the spatial covariance matrix is used to generate instantaneously decorrelated signals or bases $\mathbf{y}(n)$, as shown in Fig. 3.9. The space-time covariance matrix of the output signal is presented in Fig. 3.10. As expected, the instantaneous spatial covariance matrix, represented by red crosses, is diagonal and equivalent to the computed eigenvalues. Diagonality at the zero-lag implies that the inner product between any pair of signals (except with itself) is always zero. This example suggests that the rank of the EVD is likely to be 2 or even 3. The multi-channel sensor data $\mathbf{x}(n)$ can be reconstructed from the bases $\mathbf{y}(n)$ using the eigenvectors.

The PEVD generates a polynomial eigenvector that filters the sensor signals $\mathbf{x}(n)$ to generate strongly decorrelated outputs or bases $\mathbf{y}(n)$. As shown in Fig. 3.11, unlike the EVD case, the rectangular pulse signal, subjected to bulk delays, can be recovered using a PEVD. The space-time covariance matrix of the strongly decorrelated outputs has a

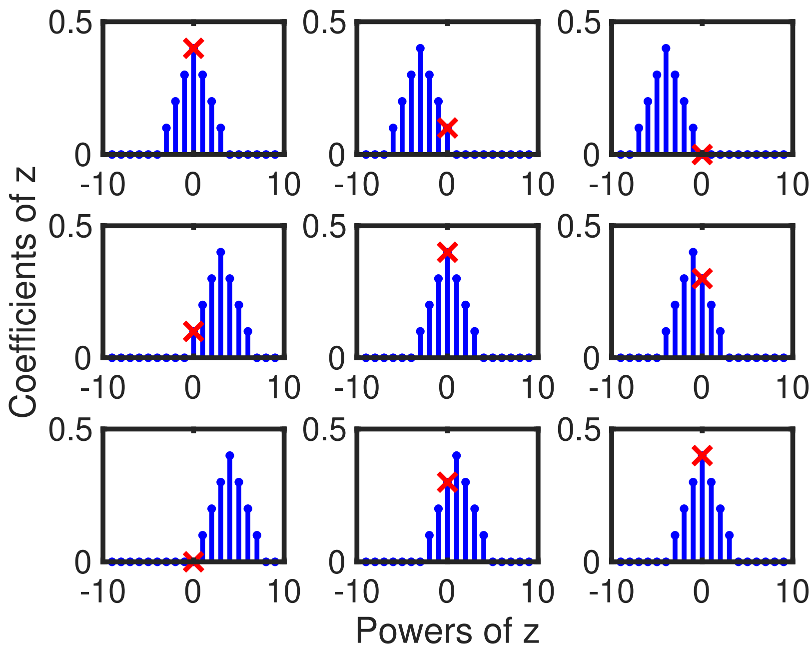


Figure 3.8: Space-time covariance matrix $\mathcal{R}_{\mathbf{xx}}(z)$, with the (p, q)-th subplot representing the correlation sequence between the p-th and q-th sensor. Red crosses represent the instantaneous correlation at $\tau = 0$.

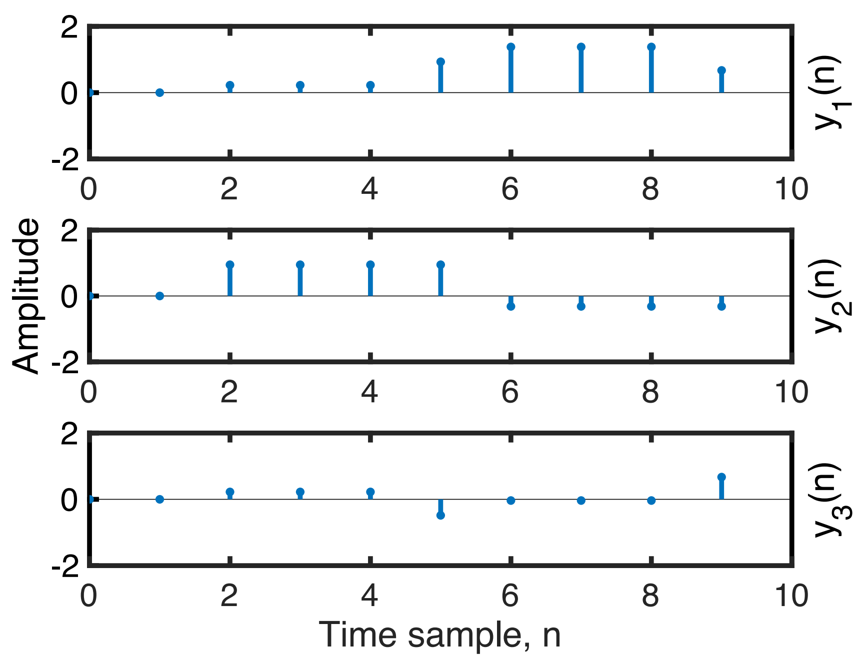


Figure 3.9: Instantaneously decorrelated output $\mathbf{y}(n)$ processed using eigenvector $\mathbf{U}$.

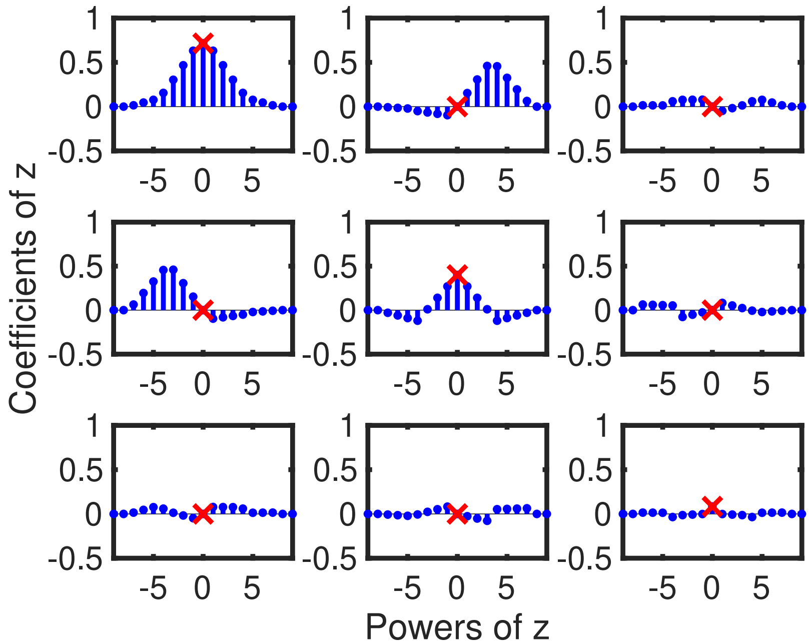


Figure 3.10: Space-time covariance matrix of instantaneously decorrelated output $\mathcal{R}_{\mathbf{yy}}(z)$, with the (p, q)-th subplot representing the correlation sequence between the p-th and q-th output in Fig. 3.9. Red crosses represent the instantaneous correlation at $\tau = 0$.

rank of 1 and a triangular envelope, as observed in Fig. 3.12.

3.2.2 Summary of Key Findings

In terms of geometric interpretation, the bases generated by an EVD span a function space in Hilbert space, e.g., $\text{span}(\{\mathbf{y}_1(n), \ldots, \mathbf{y}_Q(n)\})$, so any function can be constructed using the linear combination of the bases. However, the PEVD bases belong to shift-invariant spaces, e.g. $\text{span}(\{\mathbf{y}_1(n-\tau), \ldots, \mathbf{y}_Q(n-\tau)\})$, $\tau \in \mathbb{Z}$ within the range of the time lags imposed for decorrelation [104]. The shift-invariance outputs necessarily satisfy strong decorrelation since the cross-correlation over the range of lags between any processed pairs is close to zero. Any function can now be represented using the linear combination of the bases and their shifted versions. Well-known examples of shift-invariant space include the space spanned by shifted sinc functions used for the signal reconstruction of bandlimited signals, e.g., $\text{span}(\{\text{sinc}(n-\tau)\})$, $\tau \in \mathbb{Z}$ in the generalized sampling theorem [105, 106] and wavelet transforms [18, 104].

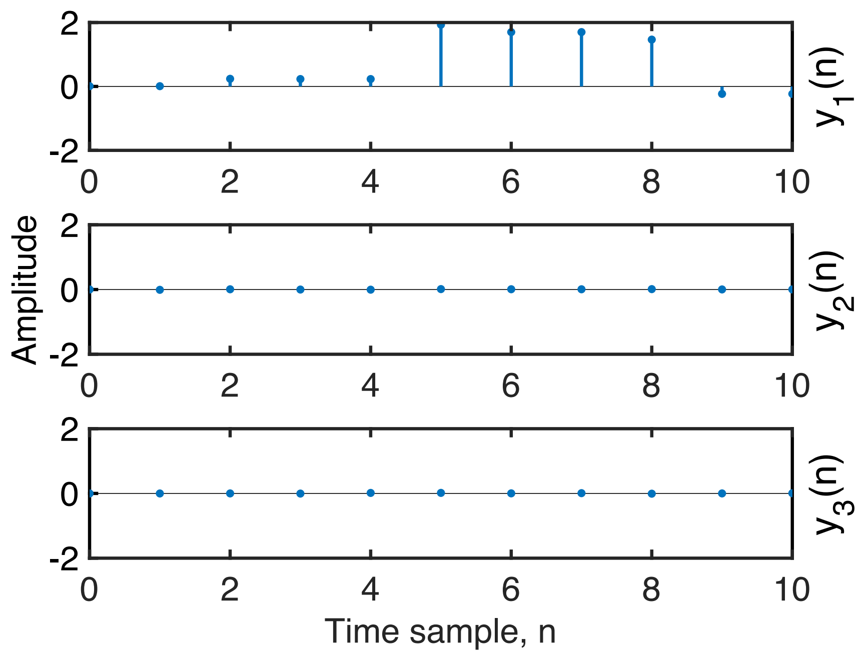


Figure 3.11: Strongly decorrelated output $\mathbf{y}(n)$ processed using eigenvector $\boldsymbol{\mathcal{U}}(z)$.

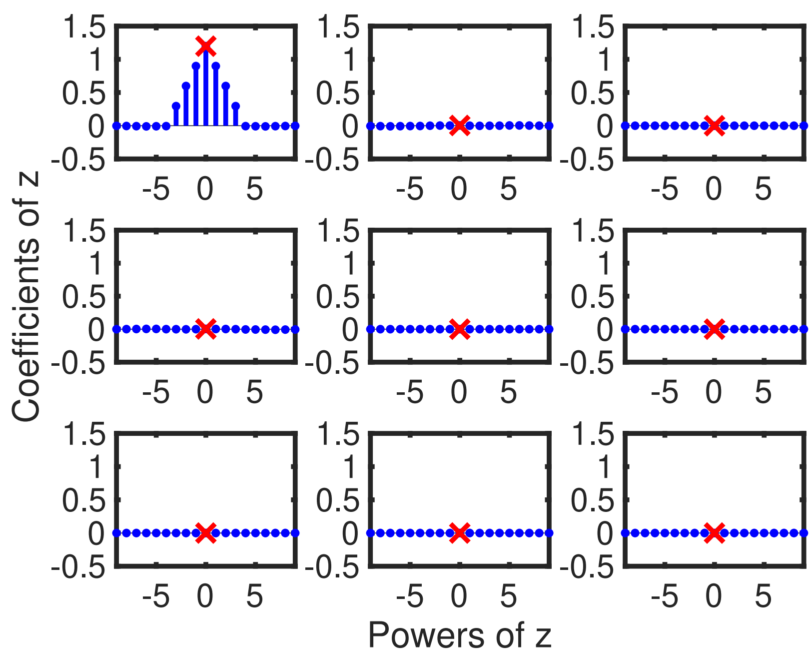


Figure 3.12: Space-time covariance matrix of strongly decorrelated output $\mathcal{R}_{\mathbf{yy}}(z)$, with the (p, q)-th subplot representing the correlation sequence between the p-th and q-th output in Fig. 3.11. Red crosses represent the instantaneous correlation at $\tau = 0$.

The shift-invariance property of the basis functions generated by the PEVD allows exact signal representation using fewer basis functions than its EVD counterpart. This enables the PEVD to achieve more significant signal compression with fewer components, as will be seen in Chapter 5.

The sinusoidal illustrative example also demonstrates that the EVD generates instantaneously decorrelated outputs using unitary matrices and provides a rank-2 decomposition because of sine and cosine components, although there was only one source. The PEVD, on the other hand, applies delays and unitary matrices for strong decorrelation, and this compacts the signals into a single sinusoidal output, giving a rank-1 decomposition. The analytical proofs for the results observed in the sinusoidal and broadband examples are provided in Appendix A.

3.3 Second-order Sequential Best Rotation with Householder Transformation

3.3.1 Review of Numerical Methods for EVD of Hermitian Matrix

The eigenvalue decomposition of a Hermitian or complex symmetric matrix, $\mathbf{R} = \mathbf{R}^{\mathrm{H}}$, where $\mathbf{R} \in \mathbb{C}^{Q\times Q}$, can be expressed as [11]

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} = \mathbf{W}^{\mathrm{H}}\boldsymbol{\Lambda}\mathbf{W}\ , \tag{3.9}$$

where the rows of the unitary matrix $\mathbf{W}$ contain the orthonormal eigenvectors with their associated real-valued distinct eigenvalues lying on the diagonal of $\boldsymbol{\Lambda}$, and $\mathbf{U} = \mathbf{W}^{\mathrm{H}}$. Since similarity transformations do not alter the eigenvalues, in practice, $\boldsymbol{\Lambda}$ is obtained by applying a sequence of unitary matrices.

In numerical diagonalization such as the Jacobi method, the unitary matrix, $\mathbf{W}$, can be constructed using a sequence of L unitary matrices $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_L$ with each of them satisfying

$$\mathbf{W}_\ell\mathbf{W}_\ell^{\mathrm{H}} = \mathbf{W}_\ell^{\mathrm{H}}\mathbf{W}_\ell = \mathbf{I}\ , \tag{3.10}$$

where $\mathbf{I}$ is the identity matrix. The construction of $\mathbf{W}_\ell$ at the ℓ-th iteration aims to reduce the off-diagonal elements in a systematic manner based on its Frobenius-norm,

$$\text{off}(\mathbf{R}) = \sum_{p=1}^{Q} \sum_{\substack{q=1 \\ p \neq q}}^{Q} r_{pq}^2 \,. \tag{3.11}$$

The unitary matrix, $\mathbf{W}_\ell$, $\ell = 1, 2, \ldots, L$, can take the form of, for example, a Givens rotation,

$$\mathbf{G}_{(p,q)}(\theta, \phi) = \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cos\theta & \mathbf{0} & \sin\theta e^{\mathrm{j}\phi} & \vdots \\ \vdots & \mathbf{0} & \mathbf{I}_{q-p-1} & \mathbf{0} & \vdots \\ \vdots & -\sin\theta e^{-\mathrm{j}\phi} & \mathbf{0} & \cos\theta & \vdots \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{I}_{Q-q} \end{bmatrix} \in \mathbb{C}^{Q\times Q} \,, \tag{3.12}$$

where $\mathbf{0}$ is the zero matrix and $\mathrm{j} = \sqrt{-1}$, designed to zero the off-diagonal element pair on the p-th and q-th row or column such that

$$\mathbf{G}_{(p,q)}(\theta, \phi) \begin{bmatrix} \ddots & \vdots & & \vdots & \iddots \\ \cdots & r_{p,p} & \cdots & r_{p,q} & \cdots \\ & \vdots & \ddots & \vdots & \\ \cdots & r_{q,p} & \cdots & r_{q,q} & \cdots \\ \iddots & \vdots & & \vdots & \ddots \end{bmatrix} \mathbf{G}^{\mathrm{H}}_{(p,q)}(\theta, \phi) = \begin{bmatrix} \ddots & \vdots & & \vdots & \iddots \\ \cdots & r'_{p,p} & \cdots & 0 & \cdots \\ & \vdots & \ddots & \vdots & \\ \cdots & 0 & \cdots & r'_{q,q} & \cdots \\ \iddots & \vdots & & \vdots & \ddots \end{bmatrix} \,, \tag{3.13}$$

where $r_{p,q}$ is the p-th row, q-th column matrix element and $[.]'$ denotes the resulting sub-matrix after multiplications. The values in (3.12) satisfying (3.13) are $\phi = \arg(r_{p,q})$ and $\tan 2\theta = \frac{2|r_{p,q}|}{r_{p,p} - r_{q,q}}$. Furthermore, the diagonal elements, $r'_{p,p}, r'_{q,q}$ can be ordered in decreasing magnitude if θ is computed using the four quadrant inverse tangent to avoid permutation ambiguity arising from arbitrary ordering.

The Householder reflection matrix $\mathbf{H}_p \in \mathbb{C}^{Q\times Q}, p = 1, \ldots, Q-2$, which satisfies (3.10), is designed to zero more than one off-diagonal pair on the p-th row and column of

$$\mathbf{A} = \mathbf{A}^{\mathrm{H}} = \left[\begin{array}{c|c} \mathbf{A}_1 & \mathbf{A}_2 \\ \hline \mathbf{A}_3 & \mathbf{A}_4 \end{array}\right] \in \mathbb{C}^{Q\times Q} , \tag{3.14}$$

partitioned into sub-matrices $\mathbf{A}_1 \in \mathbb{C}^{p\times p}$ which is tridiagonal, $\mathbf{A}_2 \in \mathbb{C}^{p\times(Q-p)}$ which is related to $\mathbf{A}_3$ according to

$$\mathbf{A}_3 = \mathbf{A}_2^{\mathrm{H}} = \left[\begin{array}{ccc|c} 0 & \ldots & 0 & a_{p+1,p} \\ 0 & \ldots & 0 & a_{p+2,p} \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & a_{Q,p} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{x} \end{array}\right] \in \mathbb{C}^{(Q-p)\times p} , \tag{3.15}$$

where $\mathbf{0}$ has a dimension of $(Q-p)\times(p-1)$, $\mathbf{x} = [a_{p+1,p}, a_{p+2,p}, \ldots, a_{Q,p}]^T \in \mathbb{C}^{(Q-p)\times 1}$ is the vector along the p-th column, and $\mathbf{A}_4 \in \mathbb{C}^{(Q-p)\times(Q-p)}$. The sub-matrix $\mathbf{J}_p$ aims to modify the p-th row in $\mathbf{A}_2$ and p-th column in $\mathbf{A}_3$ in (3.15) such that

$$\mathbf{A}_3' = \mathbf{J}_p\mathbf{A}_3 = \mathbf{A}_2'^{\mathrm{H}} = \left[\begin{array}{ccc|c} 0 & \ldots & 0 & \alpha \\ 0 & \ldots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{y} \end{array}\right] \in \mathbb{C}^{(Q-p)\times p} , \tag{3.16}$$

where $\mathbf{y} = [\alpha, 0, \ldots, 0]^T \in \mathbb{C}^{(Q-p)\times 1}$ is the target vector, $\alpha = \pm\|\mathbf{x}\|_2 \exp(\mathrm{j}\phi_{a_{p,p+1}})$ is the first element in $\mathbf{y}$, $\phi_{a_{p,p+1}}$ is the phase angle of $a_{p,p+1}$ and $[\cdot]'$ denotes the matrix after multiplication. The reflection sub-matrix is computed using $\mathbf{J}_p = \mathbf{I}_{Q-p} - \frac{2\mathbf{u}\mathbf{u}^{\mathrm{H}}}{\mathbf{u}^{\mathrm{H}}\mathbf{u}}$, where the Householder vector, $\mathbf{u} = \mathbf{x} + \mathbf{y}$, is a rank 1 modification to the identity matrix.

Then, the overall Householder reflection matrix $\mathbf{H}_p$ is computed such that

$$\mathbf{H}_p\mathbf{A}\mathbf{H}_p^{\mathrm{H}} = \left[\begin{array}{c|c} \mathbf{I}_p & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{J}_p \end{array}\right] \left[\begin{array}{c|c} \mathbf{A}_1 & \mathbf{A}_2 \\ \hline \mathbf{A}_3 & \mathbf{A}_4 \end{array}\right] \left[\begin{array}{c|c} \mathbf{I}_p & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{J}_p^H \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_1 & \mathbf{A}_2' \\ \hline \mathbf{A}_3' & \mathbf{A}_4' \end{array}\right] , \tag{3.17}$$

where $\mathbf{A}_1$ remains unaffected by the matrix products while $\mathbf{A}_4$ and $\mathbf{A}_4'$ contain arbitrary values which are unimportant for the discussion here. Deeper treatment of unitary transformation can be found in [11, 12, 107].

3.3.2 Numerical Methods for PEVD

The PEVD of a para-Hermitian polynomial matrix, $\mathcal{R}(z) = \mathcal{R}^{\mathrm{P}}(z) \in \mathbb{C}^{Q\times Q}$, where $\mathcal{R}^{\mathrm{P}}(z) = \mathcal{R}^{\mathrm{H}}(1/z^*)$, proposed in [30], can be expressed as

$$\mathcal{R}(z) \approx \mathcal{W}^{\mathrm{P}}(z)\mathbf{\Lambda}(z)\mathcal{W}(z) = \mathcal{U}(z)\mathbf{\Lambda}(z)\mathcal{U}^{\mathrm{P}}(z) \Longleftrightarrow \mathbf{\Lambda}(z) \approx \mathcal{W}(z)\mathcal{R}(z)\mathcal{W}^{\mathrm{P}}(z) , \tag{3.18}$$

where the rows of $\mathcal{W}(z)$ contain the eigenvectors with its corresponding eigenvalues on the diagonal matrix, $\mathbf{\Lambda}(z)$, and $\mathcal{U}(z) = \mathcal{W}^{\mathrm{P}}(z)$. Elements in $\mathcal{W}(z)$ and $\mathbf{\Lambda}(z)$ are, in general, polynomials in z. The eigenvalues of $\mathcal{R}(z)$ are preserved via similarity transformations if $\mathcal{W}(z)$ satisfies the para-unitary condition [45] given by

$$\mathcal{W}^{\mathrm{P}}(z)\mathcal{W}(z) = \mathcal{W}(z)\mathcal{W}^{\mathrm{P}}(z) = \mathbf{I} . \tag{3.19}$$

The Givens rotation and Householder reflection matrices described earlier also satisfy (3.19). They can be viewed as the degenerate case where the matrix elements are coefficients of z^0. An elementary operator for a polynomial matrix, that can change the polynomial order, is the τ-th delay matrix applied to the p-th row (column) upon a pre(post)-multiplication,

$$\mathcal{D}_{p,\tau}(z) = \begin{bmatrix} \mathbf{I}_{p-1} & 0 & 0 \\ 0 & z^{-\tau} & 0 \\ 0 & 0 & \mathbf{I}_{Q-p} \end{bmatrix} . \tag{3.20}$$

For numerical diagonalization in (3.18), $\boldsymbol{\mathcal{W}}(z)$ can comprise a sequence of L para-unitary matrix factors, $\boldsymbol{\mathcal{W}}_1(z), \boldsymbol{\mathcal{W}}_2(z), \ldots, \boldsymbol{\mathcal{W}}_L(z)$, built from elementary delay, rotation and/or reflection matrices.

A. Sign and Delay Ambiguity in Eigenvectors for PEVD

From [11], a conventional matrix with distinct eigenvalues has orthonormal left eigenvectors that lie in the columns of $\mathbf{U} = \mathbf{W}^{\mathrm{H}}$ and are unique up to arbitrary phase rotations $e^{\mathrm{j}\varphi}$. Additionally, for polynomial matrices, each eigenvector can be scaled with an arbitrary sign and be randomly delayed in z, without affecting the orthonormality of eigenvectors. This result will be exploited in Section 3.3.5. For example, if $Q = 3$, then

$$\boldsymbol{\Lambda}(z) = \begin{bmatrix} \lambda_1(z) & 0 & 0 \\ 0 & \lambda_2(z) & 0 \\ 0 & 0 & \lambda_3(z) \end{bmatrix} , \tag{3.21}$$

$$\boldsymbol{\mathcal{U}}(z) = \begin{bmatrix} \pm z^{-\tau_1}\boldsymbol{u}_1(z), & \pm z^{-\tau_2}\boldsymbol{u}_2(z), & \pm z^{-\tau_3}\boldsymbol{u}_3(z) \end{bmatrix} , \tag{3.22}$$

$$\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) = \begin{bmatrix} \pm z^{\tau_1}\boldsymbol{u}_1^*(1/z^*), & \pm z^{\tau_2}\boldsymbol{u}_2^*(1/z^*), & \pm z^{\tau_3}\boldsymbol{u}_3^*(1/z^*) \end{bmatrix}^T , \tag{3.23}$$

where λ_m is the m-th eigenvalue with corresponding eigenvector, $\boldsymbol{u}_m = [u_{1,m}, u_{2,m}, u_{3,m}]^T$ and its associated arbitrary delay, $z^{-\tau_m}$. The diagonal terms of $\boldsymbol{\mathcal{R}}(z)$ in (3.18) are

$$r_{p,p}(z) = \sum_{k=1}^{3} \lambda_k(z) u_{p,k}(z) u_{p,k}^*(1/z^*) , \tag{3.24}$$

and off-diagonal terms,

$$r_{p,q}(z) = r_{q,p}^{\mathrm{P}}(z) = \sum_{k=1}^{3} \lambda_k(z) u_{p,k}(z) u_{q,k}^*(1/z^*) . \tag{3.25}$$

Observe that (3.24) and (3.25) hold for all arbitrary signs and delays because the factors cancel out. Each sign ambiguity can also be expressed as a phase factor $e^{\mathrm{j}\varphi}$ which can be accounted for in either the eigenvector or z. Furthermore, if z is evaluated on the unit circle, the sign and delay ambiguity is equivalent to phase ambiguity, that is inherent in any power estimation problem using only second-order statistics. A deeper discussion is provided in [48].

B. Second-order Sequential Best Rotation (SBR2) Algorithm

McWhirter *et al.* [30] proposed the SBR2 algorithm for factorizing (3.18) using a series of para-unitary transformations. Consider the following 2-by-2 para-Hermitian polynomial matrix example,

$$\mathcal{R}(z) = \begin{bmatrix} 1 & -0.1z^{-1} + 0.5z^{1} \\ -0.1z^{1} + 0.5z^{-1} & 1 \end{bmatrix} . \tag{3.26}$$

For each iteration, SBR2 begins by searching for the largest off-diagonal element, $|g|$, across all powers of z as depicted in Fig. 3.13(i) and comparing against a predefined threshold, δ. The element is indexed by p-th row, q-th column and z^{τ}-plane. Shown in Fig. 3.13(ii), the dominant element is brought to the principal plane, z^0, using a delay matrix if it exceeds δ. A Givens rotation, computed based on the principal plane, is applied to the entire polynomial matrix as indicated in Fig. 3.13(iii). This zeros the two elements containing $|g|$ on the principal plane. Due to symmetry, the search is confined to half the off-diagonals and similarity transformation results in operations acting on dominant pairs rather than just elements.

The overall para-unitary matrix which diagonalizes the polynomial matrix approximately after L iterations is

$$\mathcal{W}(z) = \mathcal{W}_L(z)\mathcal{W}_{L-1}(z)\cdots\mathcal{W}_1(z) , \tag{3.27}$$

where $\mathcal{W}_\ell(z) = \mathbf{G}^{(\ell)}\mathcal{D}^{(\ell)}(z)$ is the product of rotation and delay matrices during the ℓ-th iteration and satisfies (3.19).

Naturally, the number of delay operations may increase as the algorithm progresses with more iterations. Consequently, the order of the polynomial matrix may grow, as evident in Fig. 3.13, and becomes unnecessarily large. This will increase search and computation

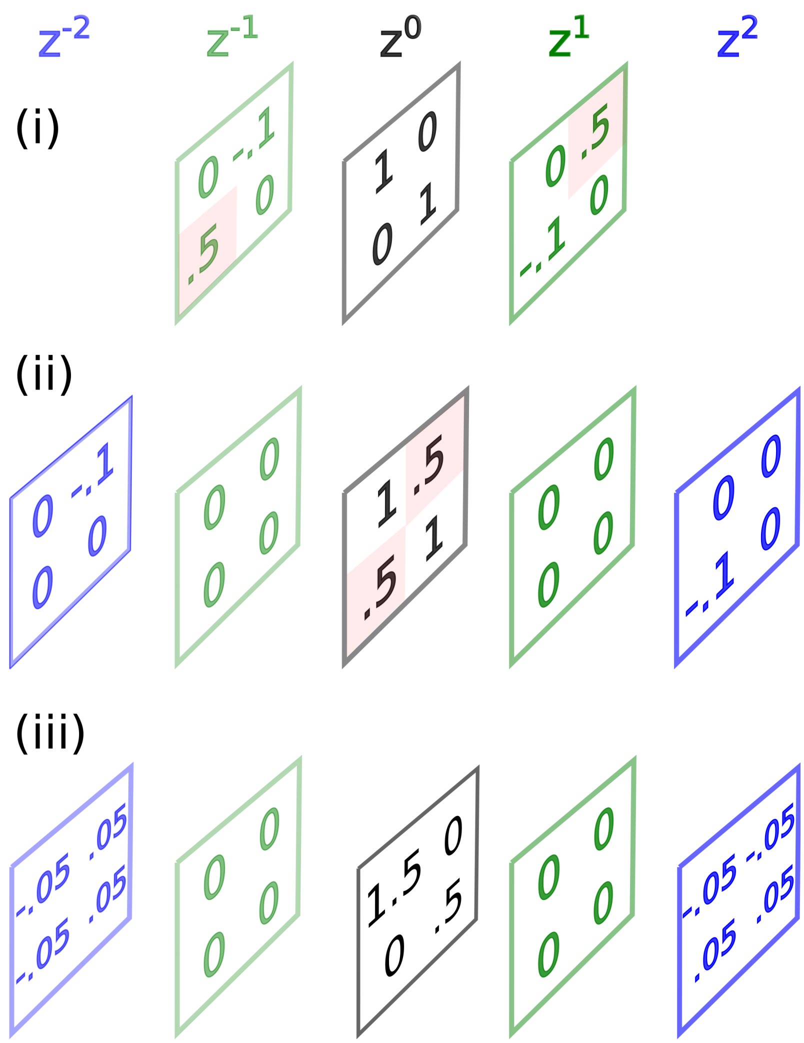


Figure 3.13: Visualization for the first iteration of SBR2 for (3.26) that demonstrates the (i) search, (ii) delay and (iii) rotate step.

time as well as storage requirements.

For practical applications, $\tilde{\boldsymbol{\mathcal{W}}}(z)$ and $\tilde{\boldsymbol{\Lambda}}(z)$, approximations of $\boldsymbol{\mathcal{W}}(z)$ and $\boldsymbol{\Lambda}(z)$ in (3.18), are usually sufficient. The reconstructed space-time covariance polynomial matrix is

$$\hat{\boldsymbol{\mathcal{R}}}(z) = \tilde{\boldsymbol{\mathcal{W}}}^{P}(z)\tilde{\boldsymbol{\Lambda}}(z)\tilde{\boldsymbol{\mathcal{W}}}(z) \; . \tag{3.28}$$

This has motivated the use of a reconstruction error measure expressed as a percentage and defined by

$$\epsilon = \frac{\sum_{\forall z} \|\hat{\boldsymbol{\mathcal{R}}}(z) - \boldsymbol{\mathcal{R}}(z)\|_F}{\sum_{\forall z} \|\boldsymbol{\mathcal{R}}(z)\|_F} \times 100\% \; , \tag{3.29}$$

where $\|.\|_F$ is the Frobenius-norm (F-norm), to indicate the accuracy of the approximation. Several methods for trimming polynomial terms with small coefficients to keep the order compact have been proposed [27, 54, 73] but the energy-based technique [74] was chosen in SBR2. For a given ratio of squared F-norm or energy permitted to be lost, $\tilde{\boldsymbol{\mathcal{R}}}(z)$ is approximated by the smallest $\tau_{\max}$ satisfying the inequality,

$$\mathrm{trim}(\boldsymbol{\mathcal{R}}(z), \mu) = \sum_{\tau=-\tau_{\max}}^{\tau_{\max}} \|\mathbf{R}(\tau) z^{-\tau}\|_F^2 \geq (1-\mu) \sum_{\forall z} \|\boldsymbol{\mathcal{R}}(z)\|_F^2 \; , \tag{3.30}$$

where μ is the trim parameter and $\mathbf{R}(\tau)$ are the matrix coefficients.

Because only one element pair is zeroed during each SBR2 iteration, all off-diagonal elements have to be iteratively processed to cycle through all elements. Consequently, more iterations and hence slower convergence is expected as the number of elements increases with the size of the matrix and polynomial order.

3.3.3 Proposed Algorithmic Enhancement Using Householder Matrices

At each iteration, the Givens rotation increases the square of the trace norm on the principal plane by $2|g|^2$ due to zeroing of the symmetric pair, where $|g|$ is the off-diagonal element with the largest magnitude across all powers of z. However, if the matrix at the principal plane is reduced to tri-diagonal form before rotation, the norm of the entire row and column, which includes the dominant pairs, becomes compactly stored in the super- and sub-diagonals. By applying the Givens rotations to the tri-diagonal matrix, more of the F-norm of the off-diagonal entries on the principal plane can be transferred onto the diagonal at each iteration.

During the ℓ-th iteration, the Householder matrix is the product of row-wise Householder reflection matrices, $\mathbf{H}_p$, defined in (3.17), recursively computed from rows $1, 2, \ldots, (Q-2)$ giving

$$\mathbf{H}^{(\ell)} = \mathbf{H}_{Q-2} \ldots \mathbf{H}_2 \mathbf{H}_1 \ , \tag{3.31}$$

which reduces the principal plane to tri-diagonal form. Each Givens rotation based on (3.12) can be computed to zero off-diagonal elements i.e. $q = p + 1, p = 1, \ldots, Q - 1$ on the principal plane so that

$$\mathbf{G}^{(\ell)} = \mathbf{G}_{(Q-1,Q)} \ldots \mathbf{G}_{(2,3)} \mathbf{G}_{(1,2)} \ . \tag{3.32}$$

The unitary term for the ℓ-th iteration fulfilling (3.19) is

$$\mathscr{W}_\ell(z) = \mathbf{G}^{(\ell)} \mathbf{H}^{(\ell)} \mathscr{D}^{(\ell)}(z) \ , \tag{3.33}$$

a product of rotation, reflection and delay (polynomial) matrices and is applied to the polynomial matrix across all z.

The proof of convergence follows directly from [30] since the proposed algorithm transfers at least as much energy as SBR2 under the worst case where there is only 1 off-diagonal pair on the principal plane to be zeroed. With more than 1 off-diagonal pair, more energy will be transferred. Consequently, there exists a minimal upper bound for the number of iterations, L, where the algorithm diagonalizes $\mathbf{\Lambda}(z)$ approximately such that the modulus square of the off-diagonals are arbitrarily small.

Although the trace norm of the principal plane increases monotonically, the same transformation applied to other z-planes may increase the magnitude of off-diagonal elements.

With this, SBR2 with Householder reduction is summarized in Algorithm 1.

Algorithm 1 SBR2 with Householder Reduction.

Inputs: $\mathcal{R}(z) \in \mathbb{C}^{Q\times Q}$, δ, maxIter, μ.
Initialize: $\ell \leftarrow 0$, $g \leftarrow 1+\delta$, $\tilde{\mathbf{\Lambda}}(z) = \mathcal{R}(z)$, $\tilde{\mathcal{W}}(z) = \mathbf{I}$.
  while ($\ell <$maxIter **and** $|g| > \delta$) **do**
    $g \leftarrow \max|r_{p,q}(z^{\tau})|, q > p, \forall\tau$. *// search for maximum off-diagonal.*
    if ($|g| > \delta$) **then**
      $\ell \leftarrow \ell + 1$.
      $\tilde{\mathbf{\Lambda}}(z) \leftarrow \mathcal{D}_q(z)\tilde{\mathbf{\Lambda}}(z)\mathcal{D}_q^{\mathrm{P}}(z)$,
      $\tilde{\mathcal{W}}(z) \leftarrow \mathcal{D}_q(z)\tilde{\mathcal{W}}(z)$ *// delay according to* (3.20).
      $\tilde{\mathbf{\Lambda}}(z) \leftarrow \mathbf{H}\tilde{\mathbf{\Lambda}}(z)\mathbf{H}^{\mathrm{H}}$,
      $\tilde{\mathcal{W}}(z) \leftarrow \mathbf{H}\tilde{\mathcal{W}}(z)$ *// reflect according to* (3.31).
      $\tilde{\mathbf{\Lambda}}(z) \leftarrow \mathbf{G}(\theta,\phi)\tilde{\mathbf{\Lambda}}(z)\mathbf{G}^{\mathrm{H}}(\theta,\phi)$,
      $\tilde{\mathcal{W}}(z) \leftarrow \mathbf{G}(\theta,\phi)\tilde{\mathcal{W}}(z)$ *// rotate according to* (3.32).
      $\tilde{\mathbf{\Lambda}}(z) \leftarrow \mathrm{trim}(\tilde{\mathbf{\Lambda}}(z),\mu)$,
      $\tilde{\mathcal{W}}(z) \leftarrow \mathrm{trim}(\tilde{\mathcal{W}}(z),\mu)$ *// trim according to* (3.30).
    end if
  end while
  return $\tilde{\mathcal{W}}(z)$, $\tilde{\mathbf{\Lambda}}(z)$.

3.3.4 Complexity Analysis of the Considered Algorithms

The overall computational complexity for the decomposition of $\mathcal{R}(z) \in \mathbb{C}^{Q\times Q}$ depends on many factors, including the number of iterations L, search strategy, zeroing and trimming method, polynomial order at each iteration T and the implementation.

Since the PEVD methods are extended from EVD techniques, relevant results from numerical linear algebra are reviewed for the computation of both the eigenvalues and eigenvectors [12, 103]. By exploiting the symmetry of the matrix, each Givens rotation similarity transform uses $12Q$ operations, while tri-diagonal reduction requires $\frac{8}{3}Q^3$ flops [12, 103]. The classical Jacobi EVD of a symmetric matrix, which typically uses 6 to 10 sweeps to achieve convergence, requires $\frac{4}{3}Q^3$ operations per sweep [103]. The proposed tri-diagonal plus Givens approach uses $\frac{8}{3}Q^3 + 12Q^2 - 12Q$ flops per sweep. Although it is more costly per sweep to use the Householder-plus-Givens approach over the Givens-only approach in Jacobi's algorithm, the proposed approach tends to take fewer sweeps or iterations to converge, requiring fewer computations overall.

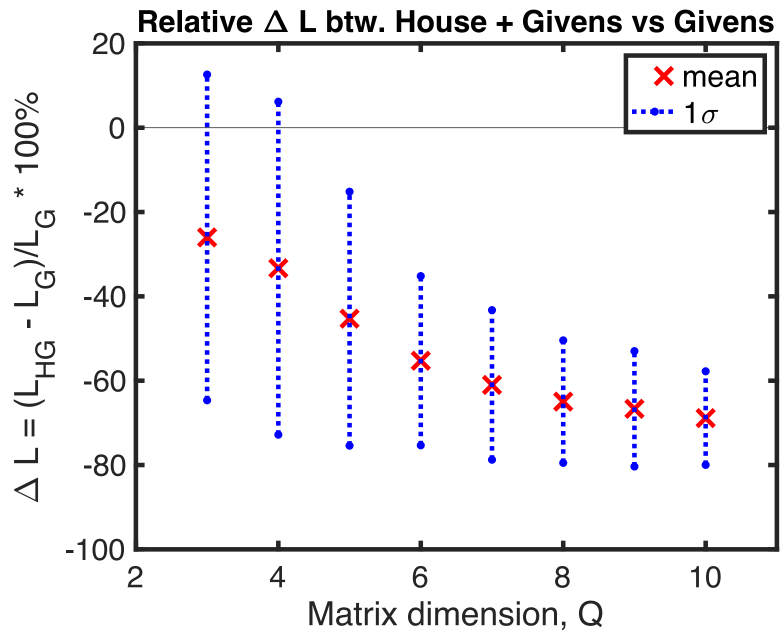


Figure 3.14: Comparison of iteration counts for Householder-plus-Givens and Givens-only in Jacobi's algorithm for EVD.

Using 1000 randomly generated symmetric matrices for each $Q = 3, \ldots, 10$, the number of iterations used for matrix diagonalization is computed for the Householder-plus-Givens L_{HG} and Givens-only L_G. The tolerable maximum off-diagonal magnitude δ is set to $\sqrt{N_1/3} \times 10^{-2}$. The normalized relative difference ΔL describes the reduction in iteration counts using the proposed approach over Jacobi's algorithm, and lower is better. As shown in Fig. 3.14, up to 70% reduction in ΔL is observed as the matrix dimension Q increases because there are more elements for Jacobi's algorithm to iterate through per sweep.

For the case of a polynomial matrix, the cost per iteration also scales with the polynomial order T. Consequently, the total complexity of each PEVD algorithm is approximately $\mathcal{O}(Q^3 T)$ for each iteration. The comparison between a Jacobi-like SBR2 algorithm and our proposed approach for PEVD is next discussed.

3.3.5 Simulations and Results

A. Numerical Examples

The results presented here will follow the convention established in [30] in which (3.18) was defined as $\mathcal{R}(z) \approx \mathcal{W}^P(z)\mathbf{\Lambda}(z)\mathcal{W}(z)$. The eigenvectors lie in the rows of $\mathcal{W}(z)$ and delays in (3.20) were applied to index q instead of p. The first example is

$$\mathcal{R}_1(z) = \begin{bmatrix} 1 & -0.4\mathrm{j}z & 0 \\ 0.4\mathrm{j}z^{-1} & 1 & 0.5z^{-2} \\ 0 & 0.5z^2 & 1 \end{bmatrix} . \tag{3.34}$$

When the final value of $|g|$ was zero (up to computational precision) using $\mu = 10^{-3}$, the convergence for this example using both methods is shown in Fig. 3.15. The convergence for the first two iterations were identical because the principal plane of the polynomial matrix had only one off-diagonal pair. From the third iteration, our method was able to transfer more energy to the diagonal as evident in the figure. Despite so, our method took one more iteration to produce the exact polynomial factors as SBR2 provided in [30]. This is because (3.34) is sparse with few non-zero polynomial elements and like conventional matrix, numerical diagonalization using the Givens method may converge at least as fast as the Householder method.

The second example is

$$\mathcal{R}_2(z) = \begin{bmatrix} 1 & 0.8z^2 - 0.4z & 0.7z \\ 0.8z^{-2} - 0.4z^{-1} & 1 & 0.5z^{-2} \\ 0.7z^{-1} & 0.5z^2 & -1 \end{bmatrix} , \tag{3.35}$$

for which our method took 17 iterations compared to 23 iterations with the original SBR2 using $\delta = 10^{-3}$ and $\mu = 10^{-4}$. As shown in Fig. 3.15, our method was able to transfer more energy at each iteration and drive down the dominant element faster. Furthermore, our method resulted in a lower reconstruction error, ϵ, from 0.0672 to 0.0478. This is

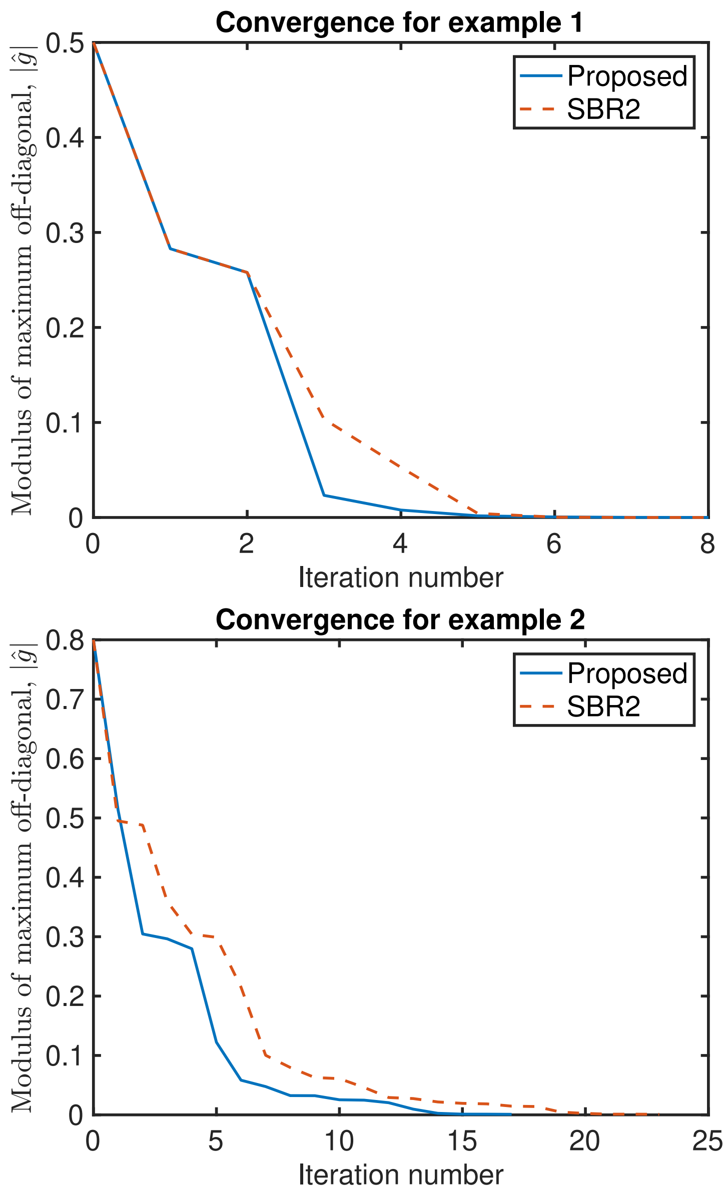


Figure 3.15: Convergence of algorithms for example 1 in (3.34) and example 2 in (3.35).

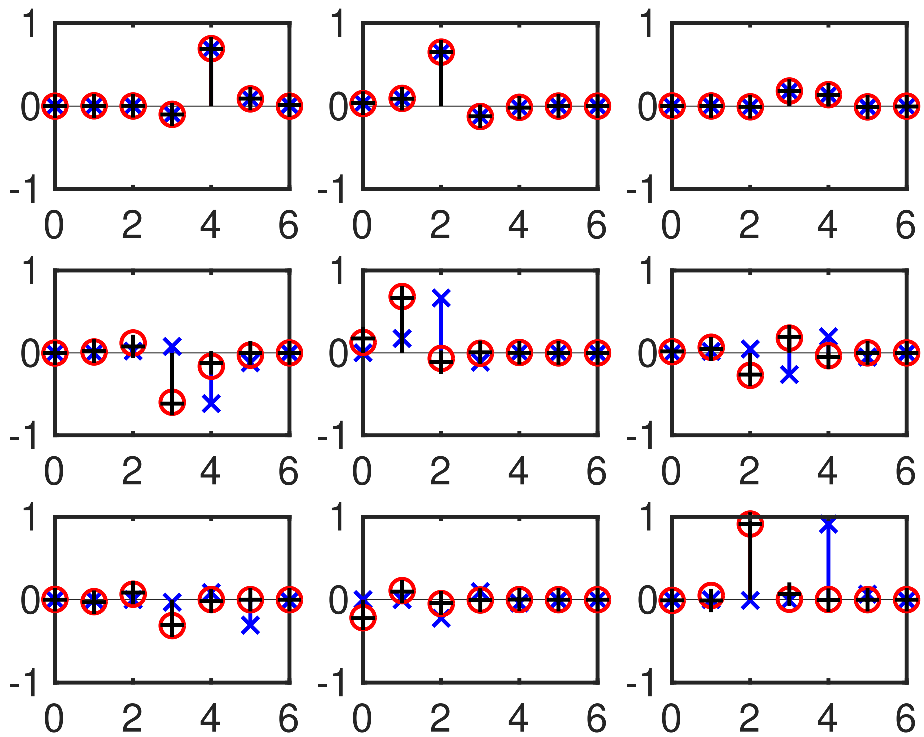


Figure 3.16: $\mathcal{W}(z)$ for (3.35) where black pluses and blue crosses indicate results computed using our proposed method with and without delay compensation, and red circles using SBR2.

because trimming at every iteration keeps the matrix factors compact at the expense of accuracy and the associated errors accumulate with more iterations. $\mathbf{\Lambda}(z)$ computed using both algorithms were exactly the same found in [30]. For $\mathcal{W}(z)$ shown in Fig. 3.16, the eigenvectors produced by both methods are comparable up to a delay ambiguity as discussed in Section 3.3.2. The compensated delay was also computed and shown as black plus signs.

Without trimming, our method converges in 30 iterations compared to 37 iterations using the original SBR2. The order for the polynomial factors were comparable with slight savings at 216 versus 224 for $\mathbf{\Lambda}(z)$ and 106 versus 110 for $\mathcal{W}(z)$. Both algorithms gave $\epsilon = 0$ (up to machine precision).

B. Decorrelation Example

The algorithms were compared with a decorrelation example involving 3 sensors and 2 sources using $\delta = \sqrt{\frac{N_1}{3\times10^4}}$ where N_1 is the square of the trace norm at the principal plane and $\mu = 10^{-4}$. Each channel was a 5th order FIR filter with coefficients drawn from a uniform distribution, $U\,[-1, 1]$. The source signals were independent, identically distributed sequences with each sample being assigned ±1 with equal probability. Each sensor output was also corrupted by zero-mean, Gaussian noise with $\sigma = 1.8$. 1000 samples were used to compute $\mathbf{R}(\tau)$, assumed to be negligibly small outside the correlation window, T. With $T = 10$, an approximated z-transform, $\mathcal{R}(z) = \sum_{\tau=-T}^{T} \mathbf{R}(\tau) z^{-\tau}$, was used in the computation.

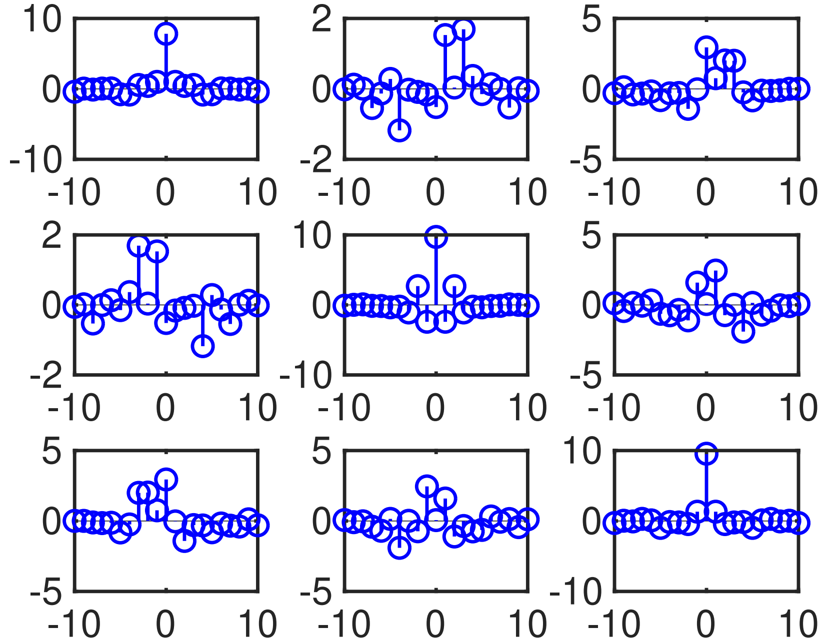
For the realization in Fig. 3.17, our method took 64 iterations while the original SBR2 took 90 iterations. Also, the resulting ϵ values were 1.4515 and 1.7722, respectively. Lower ϵ was likely due to fewer iterations and accumulation of inaccuracies arising from trimming. The order of $\mathbf{\Lambda}(z)$ was 27 for both methods while the order of $\mathcal{W}(z)$ was 33 and 35.

Without trimming, $\mu = 0$, our method took 80 iterations while SBR2 took 114 iterations and the resulting order of $\mathbf{\Lambda}(z)$ was 633 and 1148 and $\mathcal{W}(z)$ was 624 and 1138. We note, however, that the overall computational complexity depends on many factors including the number of iterations, search strategy, zeroing and trimming method, polynomial order at each iteration and the implementation.

C. Monte-Carlo Simulation

To further compare their performance, a Monte-Carlo simulation of 1000 decorrelation trials with trimming was conducted. The relative iteration difference between both methods was computed based on $(L_{\text{Proposed}} - L_{\text{SBR2}}) \times 100/L_{\text{SBR2}}\%$ and relative error difference was based on $(\epsilon_{\text{Proposed}} - \epsilon_{\text{SBR2}}) \times 100/\sum_{\forall z} \|\mathcal{R}(z)\|_F\%$. From the histogram shown in Fig. 3.18, the proposed method gave an average of 12.35% reduction in the number of iterations over SBR2 with an average reduction of 0.1% in relative ϵ. Moreover, because the distribution was asymmetric, our proposed method enjoyed a reduction in the number of

(a) $\mathcal{R}(z)$ for a single realization.



(b) $\mathbf{\Lambda}(z)$ from the PEVD of $\mathcal{R}(z)$ above.

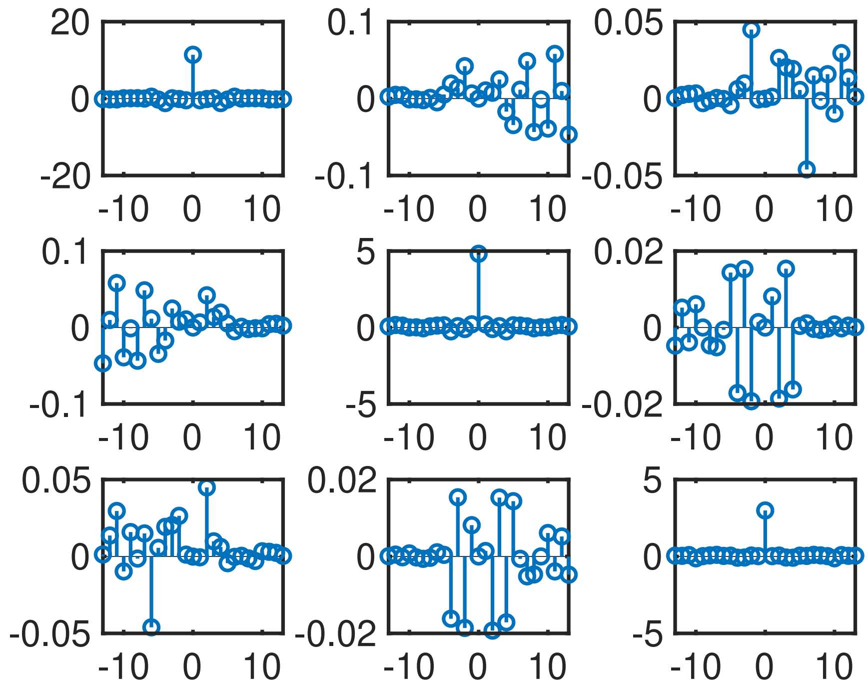


Figure 3.17: An example from the decorrelation Monte-Carlo simulation.

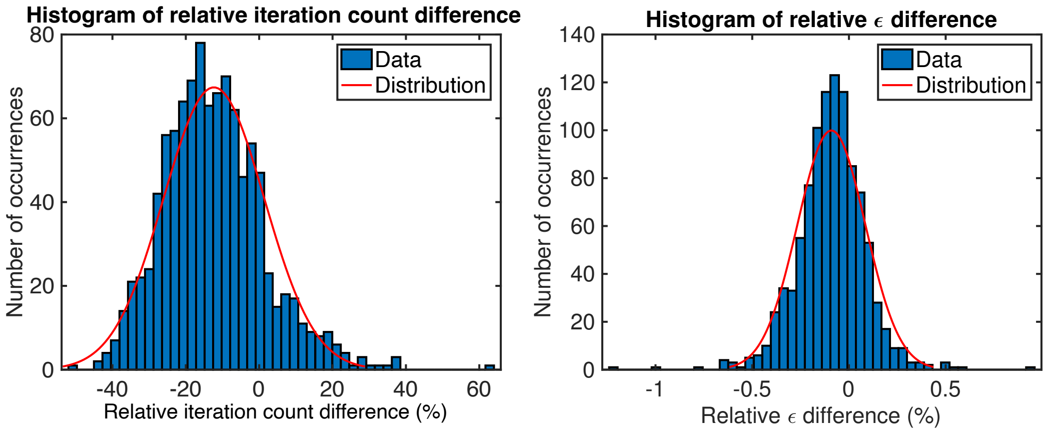


Figure 3.18: Monte-Carlo simulation results for 1000 decorrelation trials.

iterations for 82.4% of the trials. The results obtained with both methods were consistent to within 1% ϵ.

3.4 Chapter Summary and Conclusions

This chapter has highlighted the differences between EVD and PEVD and explained why PEVD is more suited for multi-channel broadband signal processing applications. The basis functions generated by the PEVD eigenvectors span a shift-invariant space, thereby allowing a more compact representation of signals compared to the EVD. The compact signal representation is further explored in Chapter 5 and the interpretation of polynomial eigenvectors will be discussed in Chapter 6.

The classical SBR2 algorithm was briefly reviewed, along with a proposed enhancement using Householder reflection to reduce the principal plane to a tri-diagonal form before rotation. This proposed method was seen able to transfer more energy to the principal diagonal at every iteration, enabling quicker convergence with an average of 12.35% reduction in iteration counts, and generally polynomial matrix factors with lower order and 0.1% improvement in reconstruction error. The proposed method enjoyed convergence speed improvements for 82.4% of the trials in the comparative simulation.

Contributions

Speech Processing Applications Using Polynomial EVD

Chapter 4

PEVD-based Speech Enhancement in Noisy Reverberant Environments

4.1 Introduction

Speech enhancement is essential for telecommunications, hearing aids, automatic speech recognition and voice-controlled systems. Enhancement algorithms aim to reduce interfering noise and reverberation while minimizing speech distortion. In this chapter, existing speech enhancement algorithms are first reviewed. Polynomial matrices are then introduced to model the spatial, spectral and temporal correlations between the speech signals received by a microphone array and polynomial matrix eigenvalue decomposition (PEVD) to decorrelate in space, time and frequency simultaneously. This leads to our proposed blind and unsupervised PEVD-based speech enhancement algorithm. Simulations and informal listening examples involving diverse reverberant and noisy environments have shown that our method can jointly suppress noise and reverberation, thereby achieving speech enhancement without introducing processing artefacts into the enhanced signal. This chapter is based closely on the published article [39].

4.2 Overview of Speech Enhancement Algorithms

The enhancement of degraded speech signals remains important in many applications ranging from human-to-human communications in telecommunications and hearing aids [108–110] to human-to-machine interaction in automatic speech recognition, voice-controlled systems and robot audition [111, 112]. The main causes of degradation are additive background noise and reverberation due to multi-path reflections in enclosed spaces [24]. Consequently, the speech signal becomes temporally smeared and contaminated by interfering signals. Furthermore, in all these applications, prior information of the target speech or the acoustic environment is not available, motivating the need for a blind, or unsupervised, approach.

The speech enhancement task which includes noise reduction [113] and dereverberation [114] will be addressed in this chapter. Existing enhancement algorithms may typically distort the speech signal and introduce processing artefacts [115–117]. Methods for controlling the trade-off between noise reduction against speech distortions are introduced in [118–120]. For instance, aggressive noise reduction might be preferred for human-to-machine applications while mobile phone and hearing aid users might prefer less speech distortion at the expense of higher residual noise.

Existing noise reduction techniques can be classified into single- and multi-channel techniques. Methods for single-channel noise reduction include spectral subtraction [121,122], statistical-based and subspace-based approaches. Statistical methods are typically based on minimizing the mean-square error (MMSE) of the clean and estimated speech spectrum [118], the log-spectrum (log-MMSE) [123] or the single-channel Wiener filter [119, 124].

In subspace methods, noisy signals are decomposed into signal and noise subspaces and enhancement is achieved by recovering the speech signal from the signal subspace. The Karhunen-Loève transform (KLT) was used in [125] and an optimal solution was derived for white noise. This approach was extended in [126] to cope with coloured noise by using a GEVD to jointly diagonalize the speech and noise covariance matrices. Subspace-based methods have also been extended to multi-channel systems [127,128] but they do not fully exploit spatial information to minimize speech distortion [113], as will be seen in Section 4.7.

Multi-microphone methods include beamformers [129–131] with optional post-filtering [132,133], the optimal MWF [119,124,134] and the multi-channel Kalman filter [135,136].

While existing multi-channel approaches may potentially achieve noise reduction without speech distortion when the noise is spatially and temporally white, this remains a practical challenge under other noise conditions in real-world scenarios [113].

Speech dereverberation approaches can be classified into speech synthesis-based, reverberation cancellation and reverberation suppression methods. In synthesis-based methods, the linear prediction coding (LPC) residual is directly computed to generate an estimated clean speech signal [137, 138]. This approach, however, is usually limited to mildly reverberant signals in order to avoid introducing artefacts [70].

In one class of reverberation cancellation approaches, the acoustic channel is first estimated using, for example, blind system identification [101, 139–141]. The estimated channel is then used in the design of inverse filter(s), such as [142] for a single-channel system and [143] for a multi-channel system based on the multi-channel inversion theorem (MINT). Because MINT is not robust to estimation errors, channel shortening techniques have been proposed [144–146]. In another class of cancellation methods, multi-channel linear prediction, such as the weighted prediction error (WPE) [147–149], is used to directly estimate the dereverberated speech signal.

The reverberation suppression methods include single-channel spectral subtraction [150] and multi-channel spatial filtering techniques such as the MWF [119, 124], which may be used in conjunction with post-filtering techniques [133]. Because these methods use noise and/or transfer function estimators, the dereverberation effectiveness depends heavily on the performance of these estimators.

A large number of contributions are based on the assumption of narrowband signal models, e.g., [125–128]. Single-channel approaches exploit temporal correlations while multichannel approaches capture spatio-temporal correlations. However, extensions to broadband signals using the STFT generally ignore the correlations between frequency bands and cannot preserve phase coherence across bands [14]. Furthermore, the microphone outputs are temporally correlated because of the speech source signal and reverberation. Consequently, an EVD or KLT, which removes correlations at a single time lag, is inadequate for decorrelating the signals completely. Polynomial matrices can simultaneously capture the correlations in space, time and frequency and are, therefore, appropriate for modelling multi-channel broadband signals.

The processing of polynomial matrices has motivated the development of a family of polynomial matrix eigenvalue decomposition (PEVD) algorithms [25, 28, 29], which are based on the SBR2 [30]. Unlike EVD or KLT, PEVD can achieve decorrelation over a suitably chosen range of time lags and is more suitable for broadband signals. It is also widely used in many multi-channel broadband signal processing applications such as blind source separation [33], source identification [34], localization [67], adaptive beamforming [31] and channel coding [35].

In this chapter, we introduce and extend PEVD to the field of speech signal processing and propose a blind and unsupervised PEVD-based speech enhancement algorithm. Preliminary studies separately focused on the task of noise reduction [36, 37] and speech dereverberation [38]. In contrast, in this chapter, we present PEVD-based speech enhancement for both noise reduction and dereverberation. In addition, we further investigate the impact of the parameter settings of the proposed algorithm and provide a thorough performance evaluation in different acoustic conditions, benchmarked against state-of-the-art baseline approaches for speech enhancement. Therefore, in supplement to the earlier studies, the novel contributions of this chapter are (i) the use of a polynomial matrix as a broadband, multi-channel signal model for speech, (ii) presentation of a novel PEVD-based speech enhancement algorithm, (iii) the investigation of the parameters of the proposed algorithm, (iv) a comprehensive evaluation and analysis of the proposed approach for realistic signals under wide-range of noise and reverberation conditions, and (v) an evaluation of our PEVD method against several comparative approaches.

4.3 Speech Enhancement Problem Formulation

The noisy and reverberant signal, $x_q(n)$, at the q-th microphone for discrete-time sample $n = 0, 1, \ldots, N$ is

$$x_q(n)=\mathbf{h}_q^{\mathrm{T}}\mathbf{s}_0(n) + v_q(n) \ , \tag{4.1}$$

where $\mathbf{s}_0(n) = \Big[s_0(n), \ldots, s_0(n-J)\Big]^{\mathrm{T}}$ is the anechoic speech, $\mathbf{h}_q = \Big[h_q(0), \ldots, h_q(J)\Big]^{\mathrm{T}}$ is the RIR from the source to the q-th microphone, assumed to be stationary and modelled using a J-th order FIR filter, $v_q(n)$ represents the additive noise at the q-th microphone and $[\cdot]^{\mathrm{T}}$ denotes the transpose operator. The noise signals are assumed to be zero-mean, and uncorrelated with each other and the source signal [113].

Using the reverberation model in [24], the early reflections represent closely spaced distinct echoes that perceptually reinforce the direct-path component and may improve speech intelligibility in certain conditions [151]. The late reflections comprise randomly distributed small amplitude components, which are commonly assumed uncorrelated with the direct-path and early reflections, and can be treated as an additive, uncorrelated noise component [152]. Consequently, (4.1) becomes

$$\begin{aligned} x_q(n) &= \mathbf{h}_{q,e}^{\mathrm{T}}\mathbf{s}_0(n) + x_{q,l}(n) + v_q(n) \\ &= \tilde{s}_q(n) + \tilde{v}_q(n) \ , \end{aligned} \tag{4.2}$$

where $\mathbf{h}_{q,e}$ is the impulse responses associated with the direct-path and early reflections, $x_{q,l}(n)$ is the late reverberant component, $\tilde{s}_q(n) = \mathbf{h}_{q,e}^{\mathrm{T}}\mathbf{s}_0(n)$ and $\tilde{v}_q(n) = x_{q,l}(n) + v_q(n)$ are the desired speech and unwanted noise at the q-th microphone. The data vector from Q microphones is $\mathbf{x}(n) = [x_1(n), \ldots, x_Q(n)]^{\mathrm{T}} \in \mathbb{R}^Q$, with $\mathbf{x}_l(n)$, $\mathbf{s}(n)$, $\tilde{\mathbf{s}}(n)$, $\mathbf{v}(n)$ and $\tilde{\mathbf{v}}(n)$ similarly defined.

4.3.1 Noise Reduction in Anechoic Environment

Without reverberation, each acoustic channel only comprises the direct-path propagation, modelled using a delay. Typically, the first channel, or the channel with the shortest propagation time from the source, can be taken as the reference. Consequently, (4.1) simplifies to

$$x_q(n) = s_q(n) + v_q(n) \ , \tag{4.3}$$

where $s_q(n)$ is an attenuated and delayed version of $s_0(n)$ at the q-th microphone. The goal of noise reduction is to recover $\mathbf{s}(n)$ from $\mathbf{x}(n)$ while keeping $\mathbf{v}(n)$ suppressed.

4.3.2 Speech Dereverberation in Noiseless Environment

In the absence of additive noise, (4.1) simplifies to

$$x_q(n) = \mathbf{h}_q^{\mathrm{T}} \mathbf{s}_0(n) \ , \tag{4.4}$$

since $v_q(n) = 0$. While retaining the early reflections is not normally perceived to be perceptually harmful, in this chapter, the evaluation of dereverberation targets the anechoic speech, $\mathbf{s}_0(n)$. Furthermore, if noise is present in the environment, residual additive noise after processing may remain but is not considered in the dereverberation evaluation.

4.4 Polynomial Matrix Decomposition

4.4.1 Motivation for Polynomial Matrices in Speech Processing

Broadband signals received by microphone arrays exhibit spatial, spectral and temporal correlations. The space-time covariance matrix is defined as [30]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau)\} \in \mathbb{C}^{Q \times Q} \ , \tag{4.5}$$

where $\mathbb{E}\{\cdot\}$ and $[\cdot]^{\mathrm{H}}$ are the expectation and Hermitian transpose operators, respectively. The $(p,q)^{\mathrm{th}}$ element of $\mathbf{R}_{\mathbf{xx}}(\tau)$, $r_{p,q}(\tau) = \mathbb{E}\{x_p(n)x_q^*(n-\tau)\}$, is computed using the correlation function between the received signal at the p-th and q-th microphone and is parameterized by the temporal lag, $\tau \in \mathbb{Z}$, and $[\cdot]^*$ is the complex conjugate operator.

If the source, $s_0(n)$, is a narrowband signal at frequency f_0 propagating in an anechoic environment, $x_q(n)$ is related to the reference signal at the first microphone, $x_1(n)$, by a constant phase shift, $\phi_q = 2\pi f_0 \tau_q$. The data vector becomes

$$\mathbf{x}(n) = [x_1(n), x_1(n)e^{-\mathrm{j}\phi_2}, \ldots, x_1(n)e^{-\mathrm{j}\phi_Q}]^{\mathrm{T}} , \tag{4.6}$$

such that the correlation between the $(p, q)^{\mathrm{th}}$ sensor pair is

$$r_{p,q}(\tau) = \mathbb{E}\{x_1(n)e^{-\mathrm{j}\phi_p}[x_1(n-\tau)e^{-\mathrm{j}\phi_q}]^*\} \tag{4.7}$$

$$= \mathbb{E}\{x_1(n)x_1(n)\}e^{-\mathrm{j}\phi_{p,q}+\mathrm{j}2\pi f_0 \tau} , \tag{4.8}$$

where $\phi_{p,q} = \phi_p - \phi_q$ is the phase difference that depends only on the array geometry and is associated with the TDoA. Since the expectation term in (4.8) is independent of τ and the array geometry is fixed such that $\phi_{p,q}$ remains constant across all lags, (4.8) can be computed at any τ for decorrelation using an EVD. Classical subspace-based approaches for the enhancement of narrowband signals compute the instantaneous spatial covariance matrix by evaluating (4.5) at $\tau = 0$, expressed as

$$\mathbf{R}_{\mathbf{xx}}(0) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\} . \tag{4.9}$$

However, for a broadband source signal like speech, (4.8) no longer holds since the source correlation now depends on τ. Therefore, correlations across different sensors and temporal lags need to be considered. Accordingly, concatenating the covariance matrix, $\mathbf{R}_{\mathbf{xx}}(\tau)$, for all values of $\tau \in \{-N, \ldots, N\}$, results in a tensor of dimension, $Q \times Q \times (2N + 1)$.

Instead of processing signals in the STFT domain, the z-transform which captures and preserves the correlations of the received signals in space, time and frequency is used. The z-transform of (4.5), denoted by $\mathbf{R}(\tau) \circ\!\!-\!\!\bullet\ \mathcal{R}(z) \in \mathbb{C}^{Q\times Q}$ and also represented using a 3D tensor, is a para-Hermitian polynomial matrix [30]

$$\mathcal{R}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau) z^{-\tau} . \tag{4.10}$$

The polynomial matrix is a matrix with polynomial elements, or equivalently, a polynomial with matrix coefficients. In the former, each element in the polynomial matrix represents the correlation function in z between a specific microphone pair. In the latter, the polynomial matrix shows how the spatial correlation between all sensor pairs changes with z.

4.4.2 Family of PEVD Algorithms

The PEVD of a para-Hermitian polynomial matrix is [48, 49]

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{\mathcal{U}}(z)\boldsymbol{\Lambda}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) , \tag{4.11}$$

where the columns of $\boldsymbol{\mathcal{U}}(z)$ are the eigenvectors and the diagonal polynomial matrix, $\boldsymbol{\Lambda}(z)$ contains the eigenvalues, and $[\cdot]^{\mathrm{P}}$ is the para-Hermitian operator defined as $\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) = \boldsymbol{\mathcal{U}}^{\mathrm{H}}(1/z^{*})$. The prefix 'para' indicates a time-reversal in this context. The decomposition in (4.11) for an analytic $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ can be factorized using analytic factors, $\boldsymbol{\mathcal{U}}(z)$ and $\boldsymbol{\Lambda}(z)$, which in most cases are infinite Laurent series. However, analyticity implies that sufficiently large order of Laurent polynomials can closely approximate (4.11) [48, 49].

The PEVD can be approximated by Laurent polynomial factors using an iterative algorithm [25, 28–30] based on similarity transforms involving L para-unitary polynomial matrices [45], $\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{U}}_L(z) \ \cdots \ \boldsymbol{\mathcal{U}}_1(z)$. At the ℓ-th iteration, the algorithm first searches for the largest off-diagonal element exceeding a predefined threshold, δ. $\boldsymbol{\mathcal{U}}_\ell(z)$ is then constructed using delay polynomial matrices and unitary matrices, which are designed to zero out the off-diagonal elements on the matrix coefficient of z^0, and applied to the entire polynomial matrix. To keep the polynomial order compact, a fraction μ, of the total Frobenius-norm squared, is truncated as detailed in [30]. After L iterations, $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ is approximately diagonalized according to [30, 153]

$$\boldsymbol{\Lambda}(z) \approx \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\mathbb{E}\{\boldsymbol{x}(z)\boldsymbol{x}^{\mathrm{P}}(z)\}\boldsymbol{\mathcal{U}}(z) , \tag{4.12}$$

where $\boldsymbol{x}(z)$ is the z-transform of $\mathbf{x}(n)$ based on (4.10).

The zeroing unitary matrix computed at each iteration can take the form of a Givens rotation in SBR2 [30], a Householder-like optimization procedure as in [25], a combination of Householder reflection and Givens rotation matrices in [29] or an eigenvector matrix in the SMD algorithm [28].

4.5 PEVD-based Speech Enhancement Algorithm

The z-transform of the space-time covariance matrix of the microphone signals obtained by applying (4.2) to (4.5) is

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{\hbar}_e(z)\boldsymbol{\hbar}_e^{\mathrm{P}}(z)\imath_{ss}(z) + \boldsymbol{\mathcal{R}}_{\boldsymbol{ll}}(z) + \boldsymbol{\mathcal{R}}_{\mathbf{vv}}(z) \tag{4.13}$$

$$= \boldsymbol{\mathcal{R}}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(z) + \boldsymbol{\mathcal{R}}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}(z) \ , \tag{4.14}$$

where $\imath_{ss}(z)$ is a 1×1 polynomial which forms a z-transform pair with the anechoic speech auto-correlation sequence, $\imath_{ss}(z) \bullet\!\!-\!\!\circ\, r_{ss}(\tau)$, computed using

$$r_{ss}(\tau) = \mathbb{E}\{s_0(n)s_0(n-\tau)\} \in \mathbb{R}^{1\times 1} \quad \forall \tau \ . \tag{4.15}$$

The channel polynomial vector is $\boldsymbol{\hbar}_e(z) = [\hbar_{1,e}(z), \ldots, \hbar_{Q,e}(z)]^{\mathrm{T}} \in \mathbb{C}^Q$, where $\hbar_{q,e}(z)$ is a 1×1 polynomial obtained by taking the z-transform of the direct-path and early reflections in the RIR from the source to the q-th microphone, i.e., $\hbar_{q,e}(z) = \sum_{i=0}^{J} h_{q,e}(i)z^{-i}$. The polynomial matrices $\boldsymbol{\mathcal{R}}_{\mathbf{vv}}(z) \in \mathbb{C}^{Q\times Q}$ and $\boldsymbol{\mathcal{R}}_{\boldsymbol{ll}}(z) \in \mathbb{C}^{Q\times Q}$ are the space-time covariance of the noise and late reverberation modelled as a spatially diffuse field [152] computed using $\mathbf{v}(n)$ and $\mathbf{x}_l(n)$ following (4.5) and (4.10), respectively. The speech component $\boldsymbol{\mathcal{R}}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(z)$, obtained from the direct-path and early reflections, is assumed uncorrelated with $\boldsymbol{\mathcal{R}}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}(z)$, which includes the late reflections and additive noise components.

Assuming stationarity within each processing frame, (4.5) is estimated using $T+1$ samples per frame according to

$$\hat{\mathbf{R}}_{\mathbf{xx}}(\tau) \approx \frac{1}{T+1}\sum_{n=0}^{T} \mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau) \ . \tag{4.16}$$

Furthermore, (4.10) can be approximated using

$$\hat{\mathcal{R}}_{\mathbf{xx}}(z) \approx \sum_{\tau=-W}^{W} \mathbf{R}_{\mathbf{xx}}(\tau) z^{-\tau} \ , \tag{4.17}$$

where W is the truncation window which determines the extent of the supported temporal correlation of the speech signals. Hence, in addition to the PEVD parameters, the frame size, T, and window length, W, can affect the performance of the proposed algorithm as investigated in Section 4.7.1. Since noise and speech are assumed uncorrelated, the PEVD gives [37]

$$\mathcal{R}_{\mathbf{xx}}(z) \approx \left[\boldsymbol{\mathcal{U}}_{\tilde{s}}(z) \;\middle|\; \boldsymbol{\mathcal{U}}_{\tilde{v}}(z) \right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_{\tilde{s}}(z) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Lambda}_{\tilde{v}}(z) \end{array} \right] \left[\begin{array}{c} \boldsymbol{\mathcal{U}}^{\mathrm{P}}_{\tilde{s}}(z) \\ \hline \boldsymbol{\mathcal{U}}^{\mathrm{P}}_{\tilde{v}}(z) \end{array} \right] \ , \tag{4.18}$$

where $\{.\}_{\tilde{s}}$ and $\{.\}_{\tilde{v}}$ are associated with the signal-plus-noise (or simply signal) and noise-only (or simply noise) subspaces.

The eigenvector polynomial matrix, $\boldsymbol{\mathcal{U}}(z)$, can be applied as a filterbank for $\mathbf{x}(z)$ so that the outputs, $\mathbf{y}(z) = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\mathbf{x}(z)$, are strongly decorrelated [28], according to

$$\mathbb{E}\{\mathbf{y}(z)\mathbf{y}^{\mathrm{P}}(z)\} = \mathbb{E}\{\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\mathbf{x}(z)\mathbf{x}^{\mathrm{P}}(z)\boldsymbol{\mathcal{U}}(z)\} \approx \boldsymbol{\Lambda}(z) \ . \tag{4.19}$$

Unlike some speech enhancement approaches, the proposed method does not use any noise estimation algorithms since the strong decorrelation property of PEVD implicitly orthogonalizes the subspaces across all time lags in the range of W. Furthermore, $\boldsymbol{\mathcal{U}}(z)$ is lossless or para-unitary by construction and has an all-pass filter frequency response [45]. This implies that $\boldsymbol{\mathcal{U}}(z)$ can only redistribute spectral power among channels and not change the total (over all subspaces) signal and noise power [30]. The PEVD algorithms in [30] also tend to sort the strongly decorrelated outputs in descending order of signal energy because of the spectral majorization property [28]. The signal subspace comprises mostly speech components, originally distributed over all microphones but now summed coherently. In contrast, the noise subspace is dominated by ambient noise and late reflections in the reverberant channels. Consequently, speech enhancement is achieved by combining components in the signal subspace of dimension 1, defined as the first channel with the largest total spectral power, and nulling components in the noise subspace, e.g., $\boldsymbol{\mathcal{U}}_{\tilde{v}}(z) = \mathbf{0}$.

The computational complexity of the proposed algorithm depends on the number of microphones Q, the order of the polynomial matrix $(2W+1)$, the number of iterations L and is approximately $\mathcal{O}(Q^3WL)$. The PEVD-based speech enhancement method is summarized in Algorithm 2.

4.5.1 Noise Reduction in Anechoic Environment Case

In an anechoic environment, (4.3) simplifies (4.13) to

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{\mathcal{R}}_{\mathbf{ss}}(z) + \boldsymbol{\mathcal{R}}_{\mathbf{vv}}(z) \ , \tag{4.20}$$

where $\boldsymbol{\mathcal{R}}_{\mathbf{ss}}(z) \bullet\!\!-\!\!\circ\ \mathbf{R}_{ss}(\tau)$ is the space-time covariance matrix of the delayed speech received by the microphones $\mathbf{s}(n)$. The $(p,q)^{\text{th}}$ element of $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ can be written as

$$\begin{aligned} r_{p,q}(\tau) &= \mathbb{E}\{s_p(n)s_q(n-\tau) + v_p(n)v_q(n-\tau)\} \\ &= \mathbb{E}\{s_0(n-\tau_p)s_0(n-\tau_q-\tau)\} + \mathbb{E}\{v_p(n)v_q(n-\tau)\} \ , \end{aligned} \tag{4.21}$$

which implies that the difference in lags between every channel pair can only be captured by PEVD in the range of W. Then, PEVD decomposes the signal information into subspaces associated with the anechoic speech and the noise signals based on (4.18).

4.5.2 Speech Dereverberation in Noiseless Environment Case

When the environment is noiseless, (4.4) simplifies (4.13) to

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{h}_e(z)\boldsymbol{h}_e^{\mathrm{P}}(z)r_{ss}(z) + \boldsymbol{\mathcal{R}}_{\boldsymbol{ll}}(z) \ . \tag{4.22}$$

Comparing with (4.14), the first term is part of the desired speech subspace. It is uncorrelated and orthogonal with the second term, which contains only the late components and forms the noise subspace. The window size, W, should be large enough to include the direct-path and early reflections in order to achieve the desired enhancement.

Algorithm 2 PEVD-based speech enhancement [37].

Inputs: $\mathbf{x}(n) \in \mathbb{R}^Q, n \in \{0, \ldots, T\}, W, \delta, \mu, L$.
$\mathbf{R}_{\mathbf{xx}}(\tau) \leftarrow \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau)\}$ *// see* (4.5)
$\mathcal{R}_{\mathbf{xx}}(z) \leftarrow \mathcal{Z}\{\mathbf{R}_{\mathbf{xx}}(\tau)\}$ *// see* (4.10)
$\mathcal{U}(z), \mathbf{\Lambda}(z) \leftarrow \mathrm{PEVD}\ \{\mathcal{R}_{\mathbf{xx}}(z), \delta, \mu, L\}$ *// any [25, 28–30]*
$\mathbf{x}(z) \leftarrow \mathcal{Z}\{\mathbf{x}(n)\}$ *// see* (4.10)
$\mathbf{y}(z) \leftarrow \mathcal{U}^{\mathrm{P}}(z)\mathbf{x}(z)$ *// speech enhancement*
$\mathbf{y}_1(z) \leftarrow \mathbf{y}(z)$ *// enhanced signal in the first channel*
return $\mathbf{y}_1(z)$.

4.6 Experimental Setup

4.6.1 Acoustic Environments and Setup

Anechoic speech signals, which are sampled at 16 kHz, are taken from TIMIT corpus [154] because it has a good phonetic balance and is widely used in the literature. Room impulse response measurements and noise recordings are taken from the complete ACE corpus [155]. The T_{60} of the rooms in the ACE corpus range from 0.332 s to 1.22 s. The $Q = 3$ channel 'mobile' array was used in most experiments. The $Q = 8$ channel linear array, with microphones spaced by 60 mm, was used for the study on the use of a different number of microphones. In the direct-path-only experiments, the propagation delays were drawn from the discrete uniform distribution, $U(1, 1000)$, and ordered such that $\tau_1 < \tau_2 < \tau_3$.

Babble, car and factory noise from the Noisex database [156], as well as restaurant, residential traffic and city street noise from the International Sound Effects (SoundFx) library [157], and white noise, were used.

The noise recordings in ACE were used directly. With the Noisex and SoundFx noise signals, diffuse noise signals were produced using [158]. In each trial, sentences from a randomly selected speaker were concatenated to have 8 to 10 s duration. The anechoic speech signals were then convolved with the impulse responses at each microphone channel before being corrupted by additive noise using [159], implemented in [160]. The signal-to-noise ratio (SNR) ranged from -10 dB to 20 dB. For each Monte-Carlo simulation, 50 trials were conducted.

4.6.2 Speech Enhancement Comparative Algorithms

The proposed PEVD method was compared against two versions of the MWF, two subspace approaches, the generalized weighted prediction error (GWPE), and two joint approaches that can suppress both noise and reverberation. Both MWFs are based on the concatenation of a minimum variance distortionless response (MVDR) beamformer followed by a single-channel Wiener filter [133]. The first is a practical MWF which uses a relative transfer function (RTF)-based speech estimator and a noise estimator based on the parameters used in [135]. The second is the oracle multi-channel Wiener filter (OMWF) which provides an ideal performance upper-bound since it uses complete prior knowledge of the clean speech signal, based on [119] where the filter length is 80. The PEVD subspace approach was also compared against the subspace method for coloured noise (COLSUB) which uses a GEVD [126] and the multi-channel subspace (MCSUB) method [127]. The GWPE dereverberation algorithm [147], and integrated methods for noise reduction and dereverberation such as the weighted power minimum distortionless response (WPD) [149] and the integrated sidelobe canceller linear prediction Kalman filter (ISCLP) [136], were also included as benchmark approaches. The published code was used for ISCLP while the published parameters and ground truth DoAs from ACE were used to compute the steering vectors for WPD [149] to avoid signal direction mismatch errors.

For all experiments, the PEVD parameters, chosen following [36–38], were $\delta = \sqrt{N_1/3} \times 10^{-2}$ where N_1 is the square of the trace-norm of $\mathbf{R}_{\mathbf{xx}}(0)$, $\mu = 10^{-3}$ and $L = 500$ iterations. In all experiments, except for those investigating the effects of varying T and W, $T = W = 1600$ samples were used. With this parameter selection, correlations within 100 ms, which were assumed to include the direct-path and early reflection components, were captured and used by the algorithm. The source code for our method and experiments is available [161].

4.6.3 Evaluation Measures

For the noise reduction evaluation, the segmental signal-to-noise ratio (SegSNR) [162, 163], and frequency-weighted segmental signal-to-noise ratio (FwSegSNR) [163, 164] are used. To measure the speech quality and intelligibility and to account for processing artefacts, short-time objective intelligibility (STOI) [165] and perceptual evaluation of speech quality (PESQ) [166] are used. A key measure of dereverberation is the direct-to-reverberant ratio (DRR) but the modified impulse responses after processing are generally unavailable. Instead, the normalized signal-to-reverberant ratio (NSRR), which is a signal-based measure shown to be equivalent to DRR under certain conditions [167], and the Bark spectral distortion (BSD) [24], are used.

These measures are computed for the signals before and after enhancement using the proposed and benchmark algorithms and the improvement Δ is reported. Positive Δ values show improvements in all measures except ΔBSD, for which a negative value indicates a reduction in spectral distortions.

4.7 Results and Discussions

Monte-Carlo simulations have been conducted to understand 1) the impact of varying the parameters of the proposed PEVD-based approach and; 2) the effectiveness of the proposed approach in comparison to other methods for speech enhancement, including the specific cases of noise reduction in anechoic environments and dereverberation in noiseless environments. Listening examples are also available [161].

4.7.1 Parameter Selection for the PEVD-based Algorithm

These parameter values have also been confirmed to exhibit similar trends for different scenarios including noise types and rooms. However, for brevity, we present in this chapter only the results for white noise in Lecture Room 2.

A. Frame Size and Window Length

With a fixed frame size, $T = 1600$ samples, the window length W was varied from 0 to 1600 samples. When $W = 0$, PEVD is equivalent to an EVD applied to the instantaneous spatial covariance matrix in (4.9), and this can only decorrelate the microphone signals at a single time lag. As W increases, correlations between microphones at other lags are computed and used in the PEVD algorithm. In the presence of white noise in Lecture Room 2, Fig. 4.1 highlights the limitation of using only the instantaneous lag when $W = 0$ as indicated by the lowest scores in all measures while the scores substantially improve as $W > 0$. When $W \geq 5$, the improvements are negligible because white noise in Fig. 4.1 does not exhibit any temporal correlations but is still advantageous to capture the speech correlations.

The marginal improvement in all measures across all SNRs with frame size T is shown in Fig. 4.2. Similar results are observed for intermediate values of T. This is expected since larger frame sizes will provide a better estimate of the second-order statistics for the PEVD algorithm, assuming stationarity.

B. Number of Microphones

The impact of changing the number of microphones Q in the range 2 to 8 was investigated using the 8 channel linear array. For white noise in Lecture Room 2, Fig. 4.3 shows that PEVD can reduce noise and reverberation without sacrificing speech intelligibility and quality even with 2 microphones as indicated by the improvement in all metrics. Generally, PEVD performs better with the number of channels as expected, with very similar results obtained for intermediate values of Q.

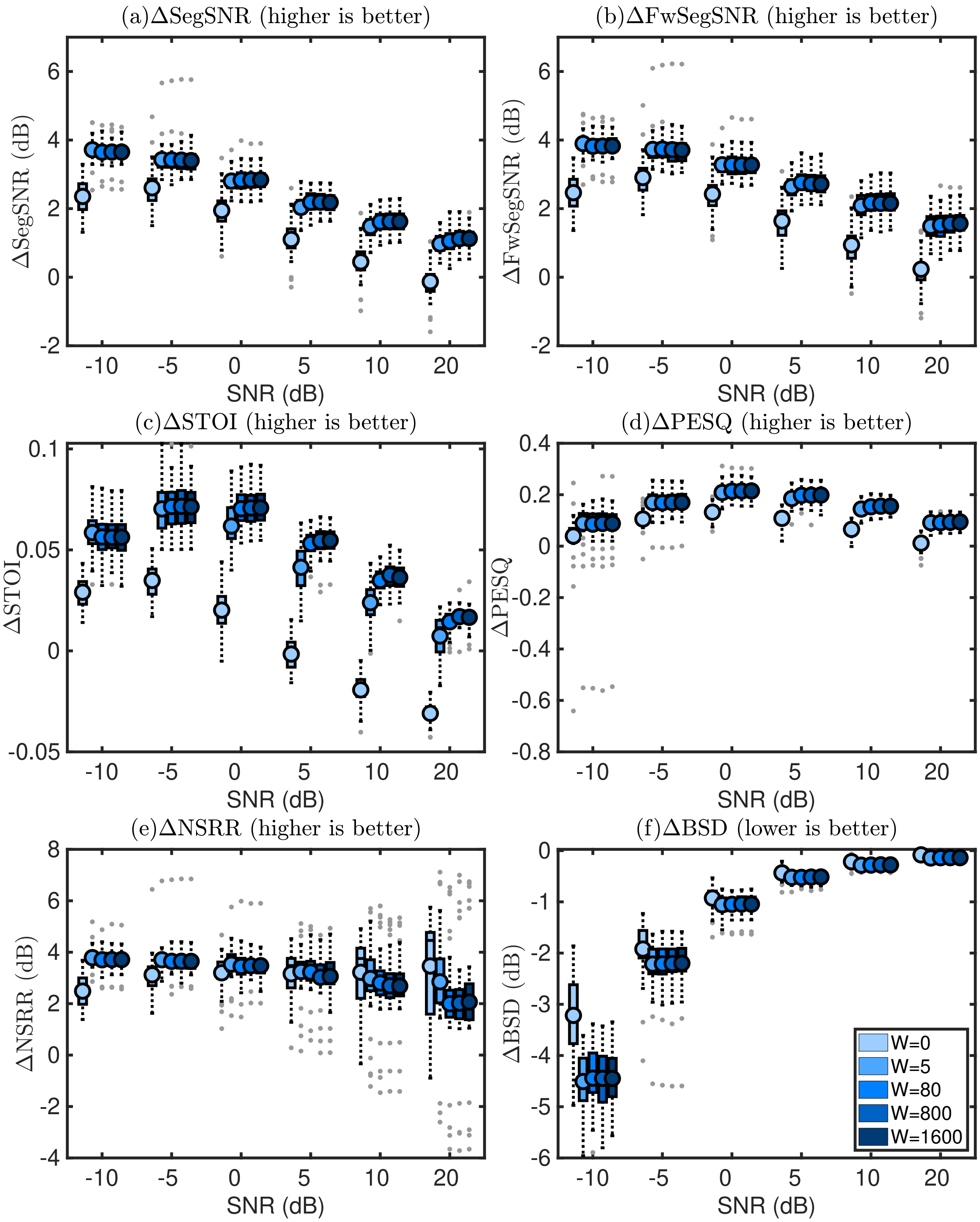


Figure 4.1: Speech enhancement for white noise in Lecture Room 2 using different window length W for the approximation of the z-transform.

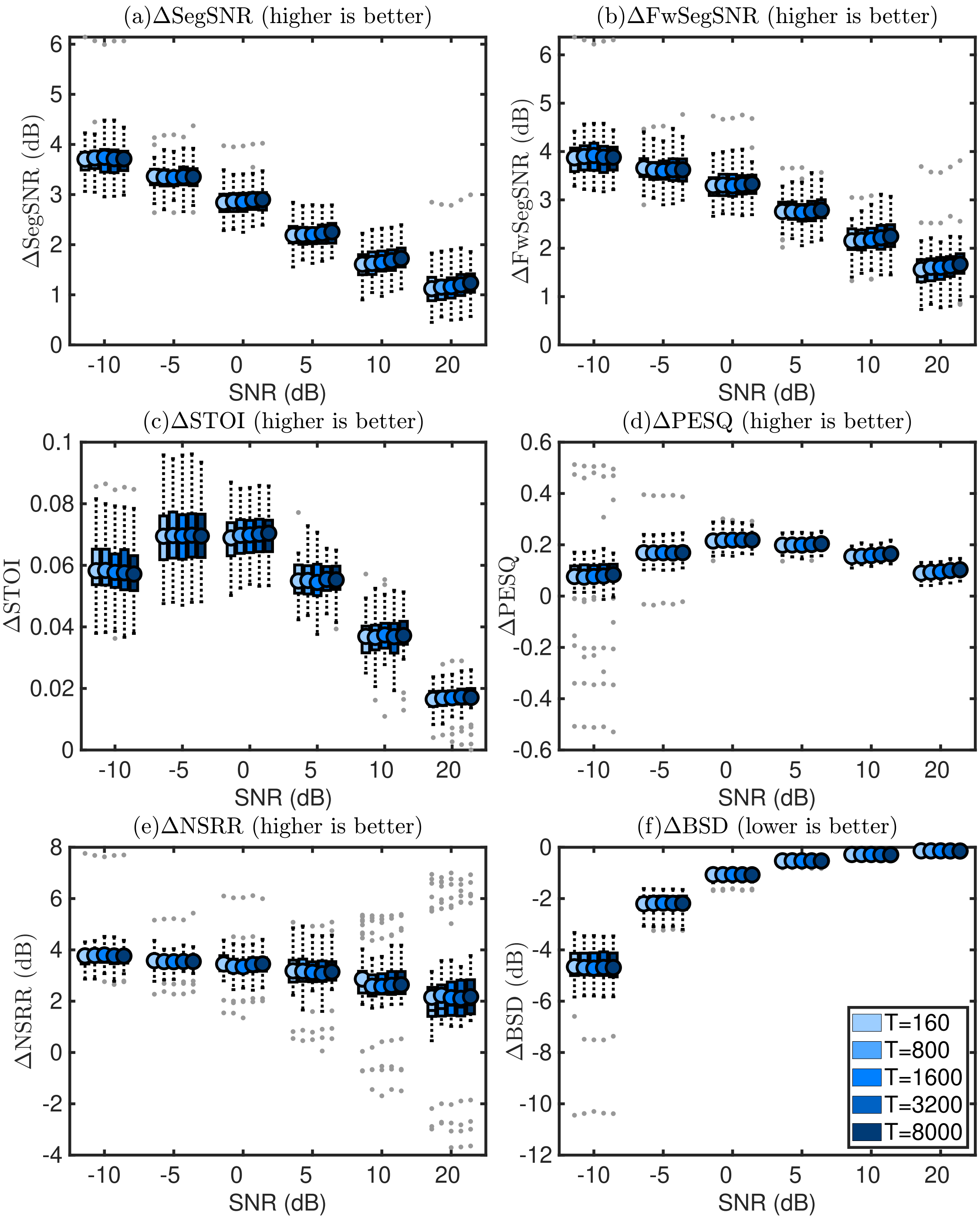


Figure 4.2: Speech enhancement for white noise in Lecture Room 2 using different frame length T for the computation of the space-time covariance matrix.

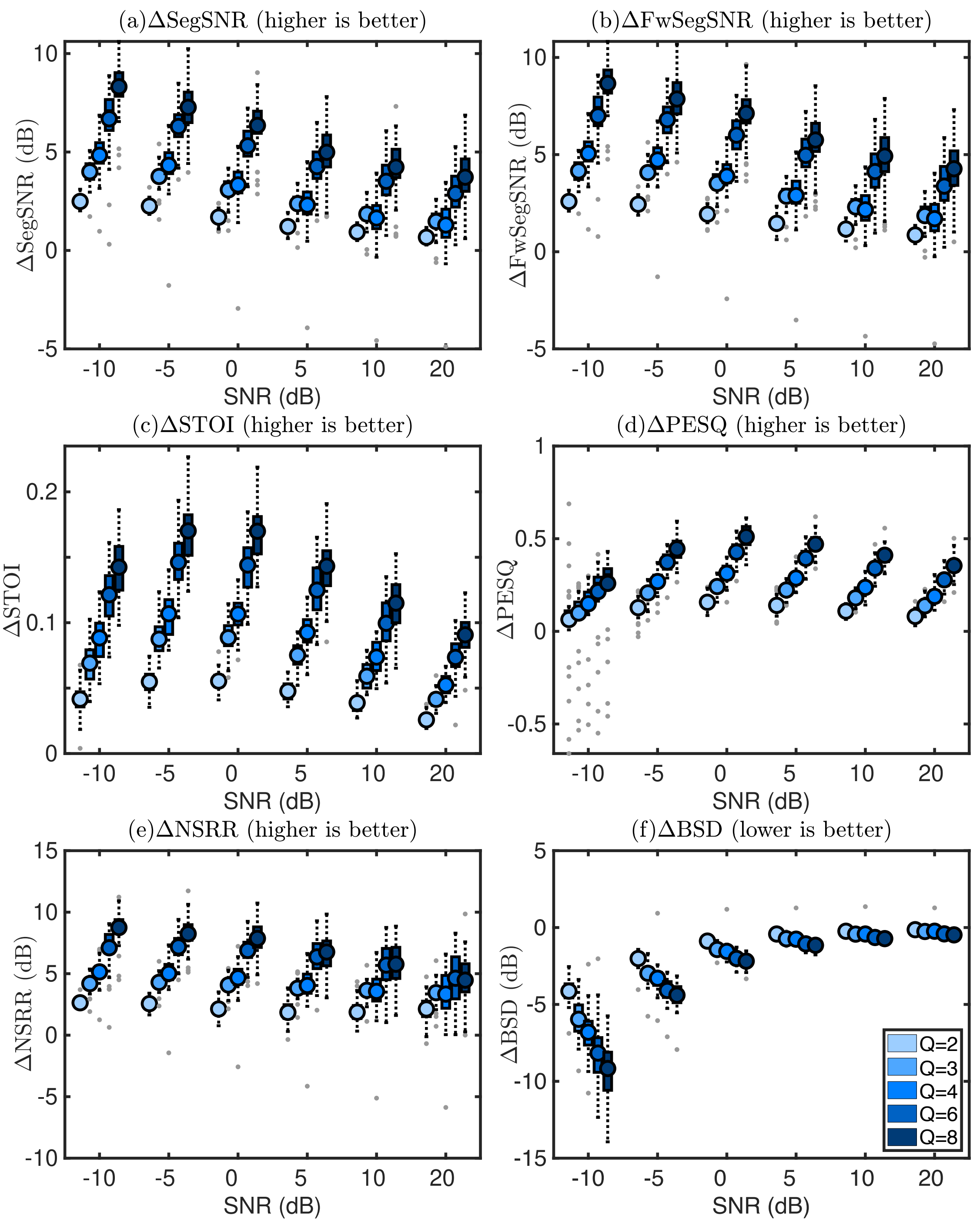


Figure 4.3: Speech enhancement results for white noise in Lecture Room 2 using different numbers of microphones Q.

4.7.2 Comparison of Speech Enhancement Algorithms

A. Noise Reduction of Anechoic Speech in Noise

The illustrative example based on a clean speech corrupted by 0 dB babble noise in an anechoic environment is shown in Fig. 4.4. Overall, the energy of the babble noise has been significantly reduced after the PEVD-based enhancement, as reflected by an improvement in ΔSegSNR and ΔFwSegSNR, respectively in Table 4.1. Although COLSUB [126], which uses a GEVD, improved in both ΔSegSNR and ΔFwSegSNR, the structures of the babble noise and speech signals are lost for example at 3 s and processing artefacts are observed in the listening examples [161]. These artefacts may have led to lower ΔPESQ and ΔSTOI for COLSUB compared to PEVD.

For this example, Table 4.1 shows that OMWF performed best in all measures as expected because it uses prior knowledge of the clean speech signal. Despite being a completely blind approach without using any noise estimator, PEVD is the best performing algorithm after OMWF in terms of ΔFwSegSNR, ΔSTOI and ΔPESQ. This example demonstrates that the PEVD approach can perform comparably to an oracle algorithm. Enhancement using MCSUB or WPD offers some improvement in some measures while MWF, GWPE or ISCLP worsens the scores across all measures. Note that in practice when noise and/or RTF estimators are used, the performance of MWF degrades significantly from OMWF.

The $\mathcal{R}_{\mathbf{xx}}(z)$ for the above example used in PEVD is presented in Fig. 4.5, with the $(p,q)^{\text{th}}$ subplot representing the z-transform of $r_{p,q}(\tau)$, p and q correspond, respectively, to the row and column indices of Fig. 4.5. An EVD, which is applied to the instantaneous covariance matrix corresponding to the coefficient of z^0, can only decorrelate the signals at a single time lag [125, 127]. This is inadequate for this example because the cross-correlations on the off-diagonals have peaks at other time lags. Instead, the PEVD can impose decorrelation over a range of time lags as shown in Fig. 4.6, where every off-diagonal element has a magnitude less than 0.58% of the trace-norm of $\mathbf{R}_{\mathbf{xx}}(0)$. Because of trimming, the polynomial order of $\mathbf{\Lambda}(z)$ in Fig. 4.6 is much smaller than that of $\mathcal{R}_{\mathbf{xx}}(z)$

Table 4.1: Noise reduction performance of an anechoic speech in 0 dB diffuse babble noise example, with measured scores (SegSNR=-8.93 dB, FwSegSNR=-6.86 dB, STOI=0.674 and PESQ=1.63).

Algorithm	ΔSegSNR	ΔFwSegSNR	ΔSTOI	ΔPESQ
MWF	-2.37 dB	-2.91 dB	-0.122	-0.22
OMWF	**5.37 dB**	**4.31 dB**	**0.104**	**0.41**
COLSUB	5.23 dB	3.92 dB	0.024	0.14
MCSUB	1.96 dB	-0.60 dB	0.008	0.17
PEVD	4.36 dB	4.22 dB	0.097	0.34
GWPE	-0.04 dB	-0.23 dB	-0.009	-0.01
WPD	0.60 dB	-0.03 dB	-0.004	0.02
ISCLP	-7.65 dB	-7.99 dB	-0.167	-0.33

in Fig. 4.5.

The results of Monte-Carlo simulations involving 50 trials of anechoic speech with -10 dB to 20 dB babble noise are plotted in Fig. 4.7. OMWF is the best performing algorithm which can consistently provide improvement in all metrics over the entire range of SNRs. While COLSUB performs up to 2 dB better than OMWF in ΔSegSNR and ΔFwSegSNR at lower SNRs, the ΔSTOI and ΔPESQ are negative. After OMWF, PEVD is the best performing algorithm in ΔSTOI and ΔPESQ and one of the best in ΔSegSNR and ΔFwSegSNR. The noise reduction performances of MWF, MCSUB, GWPE, WPD and ISCLP are limited in all measures across all SNRs.

All algorithms were tested for an extensive range of noise types with results given in Fig. 4.8, for the example of restaurant noise. Noise reduction for COLSUB is achieved at the cost of reduced speech intelligibility and quality. The OMWF, which uses oracle knowledge of the clean speech, performs best across all measures and can simultaneously reduce noise and improve speech intelligibility and quality. PEVD comes close as second best, despite being completely blind, just like the other approaches. Most approaches do not improve and may even reduce the measures.

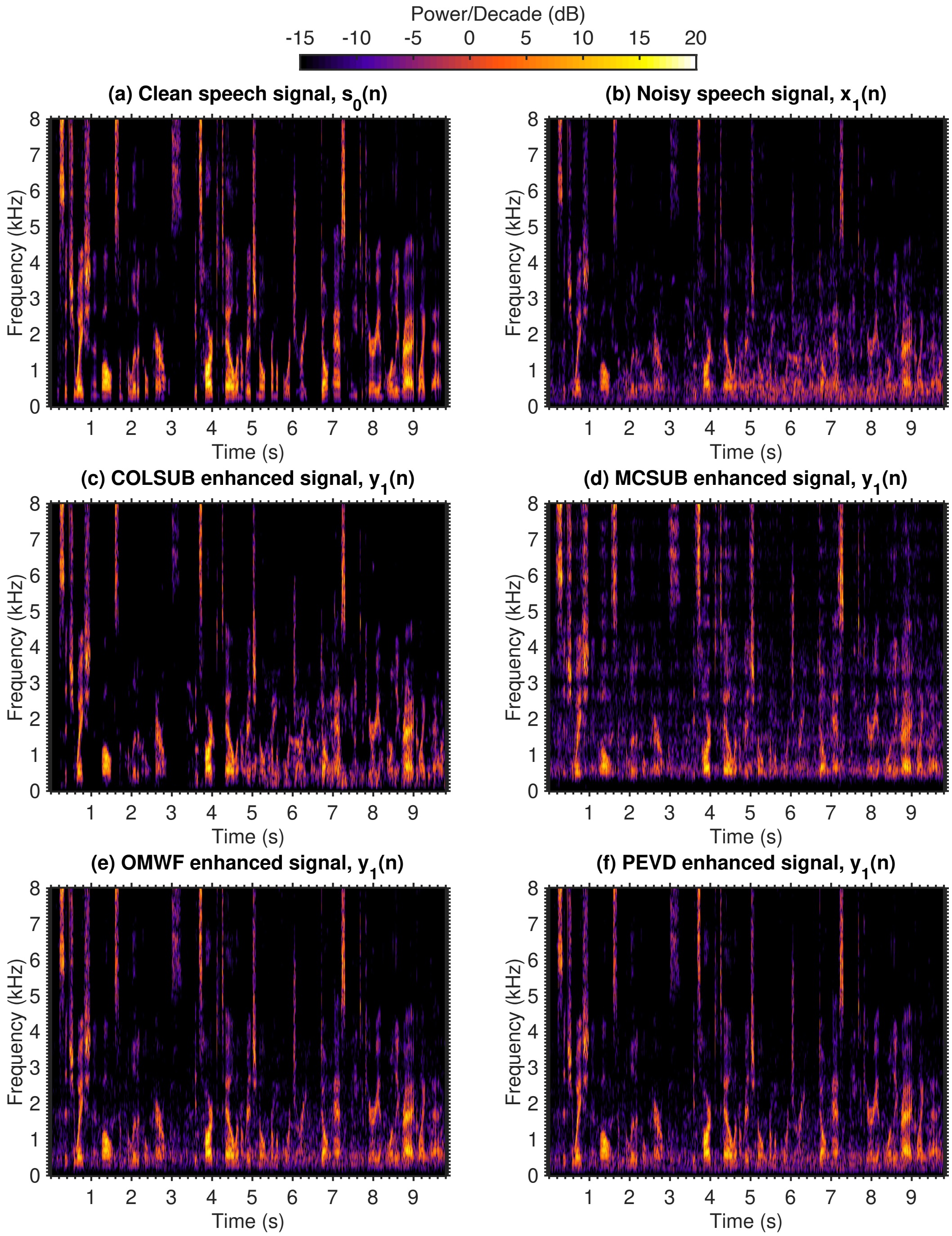


Figure 4.4: Normalized spectrograms of anechoic speech in 0 dB babble noise example and the processed signals.

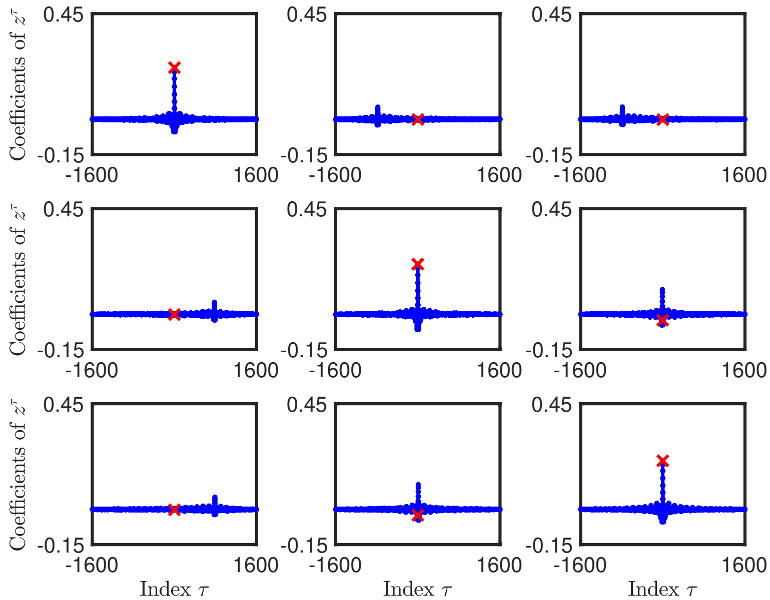


Figure 4.5: Space-time covariance, $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ of the noisy speech example in Fig. 4.4 with the instantaneous covariance matrix marked by red cross signs. The $(p, q)^{\text{th}}$ subplot is the z-transform of the element $r_{p,q}(\tau)$.

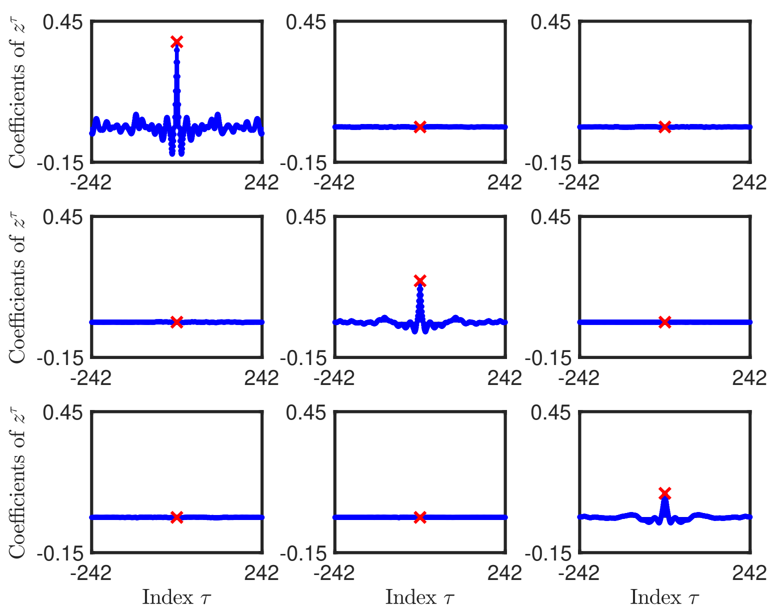


Figure 4.6: Eigenvalue polynomial matrix, $\boldsymbol{\Lambda}(z)$ of the space-time covariance matrix $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ in Fig. 4.5.

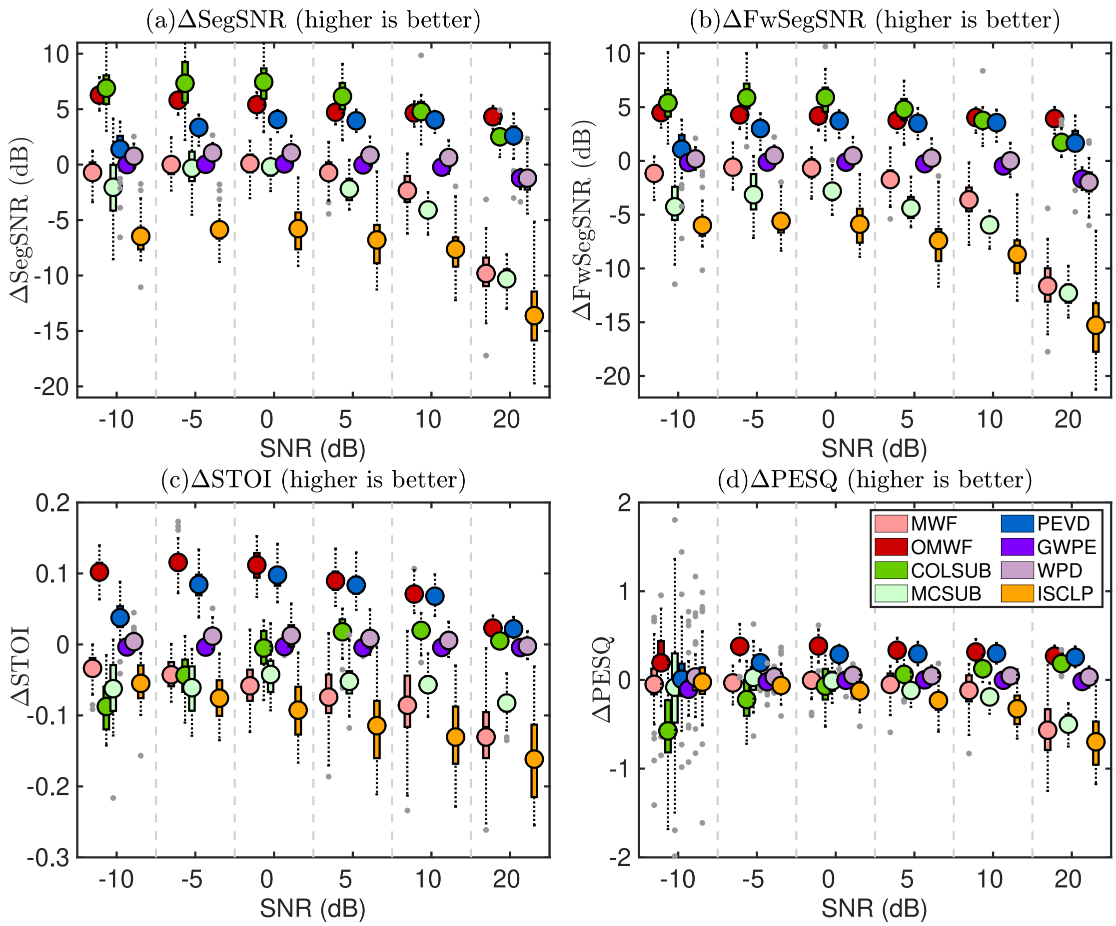


Figure 4.7: Noise reduction results for Noisex babble noise.

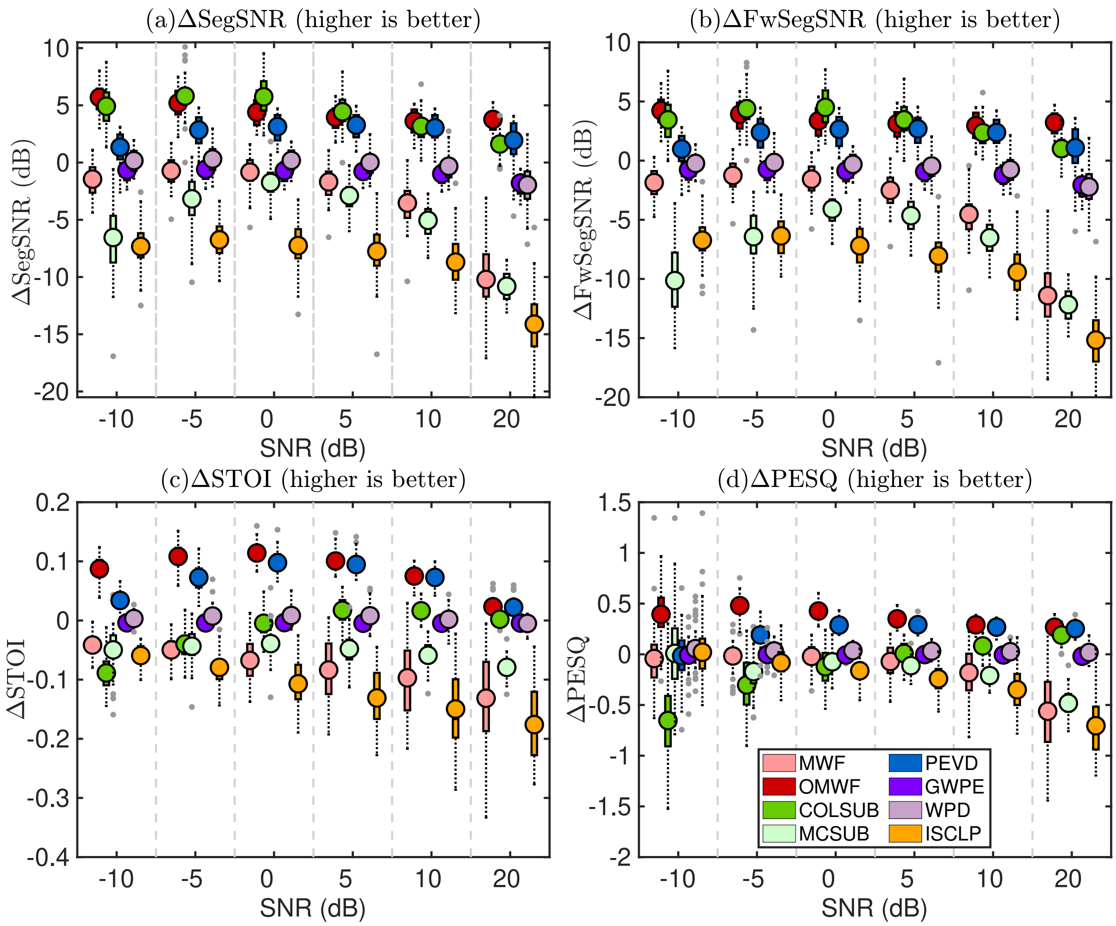


Figure 4.8: Noise reduction results for SoundFx restaurant noise.

B. Speech Dereverberation in the Absence of Noise

An illustrative result for a single reverberant speech example in Lecture Room 2 without background noise is shown in Fig. 4.9. The spectrograms show qualitatively that both GWPE and PEVD can suppress reverberation while retaining the overall speech structure. However, GWPE seems to have applied a more aggressive suppression which results in a cleaner spectrogram, as supported by the lowest ΔBSD in Table 4.2, indicating lower spectral distortions than PEVD. PEVD outperforms all algorithms in ΔNSRR and WPD, which uses the steering vector computed from the ground truth DoA, ranks second. Based on ΔSTOI and ΔPESQ, GWPE performed best, followed by WPD, ISCLP, PEVD and OMWF. The other algorithms do not provide significant improvements.

To understand the processing involved, the difference between the reverberant and processed signals are plotted in red along with the reverberant signal in blue. The PEVD improvement in Fig. 4.9(f) follows the temporal envelope changes of the reverberant signal more closely while the GWPE improvement has higher amplitudes for example at roughly 1 s and 4 s, which suggests more aggressive processing. Listening examples for GWPE indicate the removal of most of the early but not the direct-path and some late reverberant components, as also observed in [147]. Listening examples for PEVD, on the other hand, indicate that the direct-path and some early reflection components are retained in the enhanced signal in the first channel, as expected for reasons given in Section 4.5.2. The late reverberant components, which are absent in the enhanced signal, are observed in the second and third channels because of orthogonality [161].

Across all rooms, due to the removal of late reflections, PEVD ranks second in ΔNSRR after WPD as shown in Fig. 4.10. Moreover, GWPE processes the signals aggressively as discussed above and therefore leads to the best improvement in ΔBSD, ΔSTOI and ΔPESQ. The overall performance of WPD and ISCLP are comparable and is followed by PEVD. In this noise-free scenario, speech dereverberation using OMWF is equivalent to a Wiener deconvolution. Averaged across all rooms, ΔNSRR and ΔBSD for OMWF are worsened by -0.94 dB and 0.14 dB, while ΔSTOI and ΔPESQ are increased by 0.015 and

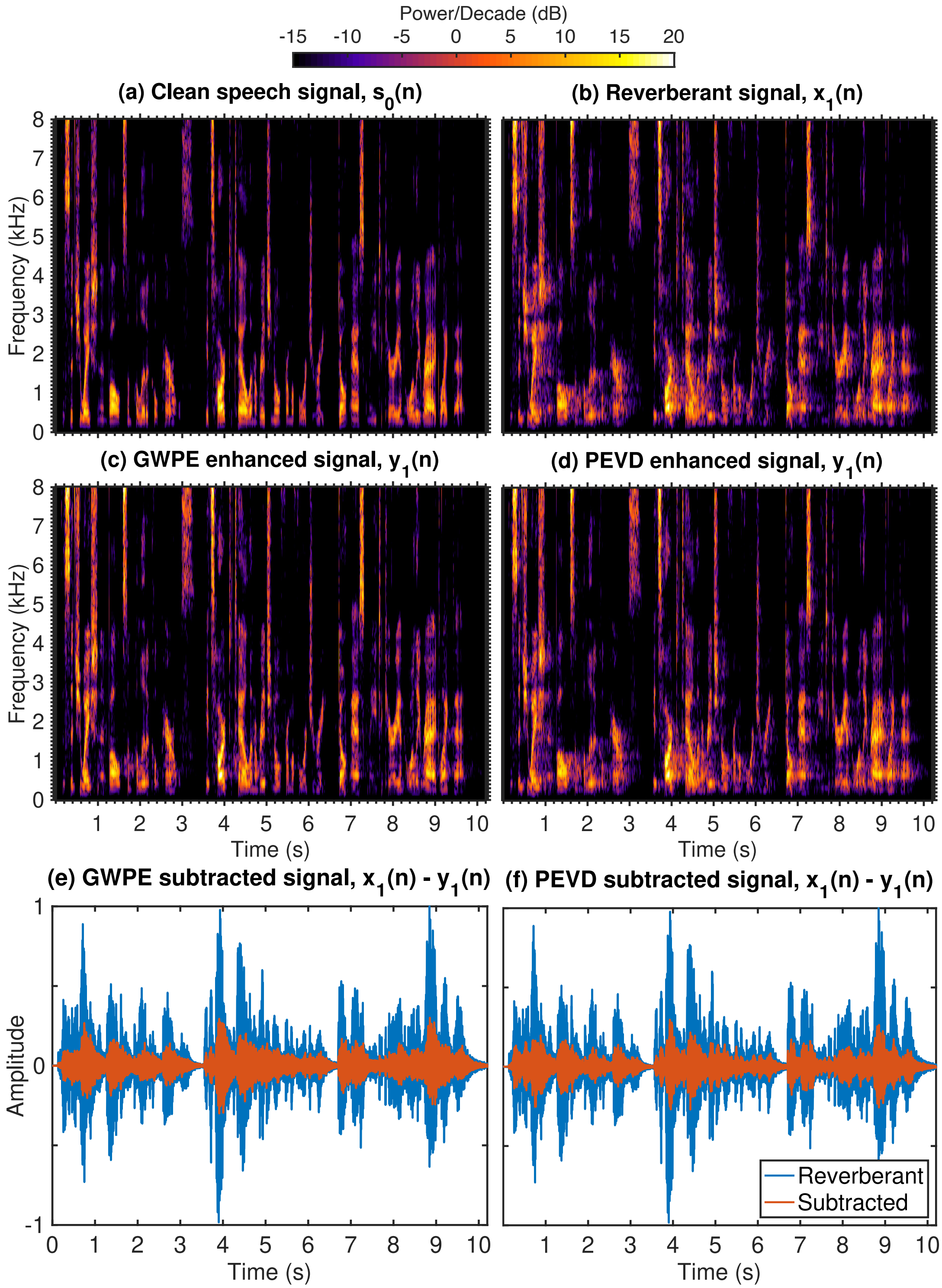


Figure 4.9: Normalized spectrograms of reverberant speech example and the improvement, $x_1(n) - y_1(n)$, is scaled and time aligned with $x_1(n)$.

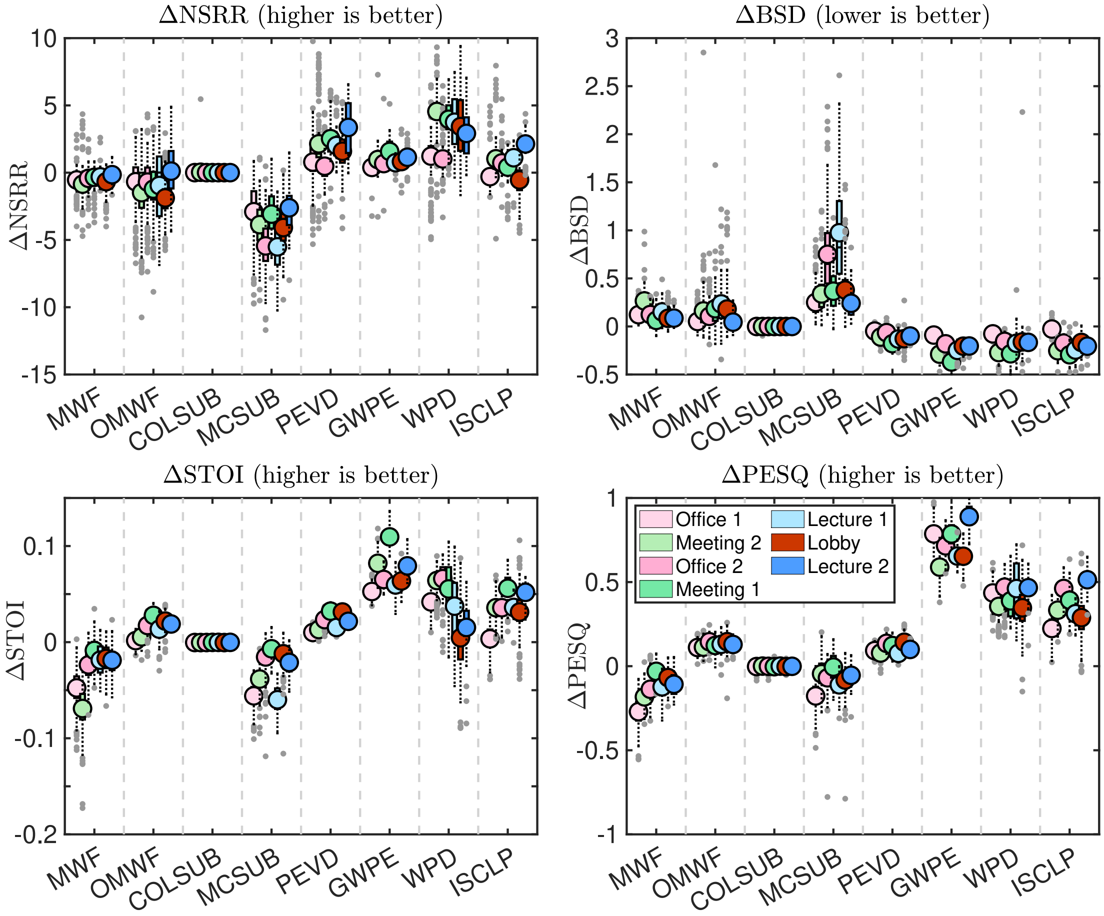


Figure 4.10: Speech dereverberation results in noiseless environments.

Table 4.2: Dereverberation performance of a noiseless, reverberant example in Lecture Room 2, with measured scores (NSRR=-6.03 dB, BSD=0.38 dB, STOI=0.810, PESQ=2.01).

Algorithm	ΔNSRR	ΔBSD	ΔSTOI	ΔPESQ
MWF	-1.06 dB	0.27 dB	-0.057	-0.26
OMWF	0.10 dB	0.04 dB	0.009	0.16
COLSUB	0.00 dB	0.00 dB	0.000	0.00
MCSUB	-3.20 dB	0.28 dB	-0.028	0.01
PEVD	**5.79 dB**	-0.12 dB	0.018	0.12
GWPE	0.68 dB	**-0.25 dB**	**0.091**	**0.70**
WPD	4.02 dB	-0.23 dB	0.071	0.40
ISCLP	0.29 dB	-0.22 dB	0.040	0.41

0.13 respectively. This is expected because the FIR filter of 80 taps, chosen following the published parameters in [119], is insufficient to invert with high accuracy the room impulse responses which are thousands of samples long. COLSUB also offers no improvement while MWF and MCSUB tend to worsen all metrics.

C. Speech Enhancement of Reverberant Speech in Noise

Results for reverberant speech corrupted by ACE babble noise in the strongly reverberant Lecture Room 2 are shown in Fig. 4.11. For SNR≤10 dB, a trade-off between noise reduction and speech intelligibility and quality, is observed for COLSUB which shows negative ΔSTOI and ΔPESQ. On the other hand, OMWF, which is designed to minimize speech distortion using knowledge of the clean speech, performs the best in ΔSTOI and ΔPESQ but not in ΔSegSNR and ΔFwSegSNR. This also reflects the fact that speech intelligibility may not necessarily be affected by noise levels, up to some limit, compared to speech. Using prior knowledge of DoA to compute the steering vector, WPD provides further improvement in all measures over GWPE. At 20 dB SNR, algorithms targeting reverberation such as GWPE and joint approaches like WPD and ISCLP, perform better than the noise reduction approaches. Fig. 4.11 also shows that PEVD performs best in ΔNSRR and ΔBSD, ranks second in ΔSegSNR, ΔFwSegSNR and ΔSTOI, and one of the best in ΔPESQ for all SNRs.

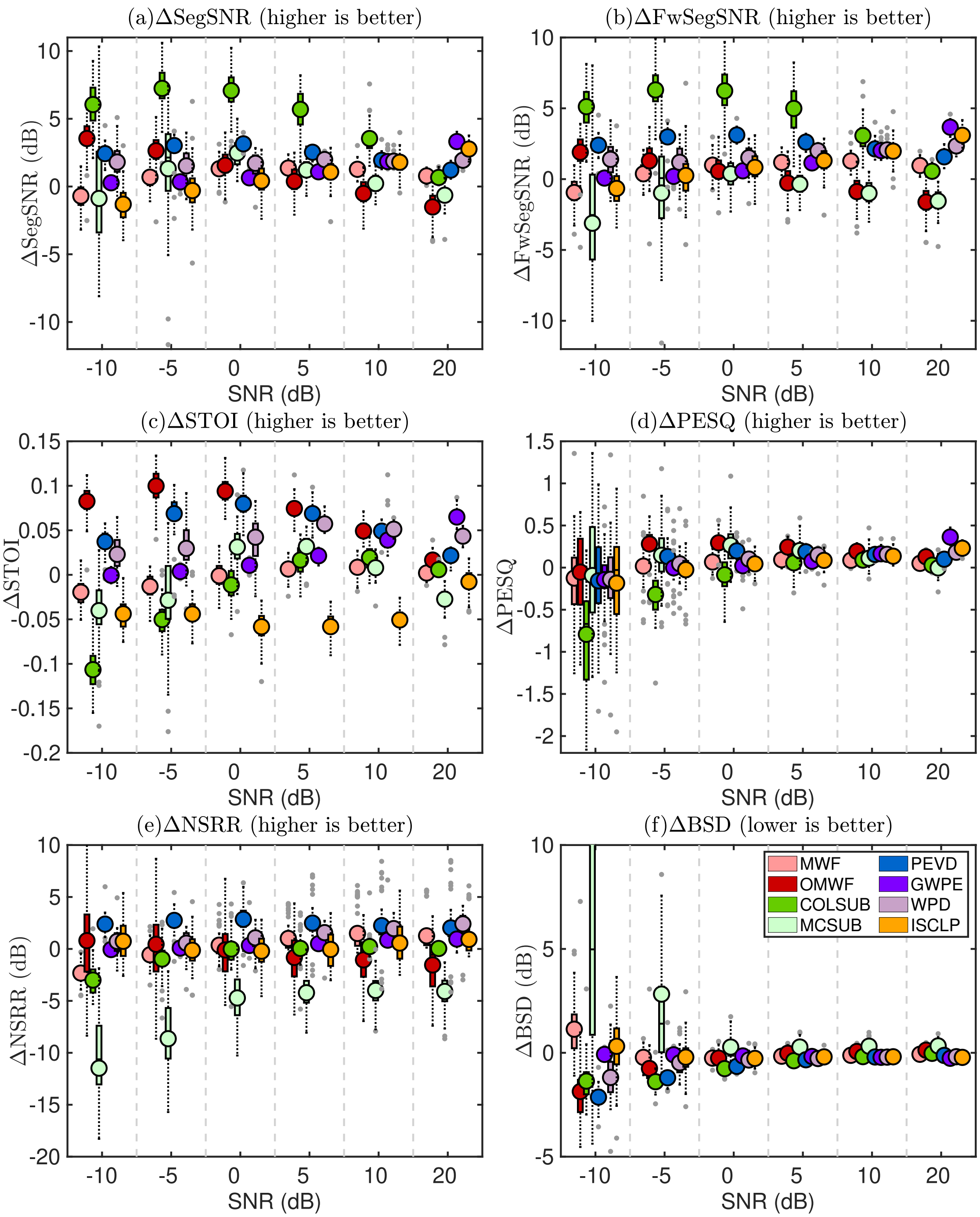


Figure 4.11: Speech enhancement results for babble noise in Lecture Room 2.

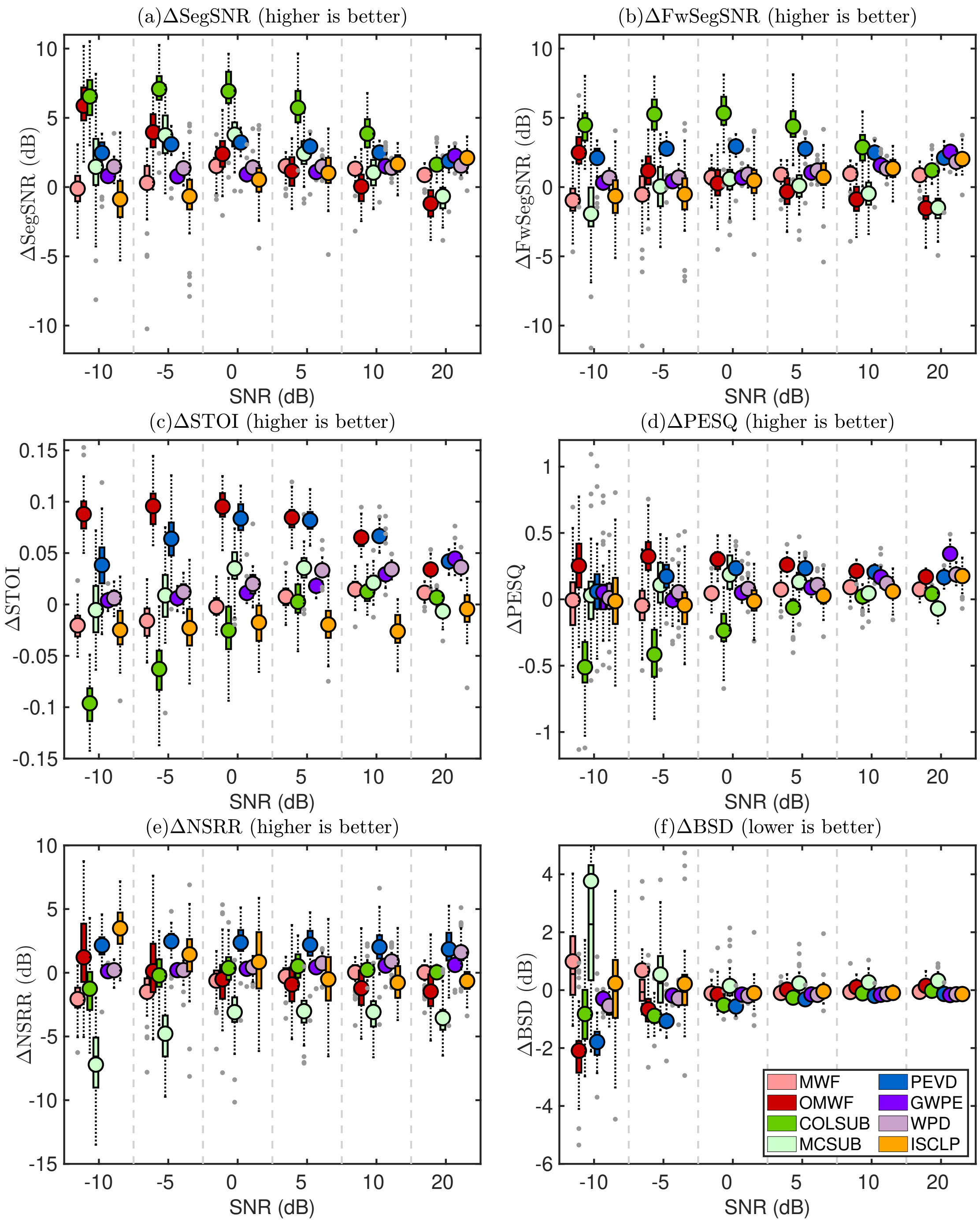


Figure 4.12: Speech enhancement results for fan noise in Office 1.

Results for the ACE fan noise in the scenario for Office 1 in Fig. 4.12 also show that noise reduction algorithms perform better at low SNRs, i.e. below 5 dB, while dereverberation algorithms perform better at high SNRs, for example at 20 dB. Across all noise and reverberation conditions, PEVD can consistently suppress both noise and reverberation and improve both speech intelligibility and quality. In addition, PEVD can offer further improvement by up to 5 dB in ΔFwSegSNR, 0.1 in ΔSTOI and -2 dB in ΔBSD over state-of-the-art joint approaches, WPD and ISCLP. Furthermore, listening examples in [161] provide supporting evidence that our PEVD-based approach does not introduce any processing artefacts into the enhanced signal.

Comprehensive testing over the complete ACE corpus has been performed with summary results given in Table 4.3, for the example of the most and least reverberant rooms in 0 dB babble noise. Results indicate that PEVD can consistently provide the best dereverberation and one of the best noise reduction performances. Only OMWF outperforms PEVD in ΔSTOI. However, OMWF requires oracle prior information, which is unavailable in practice. In contrast, the proposed PEVD approach is blind, using only the microphone signals. The source code and listening examples corresponding to the results are available [161].

Table 4.3: Enhancement performance evaluated on the ACE corpus for 0 dB babble noise, measured using (mean±standard deviation). ΔFwSegSNR, ΔNSRR and ΔBSD are measured in dB. The proposed PEVD approach consistently performs among the best.

Room (T_{60})	Office 1 (0.332 s)				Lecture Room 2 (1.22 s)			
Algorithm	ΔFwSegSNR	ΔSTOI	ΔNSRR	ΔBSD	ΔFwSegSNR	ΔSTOI	ΔNSRR	ΔBSD
MWF	-0.96 ± 1.30	-0.020 ± 0.015	-2.07 ± 1.30	0.99 ± 1.70	-0.95 ± 1.10	-0.020 ± 0.015	-2.32 ± 0.99	1.14 ± 1.30
OMWF	2.49 ± 1.90	**0.088 ± 0.022**	1.21 ± 3.30	**-2.09 ± 1.40**	1.90 ± 1.50	**0.083 ± 0.018**	0.80 ± 4.10	-1.86 ± 1.70
COLSUB	**4.49 ± 1.50**	-0.096 ± 0.024	-1.25 ± 2.20	-0.82 ± 1.00	**5.12 ± 1.60**	-0.107 ± 0.025	-3.00 ± 1.50	-1.36 ± 1.00
MCSUB	-1.93 ± 2.70	-0.005 ± 0.029	-7.21 ± 3.10	3.77 ± 6.50	-3.11 ± 6.30	-0.040 ± 0.038	-11.48 ± 6.10	17.80 ± 44.00
PEVD	2.11 ± 1.00	0.038 ± 0.030	2.15 ± 1.20	-1.79 ± 0.81	2.40 ± 0.60	0.037 ± 0.013	**2.39 ± 0.80**	**-2.13 ± 0.66**
GWPE	0.32 ± 0.18	0.004 ± 0.005	0.12 ± 0.19	-0.30 ± 0.17	0.08 ± 0.11	-0.001 ± 0.004	-0.06 ± 0.16	-0.08 ± 0.11
WPD	0.67 ± 0.41	0.006 ± 0.008	0.20 ± 0.50	-0.54 ± 0.36	1.40 ± 1.30	0.023 ± 0.019	0.73 ± 1.30	-1.19 ± 1.20
ISCLP	-0.65 ± 1.90	-0.025 ± 0.024	**3.49± 1.70**	0.24 ± 2.00	-0.66 ± 1.40	-0.044 ± 0.016	0.74 ± 2.00	0.31 ± 1.50

4.8 Chapter Summary and Conclusions

In this chapter, multi-channel broadband speech signals were modelled using polynomial matrices in order to capture the spatial, spectral as well as temporal correlations between microphones. It was shown that the proposed PEVD approach can decorrelate the microphone signals in space, time and frequency simultaneously. A novel speech enhancement algorithm based on the PEVD was proposed. The proposed algorithm achieves significant noise reduction and dereverberation using a weighted combination of the signal subspace. Comparative simulations under diverse acoustic conditions have indicated that the proposed method improves noise reduction metrics, speech intelligibility and quality scores, and dereverberation measures. Despite being a blind and unsupervised algorithm, the approach does not use any channel or noise estimation algorithms and does not introduce any processing artefacts into the enhanced signal as observed in the examples at [161].

Chapter 5

Signal Compaction Using PEVD for Spherical Arrays with Applications

5.1 Introduction

This chapter focuses on the compact representation of speech signals acquired by spherical microphone arrays. The chapter first provides an exposition of the problem, followed by a review of the processing and signal representation afforded by spherical harmonic transform (SHT) and PEVD for spherical microphone arrays. The proposed approach to combine SHT and PEVD is next presented and subsequently used for speech enhancement and source separation. In the proposed framework for signal representation, the diagonality factor improves by up to 7 dB over the microphone signal representation with a significantly lower computation cost. Moreover, when applying this framework to speech enhancement and source separation, the proposed method improves the STOI and source-to-distortion ratio (SDR) metrics by up to 0.2 and 20 dB, respectively. This chapter is closely based on the article [42].

5.2 Overview and Literature Review

In multi-channel signal processing involving spatially separated sensors, the received signals often contain a high level of data redundancy. Many compact signal representation techniques such as subband coding have been developed to improve storage and processing efficiency [18]. The processing of these compacted signals offers computational advantages and has been widely used for data compression [90], noise reduction [168], and speech and image coding [8, 169]. This chapter focuses on the compact representation of speech signals measured by spherical microphone arrays.

A data-adaptive signal representation using an infinite-order principal component filter bank (PCFB) has been shown to be optimal in terms of the mean-square error in signal reconstruction and the coding gain in data compression [170]. A PCFB generally requires a frequency-dependent switching of channels [90], whereby the switching function is not analytic and therefore cannot be well approximated with a finite degree PCFB [45, 64]. If the order of the PCFB is constrained to zero-th order, the KLT, i.e. an EVD, gives the optimal solution [170].

An alternative approach uses polynomial matrices, which can simultaneously capture the correlations in space, time and frequency and is, therefore, appropriate for modelling multi-channel broadband signals. The processing of polynomial matrices has motivated the development of polynomial matrix eigenvalue decomposition (PEVD) algorithms in the z-domain based on the SBR2 [25, 30] and SMD [28, 29], and those in the DFT-domain [60, 171]. Unlike the KLT, a PEVD can spatially decorrelate signals for a suitably chosen range of time lags and is more suitable for broadband signals [28]. It has been found useful in multi-channel broadband signal processing applications such as blind source separation [33], source identification [34], localization [67], adaptive beamforming [31] and channel coding [35].

The use of the PEVD for speech enhancement has been proposed in [39] for arbitrary arrays. It has been shown to improve noise reduction metrics, dereverberation measures, speech intelligibility and speech quality scores in diverse acoustic environments without introducing noticeable artifacts. This is due to the fact that PEVD algorithms [25, 28–30] — while using z-domain notation — operate in the time domain, and therefore maintain spectral coherence of the signals being processed. However, the computational complexity of such PEVD algorithms scales at best cubically with the number of signals used for processing [172].

Spherical array processing [131, 173] has gained significant attention owing to its applicability to hearing aids, sound field decomposition and reproduction for personal sound zones, augmented and virtual reality [174–176], and robot audition [112]. The microphone array in these applications is commonly modelled as a spherical array along a rigid sphere. Spherical microphone arrays enable a compact spatial representation of the sound field in the spherical harmonic (SH) domain using eigenbeam signals. Because the eigenbeam signals are computed using only the array geometry and are decoupled from the sensor arrangements, the beam pattern can be designed to be rotationally invariant, non-data-adaptive, and independent of the number of microphones, making the processing scalable [177]. This is why beamforming using spherical arrays [177–180] has been proposed for speech enhancement [181, 182], localization and tracking [183–185].

In anechoic environments, the 3D sound field can be perfectly represented by an infinite number of spatially orthogonal SHs (of infinite order) [131]. In practice, a finite number of microphones limiting the SH order leads to an approximate representation such that the 3D sound field cannot be accurately reconstructed with high spatial resolution. Furthermore, reverberation arising from multi-paths in enclosed spaces causes target source components in eigenbeam signals from different directions to arrive at other times [24]. Consequently, the SHT is not able to decorrelate a reverberant sound field for non-zero time delays [174].

In this chapter, the representation of microphone signals on a spherical array using SH is first compared to a PEVD representation. This will demonstrate the inability of SH to capture temporal correlations, and the challenge of scaling with the number of signals used for PEVD processing. We then propose to combine both approaches by applying a PEVD to a subset of eigenbeam signals generated from preprocessing by a SHT. This capitalizes on the compact and scalable representation offered by SH, and also exploits the ability of the PEVD to strongly decorrelate signals, i.e., to generate outputs that are mutually decorrelated at all lags. Preliminary studies have demonstrated that good speech enhancement [40] and source separation performance [41] can be achieved through a compression via eigenbeam selection. By analysing the synergies of SHT and PEVD, we can explain the benefits that this has brought to speech enhancement and source separation.

The contributions of this chapter are (i) the comparison of different representations of multi-channel signals from a spherical microphone array signals using the SHT, KLT, and PEVD, (ii) the proposed framework that relates the microphone signals, eigenbeams and beamformer outputs through a compression matrix, (iii) the use of a PEVD to combine eigenbeams and beamformer outputs under the proposed framework, and (iv) the application of the proposed approach for speech enhancement and source separation.

5.3 Problem Formulation and Overview

5.3.1 Signal Model

The noisy and reverberant speech signal arriving at the q-th microphone on the spherical array at sample index n, is

$$x_q(n) = \sum_{p=1}^{P} h_{p,q}(n) * s_p(n) + v_q(n), \quad q = 1, \ldots, Q \, , \tag{5.1}$$

where $h_{p,q}(n)$ is the time-invariant RIR from the p-th source to the q-th microphone, $s_p(n)$ is the p-th source signal, $v_q(n)$ is the additive noise assumed uncorrelated with the speech component, and $*$ denotes the linear convolution operator. The model in (5.1) accounts for P sources contributing to Q microphone signals. The data vector of the microphone signals is $\mathbf{x}(n) = \Big[x_1(n), \ldots, x_Q(n)\Big]^{\mathrm{T}}$ with $\mathbf{v}(n)$ similarly defined, whereby $[\cdot]^{\mathrm{T}}$ denotes the transpose operator.

To express the microphone signals in terms of the array geometry explicitly, (5.1) can be written as $x(n, \mathbf{r}_q)$, where $\mathbf{r}_q = (r, \theta_q, \phi_q)$ is expressed in spherical polar coordinates, r is the radius of the sphere, θ_q and ϕ_q respectively, are the elevation and azimuth angles of the q-th microphone from the array centre measured downwards from the z-axis and from the x-axis towards the y-axis. Accordingly, the data vector can also be written as $\mathbf{x}(n, \mathbf{r}) = \Big[x(n, \mathbf{r}_1), \ldots, x(n, \mathbf{r}_Q)\Big]^{\mathrm{T}}$.

5.3.2 Challenge

The number of sources, P, is generally much smaller than the number of microphones, Q. Therefore, this work aims to find a compact signal representation of the microphone signals $\mathbf{x}(n)$ by taking into account the spatial, spectral and temporal information simultaneously using a transformation $\mathbf{G}(n) \in \mathbb{R}^{Q\times Q}$ such that $y_q(n) = \sum_{p=1}^{Q} g_{p,q}(n) * x_p(n)$, where $\mathbf{y}(n) = [y_1(n), \ldots, y_Q(n)]^{\mathrm{T}}$ is maximally compact and $g_{p,q}(n)$ is the $(p,q)^{\text{th}}$ element in $\mathbf{G}(n)$. This compact representation can permit a reduction in dimensionality, which together with a potential diagonalization of the signal covariance matrix, can facilitate more efficient processing for applications such as signal enhancement or the separation of sources.

5.4 Multi-channel Array Processing

5.4.1 Spherical Harmonic Transform (SHT)

The real-valued SHT of the spatially sampled sound field is approximated by [179]

$$\chi_\ell^{(m)}(n) = \sum_{q=1}^{Q} \alpha_q x(n, \mathbf{r}_q) R_\ell^{(m)}(\mathbf{r}_q) \ , \tag{5.2}$$

where α_q is the quadrature weight for the q-th microphone and $\chi_\ell^{(m)}(n)$ is the ℓ-th order, m-th degree time-domain eigenbeam associated with the real-valued SH basis function, $R_\ell^{(m)}(\mathbf{r}_q)$, defined as

$$R_\ell^{(m)}(\mathbf{r}) = \begin{cases} \sqrt{2}(-1)^m \Im\{Y_\ell^{|(m)|}(\mathbf{r})\} & , \ m < 0 \\ Y_\ell^0(\mathbf{r}) & , \ m = 0 \ , \\ \sqrt{2}(-1)^m \Re\{Y_\ell^{(m)}(\mathbf{r})\} & , \ m > 0 \end{cases} \tag{5.3}$$

and

$$Y_\ell^{(m)}(\mathbf{r}) = \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^{(m)}(\cos\theta) e^{\mathrm{j}m\phi} \ , \tag{5.4}$$

where $P_\ell^{(m)}(\cdot)$ is the associated Legendre function, $\mathrm{j} = \sqrt{-1}$, $|\cdot|$ denotes modulus, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ extract the real and imaginary parts of a complex number. Because of the completeness and orthogonality properties of SH, any function on the sphere can be expressed as a weighted combination of SHs using [131]

$$x(n, \mathbf{r}_q) \approx \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \chi_\ell^{(m)}(n) R_\ell^{(m)}(\mathbf{r}_q) \ . \tag{5.5}$$

Alias-free spatial reconstruction of the sound field, and therefore equality in (5.5), can be achieved if $Q \geq (L+1)^2$, where L is the maximum SH order of the sound field. Dropping the sample index n for brevity, the vector of eigenbeams is $\boldsymbol{\chi} = \left[\chi_0^{(0)}, \chi_1^{(-1)}, \chi_1^{(0)}, \ldots, \chi_L^{(L)}\right]^{\mathrm{T}}$, with elements arranged in ascending SH order and degree. More compactly in matrix-vector form, $\boldsymbol{\chi} = \underline{\boldsymbol{R}}^{\mathrm{T}} \mathrm{diag}(\boldsymbol{\alpha}) \mathbf{x}(n, \mathbf{r})$, where

$$\underline{\boldsymbol{R}} = \begin{bmatrix} R_0^{(0)}(\mathbf{r}_1), & \ldots, & R_L^{(L)}(\mathbf{r}_1) \\ \vdots & \ddots & \vdots \\ R_0^{(0)}(\mathbf{r}_Q) & \ldots & R_L^{(L)}(\mathbf{r}_Q) \end{bmatrix} \in \mathbb{R}^{Q \times (L+1)^2} \ , \tag{5.6}$$

and $\mathrm{diag}(\boldsymbol{\alpha})$ creates a diagonal matrix from the vector of microphone weights, $\boldsymbol{\alpha} = \left[\alpha_1, \ldots, \alpha_Q\right]^{\mathrm{T}} \in \mathbb{R}^{Q \times Q}$.

The generation of $\mathscr{L} = (L+1)^2$ eigenbeams using all microphone signals in (5.2) is also called an eigenbeamformer [177, 186]. Modal beamforming, or the judicious linear combination of the eigenbeams (see Section 5.6 for elaboration), produces a beam pattern directed at a desired source direction using

$$\psi(n) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} w_\ell^{(m)} \chi_\ell^{(m)}(n) \ , \tag{5.7}$$

where $w_\ell^{(m)}$ is the beamformer weight associated with $\chi_\ell^{(m)}(n)$. In vectorial form, $\psi(n) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\chi}$, where $\mathbf{w} = \left[w_0^{(0)}, w_1^{(-1)}, w_1^{(0)}, \ldots, w_L^{(L)}\right]^{\mathrm{T}}$. With $\mathscr{P}$ different combinations of weights, the beamformer can generate $\mathscr{P}$ outputs

$$\boldsymbol{\psi} = \mathbf{W}^{\mathrm{T}}\boldsymbol{\chi} , \tag{5.8}$$

where $\boldsymbol{\psi} = \begin{bmatrix}\psi_1, \ldots, \psi_{\mathscr{P}}\end{bmatrix}^{\mathrm{T}}$, $\mathbf{W} = \begin{bmatrix}\mathbf{w}_1, \ldots, \mathbf{w}_{\mathscr{P}}\end{bmatrix} \in \mathbb{R}^{\mathscr{L}\times\mathscr{P}}$ and $\mathbf{w}_p$ contains the set of weights associated with the p-th output ψ_p, $p = 1, \ldots, \mathscr{P}$. Ideally, the selected beams should mainly contain the target source signal while minimizing any interfering signals and noise.

5.4.2 Polynomial Matrix Eigenvalue Decomposition

The PEVD approach considers the space-time covariance matrix [30], parameterized by a time lag τ, which is defined as

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n-\tau)\} , \tag{5.9}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator over n. Each element, $r_{p,q}(\tau)$, is computed using the cross-correlation sequence between the p-th and q-th microphone signals. Hence, $\mathbf{R}_{\mathbf{xx}}(\tau)$ contains auto- and cross-correlations on its diagonals and off-diagonals, respectively.

Classical subspace-based approaches for narrowband signals typically consider a special case of (5.9) by evaluating $\mathbf{R}_{\mathbf{xx}}(\tau)$ only at $\tau = 0$. The received signals are then decorrelated using an EVD. The resulting instantaneous spatial covariance matrix, $\mathbf{R}_{\mathbf{xx}}(0)$, does not fully capture the second-order statistics of the sensor signals, and has been shown to be not fully adequate for broadband signals such as speech [39, 67], which naturally exhibit temporal correlations especially in reverberant environments. Consequently, the proposed PEVD approach considers the decorrelation of speech signals over a range of discrete time lags. Accordingly, the concatenation of the covariance matrices in (5.9) for all values of $\tau \in \{-N, \ldots, N\}$ can be represented by a 3D-tensor of dimension $Q \times Q \times (2N+1)$.

Speech signals are typically processed using the short-time Fourier transform. However, this approach divides the broadband into multiple narrowband signals, ignoring the spectral correlation that exists between different DFT bins, and neglecting phase coherence across bands [14, 187]. As an alternative representation, the z-transform of (5.9),

$$\mathcal{R}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau) z^{-\tau} \; , \tag{5.10}$$

or $\mathbf{R}(\tau) \circ\!\!-\!\!\bullet\, \mathcal{R}(z)$, is a para-Hermitian polynomial matrix satisfying $\mathcal{R}_{\mathbf{xx}}(z) = \mathcal{R}^{\mathrm{P}}_{\mathbf{xx}}(z) = \mathcal{R}^{\mathrm{H}}_{\mathbf{xx}}(1/z^*)$, where $[\cdot]^*$, $[\cdot]^{\mathrm{H}}$ and $[\cdot]^{\mathrm{P}}$ are the complex conjugate, Hermitian and para-Hermitian operators, respectively. The PEVD of (5.10) is [48, 49]

$$\mathcal{R}_{\mathbf{xx}}(z) = \mathcal{U}_{\mathbf{x}}(z) \mathbf{\Lambda}_{\mathbf{x}}(z) \mathcal{U}^{\mathrm{P}}_{\mathbf{x}}(z) \; , \tag{5.11}$$

where the columns of $\mathcal{U}_{\mathbf{x}}(z)$ are the eigenvectors and the elements on the diagonal matrix $\mathbf{\Lambda}_{\mathbf{x}}(z)$ are the eigenvalues. To compute (5.11), iterative approaches based on the SBR2 [25, 30] and the SMD [28, 29] families of algorithms have been proposed, which are proven to converge, and encourage spectral majorization such that the eigenvalues in $\mathbf{\Lambda}_{\mathbf{x}}(z)$ are strictly ordered on the unit circle. The decomposition in (5.11) for an analytic $\mathcal{R}_{\mathbf{xx}}(z)$ gives analytic factors $\mathcal{U}_{\mathbf{x}}(z)$ and $\mathbf{\Lambda}_{\mathbf{x}}(z)$, which in most cases are infinite Laurent series [48]. Analyticity implies that an arbitrarily close approximation by Laurent polynomials in (5.11) is possible with a sufficiently large order of factors [47].

At each iteration, the SBR2 PEVD algorithm [30] first searches for the off-diagonal element with the largest magnitude. If its magnitude exceeds a predefined threshold, a delay polynomial matrix is applied to bring the element to the z^0 plane. A unitary matrix, which is designed to zero out the element, is applied to the entire polynomial matrix. A trimming procedure [30] is also used to keep the polynomial order compact. The algorithm terminates when the magnitudes of all off-diagonal elements fall below a pre-set threshold, or when a user-defined maximum number of iterations is reached.

5.4.3 Comparison of Signal Representations

For a spherical array geometry, the two designs for $\mathbf{G}(n)$ defined in Section 5.3.2 include the SHT and the PEVD. The SHT provides a compact spatial representation of the 3D sound field sampled by microphones on a spherical array while the PEVD can compactly

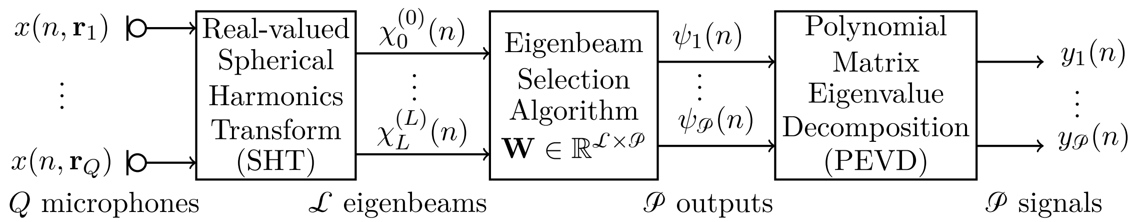


Figure 5.1: Block diagram of the proposed method which uses combinations of eigenbeams for PEVD processing.

represent space-time information in the signal eigenspace.

A. Analysis of Signal Representations

In noisy, reverberant environments, each microphone signal on a spherical array comprises the combination of many delayed and attenuated versions of the speech signal due to multi-path propagation. Consequently, microphone signals are highly correlated in space and time, as evident from a non-diagonalized space-time covariance matrix. The PEVD can diagonalize the space-time covariance matrix. Strongly decorrelated signals can be obtained using the polynomial eigenvectors as filters. Selecting only the subspaces corresponding to polynomial eigenvalues of significant magnitude leads to a signal representation based on PEVD that uses fewer signals than the number of microphone signals. When there are P independent, spatially separate sources, the ideal PEVD is expected to generate P out of Q non-zero outputs; while these outputs very likely do not match the source signals, a compaction of the data is still achieved.

For a spherical array, the SHT decomposes a 3D sound field using spatially orthogonal basis functions to generate the eigenbeam signals. Due to multi-paths, sound field components in different eigenbeams exhibit temporal correlations. Consequently, an EVD or KLT, which is typically applied to the spatial covariance matrix $\mathbf{R}(0)$, only removes instantaneous correlations i.e., at time lag $\tau = 0$, and is insufficient for processing (5.9). Instead, the PEVD can be used to achieve better compression and complete diagonalization across a range of time lags, as will be illustrated in Section 5.7.3.

B. Measures of Signal Representation Quality

I. Coding Gain. An instrumental measure of compactness is the coding gain γ, which is the ratio between the arithmetic and geometric mean of the variances of the elements in $\mathbf{y}(n)$ computed using [18, 28, 64, 90]

$$\gamma = \frac{\frac{1}{Q}\sum_{i=1}^{Q}\sigma_i^2}{\left(\prod_{i=1}^{Q}\sigma_i^2\right)^{1/Q}} = \frac{\frac{1}{Q}\mathrm{tr}\{\mathbf{R}_{\mathbf{yy}}(0)\}}{\det\{\mathbf{R}_{\mathbf{yy}}(0)\}^{\frac{1}{Q}}} , \tag{5.12}$$

where the last equality only holds for a diagonal matrix $\mathbf{R}_{\mathbf{yy}}(0)$, σ_i^2 is the variance of $y_i(n)$, i.e., the i-th element in $\mathbf{y}(n)$, $\mathrm{tr}\{\cdot\}$ and $\det\{\cdot\}$ compute the trace and determinant of a matrix. Ideally, $\mathbf{G}(n)$ in Section 5.3.2 should be a para-unitary or lossless filterbank, so the total powers of $\mathbf{x}(n)$ and $\mathbf{y}(n)$ remain the same. In this case, the numerator is invariant under para-unitary similarity transforms and maximizing γ is equivalent to minimizing the product of the variances in the denominator. Then, for $\mathbf{y}(n)$ to attain maximum coding gain, the two necessary and sufficient conditions which must be satisfied are [90]

1. strong decorrelation, such that $\mathbb{E}\{y_i(n)y_j(n-\tau)\} = 0 \ \forall\tau, i \neq j$, i.e., the elements of $\mathbf{y}(n)$ are decorrelated for all lags τ and
2. spectral majorization such that for all normalized angular frequencies Ω,

$$S_{y_{i-1}}(e^{\mathrm{j}\Omega}) \geq S_{y_i}(e^{\mathrm{j}\Omega}), \ i = 2, \ldots, Q , \tag{5.13}$$

   where $S_{y_i}(e^{\mathrm{j}\Omega})$ is the power spectral density of $y_i(n)$.

To show how strong decorrelation leads to the minimization of the product of variances, consider a 2-channel system with inputs $\mathbf{x} = [x_1(n), x_2(n)]^{\mathrm{T}}$ and outputs $\mathbf{y} = [y_1(n), y_2(n)]^{\mathrm{T}}$. If the inputs are correlated for some k, i.e., $\mathbb{E}\{x_1(n)x_2^*(n-k)\} \neq 0$, γ can be increased by para-unitary operations that preserve the total energy of the system. The para-unitary operation could consist of a delay z^{-k} and a unitary matrix computed using an EVD to transform $\mathbf{x}$ into an uncorrelated pair $\mathbf{y}$. Denoting $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{R}_{\mathbf{yy}}$, as the correlation matrices of $[x_1(n), x_2(n-k)]^{\mathrm{T}}$ and $[y_1(n), y_2(n)]^{\mathrm{T}}$, and $\sigma_{x_i}^2 = \mathbf{E}\{x_i(n)x_i^*(n)\}$ with

$\sigma^2_{y_i}$ similarly defined, the products of the input and output variances are related by

$$\sigma^2_{y_1}\sigma^2_{y_2} = \det\{\mathbf{R}_{\mathbf{yy}}\} = \det\{\mathbf{R}_{\mathbf{xx}}\} < \sigma^2_{x_1}\sigma^2_{x_2} \ , \tag{5.14}$$

since $\mathbf{R}_{\mathbf{yy}}$ is diagonal and $\det\{\mathbf{R}_{\mathbf{xx}}\} = \sigma^2_{x_1}\sigma^2_{x_2} - |r|^2$ for some non-zero cross-correlation r implies that $\det\{\mathbf{R}_{\mathbf{xx}}\} < \sigma^2_{x_1}\sigma^2_{x_2}$. The product of the variances of the outputs is smaller than the inputs, i.e., $\sigma^2_{y_1}\sigma^2_{y_2} < \sigma^2_{x_1}\sigma^2_{x_2}$, which results in $\gamma_{\mathbf{y}} > \gamma_{\mathbf{x}}$.

The second condition, spectral majorization, is also necessary to maximize the signal variances. Consider the 2-channel case when the total signal power in each channel is arranged in descending order, i.e., $\sigma^2_{x_1} \geq \sigma^2_{x_2}$ but not necessarily spectrally majorized for all Ω (5.13). This means $S_{x_1}(\mathrm{e}^{\mathrm{j}\Omega_0}) < S_{x_2}(\mathrm{e}^{\mathrm{j}\Omega_0})$ for some Ω_0. A frequency-dependent unitary matrix can be designed to reassign the PSD based on

$$\mathbf{T}(\mathrm{e}^{\mathrm{j}\Omega}) = \begin{cases} \mathbf{I} & , \quad S_{x_1}(\mathrm{e}^{\mathrm{j}\Omega}) \geq S_{x_2}(\mathrm{e}^{\mathrm{j}\Omega}) \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & , \quad S_{x_1}(\mathrm{e}^{\mathrm{j}\Omega}) < S_{x_2}(\mathrm{e}^{\mathrm{j}\Omega}) \end{cases} , \tag{5.15}$$

so $\mathbf{y}(\mathrm{e}^{\mathrm{j}\Omega}) = \mathbf{T}(\mathrm{e}^{\mathrm{j}\Omega})\mathbf{x}(\mathrm{e}^{\mathrm{j}\Omega})$ is spectrally majorized according to $S_{y_1}(\mathrm{e}^{\mathrm{j}\Omega}) \geq S_{y_2}(\mathrm{e}^{\mathrm{j}\Omega})\ \forall\Omega$. Redistributing the PSDs, $S_{y_1}(\mathrm{e}^{\mathrm{j}\Omega}) \geq S_{x_1}(\mathrm{e}^{\mathrm{j}\Omega})$ and $S_{y_2}(\mathrm{e}^{\mathrm{j}\Omega}) \leq S_{x_2}(\mathrm{e}^{\mathrm{j}\Omega})\ \forall\Omega$, gives

$$\sigma^2_{y_1} = \sigma^2_{x_1} + \delta, \quad \sigma^2_{y_2} = \sigma^2_{x_2} - \delta \ , \tag{5.16}$$

where $\delta > 0$ represents the power of components being reassigned. The product of the output variances is

$$\sigma^2_{y_1}\sigma^2_{y_2} = (\sigma^2_{x_1} + \delta)(\sigma^2_{x_2} - \delta) = \sigma^2_{x_1}\sigma^2_{x_2} + \delta(\sigma^2_{x_2} - \sigma^2_{x_1}) - \delta^2 \ , \tag{5.17}$$

which is always smaller than the input variances $\sigma^2_{x_1}\sigma^2_{x_2}$ since $(\sigma^2_{x_2} - \sigma^2_{x_1}) < 0$. Because $\mathbf{T}(\mathrm{e}^{\mathrm{j}\Omega})$ is (para-)unitary by construction, the sum of variances or signal power in the system remains, i.e., $\sigma^2_{x_1} + \sigma^2_{x_2} = \sigma^2_{y_1} + \sigma^2_{y_2}$. Further, the numerator of (5.12) is invariant to the processing, and the maximization of γ is equivalent to the minimization of the product of the variances. Therefore, $\sigma_{y_1}^2\sigma_{y_2}^2 < \sigma_{x_1}^2\sigma_{x_2}^2$ implies that $\gamma_{\mathbf{y}} > \gamma_{\mathbf{x}}$.

Strong decorrelation and spectral majorization are necessary for maximizing coding gain, becoming sufficient conditions for optimality when both are simultaneously satisfied [90]. PEVD algorithms satisfy strong decorrelation and spectral majorization, for which the latter has been proven to be enforced for SBR2 [188].

II. Reconstruction Error. The quality of the signal representation can be quantified by the reconstruction error ε, expressed as a percentage, using

$$\varepsilon = \frac{\sum_{q=1}^{Q}(\hat{x}(n,\mathbf{r}_q) - x(n,\mathbf{r}_q))^2}{\sum_{q=1}^{Q}(x(n,\mathbf{r}_q))^2} \times 100\% \ , \tag{5.18}$$

that is, the total mean square error between the received x and reconstructed microphone signals $\hat{x}$, normalized by the total energy of the received signals in Q microphones.

III. Diagonality Factor. Since with iterative PEVD algorithms, the diagonalization in (5.11) is only approximate, the total on- and off-diagonal energies E_{on} and E_{off} in the ideally diagonalized eigenvalue term $\mathbf{R}(\tau) \approx \mathbf{D}(\tau)$, where $\mathbf{D}(\tau) \circ\!\!-\!\!\bullet\ \mathbf{\Lambda}(z)$, is [30]

$$E_{\text{on}} \triangleq \sum_{\forall\tau}\sum_{i=1}^{Q}|r_{ii}(\tau)|^2, \quad E_{\text{off}} \triangleq \sum_{\forall\tau}\|\mathbf{R}(\tau)\|_{\text{F}}^2 - E_{\text{on}} \ , \tag{5.19}$$

where $r_{i,i}(\tau)$ is the i-th diagonal element of $\mathbf{R}(\tau)$, and $\|\cdot\|_{\text{F}}$ is the Frobenius norm. Using (5.19), the diagonality factor δ, which measures the level of diagonalization is

$$\delta = 5\log_{10}\left(\frac{E_{\text{on}}}{E_{\text{off}}}\right) \ , \tag{5.20}$$

expressed in dB. The unusual scaling by 5 accounts for the fact that the terms of $\mathbf{R}(\tau)$ in (5.20) are already of a quadratic nature [64]. The lower and upper bounds, $0 \leq E_{\text{on}}/E_{\text{off}} < \infty$, correspond to the case when only the off-diagonal elements are non-zero and the case when the polynomial matrix is fully diagonalized, respectively.

5.5 PEVD in the Spherical Harmonic Domain

5.5.1 Motivation for Proposed Approach

While PEVD addresses the optimum coding gain problem in multi-channel systems [64], it is data-dependent and, in practice, likely to face numerical difficulties for a large number of sensors or channels Q [189], such as the spherical microphone array here. The space-time covariance matrix in (5.10), on which the PEVD operates, contains $Q^2(2N+1)$ elements and give rise to a computational complexity that is at least $\mathcal{O}(Q^3N)$ due to matrix multiplications applied to every lag [172]. Hence, it will be computationally advantageous to reduce Q if the accuracy of the signal representation is not significantly affected.

In contrast to the PEVD, the SHT scales well w.r.t. the spatial dimension. Further, for a sufficiently large SH order L, instead of using Q microphones, a lower dimension of $\mathcal{L}$ eigenbeams can be used to represent the sound field. In terms of compaction, particularly for highly directional sources, the SHT outputs can be linearly combined via a beamforming that further compacts the source's power, and is therefore widely used for dimensionality reduction of spherical microphone array data [174, 190, 191]. This compaction is difficult to assess via the coding gain, since the SHT is not a unitary transform or otherwise norm-preserving. We can state, though, that the SHT is sub-optimal w.r.t. the coding gain in (5.12), as it only operates instantaneously. Consequently, it does not target decorrelation over a range of time lags and cannot remove correlations for other noise types and reverberation exhibiting temporal correlations.

With a view to combine the benefits of SHT and PEVD, we propose to utilize a preprocessor comprising the SHT together with a beamformer prior to applying a PEVD. This preprocessor performs a spatial decomposition, with the aim of compacting as much energy as possible into as few outputs as necessary. To achieve optimal compaction in terms of the coding gain, the PEVD is applied to the reduced number of SHT outputs, thereby achieving strong decorrelation at a lower computation cost. Our method is summarized in Fig. 5.1 and Algorithm 3.

5.5.2 Proposed Algorithm

The space-time covariance matrix of the modal beamformer outputs $\boldsymbol{\psi}(n)$ is

$$\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau) = \mathbb{E}\{\boldsymbol{\psi}(n)\boldsymbol{\psi}^{\mathrm{T}}(n-\tau)\} \ . \tag{5.21}$$

Using (5.2), (5.8) and (5.10), the z-transform of $\boldsymbol{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau)$ is

$$\begin{aligned} \boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z) &= \mathbf{W}^{\mathrm{T}}\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z)\mathbf{W} \\ &= \mathbf{W}^{\mathrm{T}}\underline{\boldsymbol{R}}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{\alpha})\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)\mathrm{diag}(\boldsymbol{\alpha})\underline{\boldsymbol{R}}\mathbf{W} \\ &= \mathbf{C}^{\mathrm{T}}\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)\mathbf{C} \ , \end{aligned} \tag{5.22}$$

where $\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z) \bullet\!\!-\!\!\circ \ \mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau)$ is the space-time covariance matrix of the eigenbeam signals. The compression matrix $\mathbf{C} = \mathrm{diag}(\boldsymbol{\alpha})\underline{\boldsymbol{R}}\mathbf{W}$ depends on the microphone weights $\boldsymbol{\alpha}$, eigenbeam weights $\mathbf{W}$, and the SH order, L. The design and impact of $\mathbf{C}$ will be discussed in Section 5.6.

Assuming stationarity, (5.21) can be estimated using

$$\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau) \approx \frac{1}{T}\sum_{n=0}^{T-1}\boldsymbol{\psi}(n)\boldsymbol{\psi}^{T}(n-\tau) \ , \tag{5.23}$$

where T is the frame size and (5.22) is approximated by

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z) \approx \sum_{\tau=-W}^{W}\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau)z^{-\tau} \ . \tag{5.24}$$

The estimation error incurred in (5.23) depends on both the ground truth $\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau)$ and frame size T [192]. In turn, the time support W, over which the estimation is performed, can be optimized to balance truncation and estimation errors [193].

The PEVD of the modal beamformer outputs in (5.22) is

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z) \approx \boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}(z)\boldsymbol{\Lambda}_{\boldsymbol{\psi}}(z)\boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}^{\mathrm{P}}(z) \ . \tag{5.25}$$

The para-unitary eigenvector filterbank, $\boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}(z) \bullet\!\!-\!\!\circ \mathbf{U}_{\boldsymbol{\psi}}(n)$, satisfies

$$\boldsymbol{\mathcal{U}}^{\mathrm{P}}_{\boldsymbol{\psi}}(z)\boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}(z) = \boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}_{\boldsymbol{\psi}}(z) = \mathbf{I} \; . \tag{5.26}$$

Thus, $\boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}(z)$ can only redistribute spectral power among channels and not change the total signal powers. The generated outputs are also spectrally majorized and sorted in descending order of their signal energy [30]. Consequently, in the case when the spectrally majorized solution corresponds to the single target source, this signal can be extracted from the first output in $\mathbf{y}(n)$ using

$$y_1(n) = \sum_{k=0}^{K} \mathbf{u}_1^{\mathrm{H}}(-k)\boldsymbol{\psi}(n-k) \; , \tag{5.27}$$

where $\mathbf{u}_1(n)$ represents the first column of $\mathbf{U}_{\boldsymbol{\psi}}(n)$ and is of order K. The filtered outputs $\mathbf{y}(n)$ are strongly decorrelated due to diagonalization achieved by $\boldsymbol{\mathcal{R}}_{\mathbf{yy}}(z) \approx \boldsymbol{\Lambda}_{\boldsymbol{\psi}}(z)$.

5.5.3 Modal Beamformer Designs for Eigenbeams

The spatial dimension of $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ is Q. For a sufficiently large SH order L, the rank of the SH basis matrix $\underline{\boldsymbol{R}}$ is $\mathcal{L}$. Consequently, the SHT generates $\mathcal{L}$ eigenbeam signals and its space-time covariance polynomial matrix, $\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z)$ in (5.22), has a spatial dimension of $\mathcal{L}$. Furthermore, if prior information such as DoAs is available, $\mathbf{W}$ in (5.22) may no longer be a square matrix, e.g., $\mathcal{P} \leq \mathcal{L}$, and contains modal beamformer weights as its elements. The resulting $\boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z)$ has dimension $\mathcal{P}$, thereby achieving further compression. The design of $\mathbf{W}$ depends on the application, and this will be demonstrated for speech enhancement and source separation in Section 5.6.

5.6 Application Examples

The proposed compaction system is used for two illustrative application examples. The first is a speech enhancement application using the proposed system to enhance the signal of a single speaker in noise, where the beamformer operates akin to a KLT to compact

Algorithm 3 PEVD processing in the SHT domain.

Inputs: $\mathbf{x}(n) \in \mathbb{R}^Q, \alpha, L, \mathbf{W}, T, W$.
$\boldsymbol{\chi}(n) \leftarrow \mathbf{x}(n)$ // *eigen-beamforming, see* (5.2)
$\boldsymbol{\psi}(n) \leftarrow \mathbf{W}^{\mathrm{T}}\boldsymbol{\chi}(n)$ // *modal beamforming*
$\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau) \leftarrow E\{\boldsymbol{\psi}(n)\boldsymbol{\psi}^{\mathrm{T}}(n-\tau)\}$ // *see* (5.23)
$\boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z) \leftarrow \mathcal{Z}\{\mathbf{R}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\tau)\}$ // *see* (5.24)
$\boldsymbol{\mathcal{U}}_{\boldsymbol{\psi}}(z), \boldsymbol{\Lambda}_{\boldsymbol{\psi}}(z) \leftarrow \mathrm{PEVD}\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\psi}\boldsymbol{\psi}}(z)\}$ // *use [25, 28–30]*
$y_1(n) \leftarrow \mathrm{filter}\{\mathbf{u}_1^{\mathrm{P}}(n), \boldsymbol{\psi}(n)\}$ // *extract target, see* (5.27)
return $y_1(n)$.

energy. The second application is source separation of two speakers. The target speaker direction is assumed to be known to inform the construction of several beamformers which coherently combine the target speaker but not the second source. This enables the extraction of the target speaker as the principal component in the PEVD stage.

5.6.1 Single Source Speech Enhancement

The problem of unsupervised speech enhancement for a single speaker is now discussed. A significant amount of data redundancy is expected when sampling the sound field of a single source by several microphones at different spatial positions. In fact, the signal subspace due to a single speech source in $\boldsymbol{\Lambda}_{\mathbf{x}}(z)$ or $\boldsymbol{\Lambda}_{\boldsymbol{\psi}}(z)$, is expected to be of rank one, thereby suggesting an opportunity for compression. If the direction of the source is not known, the full set of eigenbeams generated from the SHT can be chosen up to the L-th order SH for PEVD processing. This is equivalent to using $\mathbf{W} = \mathbf{I}$ for the modal beamformer, where $\mathbf{I} \in \mathbb{R}^{\mathcal{L}\times\mathcal{L}}$ is the identity matrix and $\mathbf{C} \in \mathbb{R}^{Q\times\mathcal{L}}$ only depends on the SH order approximation if all microphones are used for processing. The speech enhancement performance using different sets of complete eigenbeam signals will be investigated in Section 5.7.3.

5.6.2 Separation of Multiple Sources

The problem of separating multiple directional sources in a reverberant environment is next considered. Applying a PEVD for data compaction directly to the microphone signals is prohibitive in terms of computational complexity; further, due to spectral majorization, the PEVD may extract a mix of signal components that is counterproductive when performing source separation. For this reason, we assume that each target source direction is either known or can be estimated [68, 194], such that only a subset of eigenbeam signals is required, and can be further compressed by a modal beamformer. This reduces the input dimension to the PEVD stage, and also aligns the target signal while incoherently combining the secondary, interfering source(s). Thus, the PEVD becomes computationally viable and can compact the target source in its principal eigenvectors.

With $\mathscr{P}$ beamformer outputs for each target source, $\mathbf{C} \in \mathbb{R}^{Q \times \mathscr{P}}$ represents the dimensionality reduction. Consequently, the spectrally majorized outputs generated by the PEVD enable the extraction of the target source in the first channel. This process can be repeated for all sources for which directions are known. The source separation performance will be demonstrated in Section 5.7.3.

5.7 Simulations and Results

To demonstrate the benefits of the proposed framework, first a comparison of signal representations using theoretical and realistic examples generated from measured data illustrates the achievable compression and practical benefits. The proposed framework is then used for blind speech enhancement and informed source separation using the approaches discussed in Section 5.6. Listening examples are available at [195].

5.7.1 Experimental Setup

The TIMIT corpus [154] provides 16 kHz anechoic speech signals. For each speech source, short utterances from the same randomly selected speaker were concatenated to generate signals of 8 to 10 s duration. In experiments involving simulated rooms, the SMIRgen tool in [196] was used to generate the RIRs for a 32 microphone spherical array. The room setup is shown in Fig. 5.2, and the room reverberation time T_{60} [24] was varied between 0 s, 0.3 s and 0.7 s. Spatially and temporally white Gaussian noise was used to simulate sensor noise. In experiments using measured RIRs, Lecture Room 2 with $T_{60} = 1.22$ s and babble noise signals, which were recorded using the 32-channel Eigenmike spherical

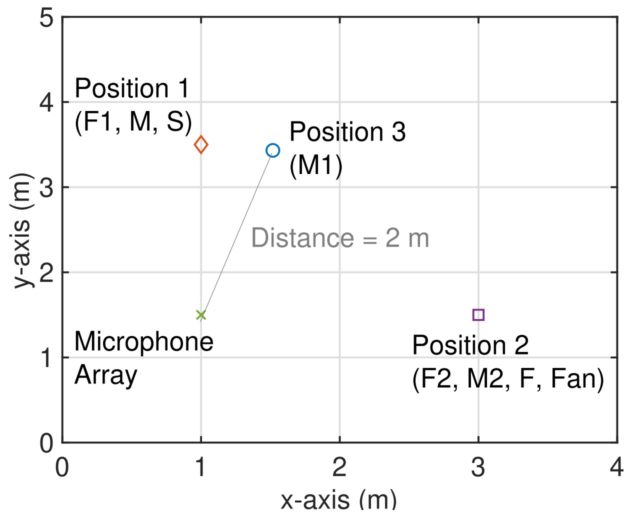


Figure 5.2: Setup of the simulated room for experiments.

array [197], were taken from the ACE corpus [155].

In each experiment, 50 trials were conducted for which the speech signals were varied. For each trial, each anechoic speech signal was convolved with the RIRs before adding noise to generate the microphone signals at a specified input SNR using [160]. The SNR ranged from -10 dB to 20 dB.

The SMD algorithm was used to compute a PEVD using $\rho = N_1/\sqrt{Q} \times 10^{-2}$ for the maximum off-diagonal column-norm, where $Q = 32$ is the number of signals used for processing, $N_1 = \mathrm{tr}\{\mathbf{R}(0)\}$ is the total energy in those Q signals, a trim factor $\mu = 10^{-3}$, 500 iterations and $T = W = 1600$ samples. The quadrature weights $\boldsymbol{\alpha}$ are computed using [198]. The three tests are now described.

A. Compression and Signal Representation Comparison

The microphone signals were directly processed using the KLT and PEVD. Additionally, SHT was applied in order to generate eigenbeam signals (SHT), with subsequent processing of eigenbeams using KLT (SHT+KLT) and PEVD (SHT+PEVD).

B. Single Source Speech Enhancement

The source was located at Position 2 in Fig. 5.2. The proposed approach using the complete set of eigenbeams for $L = 1$ (PEVD L1) and $L = 2$ (PEVD L2) are compared against the PEVD-based algorithm which uses the raw microphone signals (RAW PEVD) [39]. Consequently, 4 and 9 eigenbeam signals are passed as inputs to the PEVD algorithm. Eigenbeams $\chi_0^{(0)}$ and $\chi_1^{(1)}$ are also evaluated because the former can provide some noise reduction [131] and the latter represents a dipole directed at the source for the simulated room. To compare the ability of the PEVD to strongly decorrelate signals, the KLT was also applied to the single eigenbeam $\chi_0^{(0)}$ in the time-domain using [125] (KLT$\{\chi_0^{(0)}\}$) since spatial and temporal decorrelation are separately achieved by using the SHT and KLT, respectively.

C. Source Separation

For each target source, $\mathscr{P} = 4$ instead of 32 microphone signals were used as inputs to the PEVD to demonstrate its effectiveness while reducing computational complexity. The modal signals for the source at Position 1 were $\chi_1^{(-1)}, \chi_3^{(-3)}, \chi_3^{(-1)}$ and the modified hyper-cardioid (MHCARD) [41] beam directed at $(\frac{\pi}{2}, \frac{\pi}{2})$. For the source at Position 2, the modal signals were $\chi_1^{(1)}, \chi_3^{(1)}, \chi_3^{(3)}$ and the MHCARD directed at $(\frac{\pi}{2}, 0)$. The proposed approach was compared against well-known approaches such as the third-order hyper-cardioid modal beamformer optimized for maximum directivity index (MaxDir) [177], auxiliary function-based independent vector analysis (AuxIVA) [199], independent low-rank matrix analysis (ILRMA) [6] and fast multi-channel non-negative matrix factorization (FastMNMF) [200], implemented in [201]. During experimentation, it was found that the independent component analysis (ICA) and multi-channel non-negative matrix factorization (MNMF)-based algorithms did not perform well when all 32 microphones were used. Therefore, signals from two microphones closest to each source were chosen for these methods in the following results.

5.7.2 Evaluation Metrics

To measure the quality of the signal representation, the reconstruction error ε, coding gain γ and diagonality factor δ, defined in Section 5.4.3, are computed. The PEVD complexity factor β relative to Q microphones, estimated using $\beta = \left(\frac{\mathcal{L}}{Q}\right)^3$, quantifies the computational savings.

Speech enhancement is popularly evaluated using SegSNR and FwSegSNR for noise reduction [164], NSRR and BSD for dereverberation [167], STOI for speech intelligibility [165] and PESQ [166] for quality.

To evaluate the source separation performance, SDR, source-to-interference ratio (SIR), and source-to-artefacts ratio (SAR) are used to measure the overall source separation ability, interference rejection and processing artifacts, respectively [202].

The metrics are computed for the microphone signals and the processed signals, and their difference Δ is reported for each measure. Positive Δ values indicate improvements in all measures except ΔBSD, for which a negative value implies a reduction in spectral distortions.

5.7.3 Experiments and Discussion

A. Theoretical Example for Compression

To demonstrate the achievable compaction of the various signal representations, a single uncorrelated source is considered, i.e., $P = 1$, illuminating a spherical array with $Q' = 36$ microphones, of which initially only $Q = 25$ elements are utilized. The propagation environment is anechoic, and the array is assumed to be sufficiently small to not suffer from attenuation loss across the array. Thus, the array signal can be expressed as in (5.1), with $\mathbf{h}_1(n) \in \mathbb{R}^Q$ consisting of fractional delay filters [16, 203], and a normalization such that, for $\mathbf{h}_1(n) \circ\!\!-\!\!\bullet\, \boldsymbol{\hbar}_1(z)$, $\boldsymbol{\hbar}_1^{\mathrm{P}}(z)\boldsymbol{\hbar}_1(z) = 1$. With a source power of σ_s^2, and spatially and temporally uncorrelated noise of power σ_v^2 corrupting each microphone, the SNR at each microphone is $\rho = \sigma_s^2/\sigma_v^2$.

Table 5.1: RESULTS FOR CODING GAIN EXAMPLE.

Processing	γ [dB]
Unprocessed microphones	0
KLT on microphones	14.6
PEVD on microphones	18.3
Eigenbeams (SHT on microphones)	10.9
KLT on Eigenbeams	14.9
PEVD on Eigenbeams	19.1
Optimal	**18.7**

The ideal PEVD uses a para-unitary polynomial matrix $\boldsymbol{\mathcal{U}}(z) = [\boldsymbol{\hbar}_1(z)\boldsymbol{\mathcal{H}}(z)]^{\mathrm{P}}$, such that $\boldsymbol{\mathcal{H}}^{\mathrm{P}}(z)\boldsymbol{\hbar}_1(z) = \mathbf{0}$, with $\boldsymbol{\mathcal{H}}(z) : \mathbb{C} \to \mathbb{C}^{Q\times(Q-1)}$ generated via a para-unitary matrix completion method [31]. Then, the polynomial eigenvalues are

$$\lambda_m(z) = \begin{cases} Q\mathcal{S}(z) + \sigma_v^2 & , \ m = 1 \\ \sigma_v^2 & , \ m > 1 \end{cases} , \tag{5.28}$$

where $\mathcal{S}(z) \bullet\!\!-\!\!\circ\, S(\tau)$ and $S(\tau)$ is the auto-correlation function of the source signal. With $\mathcal{S}(z) = \sigma_s^2$, the optimal coding gain can be shown to take the form

$$\gamma_{\mathrm{opt}} = \frac{\sigma_s^2 + \sigma_v^2}{\sqrt[Q]{(Q\sigma_s^2 + \sigma_v^2)(\sigma_v^2)^{Q-1}}} = \frac{1+\rho}{\sqrt[Q]{1+Q\rho}} \, . \tag{5.29}$$

With $Q = 25$ and $\rho = 100$, i.e. an SNR of 20 dB, $\gamma_{\mathrm{opt}} = 18.7$ dB. A KLT, performing an optimum instantaneous spatial decorrelation, achieves a coding gain of $\gamma_{\mathrm{KLT}} = 14.6$ dB, while an approximate PEVD via the SMD algorithm applied to the microphone data yields $\gamma_{\mathrm{SMD}} = 18.3$ dB because of diagonalization achieved over a range of time lags. The approximate nature of the SMD algorithm leads to a small loss in coding gain compared to γ_{opt} shown in Table 5.1, but provides significantly better coding gain than the KLT. Fig. 5.3(a) illustrates the variances of the microphone signals, $(\sigma_s^2 + \sigma_v^2)$, and the outputs of the KLT and PEVD, demonstrating the effect of compaction with the KLT packing the source signal into approximately six and the PEVD into approximately two outputs before the noise floor of σ_v^2 is reached.

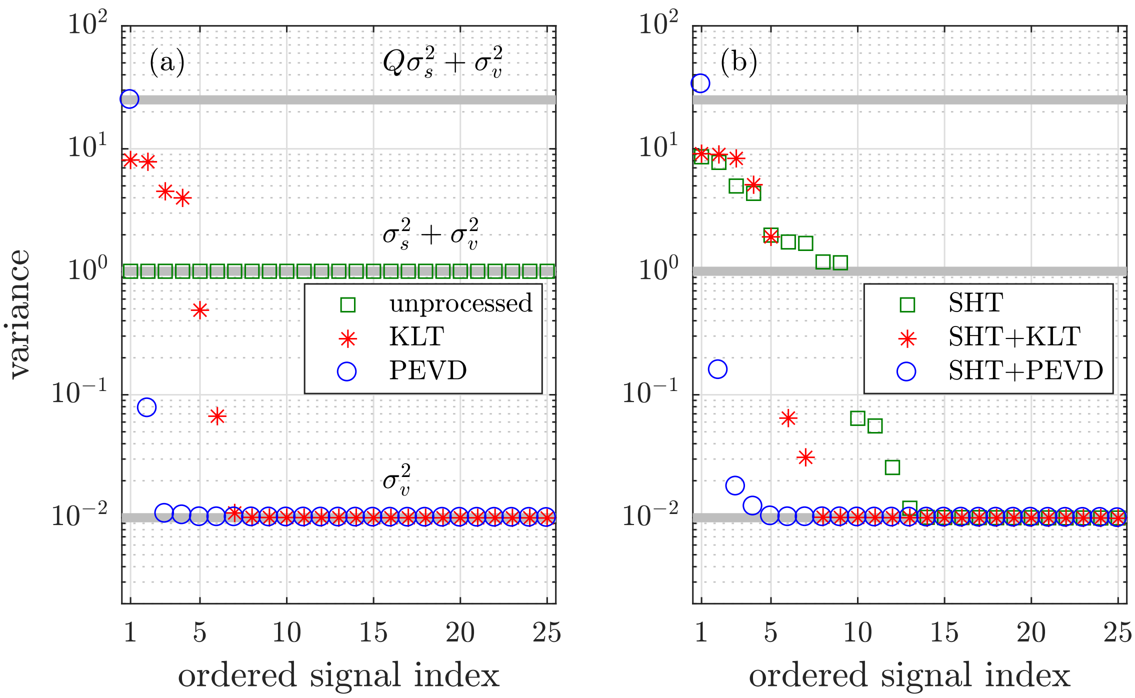


Figure 5.3: Variances achieved by various data compaction approaches based on (a) unprocessed microphone signals and (b) eigenbeams produced by the SHT. Grey horizontal lines show the upper and lower variances provided by the ideal PEVD in (5.28), and the variance of the unprocessed microphone signals.

Since the SHT is generally not orthonormal, $Q' = 36$ microphone signals were used to generate a reduced set of $Q = 25$ eigenbeams, i.e., a 4th order SH representation. This can be achieved by a transform matrix $\mathbf{S} \in \mathbb{R}^{Q \times Q'}$, with $\mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{I}$ while noting that $\mathbf{S}^{\mathrm{T}}\mathbf{S} \neq \mathbf{I}$. If the input is spatially uncorrelated and $\mathbf{x}(n) = \mathbf{S}\mathbf{x}'(n)$, then $\mathbb{E}\{\|\mathbf{x}(n)\|_2^2\} = \frac{Q}{Q'}\mathbb{E}\{\|\mathbf{x}'(n)\|_2^2\}$. For spatially correlated data, such as the above source signal, no input-output relation for the energy can be stated: in the best case, the signal $\mathbf{x}'(n)$ occupies the row space of $\mathbf{S}$; in the worst case, it will lie in the nullspace of $\mathbf{S}$ resulting in zero power for $\mathbf{x}(n)$.

Therefore, due to the incomplete preservation of power in the SHT, the coding gain cannot yield a direct comparison. Ignoring this, Table 5.1 shows that the SH signal representation yields a coding gain of 10.9 dB over the unprocessed microphone signals. A KLT or PEVD operating on the eigenbeams increases this coding gain further. Fig. 5.3(b) shows that the SHT+PEVD coding gain exceeds the optimum value according to (5.29), because the best-compacted signal breaks the upper limit in (5.28). Importantly, though, Fig. 5.3(b) illustrates three key aspects of our proposed approach: (i) without being data adaptive, the SHT still manages to compress the signal into approximately 12 eigenbeams; (ii) processing this reduced set of eigenbeams further by a KLT or PEVD yields similar compression when compared to the results in Fig. 5.3(a); thus, (iii) particularly the PEVD can operate on a significantly reduced spatial dimension and therefore at a much reduced computational cost after SHT preprocessing.

B. Signal Representation Example Using Measured Data

To compare the signal representation of speech signals captured by spherical microphone arrays in practice, an illustrative example is generated using RIRs and 10 dB babble noise measured by the 32-element Eigenmike in ACE Lecture Room 2. A portion of the space-time covariance matrix corresponding to the first 4 microphones in Fig. 5.4 shows that correlations exist at non-zero lags. The PEVD is used to generate outputs which are spatially decorrelated over a range of time lags, resulting in a space-time covariance matrix equivalent to $\mathbf{\Lambda}_{\mathbf{x}}(z)$ in Fig. 5.5.

With $Q = 32$ microphones, the sound field can be approximated up to SH order 4. Note

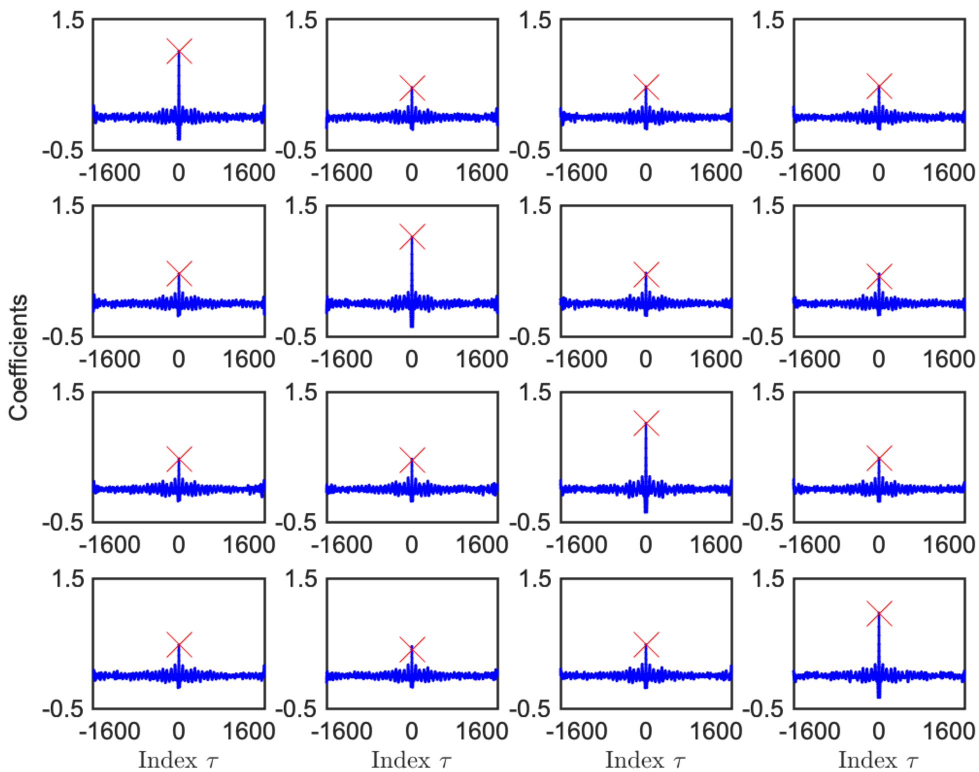


Figure 5.4: Portion of $\mathbf{R}_{\mathbf{xx}}(\tau)$ corresponding to the first 4 microphone signals from the Eigenmike with $\mathbf{R}_{\mathbf{xx}}(0)$ marked by red crosses. The (p, q)-th subplot represents the correlation sequence between the p-th and q-th microphone signals.

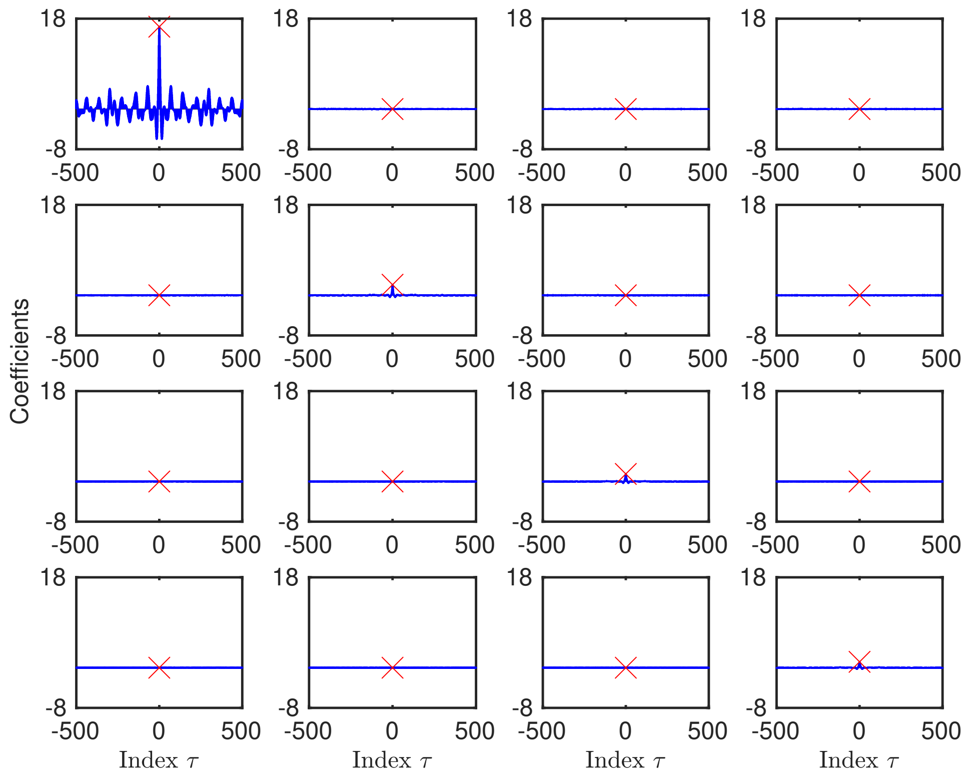


Figure 5.5: Visualization of the first 4 principal eigenvalues in $\mathbf{\Lambda}_{\mathbf{x}}(\tau) \circ\!\!-\!\!\bullet\, \mathbf{\Lambda}_{\mathbf{x}}(z)$ computed from $\mathcal{R}_{\mathbf{xx}}(z)$ in Fig. 5.4 with $\mathbf{\Lambda}_{\mathbf{x}}(0)$ marked by red crosses.

Table 5.2: Signal representation using SHT for 10 dB babble noise Lecture Room example in Fig. 5.4.

L	$\mathcal{L}$	Complexity, β	Error, ε[%]	δ [dB]
0	1	-	31.7	-
1	4	0.002	13.3	5.16
2	9	0.022	8.4	5.00
3	16	0.125	5.4	4.93
4	25	0.477	2.3	4.87

that in the theoretical example of compression achieved by different signal representations in Section 5.7.3, the number of channels before and after processing is the same, i.e. 25 channels. Practical microphone arrays like the Eigenmike, however, are usually spatially over-sampled as they use more microphones than $(L+1)^2$, represented by an over-determined non-square (tall) compression matrix $\mathbf{C} \in \mathbb{R}^{Q\times(L+1)^2}$.

While the complexity β increases with L, the reconstruction error ϵ generally decreases as seen in Table 5.2. For this example, even with $L = 2$, ε is 8.4%, indicating that 9 eigenbeams can capture most of the spatial information. This suggests that the raw microphone signals may be highly redundant, and that the SHT can offer a significant level of compression for an order-limited sound field. For a small number of $\mathcal{L}$ signals, β is also significantly lower and is omitted for $\mathcal{L} = 1$ since the PEVD is a multi-channel algorithm.

When L increases from 1 to 4, δ reduces from 5.16 dB to 4.87 dB indicating that there is a decrease in the on-diagonal energy E_{on} relative to the off-diagonal energy E_{off}, i.e., the matrix for larger L is less diagonal. This is expected because most higher-order eigenbeams do not contain the target speech component and will not exhibit correlations between many pairs of eigenbeam signals. Some higher-order eigenbeams, however, may exhibit temporal correlations with lower-order eigenbeams pointing in the same direction, as seen from the peaks between $\chi_0^{(0)}(n)$ and $L = 1$ in Fig. 5.6 occurring at $\tau \neq 0$. Since SH are spatially orthogonal, the eigenbeams may be instantaneously decorrelated, as shown by the red crosses in Fig. 5.6. To reduce the correlations between eigenbeams across the range of τ, a PEVD is used and leads to $\mathbf{\Lambda}_{\boldsymbol{\chi}}(z)$ in Fig. 5.7.

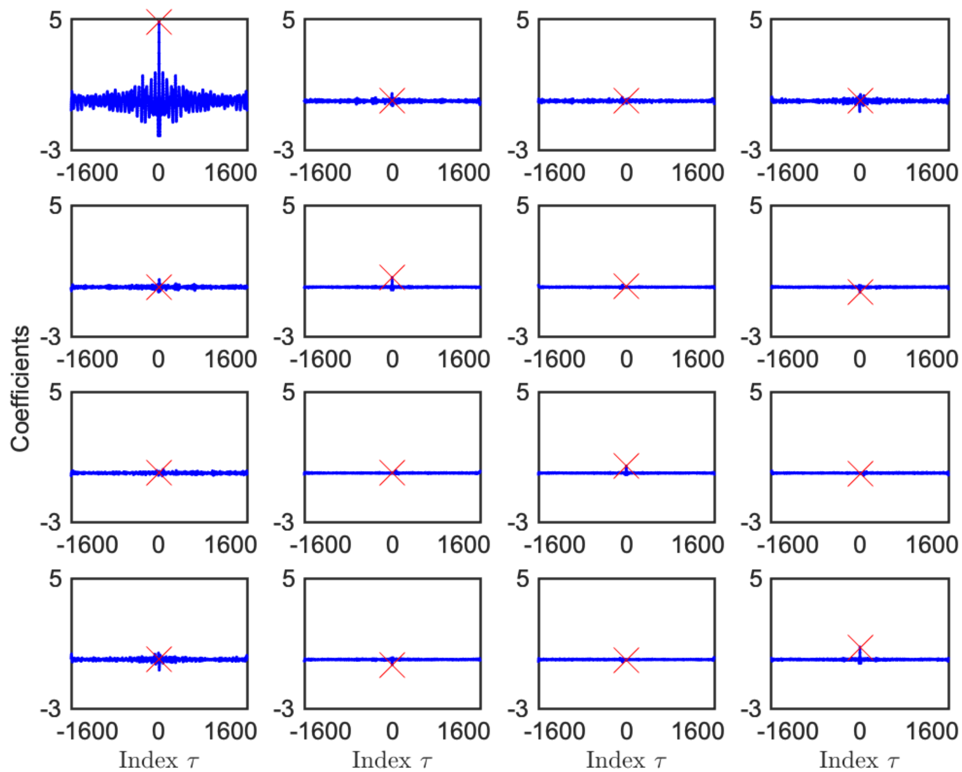


Figure 5.6: $\mathbf{R}_{\chi\chi}(\tau) \circ\!\!-\!\!\bullet\, \mathcal{R}_{\chi\chi}(z)$ of eigenbeams for $L = 2$ computed from signals used to generate Fig. 5.4, with $\mathbf{R}_{\chi\chi}(0)$ marked by red crosses. The (p, q)-th subplot represents the correlation sequence between the p-th and q-th eigenbeam signals.

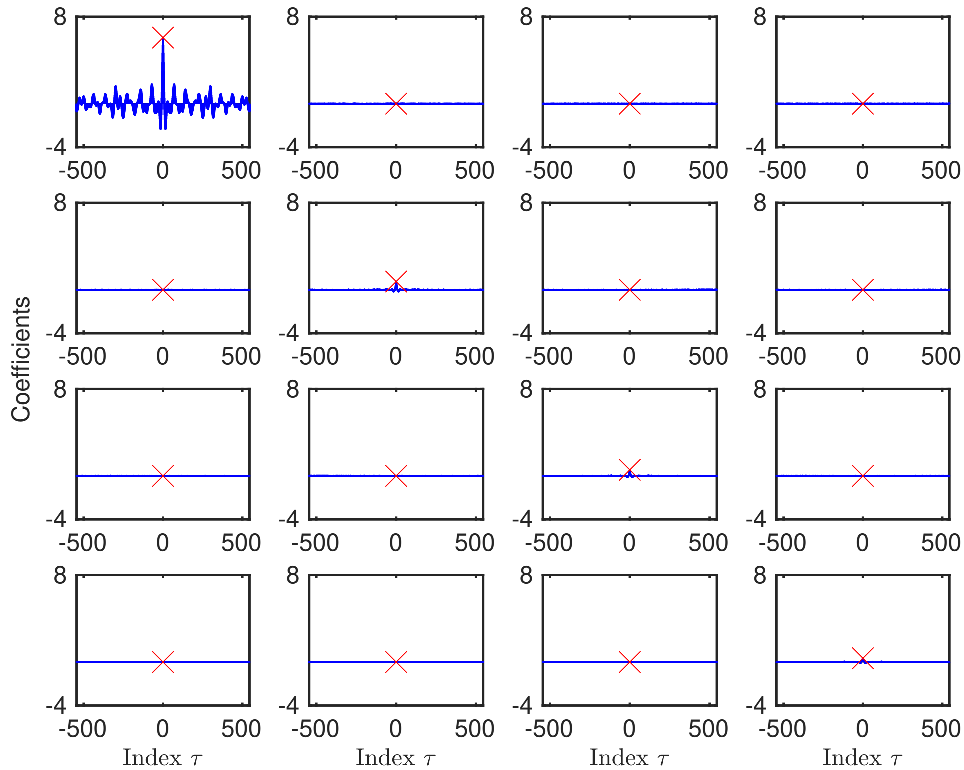


Figure 5.7: $\boldsymbol{\Lambda}_{\boldsymbol{\psi}}(\tau) \circ\!\!-\!\!\bullet\ \boldsymbol{\Lambda}_{\boldsymbol{\psi}}(z)$ of eigenbeam signals in Fig. 5.6 with $\mathbf{R}_{\psi\psi}(0)$ marked by red crosses.

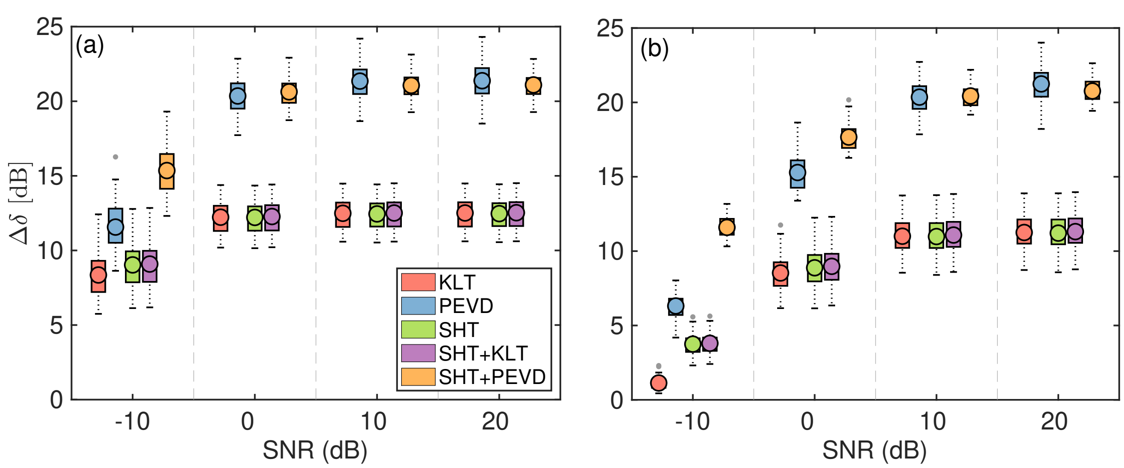


Figure 5.8: Diagonality factor improvement for (a) white noise in simulated room with $T_{60} = 0.3\,\text{s}$ and (b) babble noise in ACE Lecture Room 2.

Table 5.3: Speech enhancement results for a single example in 0 dB white noise.

Algorithm	ΔFwSegSNR	ΔSTOI	ΔPESQ	ΔNSRR	ΔBSD
$\chi_0^{(0)}$	4.86 dB	0.055	0.42	7.69 dB	-1.53 dB
KLT$\{\chi_0^{(0)}\}$	5.56 dB	0.054	**0.51**	**10.8 dB**	-1.65 dB
$\chi_1^{(1)}$	0.89 dB	0.122	0.44	1.08 dB	-0.65 dB
PEVD L1	5.72 dB	0.110	0.47	7.98 dB	-1.68 dB
PEVD L2	**5.92 dB**	**0.125**	**0.51**	7.78 dB	**-1.71 dB**
RAW PEVD	5.59 dB	0.119	0.49	8.13 dB	-1.62 dB

The diagonality factor δ defined in Section 5.4.3 is computed for 50 trials. The measures for the 32 microphones are computed and compared with the processed signals. The results for white noise in the simulated room and babble noise in Lecture Room 2 are shown in Fig. 5.8. In both cases, the difference in diagonality, $\Delta\delta$ is highest by up to 22 dB for the PEVD-based methods because PEVD can diagonalize the space-time covariance matrix over a range of time lags. At -10 dB SNR, however, the proposed SHT+PEVD improves diagonalization by 7 dB over PEVD using the microphone signals directly. SHT processing alone provides a substantial improvement in $\Delta\delta$ by up to 12 dB.

C. Single Source Speech Enhancement

Table 5.3 compares the speech enhancement performance for a single example using the methods presented in Section 5.7.1. In this simulated setup, the PEVD-based algorithms perform similarly, with PEVD L2 performing best in most metrics. PEVD L1 gives a greater improvement in ΔFwSegSNR and ΔBSD by 5.72 dB and -1.68 dB compared to RAW PEVD, which uses all microphone signals directly. After PEVD L2, the $\chi_1^{(1)}(n)$ eigenbeam directed at the source gives the best ΔSTOI. Applying the KLT to $\chi_0^{(0)}(n)$ improves all measures except for ΔSTOI.

Speech enhancement performance for 50 Monte-Carlo trials in the simulated room is shown in Fig. 5.9. Across all measures and SNRs, the PEVD-based algorithms (RAW PEVD, PEVD L1, PEVD L2) are ranked first or second after KLT$\{\chi_0^{(0)}(n)\}$, which is optimal for white noise, or the $\chi_1^{(1)}(n)$ eigenbeam, which is pointing directly at the source. The

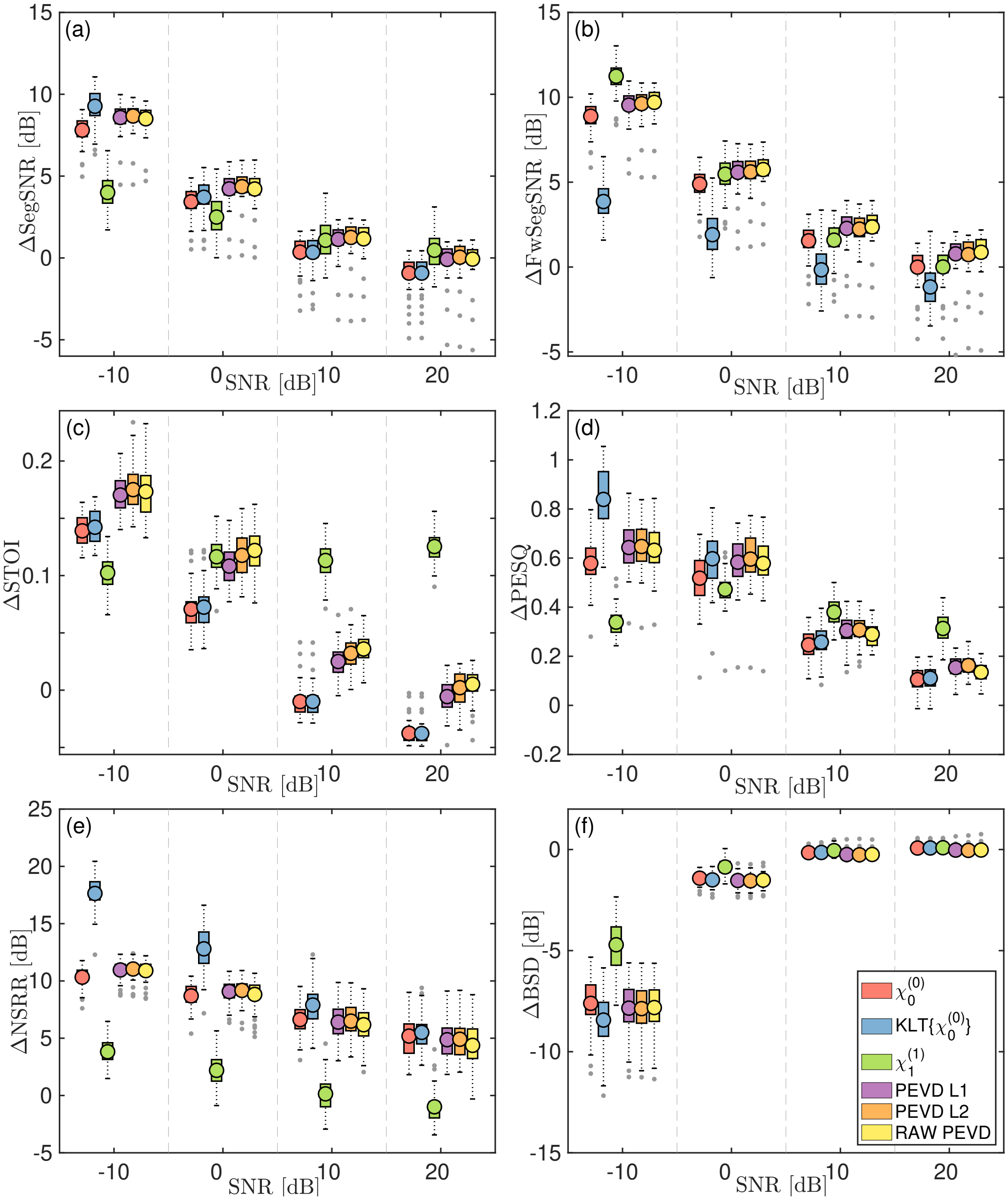


Figure 5.9: Speech enhancement results for white noise in a simulated room with $T_{60} = 0.3\,\text{s}$.

two PEVD-based algorithms perform comparably well, even though different numbers of channels are used for processing. This shows that processing the eigenbeams instead of the raw microphone signals for speech enhancement is effective and computationally advantageous, and is without the need of DoA information.

When recorded signals from the ACE corpus are used, Fig. 5.10 shows that RAW PEVD provides the greatest improvement across all metrics for SNR$\geq$ 0 dB and is closely followed by PEVD L2 and PEVD L1 even without using any DoA information. The $\chi_0^{(0)}(n)$ eigenbeam shows some improvement and the use of KLT does not offer further improvement, and may even introduce processing artifacts when the noise is not white. This can be seen from a slight reduction in ΔSTOI and ΔPESQ in Fig. 5.10(c) and Fig. 5.10(d) for babble noise. The $\chi_1^{(1)}(n)$ eigenbeam does not perform well because it is not directly pointing at the source and may only pick up weaker reverberant components along with noise.

At -10 dB SNR, the PEVD algorithms which use the eigenbeams, PEVD L1 and PEVD L2, perform better than RAW PEVD. In very noisy environments, the generated eigenbeams take advantage of the noise reduction offered by the signal independent SH domain processing, which is not available to RAW PEVD. At other SNRs, they perform comparably even though PEVD L1 and PEVD L2 use 4 and 9 channels respectively compared to the 32 channels used in the RAW PEVD approach.

D. Separation of Multiple Sources

Table 5.4 summarizes the results for a scenario involving the separation of 2 female speakers in an anechoic room. FastMNMF provides greatest ΔSIR of 35.2 dB and is closely followed by ILRMA, AuxIVA and PEVD. PEVD outperforms other algorithms in ΔSDR, ΔSAR, ΔSTOI and ΔPESQ by 21.8 dB, 16.4 dB, 0.24 and 1.39, respectively because it does not introduce processing artifacts. This is also observed in the listening examples [195] which further indicate that non-target speech signals are attenuated but remain intelligible.

These findings are also observed in Fig. 5.11 for 50 trials involving source separation with two female speakers: F1 and F2, two male speakers: M1 and M2, male and female

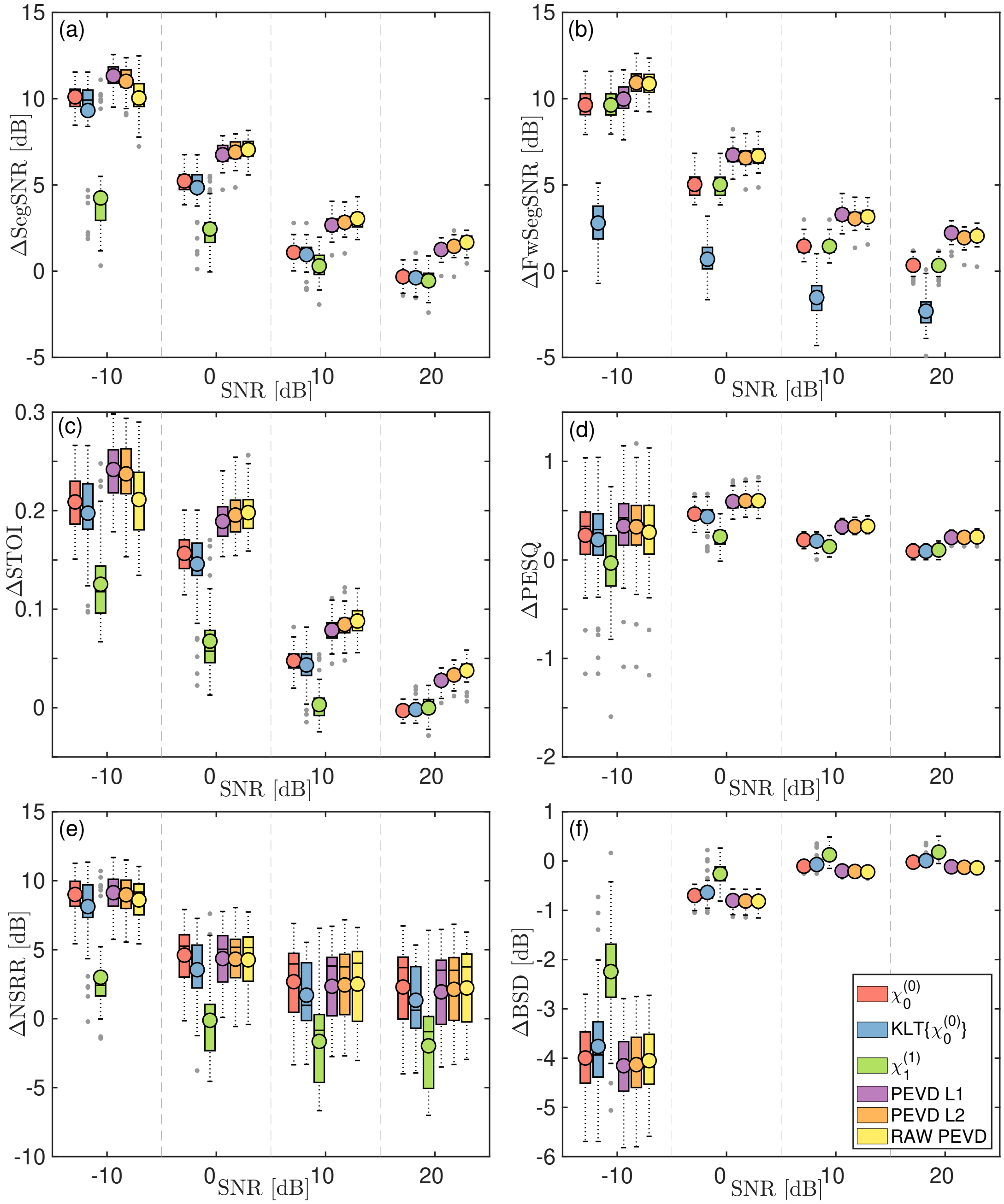


Figure 5.10: Speech enhancement results for babble noise in ACE Lecture Room 2 with $T_{60} = 1.22\,\text{s}$.

Table 5.4: Source separation results for a single female speaker example consisting of a 2 female talker scenario in an anechoic environment.

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
MHCARD	16.9 dB	17.8 dB	13.4 dB	0.21	0.93
PEVD	**21.8 dB**	25.3 dB	**16.4 dB**	**0.24**	**1.39**

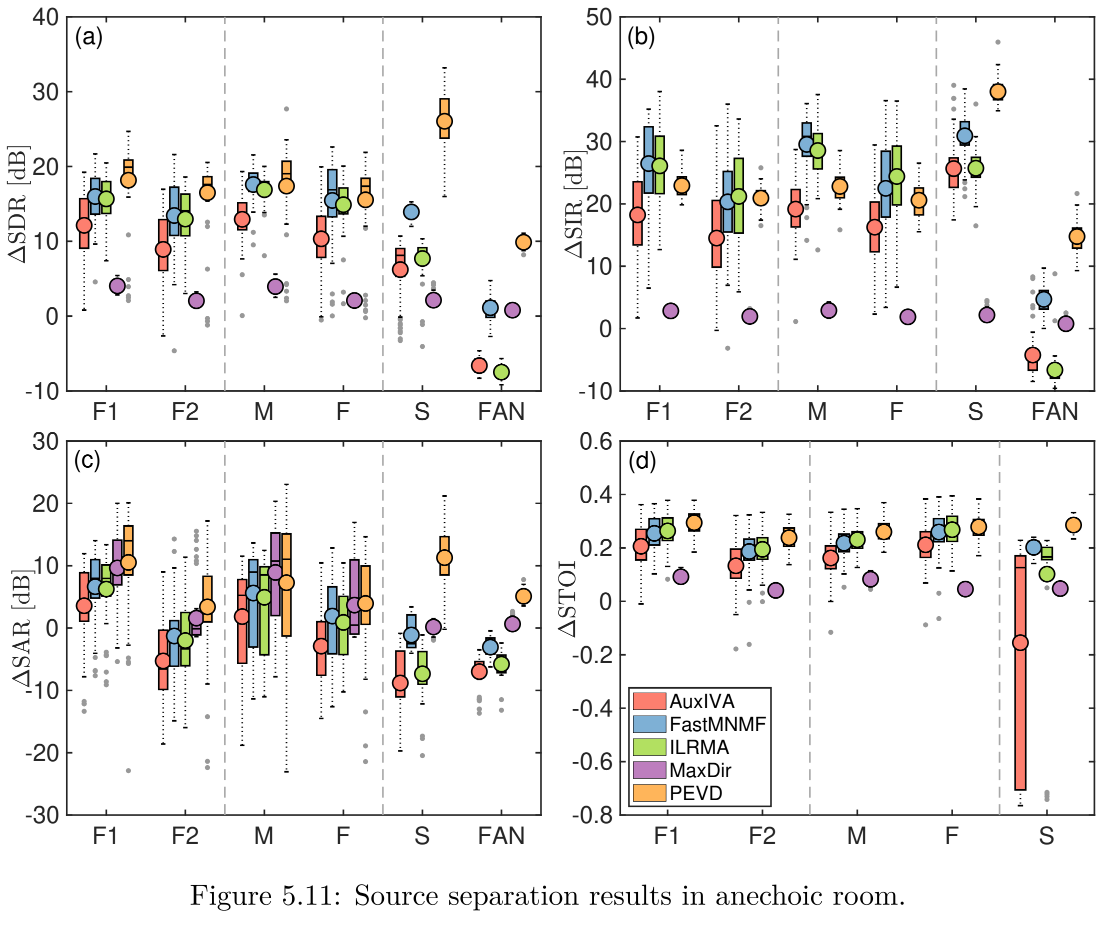


Figure 5.11: Source separation results in anechoic room.

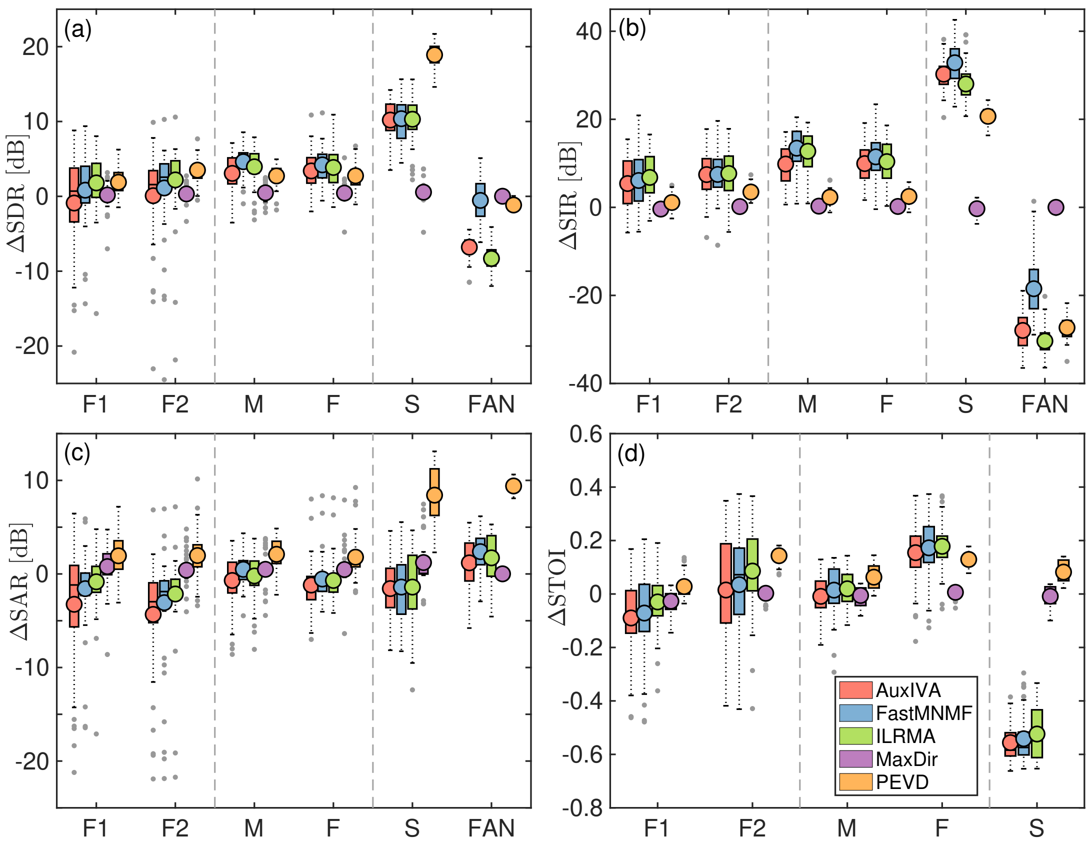


Figure 5.12: Source separation results in room with $T_{60} = 0.7\,\mathrm{s}$.

speakers: M and F, a single speaker and localized fan noise: S and FAN using the setup in Fig. 5.2. When different source types are used in an anechoic room, the source separation of the PEVD is comparable with FastMNMF and ILRMA and is better than AuxIVA in the majority of cases.

Results for the room with $T_{60} = 0.7\,\mathrm{s}$ in Fig. 5.12 indicate that the PEVD is comparable to ICA-based and MNMF-based methods in ΔSDR, worse in ΔSIR but best in ΔSAR and ΔSTOI in most cases. For the S and FAN scenario, PEVD performs better in SDR and SAR for both sources as it does not rely on source density functions which may not model fan noise well. Listening examples in [195] further support the observations that the PEVD does not introduce distortions because SMD is a time-domain method which preserves spectral coherence. We have previously demonstrated that it achieves a good compaction, resulting in good performance in the above applications.

5.8 Chapter Summary and Conclusions

For a spherical microphone array, signal representations using SHT, KLT and PEVD are compared. While the PEVD achieves close to optimum signal compaction, it is computationally expensive. In contrast, the SHT achieves sub-optimal data compaction, but is data-invariant and scales well with the number of microphones. Therefore, using the PEVD to spatially decorrelate signals over a range of time shifts while managing computational complexity, we propose to combine the SHT and PEVD approaches. We first apply the SHT and obtain a number of eigenbeams that is smaller than the number of input signals. If DoA information is available, further computational complexity saving is expected with even fewer beamformed signals for PEVD processing.

The proposed framework for signal representation demonstrates that the diagonality factor improves on average by 7 dB over microphone signal representations. When exploiting the framework for speech enhancement and source separation, the method improves STOI and SDR by up to 0.2 and 20 dB, respectively. Informal listening examples also indicate that the method does not introduce any audible artifacts [195].

Chapter 6

Fixed Beamformer Design Using PEVD

6.1 Introduction

In this chapter, fixed beamformers designed by the PEVD algorithm, which are suited for arbitrary arrays and do not require geometry information for training, will be introduced. Although fixed beamformer design generates a specific spatial response and may be limited in dynamic acoustical environments, they are still widely used because of their simplicity, effectiveness and low complexity. These polynomial eigenvector filterbanks, which can be interpreted as spatial filters, are then analyzed and visualized. The proposed fixed beamformers are lightweight, with an average filter length of 114 and will be shown to perform comparably to classical data-dependent MVDR and LCMV beamformers. This chapter is closely based on the work [43].

6.2 Overview of Array Processing and Beamforming

Array processing is widely used in many speech applications involving multiple microphones. These applications include automatic speech recognition [204], robot audition [112], telecommunications and hearing aids [205]. Beamforming, or the design of a spatio-temporal filter for the array, combines signals from different microphones to extract the desired signal arriving from a specific direction [206]. Processing these extracted signals instead of the microphone signals usually improves the application performance.

The beamformer can be designed to be fixed or adaptive [207, 208]. In fixed or data-independent designs, the array can have a specific spatial response by proper selection of the filter weights. Although fixed beamformers are limited in dynamic acoustical environments, they are still in use due to their simplicity, effectiveness and low complexity [204, 209], which is particularly important for on-device processing with limited power and computation.

Data-dependent or adaptive beamformers usually rely on the statistical properties of the signals [22]. They can be designed to adapt to time-changing acoustical environments based on statistically optimal criteria such as maximum SNR, minimum mean square error (MMSE) and LCMV [1, 210]. These have led to well-known beamformers such as the MWF [119, 124], MVDR [130, 207, 211] and generalized sidelobe canceller (GSC) [129, 212]. In most approaches, the microphone signals are usually processed in the short-time Fourier transform domain. However, this approach ignores the correlation between different DFT bins and phase coherence across bins [14].

An alternative approach can be formulated in the z-domain using polynomial matrices computed from the microphone signals. Polynomial matrices can simultaneously capture the space, time and frequency correlations and are suitable for modelling multi-channel broadband signals. The polynomial matrices are processed using an iterative PEVD algorithm such as SBR2 [30, 53] and SMD [28, 29] in the time-domain or [60] in the frequency-domain. PEVD algorithms have been found useful for speech enhancement [39, 40], source separation [33, 41], source localization [68, 100, 213] and channel coding [64].

In [31, 214], polynomial MVDR and GSC have been developed by formulating and solving the well-known optimization problem in the z-domain using polynomial matrix techniques. In this chapter, we will investigate fixed beamformers designed by PEVD for arbitrary arrays. During training, the learnt PEVD filterbanks are stored in a look-up table, and the entries are later retrieved and used directly for testing. The novel contributions of this chapter are (i) an analysis of the fixed beamformer design using PEVD, (ii) a visual interpretation of the eigenvectors generated by the PEVD, and (iii) a comparison of the proposed approach with classical beamformers including MVDR and LCMV.

6.3 Problem Formulation

6.3.1 Signal Model

The received signal at the q-th microphone for sample index n is

$$x_q(n) = \sum_{p=1}^{P} h_{p,q}(n) * s_p(n) + v_q(n) \ , \tag{6.1}$$

where $h_{p,q}(n)$ represents the time-invariant RIR from the p-th source to the q-th microphone, $s_p(n)$ is the p-th source signal, $v_q(n)$ represents the additive noise at the q-th microphone, and $*$ denotes the linear convolution operator. The noise signals are assumed to be zero-mean, uncorrelated with each other and the source signals [113]. The data vector over Q microphones is $\mathbf{x}(n) = [x_1(n), \ldots, x_Q(n)]^{\mathrm{T}} \in \mathbb{R}^Q$, where $[\cdot]^{\mathrm{T}}$ is the transpose operator.

6.3.2 Polynomial Matrix Eigenvalue Decomposition

The space-time covariance matrix [30, 48], parameterized by time lag $\tau \in \mathbb{Z}$, is computed using

$$\mathbf{R}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n-\tau)\} \ , \tag{6.2}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator over n. Each element, $r_{p,q}(\tau)$, is the correlation sequence between the p-th and q-th microphone signals. This produces auto- and cross-correlation sequences on the diagonals and off-diagonals, respectively.

The z-transform of (6.2),

$$\mathcal{R}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}(\tau) z^{-\tau} \ , \tag{6.3}$$

also denoted by $\mathbf{R}(\tau) \circ\!\!-\!\!\bullet \ \mathcal{R}(z)$, is a para-Hermitian polynomial matrix which satisfies $\mathcal{R}(z) = \mathcal{R}^{\mathrm{P}}(z) = \mathcal{R}^{\mathrm{H}}(1/z^*)$, where $[\cdot]^*$, $[\cdot]^{\mathrm{H}}$, $[\cdot]^{\mathrm{P}}$ are the complex conjugate, Hermitian and para-Hermitian operators respectively.

The para-Hermitian matrix EVD of $\boldsymbol{\mathcal{R}}(z) \in \mathbb{C}^{Q\times Q}$ in (6.3) is [48, 49]

$$\boldsymbol{\mathcal{R}}(z) = \boldsymbol{\mathcal{U}}(z)\,\boldsymbol{\Lambda}(z)\,\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\ , \tag{6.4}$$

where the columns of $\boldsymbol{\mathcal{U}}(z) \in \mathbb{C}^{Q\times Q}$ are the polynomial eigenvectors and the elements on the diagonal matrix $\boldsymbol{\Lambda}(z) \in \mathbb{C}^{Q\times Q}$ are the polynomial eigenvalues. Iterative PEVD algorithms based on the SBR2 [30, 53] and SMD [28, 29] are used to approximate (6.4) by Laurent polynomial factors, which also encourage spectrally majorization such that the eigenvalues in $\boldsymbol{\Lambda}(z)$ are strictly ordered on the unit circle. Diagonalization of (6.4) is achieved using

$$\boldsymbol{\Lambda}(z) = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\,\boldsymbol{\mathcal{R}}(z)\,\boldsymbol{\mathcal{U}}(z)\ . \tag{6.5}$$

6.4 Beamformer Designs Using Polynomial EVD

6.4.1 Polynomial Eigenvector Beamformer Design

In most applications involving physical signals, there is often coupling in space, time and frequency because of the wave equation [1], motivating the need for a decoupling approach such as the PEVD. The eigenvector $\boldsymbol{\mathcal{U}}(z)$ performs a filter-and-sum operation on the microphone inputs to generate Q strongly decorrelated signals, i.e., output signals are not spatially correlated over a range of time lags. The q-th column of $\boldsymbol{\mathcal{U}}(z)$, denoted by $\boldsymbol{u}_q(z)$, represents the q-th beamformer generated by the PEVD.

The acoustic channel in (6.1) is more compactly represented by $\mathbf{H}(n) \circ\!\!-\!\!\bullet\ \boldsymbol{\mathcal{H}}(z) \in \mathbb{C}^{P\times Q}$, where each element is $h_{p,q}(n)$. As the sources propagate through the acoustic channels, the space-time polynomial matrix of the microphone signals, corrupted by spatially and temporally uncorrelated noise with equal power σ_v^2, is

$$\boldsymbol{\mathcal{R}}_{\mathbf{x}}(z) = \boldsymbol{\mathcal{H}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{R}}_{\mathbf{s}}(z)\boldsymbol{\mathcal{H}}(z) + \sigma_v^2\mathbf{I}\ , \tag{6.6}$$

where $\boldsymbol{\mathcal{R}}_{\mathbf{x}}(z) \in \mathbb{C}^{Q\times Q}$ and $\boldsymbol{\mathcal{R}}_{\mathbf{s}}(z) \in \mathbb{C}^{P\times P}$ are the space-time covariance matrices of the microphone and source signals, and $\mathbf{I}$ is the identity matrix.

Consider the case when source signals are generated as uncorrelated, zero-mean unit variance Gaussian processes $g_p(n) \in N(0,1)$ with the following cross-correlation sequence [28]

$$r_{m,l}(\tau) = \mathbb{E}\{g_m(n)g_l(n-\tau)\} = \delta(m-l)\delta(\tau)\ . \tag{6.7}$$

This implies that $\boldsymbol{\mathcal{R}}_{\mathbf{s}}(z) = \mathbf{I}$. Further, assume that sources satisfy spectral majorization such that [28, 64]

$$\gamma_{g_{m-1}}(e^{\mathrm{j}\Omega}) \geq \gamma_{g_m}(e^{\mathrm{j}\Omega}) \quad \forall\Omega,\ m = 2,\ldots,P\ , \tag{6.8}$$

where $\gamma_{g_m}(e^{\mathrm{j}\Omega})$ is the power spectral density of $g_m(n)$ at frequency Ω. Applying the PEVD and rearranging, (6.6) becomes

$$\boldsymbol{\Lambda}(z) - \sigma_v^2\mathbf{I} = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{H}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{H}}(z)\boldsymbol{\mathcal{U}}(z)\ . \tag{6.9}$$

This implies that the right hand side term in (6.9) must also achieve diagonalization, indicating that $\boldsymbol{\mathcal{U}}(z)$ might decorrelate $\boldsymbol{\mathcal{H}}(z)$. The eigenvector filterbank generated by the iterative PEVD algorithms satisfy the para-unitary or lossless condition [45] such that

$$\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{U}}(z)\boldsymbol{\mathcal{U}}^{\mathrm{P}}(z) = \mathbf{I}\ . \tag{6.10}$$

The first consequence of (6.10) implies that $\boldsymbol{\mathcal{U}}(z)$ cannot change the total power over the subspaces but can only redistribute spectral power among the microphones [30, 39]. This gives $\boldsymbol{\mathcal{U}}(z)$ a stable and all-pass characteristic [30]. The second consequence is the orthogonality such that the inner product between different columns of $\boldsymbol{\mathcal{U}}(z)$ equals 0.

For this choice of source signals, i.e. white Gaussian, the PEVD filterbank $\boldsymbol{\mathcal{U}}(z)$ is designed to primarily target only the acoustic channels $\boldsymbol{\mathcal{H}}(z)$. With a different selection of source signals, another set of $\boldsymbol{\mathcal{U}}(z)$ will be generated to achieve diagonalization or decorrelation in space, time and frequency simultaneously as shown by

$$\boldsymbol{\Lambda}(z) - \sigma_v^2\mathbf{I} = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{H}}^{\mathrm{P}}(z)\boldsymbol{\mathcal{R}}_{\mathbf{s}}(z)\boldsymbol{\mathcal{H}}(z)\boldsymbol{\mathcal{U}}(z)\ . \tag{6.11}$$

6.4.2 Broadband Steering Vectors and Beam Patterns

To understand the spatio-spectral processing of the designed beamformer, the response of the array to a plane wave is computed in Section 6.5. The broadband steering vector, which implements fractional delays, is defined as [16]

$$\boldsymbol{a}_{\phi}(z) = [1, a_2(z), \ldots, a_Q(z)]^{\mathrm{T}} \ , \tag{6.12}$$

where $a_q(z) = \sum_n a_q(n)z^{-n}$, $a_q(n) = \mathrm{sinc}(nT_s - \Delta\tau_q)$, $\Delta\tau_q$ is the time delay from the source to the q-th microphone relative to the first reference microphone and T_s is the sampling period. The time delays are embedded in the angles of arrival ϕ and can be modelled explicitly using the array geometry [100]. In particular, a narrowband source signal with frequency Ω_0 arriving at the array is $e^{\mathrm{j}\Omega_0}\mathbf{a}_{\phi}(\Omega_0)$, where $\mathbf{a}_{\phi}(\Omega_0) = [1, \ldots, e^{-\mathrm{j}\Omega_0\Delta\tau_Q}]^{\mathrm{T}}$ is expressed in terms of phase shifts from the first reference microphone.

The broadband array response can be computed using

$$\mathcal{B}(\phi, z) = \boldsymbol{\mathcal{U}}^{\mathrm{P}}(z)\boldsymbol{a}_{\phi}(z) \ . \tag{6.13}$$

The array response at frequency Ω can be obtained by evaluating (6.13) on the unit circle such that

$$\mathbf{B}(\phi, \Omega) = \mathcal{B}(\phi, z)|_{z=e^{\mathrm{j}\Omega}} \ . \tag{6.14}$$

It is also common to reparameterize ϕ in (6.14) using wavenumber $\mathbf{k}$ for a more general representation, i.e., $\mathbf{B}(\mathbf{k}, \Omega)$. In this case, (6.14) is termed as the frequency-wavenumber response function [1]. In this chapter, only the azimuth angle ϕ is considered, and (6.14) will be sufficient but this can straightforwardly be extended to consider the general case.

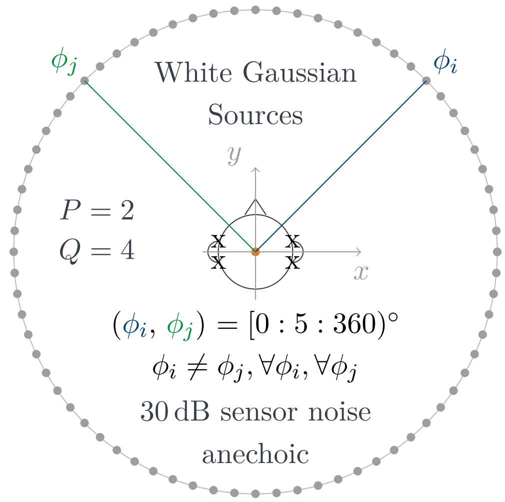


Figure 6.1: Training setup using White Gaussian sources and simulated far-field RIRs.

6.5 Simulation and Results

6.5.1 Setup and Evaluation Metrics

Spatially and temporally white Gaussian noise was used as the source signals for training the filterbanks and simulating sensor noise. Pink noise from the Noisex database [156] was also used for training. Anechoic speech signals sampled at 16 kHz were taken from the TIMIT corpus [154]. For each speaker, short utterances were concatenated to generate signals of 8 to 10 s. During both training and testing, each source signal was convolved with the RIRs and summed before adding 30 dB SNR sensor noise.

Anechoic RIRs were simulated using [215] for a single source located at different angular positions $0° \leq \phi < 360°$ in steps of $5°$ relative to the x-axis in Fig. 6.1, which shows a $Q = 4$ free-field hearing aid array. Different combinations of $P = 2$ sources at positions $(\phi_i, \phi_j), i \neq j$, and the SMD algorithm [28] were used to train the fixed beamformers. To guarantee spectral majorization in (6.8) so that the PEVD outputs are ordered in a desired sequence as explained in Section 6.4.1, the signal power of source S1 at ϕ_i was set to be 10 dB higher than S2 at ϕ_j. When both sources are white Gaussian, the filterbank has been designed to primarily target the acoustic channel, as explained in Section 6.4.1. For each azimuth pair, the learnt PEVD filterbank was stored in a look-up table, the entries

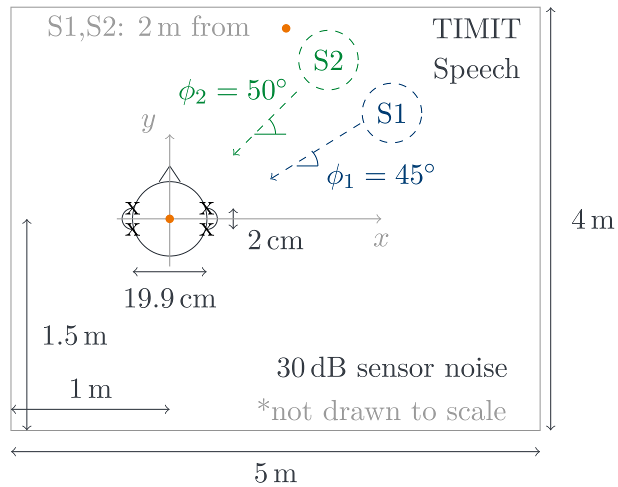


Figure 6.2: Testing setup using speech sources 1 and 2 and simulated reverberant RIRs.

of which were later retrieved and directly used for testing.

The testing setup in Fig. 6.2 shows that S1 and S2 were positioned at $\phi_1 = 45°$ and $\phi_2 = 50°$ from the x-axis, 2 m from the hearing aid array. The speech sources were adjusted to the same power levels at the source positions using [159] implemented in [160]. RIRs for a 5 m $\times$ 4 m $\times$ 6 m room with a reverberation time T_{60} of 300 ms were generated using [216].

For performance evaluation, STOI [165] and SIR [202] are used to measure speech intelligibility and interference rejection. Although each PEVD filterbank consists of 4 beamformers, results for only the first 2 are presented for brevity.

6.5.2 Experiments and Discussions

Experiment 1: Beam Patterns for White Noise Sources

The first white noise source is 10 dB higher in power than the second white source. The PEVD beamformers are computed from these signals for different combinations of source angles. The minimum, maximum and mean lengths of the filters are 17, 372 and 114, respectively. The histogram of the filter lengths is shown in Fig. 6.3. To demonstrate the effectiveness of the proposed approach for two sources closely spaced 5° apart and for

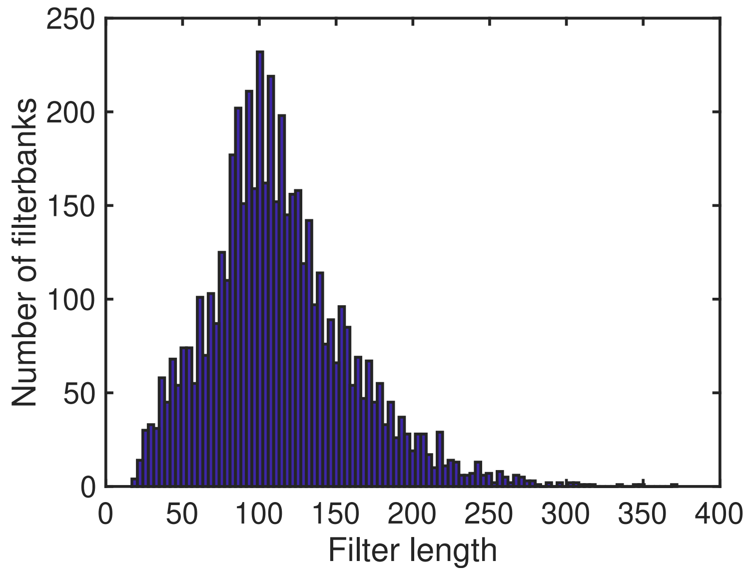


Figure 6.3: Histogram of filter lengths for the trained PEVD filterbanks.

brevity, only analysis for the $\phi_1 = 45°$ and $\phi_2 = 50°$ scenario is presented.

The beam patterns for the PEVD beamformers are shown in Fig. 6.4. The first beamformer $\boldsymbol{u}_1(z)$ in Fig. 6.4(i) has a response that resembles a delay-and-sum beamformer, with unit gain in the look direction $\phi_1 = 45°$ for all frequencies. This is expected since the SMD algorithm [28] uses a series of delays to maximize the auto-correlation sequences on the diagonal eigenvalue polynomial matrix while enforcing an ordering from the largest to smallest energy. The second output $\boldsymbol{u}_2(z)$ in Fig. 6.4(ii) shows a deep null steered in the direction of the first source position at $\phi_1 = 45°$ across all frequencies. The first and second beamformers also exhibit orthogonality characteristics in space and z-domain, e.g., unit gain and a null at $\phi_1 = 45°$ for PEVD beamformers 1 and 2, respectively.

For reference, MVDR and LCMV beamformers for $\{45°, 50°\}$ are also presented in Fig. 6.5 and Fig. 6.6, respectively. The beamformer weights are calculated using the statistics estimated from the received signals in the array and processed using [217]. The ground truth steering vectors are provided to ensure unit gain in the look directions as shown in Fig. 6.5. For the LCMV, hard constraints, i.e., unit gain in the target direction and zero gain in the interferer direction for all frequencies, are used to generate the beam patterns

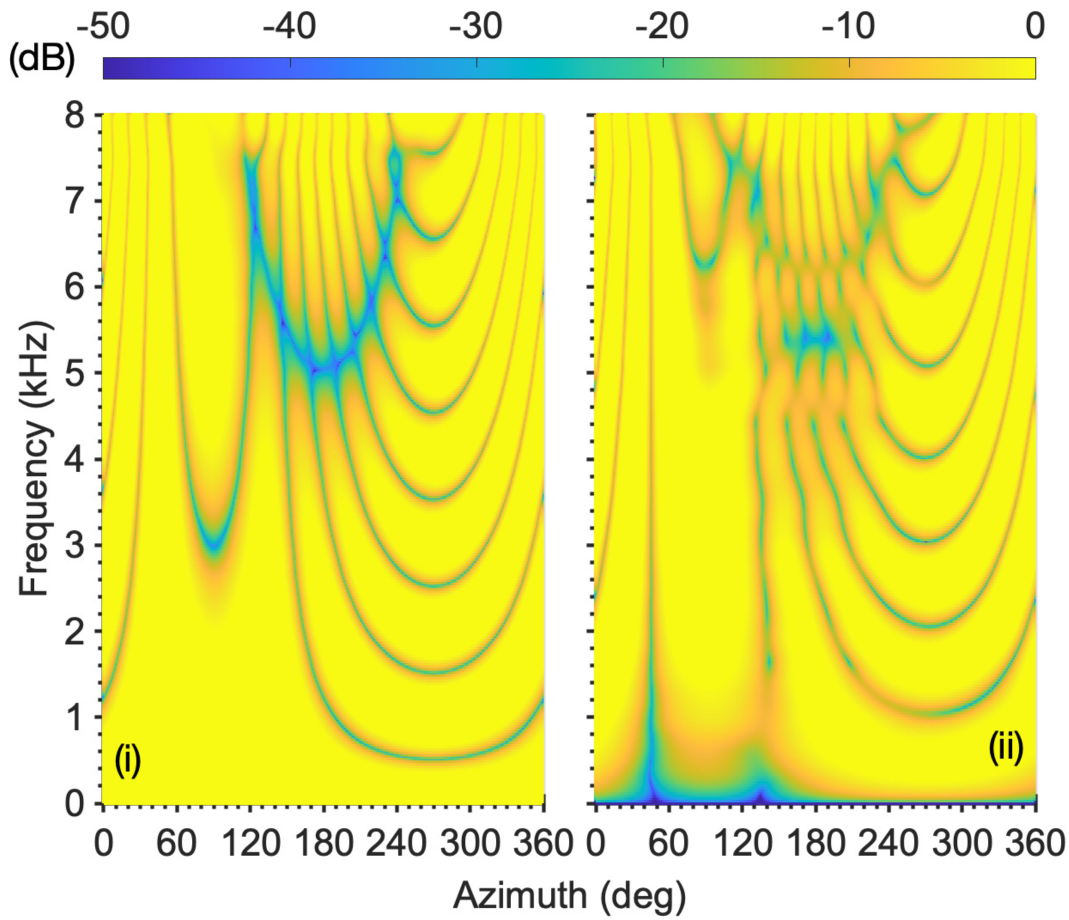


Figure 6.4: PEVD filterbank for $\{45°, 50°\}$ trained on 2 white noise sources. (i) The first filterbank $\boldsymbol{u}_1(z)$ has a response that resembles a delay-and-sum beamformer with a look direction at 45°. (ii) The second filterbank $\boldsymbol{u}_2(z)$ is a cancellation beamformer which steers a null at 45°.

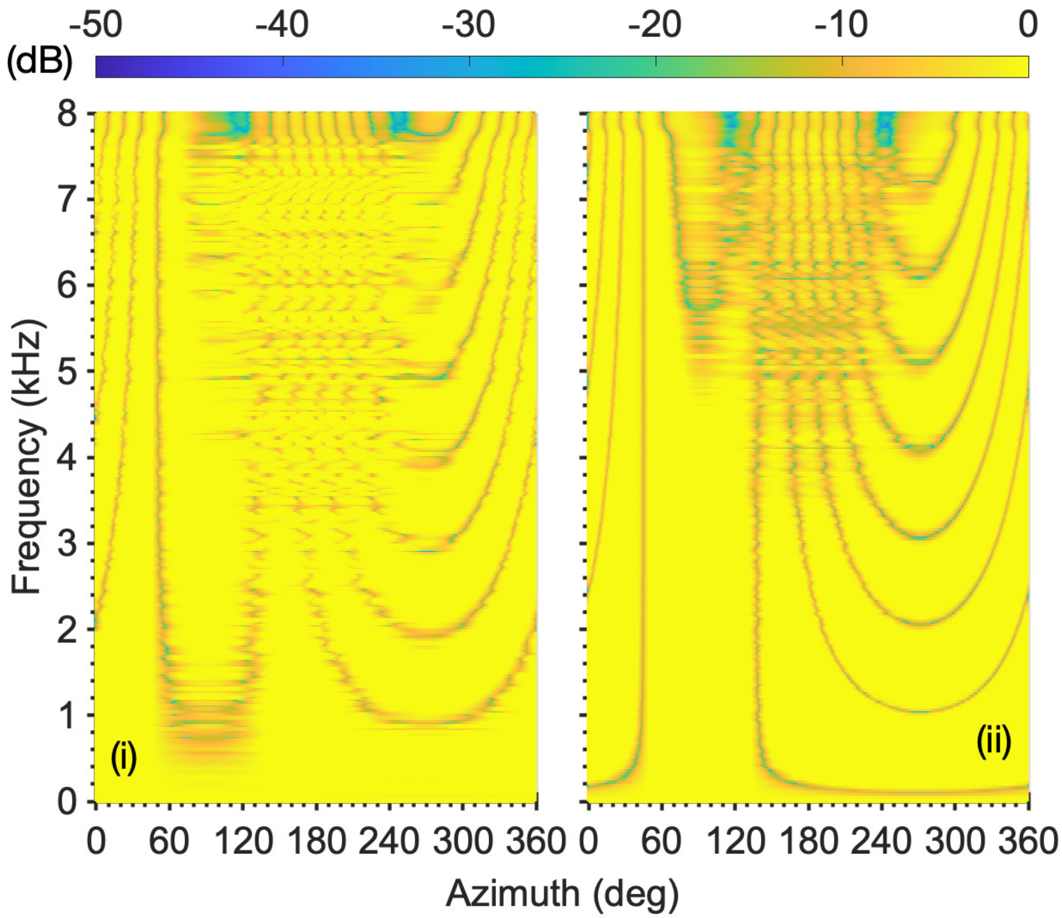


Figure 6.5: MVDR beamformer for $\{45°, 50°\}$ applied to 2 white noise sources. (i) MVDR has a look direction at 45°. (ii) MVDR has a look direction at 50°.

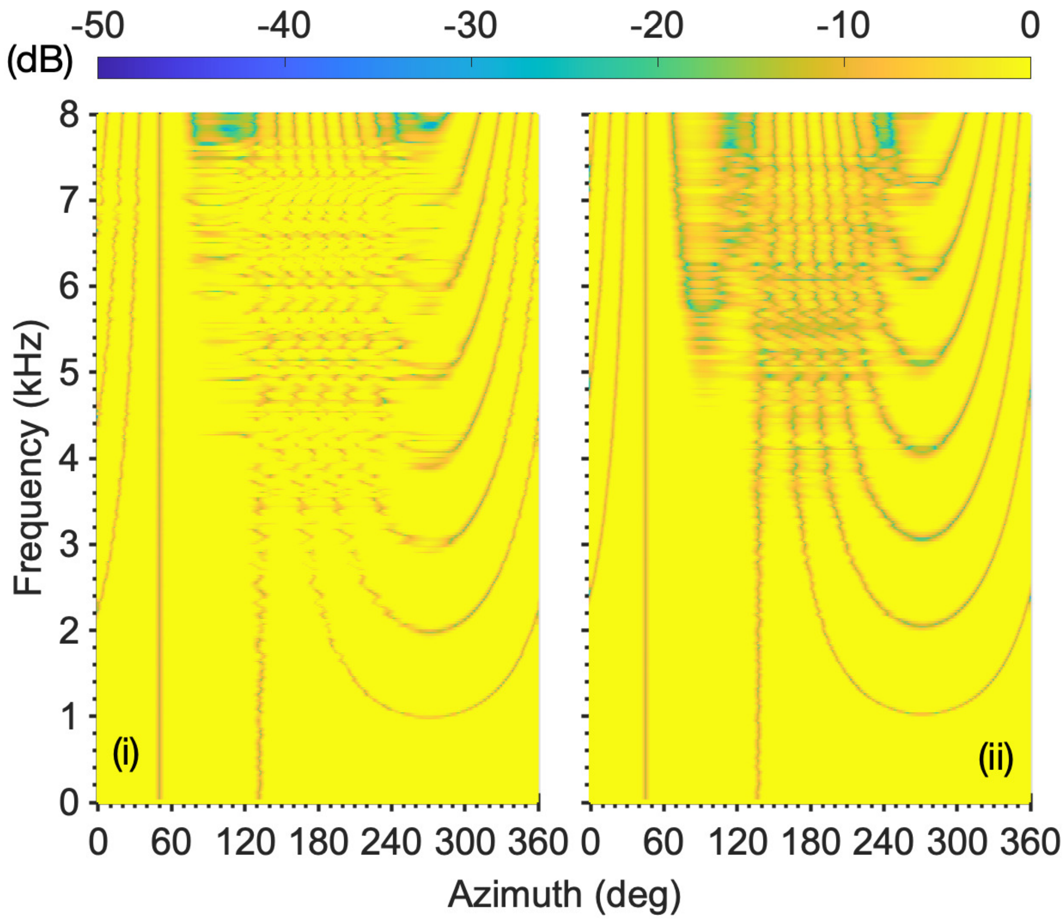


Figure 6.6: LCMV beamformer for $\{45°, 50°\}$ applied to 2 white noise sources. (i) LCMV has a look direction at 45° and a null in 50°. (ii) LCMV has a look direction at 50° and a null in 45°.

Table 6.1: Source separation of two speakers in an anechoic environment. PEVD is trained on two white source signals while PEVD$'$ is trained on one white and one pink source signals.

	S1 ($\phi_1 = 45°$)		S2 ($\phi_2 = 50°$)	
Algorithm	STOI	SIR (dB)	STOI	SIR (dB)
Received	0.809	0.240	0.640	-0.358
PEVD $\{45°, 50°\}$	0.811	0.206	**0.844**	15.394
PEVD $\{50°, 45°\}$	**0.932**	16.943	0.644	-0.111
MVDR	0.922	13.727	0.826	12.077
LCMV	0.856	**20.226**	0.796	**23.164**
PEVD$'$ $\{45°, 50°\}$	0.806	0.195	0.369	11.953
PEVD$'$ $\{50°, 45°\}$	0.366	12.392	0.636	-0.109

in Fig. 6.6. Compared to the second LCMV in Fig. 6.6(ii), the PEVD beamformer $\boldsymbol{u}_2(z)$ in Fig. 6.4(ii) has additionally provided some attenuation around the low frequencies near the null direction $\phi_1 = 45°$. Note that this direction has unit gain for the beamformer $\boldsymbol{u}_1(z)$ in Fig. 6.4(i).

Experiment 2: Separation of Speech Sources

In this experiment, the source signals are two male speakers of the same signal power, each positioned at 45° and 50°. The previously trained PEVD filterbank for white sources located at $\{45°, 50°\}$ shown in Fig. 6.4 is used for this experiment without any modifications. The beamformer weights for MVDR and LCMV are calculated using the statistics estimated from the received signals in the array while the ground truth steering vectors are provided.

The beam patterns for the MVDR and LCMV are provided in Fig. 6.7 and Fig. 6.8, respectively. The PEVD beamformer $\boldsymbol{u}_2(z)$ trained on white noise signals in Fig. 6.4(ii) has a beam pattern that resembles the LCMV beamformer in Fig. 6.8(ii) — both impose null constraints on $\phi_1 = 45°$ except for the additional low frequency attenuation provided by PEVD in the neighbourhood of the $\phi_1 = 45°$.

The source separation results for an anechoic environment are shown in Table 6.1. In the PEVD filterbanks designed for $\{\phi_i, \phi_j\}$, the first beamformer $\boldsymbol{u}_1(z)$ behaves like a delay-

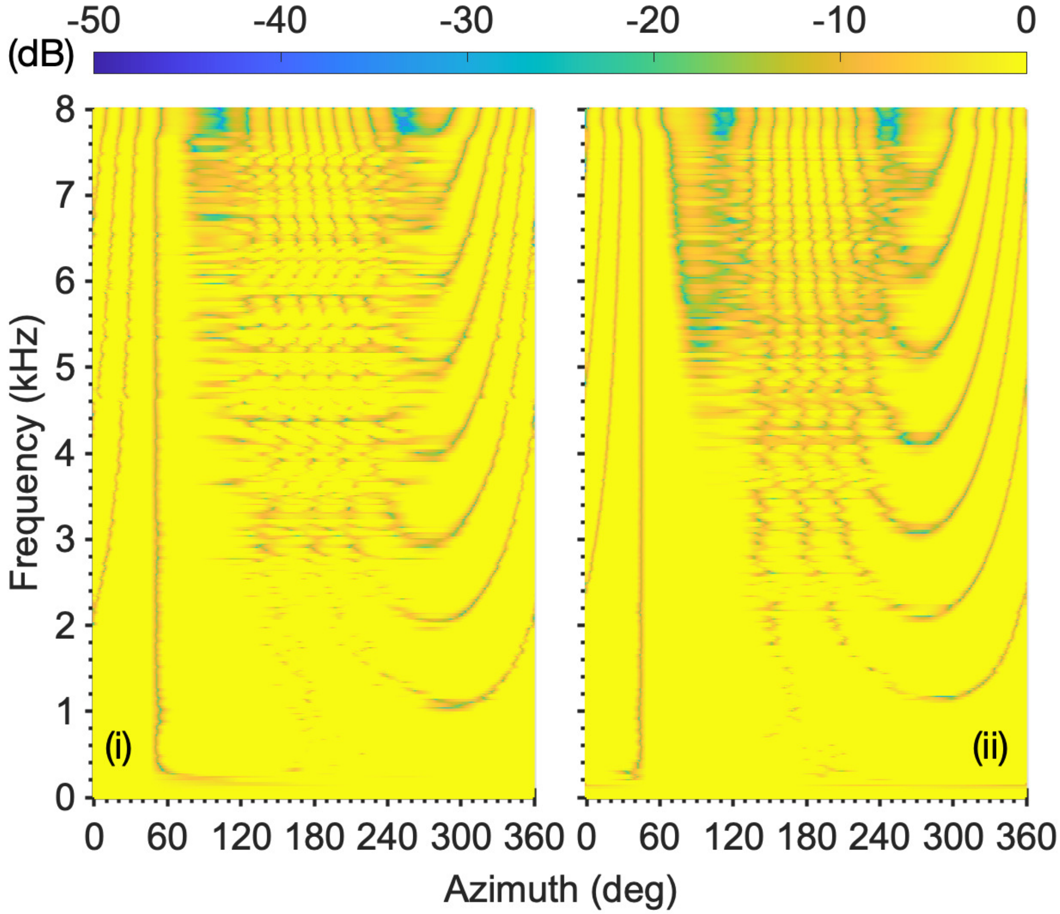


Figure 6.7: MVDR beamformer for $\{45°, 50°\}$ applied to 2 speech sources in an anechoic environment. (i) MVDR has a look direction at 45°. (ii) MVDR has a look direction at 50°.

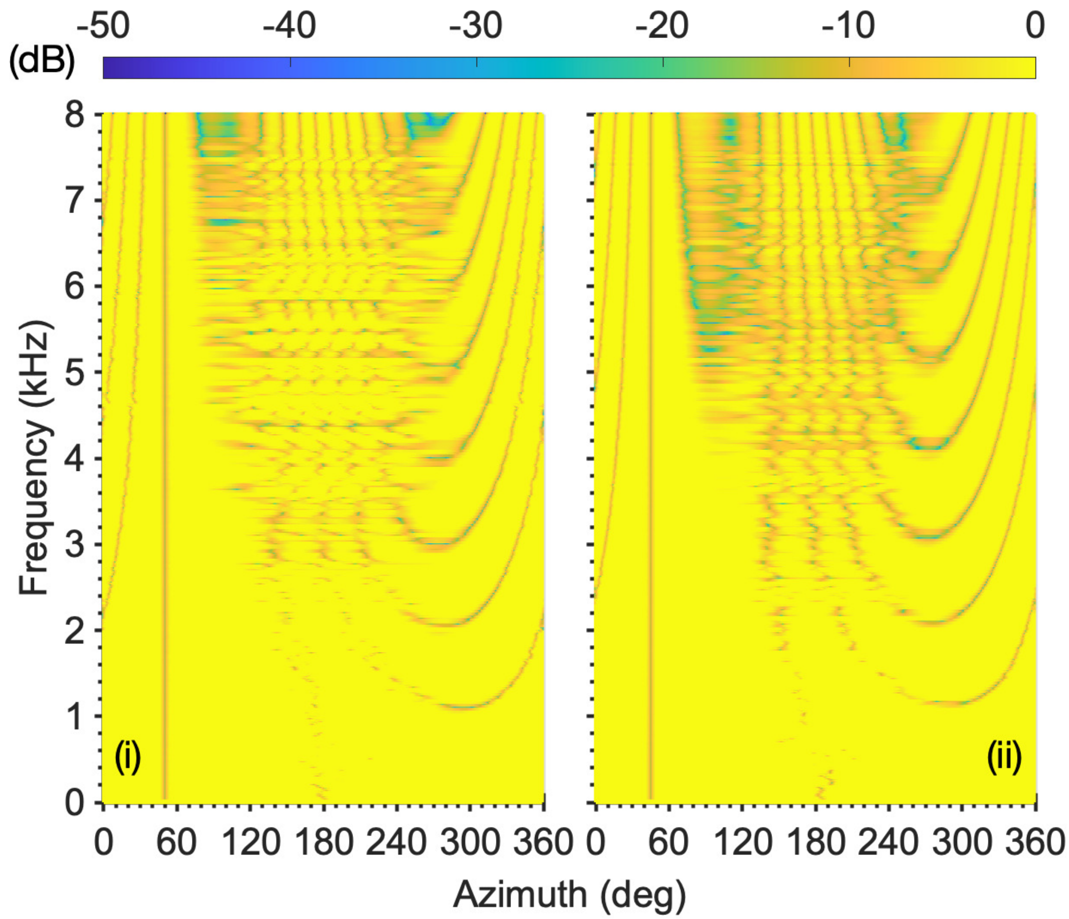


Figure 6.8: LCMV beamformer for $\{45°, 50°\}$ applied to 2 speech sources in an anechoic environment. (i) LCMV has a look direction at 45° and a null in 50°. (ii) LCMV has a look direction at 50° and a null in 45°.

Table 6.2: Source separation performance in a reverberant room with a T_{60} of 300 ms.

	S1 ($\phi_1 = 45°$)		S2 ($\phi_2 = 50°$)	
Algorithm	STOI	SIR (dB)	STOI	SIR (dB)
Received	0.698	0.298	0.566	0.055
PEVD $\{45°, 50°\}$	0.682	0.950	0.557	- 0.133
PEVD $\{50°, 45°\}$	**0.715**	1.102	0.533	- 0.502
MVDR	0.649	- 0.483	**0.582**	**0.128**
LCMV	0.638	**4.958**	0.491	- 3.940

and-sum steered towards ϕ_i while the second beamformer $\boldsymbol{u}_2(z)$ behaves like a cancellation or spatial filter nulling signals arriving from ϕ_i. Consequently, the PEVD cancellation beamformers perform the best in STOI and even outperform MVDR by 0.018 for S2. In terms of SIR, the PEVD cancellation filters perform better than MVDR by more than 3 dB but not as well as LCMV beamformers which places a deep notch in the interferer direction. However, PEVD cancellation beamformers perform better than LCMV in STOI by up to 0.15.

The results for a room with a T_{60} of 300 ms are shown in Table 6.2. In this case, the fixed beamformers designed by PEVD are used directly. The PEVD$\{50°, 45°\}$ cancellation beamformer $\boldsymbol{u}_2(z)$ that extracts S1 at $\phi = 45°$ gives the most significant improvement in STOI and SIR simultaneously. In general, PEVD beamformers do not worsen STOI and SIR scores as much as MVDR and LCMV. For example, the worst-case ΔSTOI for S1 using $\boldsymbol{u}_1(z)$ from PEVD$\{45°, 50°\}$ is -0.016 compared to -0.049 for MVDR, and -0.060 for LCMV while the worst-case ΔSIR is -0.557 dB for $\boldsymbol{u}_1(z)$ from PEVD$\{50°, 45°\}$ and -3.995 dB for LCMV, respectively. Although the ground truth steering vectors are provided, the estimation of signal statistics using the microphone signals rather than the ground truth noise covariance matrices makes the beamformers perform more like the minimum power distortionless response (MPDR) and linearly constrained minimum power (LCMP), respectively. The beam patterns for the reverberant room are also provided in [218].

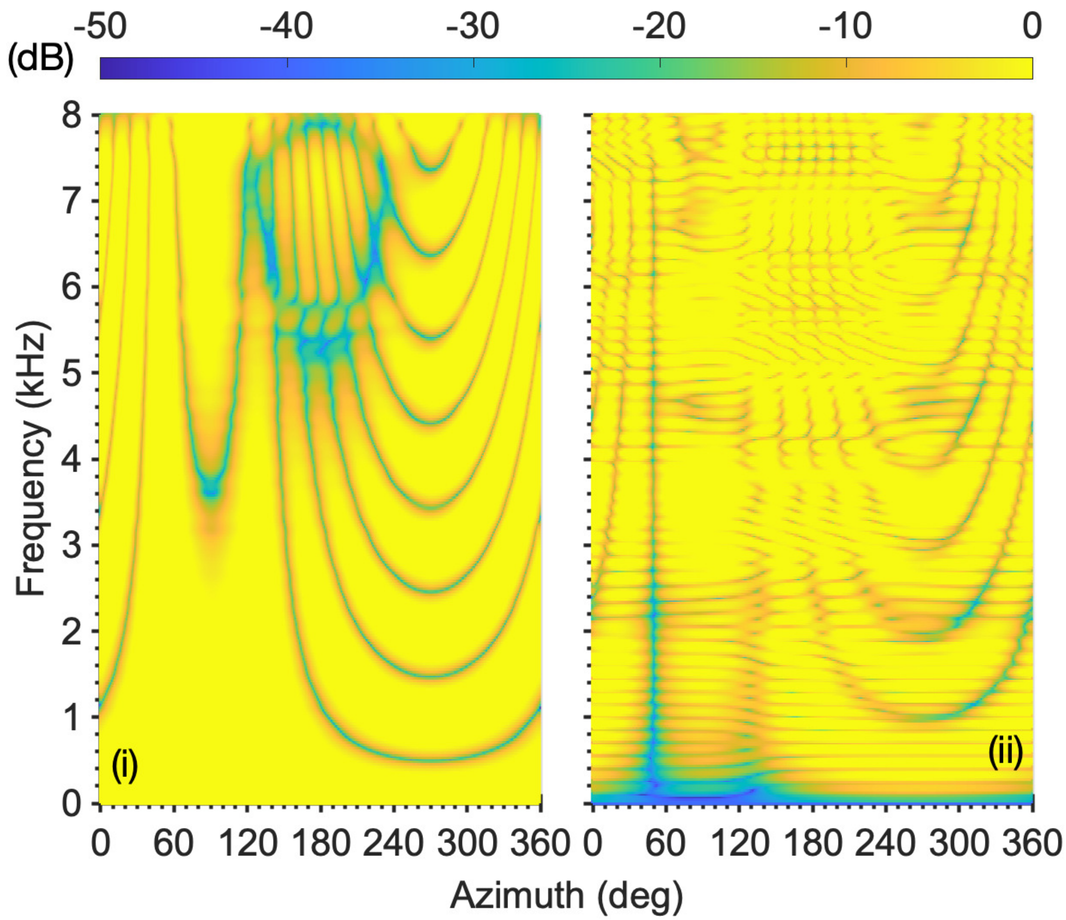


Figure 6.9: PEVD′ filterbank for $\{45°, 50°\}$ trained on 1 white and 1 pink noise sources. (i) The first filterbank $\boldsymbol{u}_1(z)$ has a response that resembles a delay-and-sum beamformer with a look direction at 45°. (ii) The second filterbank $\boldsymbol{u}_2(z)$ is a cancellation beamformer which steers a null at 45°.

Experiment 3: Different Source Signals for Training

A pink, instead of the white, noise signal is used as the second source for training and this trained PEVD filterbank is denoted by PEVD′. Pink noise is used to investigate the effect of training the filterbanks using signals with non-uniform distribution of energy with frequency, i.e., non-flat spectrum. Compared to PEVD in Fig. 6.4, the spatio-spectral characteristics of PEVD′ in Fig. 6.9 are less smooth. The source separation results in Table 6.1 show that PEVD beamformers which are trained on white noise signals targeting only spatial separation perform better than PEVD′. This suggests that using pink noise for training may not be beneficial since $\mathcal{R}_{\mathbf{s}}(z)$ in (6.11) may no longer be a scaled identity, as in the white noise case. Furthermore, the spectrally majorized outputs obtained using PEVD algorithms like the SMD, while guaranteed to be decorrelated in space, time and frequency, are unlikely to be the original source signals since the white and pink sources may not be spectrally majorized, i.e., the 10 dB higher white source signal may have smaller power at some frequencies than the pink source signal.

6.6 Chapter Summary and Conclusions

A fixed beamformer design using PEVD has been proposed. Depending on the source signals used for training, the PEVD can design spatially, and possibly spectrally, orthogonal filterbanks for arbitrary arrays. The array geometry information has been utilized for beam pattern analysis to interpret the eigenvector filterbanks generated by PEVD. For the separation of two source signals, analysis of the beam patterns has shown that the first and second PEVD filterbanks behave like filter-and-sum and cancellation beamformers, respectively. The fixed PEVD filterbanks are lightweight, with an average filter length of 114 in the experiments presented, and perform as well as, if not better than, data-dependent MVDR and LCMV beamformers even for closely spaced sources (5°). Informal listening examples are available [218].

Conclusions

Chapter 7

Conclusions

This chapter concludes the work presented in this thesis, highlights the research findings, and outlines some future research.

7.1 Summary of Thesis Contributions

The aims of the thesis are to explain and develop an understanding of polynomial matrices and PEVD algorithms and to use them for speech processing applications. To this end, the main achievements of this thesis can be broadly categorized as follows.

1. Algorithmic Enhancement to Existing PEVD Methods

Enhancement of SBR2 Using Householder Transformations [29]. The pioneering work of PEVD is the SBR2 algorithm, which uses a sequence of delay and Givens rotation matrices to diagonalize a polynomial matrix. Inspired by numerical linear algebra approaches, Householder reflections have been proposed to reduce the matrix at the principal plane to a tri-diagonal form before applying rotations. In Chapter 3, the proposed method has been shown to achieve faster convergence with an average of 12.35% reduction in iteration counts, lower-order polynomial matrix factors and 0.1% improvement in reconstruction error.

2. New Insights and Theoretical Developments for PEVD

Comparison of Subspaces Generated by EVD and PEVD. The differences between EVD and PEVD are highlighted and explained in Chapter 3, which shows that

PEVD is more suited to multi-channel broadband signal processing. Fundamentally, the subspace generated by PEVD is shift-invariant compared to the usual vector space obtained with EVD. Therefore, PEVD can provide a more compact signal representation using fewer basis components than EVD.

Beamformer Interpretation of Polynomial Eigenvectors [43]. The eigenvector polynomial matrix generated by PEVD is a matrix of stable FIR filters. The filterbank is lossless and cannot change the total power of the subspaces but can only redistribute spectral power among the channels via a filter-and-sum operation. By interpreting this filter-and-sum operation as a beamformer, the beam patterns of the eigenvectors can be computed to analyze array responses in space and frequency, as outlined in Chapter 6.

Comparison of Signal Representation using SHT and PEVD [42]. For a spherical array, the signal representations using SHT, EVD and PEVD are compared. PEVD, on the one hand, achieves close to optimum signal compaction at a high computational cost. On the other hand, SHT performs sub-optimal data compaction but is data-independent and scales well with the number of sensors. It has, therefore, been proposed to combine the SHT and PEVD approaches. In Chapter 5, it has been demonstrated that the proposed framework for signal representation enjoys a greater level of compression, measured by the diagonality factor, compared to microphone signal representations.

3. Speech Processing Applications Using PEVD

Use of Polynomial Matrices for Signal Modelling in Speech Processing Applications [36, 39]. The use of polynomial matrices to model the speech signals received by microphone arrays has first been proposed in [36] and elaborated in Chapter 4. Polynomial matrices are suited to model the naturally broadband characteristics of speech signals. Furthermore, reverberation arising from multi-path reflections in enclosed spaces causes the source signal to undergo a convolutive mixing process. These factors imply that the correlations across different sensors and time lags need to be modelled, leading to space-time covariance polynomial matrices, which are processed by PEVD.

PEVD-based Speech Enhancement in Noisy Reverberant Environments [36–40, 42]. A blind and unsupervised algorithm based on the PEVD has been proposed for speech enhancement. The approach does not rely on array geometry information and works for arbitrary and structured arrays such as linear arrays and spherical arrays, as investigated in Chapters 4 and 5, respectively. The PEVD-based algorithm achieves significant noise reduction, dereverberation and speech enhancement in noisy reverberant environments without introducing any noticeable processing artefacts.

PEVD-based Source Separation Using Informed Spherical Arrays [41, 42]. When the direction of arrival is provided, a beamformer can be computed to target the source signal spatially. In the case of spherical arrays, eigenbeam signals can also be steered in the direction of the target. While the single beamformer output may be designed to have unit gain in the target direction, undesired reverberant components and interfering signals can leak through the sidelobes and reduce separation performance. To improve the quality of the separated signals, PEVD has been proposed to combine outputs from multiple beamformers for source extraction.

Fixed Beamformer Design Using PEVD [43]. A fixed beamformer design using PEVD has been proposed. Depending on the source signals, PEVD can design spatially and possibly spectrally orthogonal filterbanks for arbitrary arrays. For separating two sources using the trained PEVD filterbank designed for the array, beam pattern analysis reveals that they behave like filter-and-sum and cancellation beamformers. The fixed beamformers are lightweight and perform as well as data-dependent MVDR and LCMV beamformers.

7.2 Research Findings

The contributions of this thesis have led to the following research findings.

1. **Algorithmic Enhancement to Existing PEVD Methods.** The proposed enhancement to SBR2 using Householder transformation continues to expand the family of PEVD algorithms in the z-domain. Between SBR2 and SMD, where the zeroing step involves a pair of matrix elements and the entire principal plane, respectively, the proposed approach is in-between by using reflection and rotation matrices. The proposed method behaves like SBR2 when there is only a pair of non-zero off-diagonal elements or only performed on a single row-column pair, and the SMD when the zeroing approach is performed several times to achieve diagonalization on the principal plane. This flexibility in the complexity-performance tradeoff could promote greater adoption of PEVD algorithms for a broader range of applications.

2. **New Insights and Theoretical Developments for PEVD.** Subspace-based approaches using an EVD or KLT have been used to solve many signal processing problems and even in an optimal sense. For multi-channel broadband signal processing, however, the EVD is shown to be inadequate and requires a PEVD for subspace decomposition. The bases generated by the PEVD eigenvectors span a shift-invariant subspace. Consequently, any function can be represented using a weighted combination of the bases and its shifted version. In contrast, the outputs from the EVD eigenvectors span a vector space and any function can be generated using a weighted combination of the outputs. This fundamental result implies that the rank of the subspace decomposition computed by the PEVD is always less than or equal to the EVD, permitting a more compact signal representation with fewer components. This has been shown for spherical arrays but can be generalized to other applications. Another significant outcome of this work is the interpretation of polynomial eigenvector filterbanks as a beamformer. This interpretation allows the analysis and explainability of PEVD algorithms.

3. **Speech Processing Applications Using PEVD.** The instantaneous spatial covariance extension to the space-time covariance matrix is a significant step towards multi-channel broadband signal models. This is necessary to capture the temporal correlations across multiple microphones involving speech signals and convolutive mixing in reverberant environments. The space-time covariance matrix can be modelled as a polynomial matrix, for which a PEVD can be used for broadband subspace decomposition. PEVD-based processing has been adopted for speech enhancement, informed source separation and fixed beamformer design. In addition to its good performance in these applications, the PEVD does not introduce distortions because the z-domain algorithms such as the SBR2 and SMD used are time-domain approaches which preserve spectral coherence. The PEVD approach is also adaptive and self-correcting in some unsupervised applications, where information on the array geometry is not required. Yet, the algorithm can optimally align and combine the signals. These advantages will hopefully motivate wider adoption of PEVD for speech processing applications.

7.3 Future Research Directions

This thesis is devoted to the theoretical and algorithmic development of PMD and, in particular, PEVD and its application to speech and acoustic signal processing. Future work should continue to focus on all three fronts in order to enjoy more widespread adoption. Accordingly, the following suggestions are made.

- **PMD Algorithmic Development.** PEVD algorithms in the frequency-domain are relatively recent [60, 77, 171], and they are fewer than the z-domain family of methods based on the SBR2 and SMD. It will, therefore, be beneficial to have more frequency-domain approaches. For the z-domain approaches such as the SBR2, the computed polynomial eigenvalues tend to be spectrally majorized because of the intrinsic maximization of the trace norm squared of the principal plane [30]. It will also be interesting to develop other constraints in the decomposition to generate different kinds of solutions. The concluding remark from [30] is still very relevant today: can numerical linear algebra approaches be extended to polynomial matrices? Addressing the question directly: the Householder approach in Chapter 3 could be relevant for polynomial QR decomposition.

- **Tensor Decomposition Framework.** Polynomial matrices are represented using 3-dimensional tensors. It will, therefore, be of theoretical interest to observe how PEVD fits into the broader tensor decomposition framework [5] and compare it with existing tensor decomposition techniques such as the multi-linear SVD and canonical polyadic decomposition (CPD).

- **Extension of Strong Uncorrelating Transform:** The SUT is used to decouple complex random variables and their complex conjugates by jointly diagonalizing the instantaneous covariance and complementary covariance. It will be interesting to model the temporal correlations captured by the space-time covariance and its complementary and use PMD to strongly uncorrelate and decorrelate simultaneously.

- **Rank of EVD vs PEVD.** A comparison between EVD and PEVD has been presented in Chapter 3 for an illustrative sinusoidal signal and a broadband pulse signal. More comprehensive proof of the findings for general functions will establish stronger theoretical foundations and understanding of the PEVD.

- **Adaptive Frame-based Processing.** In most approaches, the space-time covariance matrix is estimated while assuming stationarity. It will be of practical interest to relax the stationarity assumption and perform an adaptive estimate which opens up subspace tracking opportunities. A preliminary study is available [219].

- **Array Processing of Arrays.** The eigenbeams and modal beamformer outputs from a spherical microphone array are combined using a PEVD in Chapter 5. It will be interesting to extend the idea of using PEVD as a combiner of different beamformer outputs generated from other array geometries. An initial idea has been proposed [220].

- **Blind Source Separation.** The PEVD-based speech enhancement algorithm presented in Chapter 4 is blind and unsupervised, while the source separation approach in Chapter 5 requires DoA information. Since DoA information is not always available and may be subjected to estimation errors, it will be of practical interest to develop a blind PEVD-based source separation algorithm.

- **Other Speech Processing Applications.** PEVD lends itself well for speech processing applications involving multiple microphones, and this thesis has only presented a small number, with some recent works on voice activity detection (VAD) [221, 222].

Appendix

Appendix A

Rank of EVD vs PEVD

This appendix outlines the proofs for the illustrative examples in Section 3.2.

A.1 Sinusoidal Signal Illustrative Example

For simplicity and without loss of generality, assume that the sinusoidal source with normalized angular frequency Ω arrives at a 2-sensor system with a worst-case phase difference of $\frac{\pi}{2}$. The data vector for sample index n is

$$\mathbf{x}(n) = [\cos(\Omega n), \sin(\Omega n)]^{\mathrm{T}} \ , \tag{A.1}$$

where $[\cdot]^{\mathrm{T}}$ denotes the transpose operator, and Ω is chosen to avoid sampling at the zero-crossings for all n. The instantaneous spatial covariance matrix for $\tau = 0$ is

$$\begin{aligned}
\mathbf{R}(0) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\} &= \begin{bmatrix} \mathbb{E}\{\cos^2(\Omega n)\} & \mathbb{E}\{\cos(\Omega n)\sin(\Omega n)\} \\ \mathbb{E}\{\sin(\Omega n)\cos(\Omega n)\} & \mathbb{E}\{\sin^2(\Omega n)\} \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} \mathbb{E}\{1+\cos(2\Omega n)\} & \mathbb{E}\{\sin(2\Omega n)\} \\ \mathbb{E}\{\sin(2\Omega n)\} & \mathbb{E}\{1-\cos^2(\Omega n)\} \end{bmatrix} \\
&= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \ ,
\end{aligned} \tag{A.2}$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation operator over n that is estimated using the sample mean and is zero-valued for the sinusoid, and $[\cdot]^{\mathrm{H}}$ is the Hermitian transpose operator. Clearly, the rank of (A.2) is 2.

The time delayed version of the data vector is

$$\mathbf{x}(n-\tau) = [\cos(\Omega n - \Omega\tau), \sin(\Omega n - \Omega\tau)]^{\mathrm{T}} \ , \tag{A.3}$$

and the space-time covariance matrix is

$$\begin{aligned}
\mathbf{R}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n-\tau)\} &= \begin{bmatrix} \mathbb{E}\{\cos(\Omega n)\cos(\Omega n - \Omega\tau)\} & \mathbb{E}\{\cos(\Omega n)\sin(\Omega n - \Omega\tau)\} \\ \mathbb{E}\{\sin(\Omega n)\cos(\Omega n - \Omega\tau)\} & \mathbb{E}\{\sin(\Omega n)\sin(\Omega n - \Omega\tau)\} \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} \mathbb{E}\{\cos(\Omega\tau) + \cos(2\Omega n - \Omega\tau)\} & \mathbb{E}\{\sin(2\Omega n - \Omega\tau) - \sin(\Omega\tau)\} \\ \mathbb{E}\{\sin(2\Omega n - \Omega\tau) + \sin(\Omega\tau)\} & \mathbb{E}\{\cos(\Omega\tau) - \cos(2\Omega n - \Omega\tau)\} \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} \cos(\Omega\tau) & -\sin(\Omega\tau) \\ \sin(\Omega\tau) & \cos(\Omega\tau) \end{bmatrix} \ . \qquad (A.4)
\end{aligned}$$

The last step is obtained since the mean value of a sinusoidal sequence over time, n, is 0. When $\tau = 0$, (A.4) simplifies to (A.2), as expected. When $\tau \neq 0$, for example, $\tau = 1$ simplifies (A.4) into

$$\mathbf{R}(1) = \frac{1}{2}\begin{bmatrix} \cos(\Omega) & -\sin(\Omega) \\ \sin(\Omega) & \cos(\Omega) \end{bmatrix} \ , \tag{A.5}$$

which is a Givens rotation, known to be a rank-2 unitary matrix since its determinant is $\cos^2\Omega + \sin^2\Omega = 1 \ \forall\Omega$, as described in Chapter 2. Hence, an EVD, applied to any instantaneous spatial covariance matrix obtained by evaluating (A.4) for a specific value of τ, will give a rank-2 decomposition.

For the PEVD, by applying delays equivalent to a phase shift of $\frac{\pi}{2}$ on the off-diagonals via similarity transforms, (A.4) becomes

$$\mathbf{R}'(\tau) = \frac{1}{2}\begin{bmatrix} \cos(\Omega\tau) & -\sin(\Omega\tau - \frac{\pi}{2}) \\ \sin(\Omega\tau + \frac{\pi}{2}) & \cos(\Omega\tau) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \cos(\Omega\tau) & \cos(\Omega\tau) \\ \cos(\Omega\tau) & \cos(\Omega\tau) \end{bmatrix} \ , \tag{A.6}$$

where $\mathbf{R}'(\tau)$ is the space-time covariance matrix after processing with delay matrices.

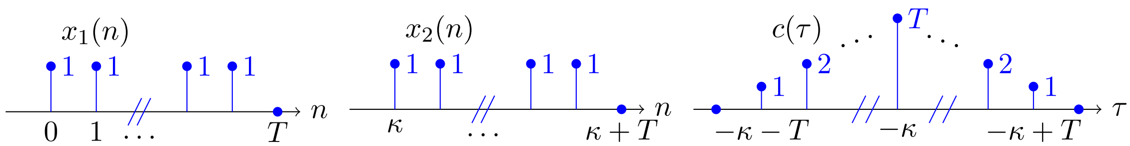


Figure A.1: Visualization of received signals, (i) $x_1(n)$, (ii) $x_2(n)$, and (iii) correlation sequence $c(\tau)$, which becomes the auto-correlation sequence when $\kappa = 0$.

Clearly, (A.6) is rank-1. This implies that applying PEVD to the space-time covariance matrix in (A.4) gives a rank-1 decomposition.

For the sinusoidal example, EVD gives a rank-2 decomposition while PEVD gives a rank-1 decomposition, demonstrating better compaction.

A.2 Broadband Pulse Signal Illustrative Example

Without loss of generality, consider a broadband pulse signal received by a two-sensor array. Let the time support of the rectangular pulse signal $s(n)$ be T, such that

$$s(n) = \begin{cases} 1\,, & 0 \le n < T \\ 0\,, & n \ge T \end{cases} . \tag{A.7}$$

For simplicity, let the first sensor be the reference channel so that the signal defined in (A.7) arrives at the sensor at $n = 0$, as shown in Fig. A.1(i). The auto-correlation sequence of the signal at the reference is a triangular sequence defined as

$$r(\tau) = \begin{cases} \tau + T\,, & \tau < 0 \\ T - \tau\,, & \tau \ge 0 \end{cases} , \tag{A.8}$$

for $\tau = -T+1, \ldots, T-1$, giving a temporal support of $2T - 1$ since $r(-T) = r(T) = 0$. Further, assume that the relative time delay between the two sensors is $\kappa \in \mathbb{Z}$, as shown in Fig. A.1, such that the cross-correlation is

$$c(\tau) = \begin{cases} \tau + \kappa + T & , \quad (-\kappa - T) \leq \tau < -\kappa \\ T - \kappa - \tau & , \quad -\kappa \leq \tau \leq (-\kappa + T) \\ 0 & , \text{otherwise} \end{cases} . \tag{A.9}$$

Observe that (A.9) reduces to (A.8) when $\kappa = 0$, i.e., simplifies to the auto-correlation sequence. Now, the space-time covariance matrix can be written as

$$\mathbf{R}(\tau) = \begin{bmatrix} r(\tau) & c(\tau) \\ c(-\tau) & r(\tau) \end{bmatrix} . \tag{A.10}$$

The first case of interest is when there is no temporal overlap between the received signals, which occurs when the relative time delay is greater than the temporal support, i.e., $\kappa \geq T$. This particular case implies that the last condition in (A.9) is satisfied for $\tau = 0$, $c(0) = 0$. The instantaneous spatial covariance matrix is

$$\mathbf{R}(0) = \begin{bmatrix} r(0) & 0 \\ 0 & r(0) \end{bmatrix} , \tag{A.11}$$

which shows that (A.11) is rank-2, already diagonal and an EVD is not needed. This is expected since the received signals are instantaneously decorrelated, e.g., the inner product between signals equals 0 because there is no time overlap.

The second case occurs when there is some time overlap between the received signals, i.e., $\kappa < T$ so that $c(\tau) \neq 0$. In this case, the covariance matrix in (A.10) is not diagonal at $\tau = 0$. An EVD is applied to factorize $\mathbf{R}(0)$ as

$$\mathbf{R}(0) = \begin{bmatrix} r(0) & c(0) \\ c(0) & r(0) \end{bmatrix} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}} . \tag{A.12}$$

where $\mathbf{U}$ and $\mathbf{\Lambda}$ are the eigenvector and eigenvalue matrices, respectively. An alternative to $\mathbf{R}(0)$ is to evaluate (A.10) at a single time lag τ satisfying $(-\kappa - T) < \tau < (-\kappa + T)$. Performing a z-transform on (A.10) gives

$$\begin{aligned} \mathcal{R}(z) &= \sum_{\tau=-T}^{T} \mathbf{R}(\tau) z^{-\tau} \\ &= \mathbf{R}(-T)z^{T} + \cdots + \mathbf{R}(-1)z^{1} + \mathbf{R}(0) + \mathbf{R}(1)z^{-1} + \cdots + \mathbf{R}(T)z^{-T} \,, \end{aligned} \tag{A.13}$$

which is a para-Hermitian polynomial matrix satisfying $\mathcal{R}(z) = \mathcal{R}^{\mathrm{P}}(z) = \mathcal{R}^{\mathrm{H}}(1/z^{*})$. In the time-domain, this property is equivalent to $\mathbf{R}(\tau) = \mathbf{R}^{\mathrm{H}}(-\tau)$, while noting that the Hermitian symmetric property at each lag is not guaranteed, i.e., $\mathbf{R}(\tau) \neq \mathbf{R}^{\mathrm{H}}(\tau)$. Further, the spatial covariance matrices are generally different at different lags, i.e., $\mathbf{R}(\tau_1) \neq \mathbf{R}(\tau_2), \tau_1 \neq \tau_2, \forall\tau_1, \forall\tau_2$. Applying similarity transforms using $\mathbf{U}$ results in

$$\mathbf{U}^{\mathrm{H}}\mathcal{R}(z)\mathbf{U} = \cdots + \mathbf{U}^{\mathrm{H}}\mathbf{R}(-1)\mathbf{U}z^{1} + \mathbf{U}^{\mathrm{H}}\mathbf{R}(0)\mathbf{U} + \mathbf{U}^{\mathrm{H}}\mathbf{R}(1)\mathbf{U}z^{-1} + \cdots \,, \tag{A.14}$$

which shows that diagonalization can only be achieved for $\mathbf{R}(0)$ on the principal plane in general. Hence, an EVD cannot completely diagonalize (A.14) over all planes of z for the range of τ.

For PEVD, (A.9) can be written as a delayed version of (A.8) so that (A.10) becomes

$$\mathbf{R}(\tau) = \begin{bmatrix} r(\tau) & r(\tau+\kappa) \\ r(\tau-\kappa) & r(\tau) \end{bmatrix} . \tag{A.15}$$

Similar to the sinusoidal example, delays $z^{-\kappa}$ can be applied to the off-diagonal elements using a similarity transform to align the signals so that (A.13) becomes

$$\mathcal{R}'(z) = \begin{bmatrix} 1 & 0 \\ 0 & z^{\kappa} \end{bmatrix} \begin{bmatrix} \mathfrak{r}(z) & \mathfrak{r}(z)z^{\kappa} \\ \mathfrak{r}(z)z^{-\kappa} & \mathfrak{r}(z) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z^{-\kappa} \end{bmatrix} = \begin{bmatrix} \mathfrak{r}(z) & \mathfrak{r}(z) \\ \mathfrak{r}(z) & \mathfrak{r}(z) \end{bmatrix} , \tag{A.16}$$

where $\mathcal{R}'(z)$ is the polynomial matrix after applying delays, and $r(\tau) \circ\!\!-\!\!\bullet\ \mathfrak{r}(z)$ forms a z-transform pair. Clearly, (A.16) has a rank of 1.

By linearly combining signals according to the eigenvector matrix computed using the EVD, the processed outputs, which reside in a vector space, are bases that are only instantaneously orthogonal, i.e. they exhibit correlations at non-zero lags.

The eigenvector polynomial matrix calculated by the PEVD uses time delays and unitary matrices. Hence, the outputs generated by the polynomial eigenvectors are strongly decorrelated bases, and they span a shift-invariant subspace.

For the broadband rectangular pulse example, an EVD cannot completely decorrelate the received signals across a range of time lags. A PEVD, however, can impose strong decorrelation and give a rank-1 decomposition.

Bibliography

[1] H. L. Van Trees, *Optimal Array Processing. Part IV of Detection, Estimation, and Modulation Theory.* John Wiley & Sons, 2002.

[2] S. Haykin, J. Litva, and T. J. Shepherd, *Radar Array Processing*, ser. Springer Series in Inform. Sciences. Springer, 1993.

[3] A. Manikas, *Beamforming Sensor Signal Processing for Defence Applications*, ser. Commun. and Signal Proc. Imperial College Press, 2015, vol. 5.

[4] F. Nekoogar, *Ultra-Wideband Communications: Fundamentals and Applications.* Pearson Edu., 2011.

[5] A. Cichocki, D. Mandic, L. D. Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, “Tensor decompositions for signal processing applications: From two-way to multiway component analysis,” *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Feb. 2015.

[6] S. Makino, *Audio Source Separation*, ser. Signals and Communication Technology. Springer-Verlag, 2018.

[7] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, “Speech enhancement,” ser. Signals and Communication Technology, J. Benesty, S. Makino, and J. Chen, Eds. Berlin, Germany: Springer-Verlag, 2005, pp. 199–228.

[8] N. S. Jayant and P. Noll, *Digital Coding of Waveforms Principles and Applications to Speech and Video.* Prentice Hall, 1984.

[9] H. Wang and M. Kaveh, “Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources,” *IEEE Trans. Acoust.,*

Speech, Signal Process., vol. 33, no. 4, pp. 823–831, Aug. 1985.

[10] M. Moonen and B. De Moor, *SVD and Signal Processing, III: Algorithms, Architectures and Applications.* Elsevier Science Publishers B. V., 1995.

[11] G. Strang, *Linear Algebra and Its Application*, 2nd ed. Academic Press, 1980.

[12] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: John Hopkins University, 1996.

[13] K. Swaminathan and P. P. Vaidyanathan, "Theory and design of uniform DFT, parallel, quadrature mirror filter banks," *IEEE Trans. Circuits Syst.*, vol. 33, no. 12, pp. 1170–11 941, Dec. 1986.

[14] A. Rao and R. Kumaresan, "On decomposing speech into modulated components," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 240–254, May 2000.

[15] W. Kellermann and H. Buchner, "Wideband algorithms versus narrowband algorithms for adaptive filtering in the DFT domain," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, 2003, pp. 1–5.

[16] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay [FIR/all pass filters design]," *IEEE Signal Process. Mag.*, vol. 13, no. 1, pp. 30–60, Jan. 1996.

[17] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[18] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding.* Upper Saddle River, NJ, USA: Prentice Hall, 1995.

[19] P. C. Hansen and S. H. Jensen, "Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: Survey and analysis," *EURASIP J. on Advances in Signal Process.*, vol. 2007, no. 1, pp. 1–24, Jan. 2007.

[20] P. Stoica and R. Moses, *Spectral Analysis of Signals.* New Jersey, USA: Prentice Hall, 2005.

[21] G. Chabriel, M. Kleinsteuber, E. Moreau, and H. Shen, "Joint matrices decompositions and blind source separation," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 34–43, May 2014.

[22] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.

[23] M. W. Shan, T. and T. Kailath, "On spatial smoothing for direction-of-arrival estimation in coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 8, pp. 806–811, Aug. 1985.

[24] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation.* Springer-Verlag, 2010.

[25] S. Redif, S. Weiss, and J. G. McWhirter, "An approximate polynomial matrix eigenvalue decomposition algorithm for para-hermitian matrices," in *Proc. Int. Symp. on Signal Process. and Inform. Technol. (ISSPIT)*, 2011, pp. 421–425.

[26] J. Corr, K. Thompson, S. Weiss, J. G. McWhirter, S. Redif, and I. K. Proudler, "Multiple shift maximum element sequential matrix diagonalisation for para-Hermitian matrices," in *Proc. IEEE/SP Workshop on Statistical Signal Process.*, 2014, pp. 844–848.

[27] Z. Wang, J. G. McWhirter, and S. Weiss, "Multichannel spectral factorization algorithm using polynomial matrix eigenvalue decomposition," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, 2015, pp. 1714–1718.

[28] S. Redif, S. Weiss, and J. G. McWhirter, "Sequential matrix diagonalisation algorithms for polynomial EVD of para-Hermitian matrices," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 81–89, Jan. 2015.

[29] V. W. Neo and P. A. Naylor, "Second order sequential best rotation algorithm with

Householder transformation for polynomial matrix eigenvalue decomposition," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2019, pp. 8043–8047.

[30] J. G. McWhirter, P. D. Baxter, T. Cooper, S. Redif, and J. Foster, "An EVD algorithm for para-Hermitian polynomial matrices," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2158–2169, May 2007.

[31] S. Weiss, S. Bendoukha, A. Alzin, F. K. Coutts, I. K. Proudler, and J. Chambers, "MVDR broadband beamforming using polynomial matrix techniques," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 839–843.

[32] J. Corr, J. Pestana, S. Weiss, I. K. Proudler, S. Redif, and M. Moonen, "Investigation of a polynomial matrix generalised EVD for multi-channel Wiener filtering," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, 2016, pp. 1354–1358.

[33] S. Redif, S. Weiss, and J. G. McWhirter, "Relevance of polynomial matrix decompositions to broadband blind signal separation," *Signal Process.*, vol. 134, pp. 76–86, May 2017.

[34] S. Weiss, N. J. Goddard, S. Somasundaram, I. K. Proudler, and P. A. Naylor, "Identification of broadband source-array responses from sensor second order statistics," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2017, pp. 1–5.

[35] S. Weiss, S. Redif, T. Cooper, C. Liu, P. D. Baxter, and J. G. McWhirter, "Paraunitary oversampled filter bank design for channel coding," *EURASIP J. on Advances in Signal Process.*, vol. 2006, no. 1, pp. 1–10, Mar. 2006.

[36] V. W. Neo, C. Evers, and P. A. Naylor, "Speech enhancement using polynomial eigenvalue decomposition," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2019, pp. 125–129.

[37] V. W. Neo, C. Evers, and P. A. Naylor, "PEVD-based speech enhancement in re-

verberant environments," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2020, pp. 186–190.

[38] V. W. Neo, C. Evers, and P. A. Naylor, "Speech dereverberation performance of a polynomial-EVD subspace approach," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2020, pp. 221–225.

[39] V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 3255–3266, Oct. 2021.

[40] V. W. Neo, C. Evers, and P. A. Naylor, "Polynomial matrix eigenvalue decomposition of spherical harmonics for speech enhancement," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Jun. 2021, pp. 786–790.

[41] V. W. Neo, C. Evers, and P. A. Naylor, "Polynomial matrix eigenvalue decomposition-based source separation using informed spherical microphone arrays," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, Oct. 2021, pp. 201–205.

[42] V. W. Neo, C. Evers, S. Weiss, and P. A. Naylor, "Signal compaction using polynomial EVD for spherical array processing with applications," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 2022 (submitted).

[43] V. W. Neo, E. d'Olne, A. H. Moore, and P. A. Naylor, "Fixed beamformer design using polynomial eigenvalue decomposition," in *Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022, pp. 1–5.

[44] T. Kailath, *Linear Systems*, 1st ed. Englewood Cliffs, NJ 07632, USA, 1980.

[45] P. P. Vaidyanathan, *Multirate Systems and Filters Banks*, 1st ed. New Jersey, USA: Prentice Hall, 1993.

[46] T. Gao, T. Q. Nguyen, and G. Strang, "On factorization of M-channel paraunitary

filterbanks," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1433–1446, Jul. 2001.

[47] S. Icart and P. Comon, "Some properties of Laurent polynomial matrices," in *IMA Int. Conf. on Math. in Signal Process.*, Dec. 2012, pp. 1–4.

[48] S. Weiss, J. Pestana, and I. K. Proudler, "On the existence and uniqueness of the eigenvalue decomposition of a para-Hermitian matrix," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2659–2672, May 2018.

[49] S. Weiss, J. Pestana, I. K. Proudler, and F. K. Coutts, "Corrections to "On the existence and uniqueness of the eigenvalue decomposition of a para-Hermitian matrix"," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6325–6327, Dec. 2018.

[50] F. Rellich, "Störungstheorie der Spektralzerlegung. I. Mitteilung. Analytische Störung der isolierten Punkteigenwerte eines beschränkten Operators," *Mathematische Annalen*, vol. 113, pp. DC–DCXIX, 1937. [Online]. Available: https://eudml.org/doc/159886

[51] M. Davies, S. Lambotharan, J. Foster, J. Chambers, and J. G. McWhirter, "Polynomial matrix QR decomposition and iterative decoding of frequency selective MIMO channels," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, 2009, pp. 1–6.

[52] J. Foster, J. G. McWhirter, M. R. Davies, and J. A. Chambers, "An algorithm for calculating the QR and singular value decompositions of polynomial matrices," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1263–1274, Mar. 2010.

[53] Z. Wang, J. G. McWhirter, J. Corr, and S. Weiss, "Multiple shift second order sequential best rotation algorithm for polynomial matrix EVD," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 844–848.

[54] Z. Wang, J. G. McWhirter, J. Corr, and S. Weiss, "Order-controlled multiple shift SBR2 algorithm for para-Hermitian polynomial matrices," in *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, 2016, pp. 1–5.

[55] A. Tkacenko, "Approximate eigenvalue decomposition of para-Hermitian systems through successive FIR paraunitary transformations," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2010, pp. 4074–4077.

[56] A. Tkacenko and P. P. Vaidyanathan, "On the spectral factor ambiguity of FIR energy compaction filter banks," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 380–385, Jan. 2006.

[57] A. Tkacenko and P. P. Vaidyanathan, "Iterative greedy algorithm for solving the FIR paraunitary approximation problem," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 146–160, Jan. 2006.

[58] M. Tohidian, H. Amindavar, and A. M. Reza, "A DFT-based approximate eigenvalue and singular value decomposition of polynomial matrices," *EURASIP J. on Appl. Signal Process.*, vol. 1, no. 93, pp. 1–16, Apr. 2013.

[59] F. K. Coutts, K. Thompson, I. K. Proudler, and S. Weiss, "An iterative DFT-based approach to the polynomial matrix eigenvalue decomposition," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, 2018, pp. 1011–1015.

[60] S. Weiss, I. K. Proudler, and F. K. Coutts, "Eigenvalue decomposition of a para-Hermitian matrix: Extraction of analytic eigenvalues," *IEEE Trans. Signal Process.*, vol. 69, pp. 722–737, 2021.

[61] D. Hassan, S. Redif, and S. Lambotharan, "Polynomial matrix decompositions and semi-blind channel estimation for MIMO frequency selective channels," *Proc. IET Signal Process.*, vol. 13, no. 3, pp. 356–366, May 2019.

[62] D. Hassan, S. Redif, and S. Lambotharan, "Polynomial GSVD beamforming for two-user frequency-selective MIMO channels," *IEEE Trans. Signal Process.*, vol. 69, pp. 948–959, 2021.

[63] R. Brandt and M. Bengtsson, "Wideband MIMO channel diagonalization in the time

domain," in *Proc. Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, 2011, pp. 1958–1962.

[64] S. Redif, J. G. McWhirter, and S. Weiss, "Design of FIR paraunitary filter banks for subband coding using a polynomial eigenvalue decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5253–5264, Nov. 2011.

[65] S. Redif, "Convolutive blind signal separation via polynomial matrix generalized eigenvalue decomposition," *IET Electron. Lett.*, vol. 53, no. 2, pp. 87–89, 2017.

[66] S. Redif, J. Pestana, and I. K. Proudler, "Analysis of broadband GEVD-based source separation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2019, pp. 8028–8032.

[67] S. Weiss, M. Alrmah, S. Lambotharan, J. G. McWhirter, and M. Kaveh, "Broadband angle of arrival estimation methods in a polynomial matrix decomposition framework," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process. (CAMSAP)*, 2013, pp. 109–112.

[68] A. O. T. Hogg, V. W. Neo, S. Weiss, C. Evers, and P. A. Naylor, "A polynomial eigenvalue decomposition MUSIC approach for broadband sound source localization," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, Oct. 2021, pp. 326–330.

[69] W. Coventry, C. Clemente, and J. Soraghan, "Enhancing polynomial MUSIC algorithm for coherent broadband sources through spatial smoothing," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2017, pp. 2448–2452.

[70] J. Benesty, M. M. Sondhi, and Y. Huang, Eds., *Springer Handbook of Speech Processing*, 1st ed. Springer-Verlag, 2008.

[71] J. Foster, "Algorithms and techniques for polynomial matrix decompositions," Ph.D. dissertation, Cardiff University, United Kingdom, 2008.

[72] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, 1st ed. Prentice Hall, 2000.

[73] J. Corr, K. Thompson, S. Weiss, I. Proudler, and J. G. McWhirter, "Shortening of paraunitary matrices obtained by polynomial eigenvalue decomposition algorithms," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2015, pp. 1–5.

[74] J. Foster, J. G. McWhirter, and J. Chambers, "Limiting the order of polynomial matrices within the SBR2 algorithm," in *IMA Int. Conf. on Math. in Signal Process.*, 2006, pp. 1–4.

[75] F. K. Coutts, K. Thompson, I. K. Proudler, and S. Weiss, "A comparison of iterative and DFT-based polynomial matrix eigenvalue decomposition," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process. (CAM-SAP)*, 2017, pp. 1–5.

[76] F. K. Coutts, K. Thompson, J. Pestana, I. K. Proudler, and S. Weiss, "Enforcing eigenvector smoothness for a compact DFT-based polynomial eigenvalue decomposition," in *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, 2018, pp. 159–163.

[77] S. Weiss, I. K. Proudler, F. K. Coutts, and J. Deeks, "Extraction of analytic eigenvectors from para-Hermitian matrix," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2020, pp. 1–5.

[78] A. Bunse-Gerstner, R. Byers, V. Mehrmann, and N. K. Nichols, "Numerical computation of an analytic singular value decomposition of a matrix valued function," *Numerische Mathematik*, vol. 60, pp. 1–39, 1991.

[79] A. G. J. Macfarlane and Y. S. Hung, "Analytic properties of the singular values of a rational matrix," *Int. J. Control*, vol. 37, no. 2, pp. 221–234, 1983.

[80] B. L. R. De Moor and S. Boyd, "Analytic properties of singular values and vectors,"

Katholieke Universiteit Leuven, Tech. Report 1989-28, 1989.

[81] I. Gohberg, P. Lancaster, and L. Rodamn, *Matrix Polynomials*, 2nd ed. SIAM, 2009.

[82] F. R. Gantmacher, *The Theory of Matrices*. Chelsea Publishing Company, New York, USA, 1959, vol. 1.

[83] J. W. Brown and R. V. Churchill, *Complex Variables and Applications*, 9th ed. McGraw-Hill, 2014.

[84] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. John Wiley & Sons, Inc., 2009.

[85] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing*, 2nd ed., ser. signal processing. Upper Saddle River, New Jersey, USA: Prentice Hall, 1999.

[86] S. Mitra, *Digital Signal Processing: A Computer-based Approach*, 4th ed. McGraw-Hill, 2011.

[87] J. Proakis and D. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*. Upper Saddle River, NJ, USA: Prentice Hall, 1996.

[88] E. M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*. Princeton University Press, 2003.

[89] A. J. Hoffman and H. W. Wielandt, “The variation of the spectrum of a normal matrix,” *Duke Math J.*, vol. 20, no. 1, pp. 37–39, Mar. 1953.

[90] P. P. Vaidyanathan, “Theory of optimal orthonormal subband coders,” *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1528–1543, Jun. 1998.

[91] G. W. Stewart, “On the early history of the singular value decomposition,” *SIAM Rev.*, vol. 35, no. 4, pp. 551–566, Dec. 1993.

[92] G. Peters and J. H. Wilkinson, "Ax= λBx and the generalized eigenproblem," *SIAM J. Num. Anal.*, vol. 7, no. 4, pp. 479–492, Dec. 1970.

[93] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data The Theory of Improper and Noncircular Signals*, 1st ed. Cambridge University Press, 2010.

[94] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[95] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[96] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008.

[97] C. Cheong Took, S. C. Douglas, and D. P. Mandic, "On approximate diagonalization of correlation matrices in widely linear signal processing," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1469–1473, Mar. 2012.

[98] G. Okopal, S. Wisdom, and L. Atlas, "Speech analysis with the strong uncorrelating transform," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 11, pp. 1858–1867, Nov. 2015.

[99] S. Weiss, A. P. Millar, and R. W. Stewart, "Inversion of para-Hermitian matrices," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2010, pp. 447–451.

[100] M. A. Alrmah, S. Weiss, and S. Lambotharan, "An extension of the MUSIC algorithm to broadband scenarios using a polynomial eigenvalue decomposition," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2011, pp. 629–633.

[101] A. W. Khong, P. A. Naylor, and J. Benesty, "A low delay and fast converging improved proportionate algorithm for sparse system identification," *EURASIP J.*

on Audio, Speech, and Music Process., vol. 2007, no. 1, pp. 1–8, Apr. 2007.

[102] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, 1st ed. Oxford, UK: Clarendon Press, 1965.

[103] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 3rd ed. Cambridge University Press, 2007.

[104] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, 2nd ed. Wellesley-Cambridge Press, 1997.

[105] A. Aldroubi and M. Unser, "Sampling procedures in function spaces and asymptotic equivalence with Shannon sampling theorem," *Numer. Funct. Anal. Optimiz.*, vol. 15, pp. 1–21, 1994.

[106] W. Chen and S. Itoh, "A sampling theorem for shift-invariant subspace," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2822–2824, Oct. 1998.

[107] A. H. Sayed, "Unitary transformations," in *Adaptive Filters.* Wiley-IEEE Press, 2008.

[108] B. Widrow and F.-L. Luo, "Microphone arrays for hearing aids: An overview," *Speech Commun.*, vol. 39, pp. 139–146, 2003.

[109] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. J. R. Liu, Eds. John Wiley & Sons, Inc., 2010, pp. 269–302.

[110] A. Kuklasinski and J. Jensen, "Multichannel Wiener filters in binaural and bilateral hearing aids—Speech intelligibility improvement and robustness to DoA errors," *J. Audio Eng. Soc. (AES)*, vol. 65, no. 1/2, pp. 8–16, 2017.

[111] C. Evers, Y. Dorfan, S. Gannot, and P. A. Naylor, "Source tracking using moving

microphone arrays for robot audition," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 6145–6149.

[112] C. Evers and P. A. Naylor, "Acoustic SLAM," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 9, pp. 1484–1498, Sep. 2018.

[113] Y. Huang, J. Benesty, and J. Chen, "Analysis and comparison of multichannel noise reduction methods in a common framework," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 5, pp. 957–968, Jul. 2008.

[114] E. A. P. Habets and P. A. Naylor, "Dereverberation," in *Audio Source Separation and Speech Enhancement*, E. Vincent, T. Virtanen, and S. Gannot, Eds. John Wiley & Sons, Inc., 2018, pp. 317–343.

[115] T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement systems," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2009, pp. 4409–4412.

[116] M. Torcoli, "An improved measure of musical noise based on spectral kurtosis," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, 2019, pp. 90–94.

[117] K. Tan and D. Wang, "Improving robustness of deep learning based monaural speech enhancement against processing artifacts," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2020, pp. 6914–6918.

[118] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[119] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[120] C. Uhle, M. Torcoli, and J. Paulus, “Controlling the perceived sound quality for dialogue enhancement with deep learning,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2020, pp. 51–55.

[121] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[122] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[123] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, 1985.

[124] J. Chen, J. Benesty, Y. Huang, and S. Doclo, “New insights into the noise reduction Wiener filter,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1218–1234, Jul. 2006.

[125] Y. Ephraim and H. L. Van Trees, “A signal subspace approach for speech enhancement,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[126] Y. Hu and P. C. Loizou, “A subspace approach for enhancing speech corrupted by colored noise,” *IEEE Signal Process. Lett.*, vol. 9, no. 7, pp. 204–206, Jul. 2002.

[127] F. Jabloun and B. Champagne, “A multi-microphone signal subspace approach for speech enhancement,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2001, pp. 205–208.

[128] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, “Speech enhancement based on the subspace method,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 497–507, 2000.

[129] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. Signal Process.*,

vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[130] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing.* Berlin, Germany: Springer-Verlag, 2008.

[131] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, ser. Springer Topics in Signal Processing. Springer, 2017.

[132] I. Cohen and B. Berdugo, "Microphone array post-filtering for non-stationary noise suppression," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Orlando, Florida, USA, May 2002, pp. 901–904.

[133] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 39–60.

[134] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum Likelihood PSD Estimation for Speech Enhancement in Reverberation and Noise," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1599–1612, Sep. 2016.

[135] W. Xue, A. H. Moore, M. Brookes, and P. A. Naylor, "Modulation-Domain Multichannel Kalman Filtering for Speech Enhancement," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 10, pp. 1833–1847, Oct. 2018.

[136] T. Dietzen, S. Doclo, M. Moonen, and T. van Waterschoot, "Integrated sidelobe cancellation and linear prediction Kalman filter for joint multi-microphone speech dereverberation, interfering speech cancellation, and noise reduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 740–754, Jan. 2020.

[137] J. B. Allen, "Synthesis of pure speech from a reverberant signal," U.S. Patent 3 786 188, 1974.

[138] N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "On the use of linear prediction

for dereverberation of speech," in *Proc. Int. on Workshop Acoust. Echo and Noise Control (IWAENC)*, 2003, pp. 99–102.

[139] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, pp. 1127–1138, Aug. 2002.

[140] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[141] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.

[142] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, Jul. 1979.

[143] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[144] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Int. on Workshop Acoust. Echo and Noise Control (IWAENC)*, Aug. 2010, pp. 1–4.

[145] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 9, pp. 1879–1890, Sep. 2013.

[146] F. Lim, W. Zhang, E. A. P. Habets, and P. A. Naylor, "Robust multichannel dereverberation using relaxed multichannel least squares," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 9, pp. 1379–1390, Sep. 2014.

[147] T. Yoshioka and T. Nakatani, "Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening," *IEEE Trans. Audio, Speech,*

Language Process., vol. 20, no. 10, pp. 2707–2720, Dec. 2012.

[148] M. Delcroix, T. Yoshioka, A. Ogawa, Y. Kubo, M. Fujimoto, N. Ito, K. Kinoshita, M. Espi, S. Araki, T. Hori, and T. Nakatani, "Strategies for distant speech recognitionin reverberant environments," *EURASIP J. on Advances in Signal Process.*, vol. 2015, no. 1, pp. 1–15, 2015.

[149] T. Nakatani and K. Kinoshita, "A unified convolutional beamformer for simultaneous denoising and dereverberation," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 903–907, Jun. 2019.

[150] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech de-reverberation," *Acta Acoustica*, vol. 87, pp. 359–366, 2001.

[151] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.*, vol. 113, no. 6, pp. 3233–3244, Jun. 2003.

[152] I. Kodrasi and S. Doclo, "Analysis of eigenvalue decomposition-based late reverberation power spectral density estimation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 6, pp. 1106–1118, Jun. 2018.

[153] P. D. Baxter and J. G. McWhirter, "Blind signal separation of convolutive mixtures," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, 2003, pp. 124–128.

[154] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium (LDC), Philadelphia, USA, Corpus LDC93S1, 1993.

[155] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of room acoustic parameters: the ACE challenge," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 10, pp. 1681–1693, Oct. 2016.

[156] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition II: NOISEX-92: A database and an experiment to study the effect of additive noise on

speech recognition systems," *Speech Commun.*, vol. 3, no. 3, pp. 247–251, Jul. 1993.

[157] S. Ideas, "International Sound Effects Library," Richmond Hill, Ont, 1999.

[158] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Am.*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.

[159] "Objective measurement of active speech level," Int. Telecommun. Union (ITU-T), Geneva, Switzerland, Recommendation P.56, Mar. 1993.

[160] D. M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," 1997. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[161] V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using PEVD," Oct. 2020. [Online]. Available: https://vwn09.github.io/pevd-enhance/

[162] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proc. Int. Conf. on Spoken Language Process. (ICSLP)*, vol. 7, 1998, pp. 2819–2822.

[163] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[164] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," in *Proc. Conf. of Int. Speech Commun. Assoc. (INTERSPEECH)*, 2006, pp. 1447–1450.

[165] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

[166] ITU-T, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” Int. Telecommun. Union (ITU-T), Recommendation P.862, Nov. 2003.

[167] P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets, “Signal-based performance evaluation of dereverberation algorithms,” *J. of Elect. and Comput. Eng.*, vol. 2010, pp. 1–5, 2010.

[168] S. Akkarakaran and P. P. Vaidyanathan, “On optimization of filter banks with denoising applications,” in *Proc. Int. Symp. on Circuits and Syst.*, Jul. 1999, pp. 512–515.

[169] S. J. Campanella and G. S. Robinson, “A comparison of orthogonal transformations for digital speech processing,” *IEEE Trans. Commun. Technol.*, vol. 19, no. 6, pp. 1045–1050, Dec. 1971.

[170] M. K. Tsatsanis and G. B. Giannakis, “Principal component filter banks for optimal multiresolution analysis,” *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1766–1777, Aug. 1995.

[171] S. Weiss, I. K. Proudler, F. K. Coutts, and J. Pestana, “Iterative approximation of analytic eigenvalues of a para-Hermitian matrix EVD,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2019, pp. 8038–8042.

[172] F. K. Coutts, J. Corr, K. Thompson, S. Weiss, I. K. Proudler, and J. G. McWhirter, “Memory and complexity reduction in para-Hermitian matrix manipulations of PEVD algorithms,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2016, pp. 1633–1637.

[173] B. Rafaely, *Fundamentals of Spherical Array Processing*, ser. Springer Topics in Signal Processing. Springer, 2015.

[174] N. Epain and C. T. Jin, “Spherical harmonic signal covariance and sound field

diffuseness," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 10, pp. 1796–1807, Oct. 2016.

[175] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3094–3103, Nov. 2005.

[176] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, 2015.

[177] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2002, pp. 1781–1784.

[178] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2002, pp. 1949–1952.

[179] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[180] G. W. Elko and J. Meyer, "Adaptive beamformer for spherical eigenbeamforming microphone arrays," in *Proc. Joint Workshop on Hands-free Speech Commun. and Microphone Arrays (HSCMA)*, 2014, pp. 52–56.

[181] Y. Peled and B. Rafaely, "Linearly-constrained minimum-variance method for spherical microphone arrays based on plane-wave decomposition of the sound field," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 21, no. 12, pp. 2532–2540, Dec. 2013.

[182] S. Braun, D. P. Jarrett, J. Fischer, and E. A. P. Habets, "An informed spatial filter for dereverberation in the spherical harmonic domain," in *Proc. IEEE Int. Conf. on*

Acoust., Speech and Signal Process. (ICASSP), May 2013, pp. 669–673.

[183] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2010, pp. 442–446.

[184] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "Eigenbeam-based acoustic source tracking in noisy reverberant environments," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, Nov. 2010, pp. 576–580.

[185] A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace psuedointensity vectors," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 1, pp. 178–192, Jan. 2017. [Online]. Available: http://dx.doi.org/10.1109/TASLP.2016.2613280

[186] J. Meyer and G. W. Elko, "Spherical microphone arrays for 3D sound recording," in *Audio Signal Processing For Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Kluwer Academic Publisher, 2004, pp. 67–89.

[187] S. Weiss and I. K. Proudler, "Comparing efficient broadband beamforming architectures and their performance trade-offs," in *Proc. IEEE Int. Conf. Digital Signal Process. (DSP)*, Jul. 2002, pp. 417–424.

[188] Z. Wang and J. G. McWhirter, "A novel insight to the SBR2 algorithm for diagonalising para-Hermitian matrices," in *IMA Int. Conf. on Math. in Signal Process.*, Dec. 2016, pp. 1–4.

[189] F. K. Coutts, I. K. Proudler, and S. Weiss, "Efficient implementation of iterative polynomial matrix EVD algorithms exploiting structural redundancy and parallelisation," *IEEE Trans. Circuits Syst. I*, vol. 66, no. 12, pp. 4753–4766, Dec. 2019.

[190] C. Hold, S. J. Schlecht, A. Politis, and V. Pulkki, "Spatial filter bank in the spherical harmonic: reconstruction and application," in *Proc. IEEE Workshop on Appl. of*

Signal Process. to Audio and Acoust. (WASPAA), New Paltz, NY, Oct. 2021, pp. 361–365.

[191] T. Deppisch, J. Ahrens, S. V. A. Garí, and P. Calamia, "Spatial subtraction of reflections from room impulse responses measured with a spherical microphone array," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, Oct. 2021, pp. 346–350.

[192] C. Delaosa, J. Pestana, N. J. Goddard, S. Somasundaram, and S. Weiss, "Sample space-time covariance matrix estimation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2019, pp. 8033–8037.

[193] C. Delaosa, J. Pestana, N. J. Goddard, S. Somasundaram, and S. Weiss, "Support estimation of a sample space-time covariance matrix," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2019, pp. 1–5.

[194] C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge: acoustic source localization and tracking," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1620–1643, Apr. 2020.

[195] V. W. Neo, "PEVD exploiting spherical microphone array processing," Apr. 2021. [Online]. Available: https://vwn09.github.io/pevd-smap

[196] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: Algorithm and applications," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sep.

[197] mh acoustics, "EM32 Eigenmike microphone array release notes (v17.0)," M. H. Acoust., NJ USA, Hardware, Oct. 2013. [Online]. Available: http://www.mhacoustics.com/sites/default/files/ReleaseNotes.pdf

[198] S. Brown and D. Sen, "Error analysis of spherical harmonic soundfield represen-

tations in terms of truncation and aliasing errors," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2013, pp. 360–364.

[199] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2011, pp. 189–192.

[200] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, "Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2019, pp. 1–5.

[201] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulation and array processing algorithms," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2018, pp. 351–355.

[202] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[203] J. Selva, "An efficient structure for the design of variable fractional delay filters based on the windowing method," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3770–3775, Aug. 2008.

[204] D. Sharma, R. Gong, J. Fosburgh, S. Y. Kruchinin, P. A. Naylor, and L. Milanović, "Spatial processing front-end for distant ASR exploiting self-attention channel combinator," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2022, pp. 7997–8001.

[205] A. H. Moore, L. Lightburn, W. Xue, P. A. Naylor, and M. Brookes, "Binaural mask-informed speech enhancement for hearing aids with head tracking," in *Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Tokyo, Japan, Sep. 2018, pp. 461–465.

[206] B. D. van Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[207] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications.* Berlin, Germany: Springer-Verlag, 2001.

[208] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing.* Wiley-IEEE Press, 2018.

[209] S. Sun, S. Zhou, Hwang, L. Xie, Q. Li, and X. Lei, “Multiple fixed beamformers with a spacial Wiener-form postfilter for far-field speech recognition,” in *Asia-Pacific Signal and Inform. Process. Assoc. Annual Summit and Conf. (APSIPA)*, 2019, pp. 633–637.

[210] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, “Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction,” *Speech Commun.*, vol. 49, no. 7–8, pp. 636–656, Aug. 2007.

[211] E. d’Olne, A. H. Moore, and P. A. Naylor, “Model-based beamforming for wearable microphone arrays,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 1105–1109.

[212] I. Cohen, S. Gannot, and B. Berdugo, “An integrated real-time beamforming and post filtering system for nonstationary noise environments,” *EURASIP J. on Appl. Signal Process.*, vol. 11, pp. 1064–1073, 2003.

[213] M. Alrmah, S. Weiss, S. Redif, S. Lambotharan, and J. G. McWhirter, “Angle of arrival estimation for broadband signals: a comparison,” in *Proc. IET Int. Conf. on Intelligent Signal Process.*, Jan. 2013, pp. 1–6.

[214] A. Alzin, F. K. Coutts, J. Corr, S. Weiss, I. K. Proudler, and J. A. Chambers, “Adaptive broadband beamforming with arbitrary array geometry,” in *Proc. IET*

Int. Conf. on Intelligent Signal Process., Dec. 2015, pp. 1–6.

[215] A. H. Moore, "Microphone array classes in the ELOBES project," Jan. 2020. [Online]. Available: https://github.com/ImperialCollegeLondon/sap-elobes-microphone-arrays

[216] E. A. P. Habets, "Room impulse response generator," Technische Universiteit Eindhoven (TU/e), Tech. Rep., 2006.

[217] R. E. Crochiere, "A weighted overlap-add method for short-time Fourier analysis/Synthesis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 1, pp. 99–102, Feb. 1980.

[218] V. W. Neo, "Fixed beamformer design using PEVD," May 2022. [Online]. Available: https://vwn09.github.io/research/pevd-beamformer-iwaenc

[219] E. d'Olne, V. W. Neo, and P. A. Naylor, "Frame-based space-time covariance matrix estimation for polynomial eigenvalue decomposition-based speech enhancement," in *Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022, pp. 1–5.

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

[221] V. W. Neo, S. Weiss, and P. A. Naylor, "A polynomial subspace projection approach for the detection of weak voice activity," in *Sensor Signal Process. for Defence Conf. (SSPD)*, Sep. 2022, pp. 81–85.

[222] V. W. Neo, S. Weiss, S. W. McKnight, A. O. T. Hogg, and P. A. Naylor, "Polynomial eigenvalue decomposition-based target speaker voice activity detection in the presence of competing talkers," in *Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022, pp. 1–5.