

國立交通大學

電控工程研究所

博士論文

基於多秩訊號模型之語音純化
波束形成與多通道後濾波



**Beamforming and Multi-Channel Post-Filtering
for Speech Enhancement Based on
Multi-Rank Signal Models**

研究生：李明唐

指導教授：胡竹生 教授

中華民國一百零二年九月

基於多秩訊號模型之語音純化
波束形成與多通道後濾波

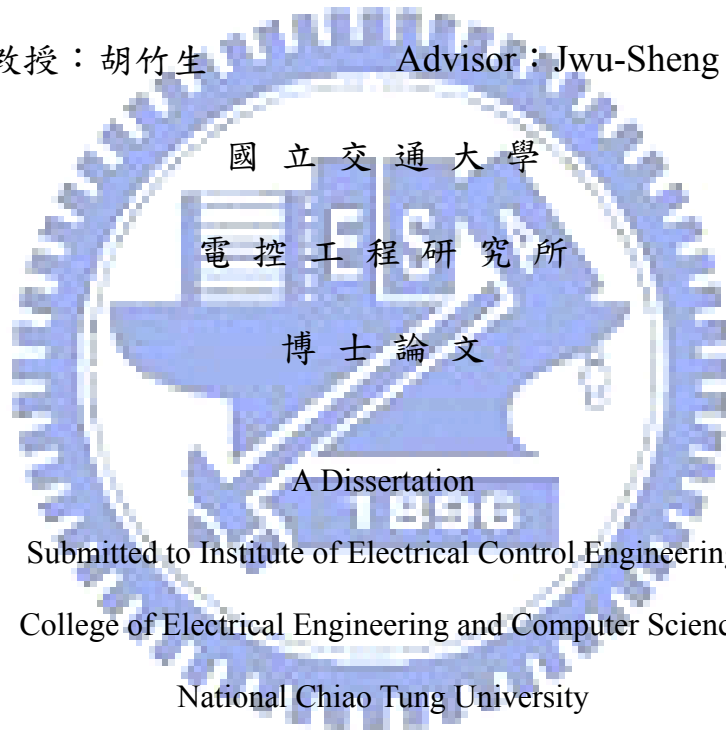
Beamforming and Multi-Channel Post-Filtering for
Speech Enhancement Based on Multi-Rank Signal Models

研究生：李明唐

Student : Ming-Tang Lee

指導教授：胡竹生

Advisor : Jwu-Sheng Hu



Submitted to Institute of Electrical Control Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Institute of Electrical Control Engineering

September 2013

Hsinchu, Taiwan, Republic of China

中華民國一百零二年九月

基於多秩訊號模型之語音純化

波束形成與多通道後濾波

研究生：李明唐

指導教授：胡竹生 博士

國立交通大學電控工程研究所博士班

摘要

波束形成(Beamforming)與多通道後濾波(multi-channel post-filtering)為麥克風陣列語音純化應用之兩項主要技術。在傳統的演算法中，聲源相對於麥克風間的空間資訊通常是以兩兩麥克風之間的延遲或是相對轉移函數來描述，以上方式描述的聲源模型侷限於單一秩(rank-1)的空間關係中。然而，實際聲源在傳遞至麥克風陣列過程中，受到區域性散射、波前不規則變動、或是迴響等影響，往往空間關係是多秩的。因此，本論文利用多秩訊號模型提出波束形成及多通道後濾波等技術，用來降低壓抑雜訊時過度刪減到目標聲源的現象，藉此提升語音品質。

在波束形成技術上，本論文引入了多秩(multi-rank)訊號模型與範數限制(norm constraint)來降低傳統最小變異無失真響應(minimum variance distortionless response)設計下對陣列不確定性的影響。本論文經由虛擬觀測(pseudo-observation)的技巧，將問題轉換到狀態空間，並使用一階、二階擴展卡爾曼濾波器(extended Kalman filter)以及非察覺型卡爾曼濾波器(unscented Kalman filter)來實現以上非線性問題。此外，本論文針對範數限制值的選擇進行完整分析。由模擬結果可看出，相較於使用對角加載(diagonal loading)，使用範數限制對於未知訊號能量、模型誤差等更具有強健性。

在多通道後濾波技術上，我們定義了一種新的空間相干量度(spatial coherence measure)並首度引入多秩訊號模型到後濾波的發展中。此相干量度透過兩功率頻譜密度(power spectral density)矩陣來描述兩聲場之間的相似度。此相干量度可用於設計一種新的多通道濾波器。基於此相干量度，本論文分析了由目標聲場及雜訊聲場造成的偏離(bias)，並提出另一套偏離補償的多通道濾波器。此多通道濾波器在偏離或是雜訊功率頻譜密度矩陣被正確估測下，等同於理想的維納濾波器(Wiener filter)。本論文透過理論以及實驗證實在使用更正確的訊號模型下，本論文提出之偏離補償後的多通道濾波器可提供較佳的語音品質。

Beamforming and Multi-Channel Post-Filtering for Speech Enhancement Based on Multi-Rank Signal Models

Graduate Student: Ming-Tang Lee

Advisor: Dr. Jwu-Sheng Hu

Institute of Electrical Control Engineering
National Chiao-Tung University

Abstract

Beamforming and multi-channel post-filtering are two major techniques in microphone array speech enhancement. In conventional algorithms, the relationship between a source to the microphones is usually described by delays or relative transfer functions (RTFs) of each microphone pair. The description of the source model is limited to the rank-1 spatial correlation. However, in the real sound propagation of the source, the spatial correlation is typically multi-rank due to local scattering, wavefront fluctuation, or reverberation. Thus, this dissertation proposes the beamforming and multi-channel post-filtering algorithms based on multi-rank signal models. The proposed algorithms can alleviate the *self-cancellation* phenomenon of the desired source during noise reduction and improve the speech quality.

For the proposed beamforming algorithms, multi-rank signal models and norm constraint are introduced into the minimum variance distortionless response (MVDR) beamforming problem for reducing the sensitivity of the design against array uncertainties. Based on the pseudo-observation method, the beamforming problems are

transformed into state spaces and solved by the first- and second-order extended Kalman filters (EKF) and the unscented Kalman filter (UKF). In addition, the selection of the norm constraint value is completely studied. The simulations show that the usage of the norm constraint is more robust to the unknown signal powers and model errors compared to the usage of the diagonal loading (DL) technique.

For the proposed multi-channel post-filtering algorithms, a novel spatial coherence measure is defined and multi-rank signal models are firstly conducted into the post-filtering development. The spatial coherence measure evaluates the similarity between the measured signal fields using power spectral density matrices. A multi-channel post-filter is proposed based on this measure. Under this measure, the bias term due to the similarity of the desired signal field and the noise field is further investigated and a solution based on bias compensation is proposed. It can be shown that the compensated solution is equivalent to the optimal Wiener filter if the bias or the noise power spectral density matrix is perfectly measured. The theoretical and empirical results demonstrate that the proposed bias compensated post-filter provides better speech quality with a more accurate signal model.

誌謝

對於本論文的完成，首先要感謝胡竹生老師對我的栽培與鼓勵。沒有老師，就沒有今天的我。當初讀博班的目標，是希望達到"理論與實作兼備"，如今學生對老師能有些許貢獻，已感到心滿意足。老師，您是造夢者，而學生是您的追隨者。您為我們編織了一個機器人的夢。"那些年，我們一起做的阿凡達"，似乎讓我們又向夢想跨進了一步。與老師腦力激盪與築夢的日子，都已成回憶。在未來的日子裡，學生將以曾經身為 XLAB 的一員而感到榮耀，也永遠記得老師所傳遞的夢。

接下來要感謝興哥，從碩班到博班這些年中，沒有你的提攜指點、經驗傳承，我在計畫執行的部分就沒辦法如此順遂。過去一同參與的比賽、一同執行的計畫、一同熬夜趕工的夜晚，歷歷在目，猶如昨日。過程中，也有幸能見到你與淑伶終成眷屬，是我們 XLAB 難得的喜事呢！祝福你們！另外也非常感謝淑伶多年來在各個方面的協助，沒有妳，XLAB 的成員們就無法無後顧之憂地專注於研究上。

在聲音組的傳承部分，要感謝工作能力超強的耕維加入，有如神助！耕維的"工作"量真的是非常地多、非常地扎實，要是沒偷學"分身術"，我還真不知怎麼達到的。辛苦了！也謝謝你在我畢業前艱困時期幫我分擔了不少工作，以後維護聲音組這個重大任務就交給你了！

博班生涯中還有許多一路走來的好夥伴們：永融，謝謝你從碩班以來的照顧，我也很榮幸有機會能與你還有昀軒一同去上海參加會議，那是我唯一一次出國會議有玩伴，我十分珍惜那段美好的回憶。一路看著你從求婚到第二個小孩出生，打從心底為你的幸福生活感到高興，也希望能盡快追隨你的腳步。阿法，感謝你多年來的熱心協助，碰到問題遇到你就覺得很安心，希望你能順利圓夢！鏗元，恭喜妳成為美女教授，妳的甜美笑容是 XLAB 的招牌阿！大師兄，好懷念你辦的幾次聯誼，你的定點投籃也約來越準了，有機會再來打個球！阿吉，能跟你一同合作開發阿凡達機器人的"遠距喚醒"功能，真的是非常榮幸。而且烤肉時只要坐在你身邊就不怕餓肚子，謝謝招待！Judo，我們 XLAB 的籃球中鋒不能沒有你，裡外配合真的是非常地好打！偷偷告訴你我在奶油廚房巧遇你的母親，她做的蛋餅特別好吃！

博班生涯中也帶過不少聲音組的同仁們：嘟嘟，恭喜你成為 Berkeley 的一員，也很高興你能一同跟我拍到博士照~祝福你在美國一切順利！蛋糕社的小蔡，未來還是同部門的工作夥伴，請多多指教阿~ 未來同公司的沛錡，我們好像是在你畢業後，打球次數增多後才變更熟的~謝謝你陪我分享心情！以後也請多多指教！星聲社長庭昭，很高興能帶到音樂同好的你，謝謝你開發了語音喚醒的功能！小提琴王子昀軒，你大概是第一個跟我共同合作最密集的學弟，我們針對"阿凡達"這個關鍵字實現的偵測器還有拍攝的影片，是我最驕傲的成果之一！很高興能跟你共事過！害羞的學文，沒想到你到台北上班後竟然有每天午餐要吃不同家店的創舉！育成，謝謝你先前再將數位麥克風陣列平台更改了一版，還有你是我第一位

帶進 CIC 工讀的學弟，校內校外都進行了不少合作呢！新文，很感謝你當初分擔了計畫的組長，看著你的快速成長，我也感到非常地欣慰！翰哥，很懷念當初一起打球的日子，你跟男哥的跳躍力都相當驚人阿~你對美食方面有獨到眼光，拖你的福大家才會一同去吃 BBQ House，那真的是相當美味的牛排呢！鳴哥，謝謝你當初辦的聯誼，尤其是輔大那一團，大家還一起出遊好幾次呢！謝謝你留給我們這麼好的回憶~ 哲宇，謝謝你一起完成中華電信的技轉，看到你最後論文的突破，也替你的成長感到喜悅~大夢、小山東，謝謝你們幫忙開發 AGC 以及分擔了產學計畫，綠島之行沒有你們就 High 不起來了~ 佑軒、小樂，謝謝你們一起改造了手機的模具，未來一年還請你們繼續幫忙耕維了！

實驗室其他非聲音組的學長學弟們：Lundy，同為星聲社當期成員，沒想到你工作後仍有玩兩個團，請成為我們星聲社的榮耀之一吧！肉鬆，自從你到工研院後就比較少打球了，當初也是帶你進 CIC 短期工讀呢，有機會吃飯時再聊一聊吧！Rodolfo, you are so kind and shining, and I really appreciate the India food on the Jinshan street. Wish you have a great life in Taipei! 阿 Him，你回到澳門後就沒什麼機會跟你見面，有機會去澳門一定會再找你玩！我還記得你背位單打時晃肩的假動作呢，下次再跟你請教一下~ 昌言，謝謝你先前一同分擔了計畫工作，也祝福您再創事業高峰！湘筑，當時妳和大家一起送的其實是我在實驗室收到的第一張生日賀卡呢！實驗室沒有妳可真是少了點歡樂氣氛呢！Makaka，你做的 App 遊戲真的很不錯呢！期待你可以繼續發展下去喔！建安，雖然最後你去了政大，你是很聰明且有想法的人，祝福你可以找到自己的一片天！Daniel, you have great talent in painting! 而且你的中文真的說得很好! Thanks for the companion at the XLAB. 德洋，很高興能有機會跟你談論到創業的想法與經驗，也祝福你可以早日實現創業夢想！震華，永融跟我都覺得你的穿搭很有品味呢！建廷，很高興在我最脆弱的時候有你可以一同分擔，謝謝你！男哥，你是我在 CIC 交棒的人，以後 Carven 就請你繼續輔佐他了！我們有相近的興趣跟家庭背景，當初 FB 成立的 XLAB basketball club 也變成了 XLAB family 了~ 有打球記得要再找我喔！罐頭，很感謝你辦的綠島行，你真的是非常貼心的人~ 期元、鳴遠，GPS 二人組，很榮幸能跟你們一起畢業~ 阿文，哪天我想練肌肉的時候一定找你，你是在球場上唯一能跟 Judo 體型抗衡的人啊！你之後如果回馬來西亞的話，記得帶我去好吃的在地美食阿~ 小綜，謝謝你規劃了小琉球之旅~ 凱翔、凱傑兩位書卷，你們很有潛力，實驗室的世代交替就交給你們了。另外還有新進的碩士學弟妹們，歡迎加入 XLAB 大家族~

另外要特別感謝在 CIC 的 Carven、Sieg 對我的照顧；一同出遊的電控同學們修哥、杰元、色鬍、宏爺、凱項、color 生、熙良哥、大伯，沒有你們我就沒有那些歡樂的旅程；一起留下來攻讀電控博班的輝哥、小叮噹、氧氣、阿賢、立岡；還有我的好友們加鵬、阿光、嚕嚕米、Mukesh、David。因為有你們，讓我的博班生活增添了不少樂趣。

最後要感謝我的家人對我的全力支持，讓我沒有後顧之憂。謝謝你們！

Contents

摘要.....	i
Abstract.....	ii
誌謝.....	iv
Contents.....	vi
List of Figures.....	viii
List of Tables.....	ix
Index.....	x
List of Notations.....	xi
Chapter 1 Introduction.....	1
1.1 Overview of Multi-Channel Speech Enhancement.....	1
1.1.1 Overview of Beamformers.....	2
1.1.2 Overview of Post-Filters.....	4
1.2 Problem Formulation and Multi-Rank Signal Models.....	5
1.2.1 Framework of Beamforming and Post-Filtering.....	5
1.2.2 Multi-rank Signal Models in Array Signal Processing.....	6
1.3 Outline of Proposed Beamformer and Post-Filter.....	8
1.3.1 Robust Adaptive Beamforming with Multi-Rank Signal Models Based on the Kalman Filter.....	8
1.3.2 Multi-Channel Post-Filtering Based on Spatial Coherence Measure.....	8
1.4 Contribution of this Dissertation.....	9
1.5 Organization of this Dissertation.....	10
Chapter 2 Robust Adaptive Beamforming with Multi-Rank Signal Models Based on the Kalman Filter.....	11
2.1 Introduction.....	11
2.2 Problem Formulation.....	13
2.2.1 Robust MVDR Beamforming with Multi-Rank Signal Models.....	13
2.2.2 Motivations of the Proposed Robust Beamforming.....	16
2.3 Proposed Robust Beamforming Based on the Soft-Constrained Pseudo-Observation Method.....	18
2.3.1 Normalized Multi-Rank Signal Model for Wide-Band Applications.....	18
2.3.2 Norm-Constrained Capon Beamforming.....	20
2.3.3 State Space Formulation Using the SCPO Method.....	23
2.4 Solutions Using Nonlinear Kalman Filters.....	26
2.4.1 Solutions Using Extended Kalman Filters.....	27
2.4.2 Solution Using the Unscented Kalman Filter.....	29
2.4.3 Initial Conditions.....	30

2.4.4 Estimation of the Parameter Matrices	31
2.5 Summary.....	33
Chapter 3 Multi-Channel Post-Filtering Based on Spatial Coherence Measure.....	35
3.1 Introduction	35
3.2 Brief Review of Zelinski and McCowan Post-Filters	36
3.3 Proposed Spatial Coherence Measure	39
3.3.1 Definition of the Spatial Coherence Measure.....	39
3.3.2 Properties of Proposed Spatial Coherence Measure.....	40
3.4 The Proposed Spatial Coherence Based Post-Filter (SCPF)	42
3.4.1 A Description of the Proposed SCPF.....	42
3.4.2 Mean Square Error Analysis of Proposed SCPF	44
3.5 The Proposed Bias Compensated SCPF (BC-SCPF)	48
3.5.1 Derivation of Proposed BC-SCPF	48
3.5.2 Comparison between BC-SCPF and McCowan Post-Filter.....	49
3.6 Summary.....	51
Chapter 4 Experimental Results.....	52
4.1 Narrow-Band Simulations of the Proposed Robust Beamformer	52
4.1.1 Analyses of the Norm Constraint Value for Half-Wavelength Spacing ..	53
4.1.2 Narrow-Band Comparisons	59
4.2 Speech Enhancement Results of the Proposed SCPF and BC-SCPF	65
4.2.1 The Experimental Setup	66
4.2.2 Speech Quality and Noise Reduction Evaluations	66
4.3 Speech Enhancement Results of the Proposed Speech Enhancement System (Beamformer + Post-Filter).....	71
4.3.1 Relaxation of the Constraints at Low Frequency Bands	71
4.3.2 Speech Quality and Noise Reduction Evaluations	75
4.4 Summary.....	78
Chapter 5 Conclusions and Potential Research Topics.....	80
5.1 Conclusions	80
5.2 Potential Research Topics	81
References.....	83
Appendix I	90
Appendix II.....	97
Appendix III	99

List of Figures

Figure 1-1	Implementation of a beamformer with a multi-channel post-filter.....	2
Figure 1-2	An example of the beam pattern.	3
Figure 3-1	System architecture of the Zelinski and McCowan post-filters.....	37
Figure 3-2	MSE comparison among the post-filters and the DS beamformer.	47
Figure 4-1	Presumed and actual angular power density functions for $\eta = 5$ and 5° central angle mismatch.....	54
Figure 4-2	Output SINR vs. input SNR.....	55
Figure 4-3	Output SINR vs. $\lambda(\omega)$ for different input SNRs.	56
Figure 4-4	Output SINR vs. $\gamma(\omega)$ for different input SNRs.....	56
Figure 4-5	Comparison between the best selected of $\lambda(\omega)$ and $\gamma(\omega)$ for different input SNR conditions	57
Figure 4-6	Comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different central angle mismatches	58
Figure 4-7	Comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different number of sensors.....	58
Figure 4-8	Comparisons of the beamformers at 0 dB input SNR.	61
Figure 4-9	Comparisons of the beamformers at 20 dB input SNR.	62
Figure 4-10	Output SINR vs. input SNR for different beamformers.....	63
Figure 4-11	Output SINR vs. input SNR for proposed Kalman filters	63
Figure 4-12	Beam patterns of the proposed Kalman-filters at 20 dB input SNR.....	64
Figure 4-13	Angular power density function for $\eta = 10$	67
Figure 4-14	Simulation environment with locally scattered sources.	67
Figure 4-15	Squared norm of the optimal weight vs. the ratio between the array spacing and the wavelength for different noise fields.....	73
Figure 4-16	The error of the distortionless constraint vs. the ratio between the array spacing and the wavelength.....	73
Figure 4-17	κ vs. frequency.....	75
Figure 4-18	Waveforms and spectrograms at SNR = 5 dB.	78

List of Tables

Table 4-1	PESQ score improvement obtained by different input SNRs	68
Table 4-2	SNRI score obtained by different input SNRs	69
Table 4-3	TNLR score obtained by different input SNRs	70
Table 4-4	PESQ score improvement with 5° angle mismatch.....	76
Table 4-5	SNRI score with 5° angle mismatch.....	76
Table 4-6	TNLR score with 5° angle mismatch	76



Index

Bias compensated spatial coherence based post-filter (BC-SCPF).....	8
Constrained Kalman filter (CKF).....	59
Delay-and-sum (DS).....	2
Diagonal loading (DL).....	4
Extended Kalman filter (EKF).....	8
First-order extended Kalman filter (FOEKF).....	26
Generalized sidelobe canceller (GSC).....	3
Mean square error (MSE).....	1
Measured signal field (MSF).....	40
Minimum mean square error (MMSE).....	2
Minimum variance distortionless response (MVDR).....	1
Multi-rank loaded sample matrix inverse (MRLSMI).....	16
Multi-rank norm-constrained Kalman filter (MRNCKF).....	8
Multi-rank sample matrix inverse (MRSMI).....	16
Norm-constrained Capon Beamforming (NCCB).....	9
Perceptual evaluation of speech quality (PESQ).....	66
Power spectral density (PSD).....	1
Quadratically constrained quadratic programming (QCQP).....	13
Recursive least square with variable loading (RLSVL).....	59
Second-order extended Kalman filter (SOEKF).....	13
Short-time Fourier transform (STFT).....	6
Signal-to-interference-plus-noise ratio (SINR).....	3
Signal-to-noise ratio (SNR).....	2
Signal-to-noise ratio improvement (SNRI).....	66
Soft-constrained pseudo-observation (SCPO).....	8
Spatial coherence based post-filter (SCPF).....	8
Super-directive (SD).....	65
Total noise reduction level (TNLR).....	66
Uniformly distributed linear array (ULA).....	32
Unsented Kalman filter (UKF).....	8

List of Notations

Common Notations:

\mathbf{a}_s	: steering vector of the desired signal
\mathbf{w}	: weight vector (beamformer)
\mathbf{x}	: microphone input vector
\mathbf{Q}	: parameter matrix of the state equations
\mathbf{R}	: parameter matrix of the measurement equations
Φ	: power spectral density matrix
$\tilde{\Phi}_s$: normalized multi-rank signal model
Γ	: coherence matrix
s	: desired signal
\hat{s}	: estimate of the desired signal
λ	: diagonal loading level
γ	: norm constraint parameter = $T - 1/M$
ϕ	: power spectral density
σ^2	: variance
T	: norm constraint value
G	: post-filter
θ	: angle
k	: frame index
ω	: frequency index
α	: forgetting factor
M	: number of sensors
N	: training size (or number of samples)

Chapter 1:

\mathbf{a}	: steering vector
\mathbf{h}	: impulse response vector of the desired signal to the microphones in the STFT domain
\mathbf{B}	: coherence loss matrix
\mathbf{i}	: interference vector
\mathbf{v}	: sensor noise vector
y	: filtered output
ρ	: normalized angular power density function
ζ	: coherence loss parameter

Chapter 2:

$\hat{\Phi}_s$: multi-rank signal model
\mathbf{e}	: innovation vector
\mathbf{f}	: nonlinear observation vector
\mathbf{v}_s	: process noise vector
\mathbf{v}_m	: measurement noise vector
\mathbf{z}	: desired response vector
ϕ_i	: elementary vector with the i -th element equal to 1. i.e., $[0, \dots, 1, \dots, 0]^T$
$\boldsymbol{\pi}$: bias vector in the innovation
$\hat{\mathbf{w}}$: estimate of the weight vector
\mathbf{F}_w	: Jacobian matrix of the nonlinear function \mathbf{f}
$\mathbf{F}_{ww}^{(i)}$: Hessian matrix of the i -th measurement equation in \mathbf{f}
\mathbf{K}	: Kalman gain
\mathbf{P}^-	: <i>a priori</i> state error covariance matrix
\mathbf{P}^+	: <i>a posteriori</i> state error covariance matrix
\mathbf{S}	: covariance matrix of the innovation \mathbf{e}
$\boldsymbol{\Lambda}$: bias matrix due to the bias vector $\boldsymbol{\pi}$
$\hat{\mathbf{w}}_\sigma$: sigma points of the weight vector
$\hat{\mathbf{z}}_\sigma$: transformed sigma points
κ	: variance parameter for the constraints
μ	: Lagrange multiplier
P	: number of measurement equations
C	: normalization factor to meet the distortionless constraint
$\mathcal{P}\{\mathbf{A}\}$: principal eigenvector of matrix \mathbf{A}

Chapter 3:

$\tilde{\mathbf{x}}$: time-aligned input vector
$\tilde{\boldsymbol{\Gamma}}_n$: coherence matrix of the time-aligned noise
ξ	: signal-to-noise ratio
β_ω	: bias in the spatial coherence measure at frequency ω
\mathcal{B}	: inner product of the coherence matrices of the desired signal field and the noise field
$\Re\{a\}$: real operator, which gives the real part of a
$\mathbf{u}_i^2\{\mathbf{A}\}$: the i -th eigenvector of matrix \mathbf{A}
$\sigma_i^2\{\mathbf{A}\}$: the i -th eigenvalue of matrix \mathbf{A}
$\mathcal{F}\{\mathbf{A}, \mathbf{B}\}$: spatial coherence measure between PSD matrix \mathbf{A} and \mathbf{B}

Chapter 4:

- η : angular spreading parameter of the Gaussian model
 c : sound velocity
 d : spacing between sensors
 f : frequency in Hz



Chapter 1

Introduction

1.1 Overview of Multi-Channel Speech Enhancement

Multi-channel speech enhancement has attracted much attention in recent years. In the real world, desired speech signals are often corrupted by background noises, speech interferences, and reverberation. For more than two microphones, there are two main categories of speech enhancement approach: beamforming and multi-channel post-filtering. Beamforming has been applied to several narrow- or wide-band signals processes, which can be defined by a *filter-and-sum* process [1] in the conventional sense. A well-known designing strategy is to preserve the signal from the direction of interest while attenuating others, which can be achieved by the minimum variance distortionless response (MVDR) algorithm [1–3]. The MVDR beamforming (or so-called Capon beamforming) is optimal in the mean square error (MSE) sense when the interference-plus-noise power spectral density (PSD) matrix can be obtained and there is no mismatch on the presumed steering vector. Typically, adaptive filtering techniques are applied to estimate the PSD matrix and additional training processes or

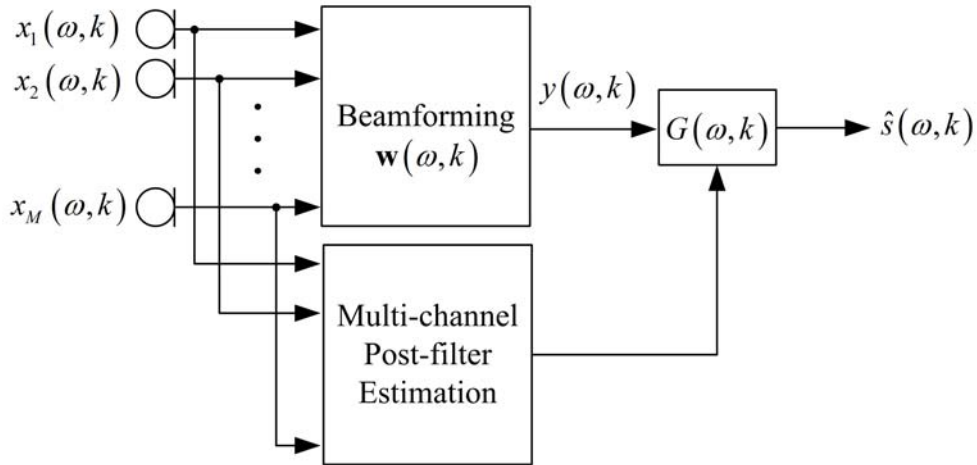


Figure 1-1 Implementation of a beamformer with a multi-channel post-filter.

a priori information of signal presence is needed for offline or online implementation [1–4]. On the other hand, the multi-channel post-filtering, which considers both the spatial information and the signal-to-noise ratio (SNR), can be designed in a more general way. Simmer *et al.* [5] show that the optimal minimum mean square error (MMSE) solution can be decomposed into an MVDR beamformer followed by a single-channel Wiener filter. This solution is also called a *multi-channel Wiener filter*. A typical implementation of a beamformer with a multi-channel post-filter is illustrated in Figure 1-1.

1.1.1 Overview of Beamformers

Beamformer is a spatial filter which can be defined by a *filter-and-sum* process. In the beamforming design, the major objective is to control the mainlobe, nulls, and the sidelobe level according to the spatial characteristics of the sound fields. Figure 1-2 gives an illustration of the beam pattern. Fixed beamformers aim to design a filter for some desired spatial response. The commonly used delay-and-sum (DS) beamformer [1] is known to be optimal in an incoherent noise field with the point source model

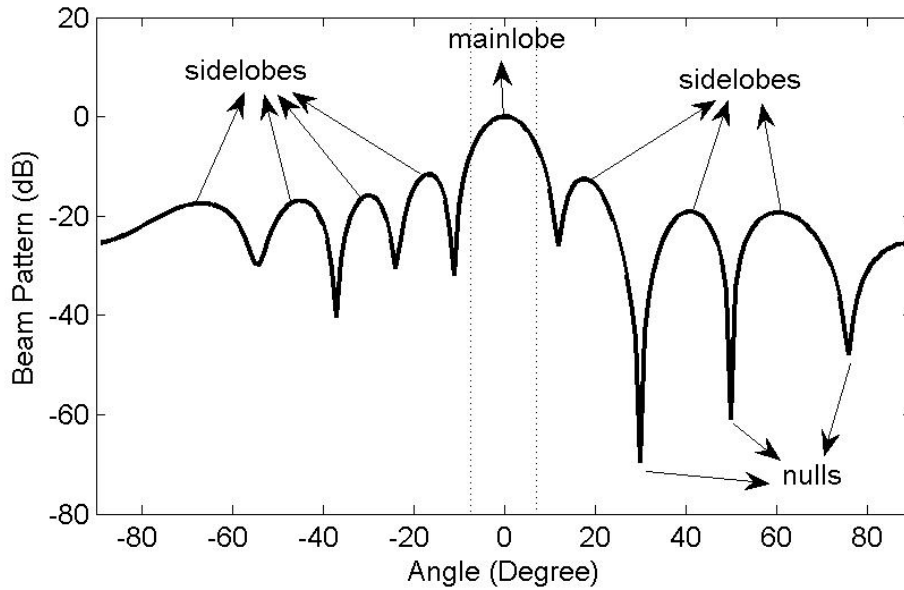


Figure 1-2 An example of the beam pattern.

assumption. In coherent fields, the spatial coherence is high at low frequencies due to the insufficient spatial sampling. Some beamformers are designed to keep the directivity [1] through frequencies such as constant directivity beamformers [6–8] and super-directive beamformers [9–11]. Besides, due to the uncertainty of steering error, some fixed beamformer aim to extend the bandwidth [1] of the mainlobe to reduce the sensitivity to the uncertainties.

When the spatial information is unknown *a priori* or time-varying, adaptive beamformers are able to adjust their filter responses according to the defined optimization problem and the training data. The most popular optimization strategy is the MVDR problem [1–3]. Cox *et al.* [2] showed that the MVDR problem is equivalent to maximize the improvement in the signal-to-interference-plus-noise ratio (SINR) or the array gain. Under this problem, a famous method called generalized sidelobe canceller (GSC) [12–17] decomposes the MVDR beamformer into two mutually orthogonal subspaces followed by a noise canceller. Several algorithms follow this structure to design the null-space (or blocking matrix) of the given fixed beamformer

[16][17], or the noise canceler [17–19].

The major problem of the MVDR beamforming is the sensitivity to the array or propagation uncertainties in the presence of the desired signal. Shahbazpanahi *et al.* [20] investigated the MVDR problem with multi-rank signal models to reduce the model errors. Rather than using multi-rank models [20–22], several robust adaptive beamforming algorithms [2][23–28] have been designed to improve the robustness against uncertainties. One of the famous techniques is called diagonal loading (DL) [2][23][24]. The DL can be explained as imposing additional spatially white noise on the input covariance matrix. The other equivalent explanation is to add a norm constraint on the MVDR problem [2]. Several robust adaptive beamforming approaches can be categorized into this family [23–27]. The major drawback of DL is that it is not clear how to decide the diagonal loading level and the level depends on the unknown signal powers.

1.1.2 Overview of Post-Filters

Most post-filtering algorithms aim to enhance the estimation of the Wiener post-filter by a more accurate estimation of SNR. The SNR estimation for speech enhancement can be implemented based on the minimum statistics for the stationary noise [29–31], or the spatially pre-processed power [32][33]. Most of them are energy-based. Given the SNR or noise estimates, some post-filtering algorithms emphasize on the designs of different objective functions, such as the MMSE estimator under spectral amplitude domains [33–37] or the psychoacoustic masking [38–40].

Alternatively, the phase information of a microphone pair has already been used in blind source separation (BSS) [41] as well as the computational auditory scene analysis (CASA) [42]. Aarabi *et al.* [43–45] provide a different view of the SNR from the phase

error perspective for the dual-channel case. In their work, the relationship between the phase error and the SNR was derived [43]. However, the idea of phase error can only be applied to the case of two-microphone. In addition to the SNR estimation, some post-filtering algorithms directly estimate the spectral densities [46–48]. Like the case of phase error, the cross-spectral density is usually defined between two microphones. For more than two microphones, the common practice is to perform average among all distinct microphone pairs [46][47]. Although this might enhance the robustness of the estimation, there is still no formal proof regarding its effectiveness. In particular, it does not consider the spatial arrangement of microphones, i.e., the advantages of using more than two microphones is not fully explored. In addition, the description of the relationship from the source to the microphones is limited to the rank-1 signal model, which is not sufficient to represent the real world source where local scattering, wavefront fluctuation, or reverberation can happen.

1.2 Problem Formulation and Multi-Rank Signal Models

1.2.1 Framework of Beamforming and Post-Filtering

Consider that a linear array with M omni-directional microphones and the observation in the m -th microphone at time instant t is given by

$$x_m^t(t) = h_m^t(t) * s^t(t) + \sum_p i_{m,p}^t(t) + v_m^t(t) \quad (1-1)$$

where $s^t(t)$ is the coherent desired signal; $i_{m,p}^t(t)$ is the p -th interference sampled in the m -th microphone; $v_m^t(t)$ is the sensor noise sampled in the m -th microphone; $h_m^t(t)$ is the impulse response of the desired signal corresponding to the m -th microphone; $*$ denotes convolution; and the desired signal and the noise are assumed to be zero mean and

mutually uncorrelated. Assuming time-invariant transfer functions, the observations are divided in time into overlapping frames by the application of a window function and analyzed using the short-time Fourier transform (STFT) and expressed in the time-frequency domain in a vector form as,

$$\begin{aligned}\mathbf{x}(\omega, k) &= \mathbf{h}(\omega) s(\omega, k) + \mathbf{i}(\omega, k) + \mathbf{v}(\omega, k) \\ &= \mathbf{s}(\omega, k) + \mathbf{n}(\omega, k)\end{aligned}\quad (1-2)$$

where ω and k are discrete frequency and frame indices respectively. $\mathbf{n}(\omega, k)$ is the total noise (i.e., interference plus noise).

A beamforming method aims to find a spatial filter \mathbf{w} to estimate the desired source by

$$y(\omega, k) = \mathbf{w}^H(\omega, k) \mathbf{x}(\omega, k) \quad (1-3)$$

A post-filtering method aims to find a gain function (or mask) in the STFT domain to suppress the undesired noise, which can be multiplied on the beamformer output as

$$\hat{s}(\omega, k) = G(\omega, k) \cdot y(\omega, k) \quad (1-4)$$

1.2.2 Multi-rank Signal Models in Array Signal Processing

Multi-rank signal models or rank relaxation has been widely used in sensor array localization [49–52], beamforming [20–22][53], or quadratic optimization problems [53][54]. One commonly used model of the signal field is a point source in a homogeneous sound field [33]. Assuming that there is no mismatch between microphones, the desired signal field can be measured using the PSD matrix as

$$\begin{aligned}\mathbf{\Phi}_s(\omega) &= \phi_s(\omega) \mathbf{\Gamma}_s(\omega) \\ &= \phi_s(\omega) \mathbf{a}_s(\omega) \mathbf{a}_s^H(\omega)\end{aligned}\quad (1-5)$$

where $\Phi_s(\omega) = E_k[\mathbf{s}(\omega, k) \mathbf{s}^H(\omega, k)]$; $\Gamma_s(\omega)$, $\mathbf{a}_s(\omega)$, and $\phi_s(\omega)$ are the coherence matrix, steering vector, and power spectral density of the desired signal respectively. A single point source is usually referred as the rank-1 signal model.

However, in practice, the rank of signal model is usually greater than 1. Typical examples are incoherently scattered signal source or signals with random fluctuating wavefronts in wireless communication, sonar, and microphone array [49–52]. Further, environmental reverberation also increases the rank. For example, in the case of incoherently scattered source, the desired PSD matrix can be expressed by [49–51]

$$\Phi_s(\omega) = \phi_s(\omega) \int_{-\pi/2}^{\pi/2} \rho(\theta, \omega) \mathbf{a}(\theta, \omega) \mathbf{a}^H(\theta, \omega) d\theta \quad (1-6)$$

where $\rho(\theta, \omega)$ is the normalized angular power density function ($\int_{-\pi/2}^{\pi/2} \rho(\theta, \omega) d\theta = 1$), and $\mathbf{a}(\theta, \omega)$ is the steering vector at direction θ . In the case of randomly fluctuating wavefronts, the desired PSD matrix can be expressed by [20]

$$\Phi_s(\omega) = \phi_s(\omega) \mathbf{B} \circ \{ \mathbf{a}_s(\omega) \mathbf{a}_s^H(\omega) \} \quad (1-7)$$

where \mathbf{B} is the M -by- M coherence loss matrix, and \circ is the Schur-Hadamard (elementwise) matrix product. Two commonly used models for the coherence loss matrix are

$$[\mathbf{B}]_{i,j} = \exp\left\{-(i-j)^2 \zeta\right\} \quad (1-8a)$$

$$[\mathbf{B}]_{i,j} = \exp\left\{-|i-j|\zeta\right\} \quad (1-8b)$$

where ζ is the coherence loss parameter. Note that both the signal models in (1-6) and (1-7) are multi-rank.

1.3 Outline of Proposed Beamformer and Post-Filter

1.3.1 Robust Adaptive Beamforming with Multi-Rank Signal Models Based on the Kalman Filter

In this dissertation, the multi-rank norm-constrained MVDR beamforming based on the Kalman filter (MRNCKF) is proposed. The multi-rank MVDR problem defined in [20] is modified with normalized signal model and norm constraint. Simulations show the superiority of using the norm constraint rather than the diagonal loading formulation. The modified problem is transformed into state space models using the soft-constrained pseudo-observation (SCPO) method in constrained Kalman filtering [55–59]. Several nonlinear Kalman filtering approaches including the first- and second-order extended Kalman filters (EKF) [60–62] and the unscented Kalman filter (UKF) [62–66] are introduced to solve the quadratically constrained problems. The settings of initial conditions and parameter matrices are also studied.

1.3.2 Multi-Channel Post-Filtering Based on Spatial Coherence Measure

In this dissertation, a novel multi-channel post-filter where multi-rank signal models can be easily adopted is proposed. A new spatial coherence measure is introduced and analyzed. Based on this measure, the new spatial coherence based post-filter (SCPF) is derived. Due to the similarity of the desired signal field and the noise field, a bias compensated (BC) solution, BC-SCPF, is proposed. It can be shown that the BC-SCPF is equivalent to the optimal Wiener filter if the bias or the noise power spectral density matrix is perfectly measured. The proposed post-filters are compared with the state-of-the-art post-filters, Zelinski post-filter [46] and McCowan

post-filter [47], theoretically and experimentally.

1.4 Contribution of this Dissertation

The contributions of this dissertation are listed below:

- 1) The norm constrained Capon beamforming (NCCB) is known to be equivalent to the DL [2][26]. In this dissertation, the simulations demonstrates that an appropriately chosen norm constraint value is more robust to unknown signal powers, small angle mismatches, and number of sensors compared to the selection of the diagonal loading level. By using the SCPO method, the multi-rank signal model and the norm constraint can be easily adopted. On-line implementations using nonlinear Kalman filters are proposed to reduce the computation burden and give a more flexible structure. The settings of initial conditions and parameter matrices are also investigated.
- 2) Compared to the Zelinski and McCowan post-filters, the assumption of homogeneous sound fields is relaxed and multi-rank signal models are firstly introduced into the post-filter design. In addition, a spatial coherence measure which describes the similarity between two signal fields is introduced and analyzed. The proposed BC-SCPF is less sensitive to the individual estimation error of the noise coherence function. Besides, compared to the noise coherence function, the bias term in the proposed method can be estimated in many ways. The speech enhancement results demonstrate the superiority of the proposed BC-SCPF across different types of noise fields with a more accurate signal model.

1.5 Organization of this Dissertation

The remainder of this dissertation is organized as follows. The robust adaptive beamforming with multi-rank models based on the Kalman filter is introduced in Chapter 2. Chapter 3 presents the multi-channel post-filtering based on spatial coherence measure. Chapter 4 shows the experimental results of the proposed beamformer and post-filter. Finally, conclusions and future works are drawn in Chapter 5.



Chapter 2

Robust Adaptive Beamforming with Multi-Rank Signal Models Based on the Kalman Filter

2.1 Introduction

MVDR beamforming aims to minimize variances of the interferences and noise while maintaining the desired array response. It is known to degrade dramatically due to even small mismatches of the desired signal model, especially when the desired signal is present in the training data. The robust MVDR beamforming aims to keep the output SINR performance against several array or propagation uncertainties. In real world environments, the spatial correlation is typically multi-rank due to local scattering, wavefront fluctuation, or reverberation. Therefore, the multi-rank signal model is able to provide a more accurate model of the sound propagation of the desired source to the microphones. In this case, if the array response provided with multi-rank signal models is known exactly, such performance degradation can be reduced. To further reduce the sensitivity of arbitrary kinds of mismatches, diagonal loading (DL) [2][23][24] technique has been a popular used approach to improve the robustness of the MVDR

beamformer. The major drawback of DL is that it is not clear how to choose the diagonal loading level given the input covariance matrix. The norm constrained Capon beamforming (NCCB) is known to be equivalent to the DL [2][26]. However, the knowledge of how to choose the norm constraint value has not been completely studied. In this dissertation, we use the norm constraint rather than the original DL formulation. Besides, the superiority of using the norm constraint value rather than using the diagonal loading level is demonstrated in the simulations (see Section 4.1.1).

In this dissertation, we conduct the constrained Kalman filtering for more flexible on-line implementations. Constrained Kalman filtering [55–59] has been widely investigated in the last decade. The approaches mainly fall into one of three categories: pseudo-observation methods (or penalty methods), projection methods, and dimension reduction methods. Among these methods, the pseudo-observation method is the most intuitive way to conduct the constraints into the state-space of the Kalman filtering by considering the constraints as additional measurement equations. In this way, several developed nonlinear Kalman filtering algorithms can be directly applied. Chen *et al.* was the first one who introduced the soft-constrained pseudo-observation (SCPO) [58][59] into the traditional MVDR problem [67]. El-Keyi *et al.* conducted the SCPO for the robust adaptive beamforming based on worst-case performance optimization [68]. In this dissertation, we also apply the SCPO for the robust adaptive beamforming with multi-rank signal models. The potential drawback is that the unconstrained problem by using SCPO can be ill-conditioned if the parameter matrices are not appropriately chosen. In this dissertation, the settings of the initial conditions and parameter matrices are studied to achieve the good performance, which also prevent the SCPO method from the ill-conditioning problem. Compared to the prior work [20], the computation of principal eigenvector can be avoided in the proposed method.

Since the robust adaptive beamforming problems with multi-rank signal models and the norm constraint belong to the quadratically constrained quadratic programming (QCQP) [69], nonlinear Kalman filtering is introduced. The most widely used nonlinear Kalman filter is the extended Kalman filter (EKF) [60–62]. Another popular method is the unscented Kalman filter (UKF) [62–66]. The EKF approximates the Jacobian and Hessian matrices (in the first- and second-order approximations) of the nonlinear functions, while the UKF approximates the probability distribution of the nonlinear transformation using *sigma points*. Theoretically, the second-order extended Kalman filter (SOEKF) gives the best approximation in the MSE sense. However, due to the approximation of the second-order errors (see Appendix I), the SOEKF is sensitive to improper initial conditions and parameter matrices. The comparison of the above nonlinear Kalman filters will be discussed in Section 4.1.2.

The remainder of this chapter is organized as follows. In Section 2.2, we briefly review the problem of the MVDR beamforming with multi-rank signal models. Section 2.3 gives a modified problem with the normalized signal model and the norm constraint, and formulates its state space model based on the SCPO method. The relationship between the diagonal loading level and the norm constraint value for the multi-rank case is also analyzed. Section 2.4 presents the solutions using the EKFs and the UKF. Finally, a summary is drawn in Section 2.5.

2.2 Problem Formulation

2.2.1 Robust MVDR Beamforming with Multi-Rank Signal Models

The well-known MVDR beamformer minimizes the output power of interference-signals-plus-noise while maintaining a distortionless constraint at the look

direction [1]. Consider the noise vector $\mathbf{n}(\omega, k)$ in the STFT domain given in (1-2), the problem can be formulated as

$$\min_{\mathbf{w}(\omega)} \mathbf{w}^H(\omega) \mathbf{\Phi}_n(\omega) \mathbf{w}(\omega) \quad \text{subject to} \quad \mathbf{w}^H(\omega) \mathbf{a}_s(\omega) = 1 \quad (2-1)$$

where $\mathbf{\Phi}_n(\omega) = E[\mathbf{n}(\omega, k) \mathbf{n}^H(\omega, k)]$ is the noise-only PSD matrix. The problem is equivalent to the maximum signal-to-interference-plus-noise ratio (SINR) beamformer [2]. The solution of the MVDR beamformer can be easily obtained by the Lagrange multiplier method as

$$\mathbf{w}_{opt1}(\omega) = \frac{\mathbf{\Phi}_n^{-1}(\omega) \mathbf{a}_s(\omega)}{\mathbf{a}_s^H(\omega) \mathbf{\Phi}_n^{-1}(\omega) \mathbf{a}_s(\omega)} \quad (2-2)$$

Note that the distortionless constraint $\mathbf{w}^H(\omega) \mathbf{a}_s(\omega) = 1$ in (2-1) constrains the signal array response on the steering vector $\mathbf{a}_s(\omega)$, where the steering vector is usually considered as a point source model, or a rank-1 signal model. Shahbazpanahi *et al.* [20] modified the distortionless constraint to a quadratic one and incorporated multi-rank signal models given in Section 1.2.2. The modified MVDR problem is given by

$$\min_{\mathbf{w}(\omega)} \mathbf{w}^H(\omega) \mathbf{\Phi}_n(\omega) \mathbf{w}(\omega) \quad \text{subject to} \quad \mathbf{w}^H(\omega) \hat{\mathbf{\Phi}}_s(\omega) \mathbf{w}(\omega) = 1 \quad (2-3)$$

where $\hat{\mathbf{\Phi}}_s(\omega)$ is the designed or estimated multi-rank signal model. The solution of the modified problem can be solved by the Lagrange multiplier, which results in the following generalized eigenvalue problem [70]:

$$\mathbf{\Phi}_n(\omega) \mathbf{w}(\omega) = \mu(\omega) \hat{\mathbf{\Phi}}_s(\omega) \mathbf{w}(\omega) \quad (2-4)$$

where the Lagrange multiplier $\mu(\omega)$ can be considered as a corresponding generalized eigenvalue. Since the PSD matrices $\hat{\mathbf{\Phi}}_s(\omega)$ and $\mathbf{\Phi}_n(\omega)$ are positive semi-definite, $\mu(\omega)$

is always real-valued and non-negative.

The solution to the minimization problem in (2-3) is the generalized eigenvector corresponding to the smallest generalized eigenvalue of the matrix pencil $\{\Phi_n(\omega), \hat{\Phi}_s(\omega)\}$. Assuming that $\Phi_n(\omega)$ is full-rank and invertible, the equation (2-4) can be rewritten as

$$\Phi_n^{-1}(\omega) \hat{\Phi}_s(\omega) \mathbf{w}(\omega) = \frac{1}{\mu(\omega)} \mathbf{w}(\omega) \quad (2-5)$$

which is the characteristic equation for the matrix $\Phi_n^{-1}(\omega) \hat{\Phi}_s(\omega)$. In this case, the smallest eigenvalue in (2-4) corresponds to the maximum eigenvalue in (2-5). Thus, the optimal weight vector of the problem in (2-3) can be expressed by

$$\mathbf{w}_{\text{opt}}(\omega) = \mathcal{P}\{\Phi_n^{-1}(\omega) \hat{\Phi}_s(\omega)\} \quad (2-6)$$

where $\mathcal{P}\{\cdot\}$ denotes the operator that yields the *principal eigenvector* of a matrix. It is known that when the noise field is incoherent (i.e., $\Phi_n(\omega) = \mathbf{I}$, where \mathbf{I} is the identity matrix), the optimal MVDR turns into the matched filter (or the delay-and-sum (DS) filter for the rank-1 signal model). Hence, the matched filter for the multi-rank case can be obtained by

$$\mathbf{w}_{\text{matched}}(\omega) = \mathcal{P}\{\hat{\Phi}_s(\omega)\} \quad (2-7)$$

For speech enhancement, the desired signal can appear with the interferences and noise. In practice, the noise PSD matrix is replaced by the input PSD matrix of the training data as

$$\hat{\mathbf{w}}_{\text{MRSMI}}(\omega) = \mathcal{P}\{\hat{\Phi}_x^{-1}(\omega) \hat{\Phi}_s(\omega)\} \quad (2-8a)$$

where

$$\hat{\Phi}_x(\omega) = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(\omega, n) \mathbf{x}^H(\omega, n) \quad (2-8b)$$

is referred as the sample matrix [1] and N is the training size. The solution of (2-8) is the multi-rank (MR) version of the well-known sample matrix inverse (SMI) beamformer [20]. However, when the desired signal exists in the training data, the MVDR beamforming is known to degrade dramatically due to the mismatches between the presumed and actual array responses to the desired signal [20]. This is the so-called *self-cancellation* phenomenon. To improve the robustness of the MVDR beamforming against mismatches, one of the most popular approaches is the diagonal loading (DL) method. It is equivalent to impose an additive noise on the covariance matrix [1][2], and the MVDR problem in (2-3) can be modified as

$$\begin{aligned} \min_{\mathbf{w}(\omega)} \quad & \mathbf{w}^H(\omega) \Phi_x(\omega) \mathbf{w}(\omega) + \lambda(\omega) \mathbf{w}^H(\omega) \mathbf{w}(\omega) \\ \text{subject to} \quad & \mathbf{w}^H(\omega) \hat{\Phi}_s(\omega) \mathbf{w}(\omega) = 1 \end{aligned} \quad (2-9)$$

where $\lambda(\omega)$ is the diagonal loading level to be determined. The solution of the modified problem (2-9) is referred as the multi-rank loaded SMI (MRLSMI) beamformer as

$$\hat{\mathbf{w}}_{\text{MRLSMI}}(\omega) = \mathcal{P} \left\{ \left(\hat{\Phi}_x(\omega) + \lambda(\omega) \mathbf{I} \right)^{-1} \hat{\Phi}_s(\omega) \right\} \quad (2-10)$$

The major drawback of MRLSMI is that it is not clear how to choose the best diagonal loading level $\lambda(\omega)$ since the optimal choice depends on the unknown signal and interference parameters [20].

2.2.2 Motivations of the Proposed Robust Beamforming

In this dissertation, the motivations of proposed robust beamforming are listed below:

- 1) The performance of the MVDR beamforming is known to degrade severely in the presence of even small mismatches between the actual and presumed array responses to the desired signal [20], especially when the desired signal “contaminates” the training data. Therefore, with multi-rank signal models, it is possible to model a more accurate array response which reduces the performance degradation due to the model mismatches.
- 2) In the original multi-rank MVDR beamforming algorithm [20], the normalization problem is not taken into account in the narrow-band applications. However, it is important in the wide-band applications since different normalization factors among frequencies will introduce frequency dependent distortion or different white noise gains [1]. This can seriously deteriorate the speech quality for multi-channel speech enhancement. Therefore, this dissertation proposes a modification of the problem based on a normalized multi-rank signal model.
- 3) The selection of the diagonal loading level $\lambda(\omega)$ depends on the unknown signal and interference parameters. Cox *et al.* [2] have shown that the DL problem in (2-9) is equivalent to the norm-constrained Capon beamforming problem. In this dissertation, the relationship between the diagonal loading level and the norm constraint value for the multi-rank case is analyzed. The simulations in Section 4.1.1 show that the optimal choice of the norm constraint value is less sensitive to unknown signal powers and small angle mismatches at high SNRs.
- 4) In [20], the computation of the principal eigenvector is needed. For the on-line implementation, we introduce the Kalman filter algorithms for more flexible designs.

- 5) The selection of initial conditions and parameter matrices is critical, especially when the system is nonlinear. Wrong settings can break down the performance of the system. In this dissertation, the initial state is suggested to be in the feasible set of the constraints, or at least close to the feasible set corresponding to chosen variance parameters. The error covariance is initialized as the null space of the initial state. Further, the selection of the parameter matrices is investigated to achieve a good performance and prevent from the ill-conditioning problem.

2.3 Proposed Robust Beamforming Based on the Soft-Constrained Pseudo-Observation Method

The soft-constrained pseudo-observation (SCPO) is one of the methods in constrained Kalman filtering [55–59]. By the SCPO method, constraints can be easily formulated into the state space as augmented measurements. In the follows, the distortionless constraint using normalized signal models for wide-band applications is proposed in Section 2.3.1. Then, the norm-constrained Capon beamforming (NCCB) for the multi-rank case is introduced in Section 2.3.2. In the sequel, the state space of the NCCB problem is formulated using the SCPO method in Section 2.3.3.

2.3.1 Normalized Multi-Rank Signal Model for Wide-Band Applications

In narrow-band applications, the normalization is immaterial since it does not affect the SINR defined as

$$\text{SINR} = \frac{\mathbf{w}^H(\omega)\mathbf{\Phi}_s(\omega)\mathbf{w}(\omega)}{\mathbf{w}^H(\omega)\mathbf{\Phi}_n(\omega)\mathbf{w}(\omega)} \quad (2-11)$$

However, the normalization is important in wide-band applications to keep the array gain consistent at the desired signal array response. In (2-3) and (2-9), it is worth to note that different powers of the designed signal models $\hat{\Phi}_s(\omega)$ lead to different normalization factors. Since the distortionless constraint is to constrain the desired array response without the consideration of signal power, normalization on the PSD matrix $\hat{\Phi}_s(\omega)$ is reasonable. Thus, we modified the distortionless constraint by using the normalized signal model $\check{\Phi}_s(\omega)$ as

$$\mathbf{w}^H(\omega)\check{\Phi}_s(\omega)\mathbf{w}(\omega)=1 \quad (2-12a)$$

where

$$\check{\Phi}_s(\omega)=\frac{M \cdot \hat{\Phi}_s(\omega)}{\text{tr}\{\hat{\Phi}_s(\omega)\}} \quad (2-12b)$$

The factor M in (2-12b) is to keep the norm of the weight vector $\|\mathbf{w}(\omega)\|$ be comparable to the conventional case. For example, consider the rank-1 signal model given in (1-5), the left hand side of the distortionless constraint in (2-12a) is

$$\begin{aligned} \mathbf{w}^H(\omega)\check{\Phi}_s(\omega)\mathbf{w}(\omega) &= \frac{M \cdot \phi_s(\omega) \cdot \mathbf{w}^H(\omega)\mathbf{a}_s(\omega)\mathbf{a}_s^H(\omega)\mathbf{w}(\omega)}{\text{tr}\{\phi_s(\omega)\mathbf{a}_s(\omega)\mathbf{a}_s^H(\omega)\}} \\ &= \frac{M \cdot \phi_s(\omega) \cdot \|\mathbf{w}^H(\omega)\mathbf{a}_s(\omega)\|^2}{\phi_s(\omega) \cdot M} \\ &= \|\mathbf{w}^H(\omega)\mathbf{a}_s(\omega)\|^2 \end{aligned} \quad (2-13)$$

This gives the same norm as the distortionless constraint given in (2-1).

2.3.2 Norm-Constrained Capon Beamforming

Cox *et al.* [2] have shown that the DL problem is equivalent to the norm-constrained Capon beamforming (NCCB) problem [23][26]. For the rank-1 signal model, the NCCB can be expressed by

$$\min_{\mathbf{w}(\omega)} \mathbf{w}^H(\omega) \Phi_x(\omega) \mathbf{w}(\omega) \quad \text{subject to} \quad \begin{cases} \mathbf{w}^H(\omega) \mathbf{a}_s(\omega) = 1 \\ \|\mathbf{w}(\omega)\|^2 = T(\omega) \end{cases} \quad (2-14)$$

where $T(\omega)$ is the designed constraint value of the squared weight vector norm. The solution of the NCCB has the diagonal loading (DL) form as

$$\hat{\mathbf{w}}_{\text{DL}}(\omega) = \frac{(\Phi_x(\omega) + \lambda(\omega) \mathbf{I})^{-1} \mathbf{a}_s(\omega)}{\mathbf{a}_s^H(\omega) (\Phi_x(\omega) + \lambda(\omega) \mathbf{I})^{-1} \mathbf{a}_s(\omega)} \quad (2-15)$$

By substituting the weight vector in (2-15) into the norm constraint, the relationship between the diagonal loading level $\lambda(\omega)$ and the norm-constrained value of the weight vector norm $T(\omega)$ can be obtained by [26]

$$\frac{\mathbf{a}_s^H(\omega) (\Phi_x(\omega) + \lambda(\omega) \mathbf{I})^{-2} \mathbf{a}_s(\omega)}{\left[\mathbf{a}_s^H(\omega) (\Phi_x(\omega) + \lambda(\omega) \mathbf{I})^{-1} \mathbf{a}_s(\omega) \right]^2} = T(\omega) \quad (2-16)$$

Now, the NCCB formulation can be extended with multi-rank signal models as

$$\min_{\mathbf{w}(\omega)} \mathbf{w}^H(\omega) \Phi_x(\omega) \mathbf{w}(\omega) \quad \text{subject to} \quad \begin{cases} \mathbf{w}^H(\omega) \check{\Phi}_s(\omega) \mathbf{w}(\omega) = 1 \\ \|\mathbf{w}(\omega)\|^2 = T(\omega) \end{cases} \quad (2-17)$$

where $\check{\Phi}_s(\omega)$ is the normalized signal PSD matrix given in (2-12b). The Lagrangian function of (2-17) can be defined by

$$J(\mathbf{w}(\omega), \lambda(\omega), \mu(\omega)) = \mathbf{w}^H(\omega) \mathbf{\Phi}_x(\omega) \mathbf{w}(\omega) + \lambda(\omega) (\|\mathbf{w}(\omega)\|^2 - T(\omega)) + \mu(\omega) (1 - \mathbf{w}^H(\omega) \check{\mathbf{\Phi}}_s(\omega) \mathbf{w}(\omega)) \quad (2-18)$$

By letting the derivatives with respect to $\mathbf{w}(\omega)$, $\lambda(\omega)$, and $\mu(\omega)$ be equal to zeros, we obtain the solution similar to (2-10) as

$$\hat{\mathbf{w}}_{\text{DL}}(\omega) = \frac{\check{\mathbf{w}}_{\text{DL}}(\omega)}{\mathcal{C}(\omega)} \quad (2-19a)$$

where

$$\check{\mathbf{w}}_{\text{DL}}(\omega) = \mathcal{P} \left\{ \left(\mathbf{\Phi}_x(\omega) + \lambda(\omega) \mathbf{I} \right)^{-1} \check{\mathbf{\Phi}}_s(\omega) \right\} \quad (2-19b)$$

and

$$\mathcal{C}(\omega) = \sqrt{\check{\mathbf{w}}_{\text{DL}}^H(\omega) \check{\mathbf{\Phi}}_s(\omega) \check{\mathbf{w}}_{\text{DL}}(\omega)} \quad (2-19c)$$

is the normalization factor to meet the distortionless constraint. Likewise, the relationship between $\lambda(\omega)$ and $T(\omega)$ with the multi-rank signal models can be derived as

$$\frac{\|\check{\mathbf{w}}_{\text{DL}}(\omega)\|^2}{\mathcal{C}(\omega)^2} = T(\omega) \quad (2-20)$$

The value of $T(\omega)$ is greater than $1/M$. According to the distortionless constraint, we have

$$\begin{aligned} 1 &= \mathbf{w}^H(\omega) \check{\mathbf{\Phi}}_s(\omega) \mathbf{w}(\omega) \\ &= \text{tr} \left(\check{\mathbf{\Phi}}_s(\omega) \mathbf{w}(\omega) \mathbf{w}^H(\omega) \right) \\ &\leq \text{tr} \left(\check{\mathbf{\Phi}}_s(\omega) \right) \|\mathbf{w}(\omega)\|^2 = M \cdot T(\omega) \end{aligned} \quad (2-21)$$

which gives the lower bound of $T(\omega)$ as

$$T(\omega) \geq \frac{1}{M} \quad (2-22)$$

Note that the trace inequality $\text{tr}(\mathbf{AB}) \leq \text{tr}(\mathbf{A}) \text{tr}(\mathbf{B})$ used in (2-21) is described in Section 3.3.1 in detail. If we assume a semi-positive diagonal loading level $\lambda(\omega)$, there is also an upper bound for $T(\omega)$. The equation (2-20) monotonically decreases to $1/M$ as $\lambda(\omega)$ approaches to ∞ . Thus, the upper bound happens when $\lambda(\omega) = 0$, where $T(\omega)$ should be smaller than

$$T(\omega) \leq \frac{\|\check{\mathbf{w}}_{\text{MVDR}}(\omega)\|^2}{\check{\mathbf{w}}_{\text{MVDR}}^H(\omega) \check{\Phi}_s(\omega) \check{\mathbf{w}}_{\text{MVDR}}(\omega)} \quad (2-23a)$$

where

$$\check{\mathbf{w}}_{\text{MVDR}}(\omega) = \mathcal{P} \left\{ \Phi_x^{-1}(\omega) \check{\Phi}_s(\omega) \right\} \quad (2-23b)$$

If the norm constraint value $T(\omega)$ is greater than the upper bound, there is no feasible solution for a semi-positive $\lambda(\omega)$.

For the rank-1 signal model, the weight vector can be decomposed into subspaces of the presumed steering vector and its null space based on the concept of GSC [71] as

$$\mathbf{w}(\omega) = \frac{\mathbf{a}_s(\omega)}{M} + \mathbf{a}_s^\perp(\omega) \quad (2-24)$$

where $\mathbf{a}_s^H(\omega) \mathbf{a}_s^\perp(\omega) = 0$ and the weight vector satisfies the distortionless constraint

$\mathbf{w}^H(\omega) \mathbf{a}_s(\omega) = 1$. In this case, the norm of the weight vector can be expressed as

$$\|\mathbf{w}(\omega)\|^2 = \frac{1}{M} + \|\mathbf{a}_s^\perp(\omega)\|^2 \quad (2-25)$$

It can be seen that constraining the norm of $\mathbf{w}(\omega)$ is equivalent to constrain the norm of $\mathbf{a}_s^\perp(\omega)$. Thus, to express the effect of the latter term in (2-25), we decompose the

threshold $T(\omega)$ as

$$T(\omega) = \frac{1}{M} + \gamma(\omega), \quad \text{where } \gamma(\omega) \geq 0 \quad (2-26)$$

In the rest of this dissertation, we discuss the selection of $\gamma(\omega)$ instead of $T(\omega)$ for better description of the scale of the norm deviations.

Compared to the selection of the diagonal loading level $\lambda(\omega)$, the selection of $\gamma(\omega)$ is less sensitive to the signal powers due to the division in (2-16) and (2-19). It is worth to note that the speech signals are nonstationary and time-varying. The insensitivity property of $\gamma(\omega)$ benefits the application of speech enhancement. It can be shown in Section 4.1.1 that the selection of norm constraint is also less sensitive to the small angle mismatches at high SNRs.

2.3.3 State Space Formulation Using the SCPO Method

The pseudo-observation method treats the set of constraint equations as additional observations, but with no measurement noise [55–59]. In this case, the constraint equations are called *perfect measurements*, and the constraints are considered as “hard constraint”. However, it is known that perfect measurements give a singular error covariance matrix, which will lead to the ill-conditioning problem in the Kalman filter. Thus, small variances of the constraint equations are used instead and it gives the “soft-constrained” solutions.

Considering the NCCB problem in (2-17), the state space model is given as follows:

State Space Formulation of the Proposed Robust Beamforming Problem

State equation

$$\mathbf{w}(\omega, k+1) = \mathbf{w}(\omega, k) + \mathbf{v}_s(\omega, k) \quad (2-27a)$$

Measurement equations

$$\begin{bmatrix} 0 \\ 1 \\ T(\omega) \end{bmatrix} = \begin{bmatrix} \mathbf{x}^H(\omega, k) \mathbf{w}(\omega, k) \\ f_1(\mathbf{w}(\omega, k)) \\ f_2(\mathbf{w}(\omega, k)) \end{bmatrix} + \begin{bmatrix} v_{m1}(\omega, k) \\ v_{m2}(\omega, k) \\ v_{m3}(\omega, k) \end{bmatrix} \quad (2-27b)$$

where

$$f_1(\mathbf{w}(\omega, k)) = \mathbf{w}^H(\omega, k) \tilde{\Phi}_s(\omega) \mathbf{w}(\omega, k) \quad (2-27c)$$

$$f_2(\mathbf{w}(\omega, k)) = \mathbf{w}^H(\omega, k) \mathbf{w}(\omega, k) \quad (2-27d)$$

A vector form of (2-27b) is expressed as

$$\mathbf{z} = \mathbf{f}(\mathbf{w}(\omega, k)) + \mathbf{v}_m(\omega, k) \quad (2-27e)$$

where $\mathbf{v}_s(\omega, k)$ and $\mathbf{v}_m(\omega, k)$ are the process and measurement noises respectively.

Typically, the noise processes $\mathbf{v}_s(\omega, k)$ and $\mathbf{v}_m(\omega, k)$ are assumed to be zero-mean and mutually uncorrelated with the covariance matrices

$$\mathbf{Q}(\omega) = E_k \left[\mathbf{v}_s(\omega, k) \mathbf{v}_s^H(\omega, k) \right] = \sigma_0^2(\omega) \mathbf{I} \quad (2-28)$$

and

$$\mathbf{R}(\omega) = E_k \left[\mathbf{v}_m(\omega, k) \mathbf{v}_m^H(\omega, k) \right] = \begin{bmatrix} \sigma_1^2(\omega) & 0 & 0 \\ 0 & \sigma_2^2(\omega) & 0 \\ 0 & 0 & \sigma_3^2(\omega) \end{bmatrix} \quad (2-29)$$

The only real measurement in (2-27b) is the input vector $\mathbf{x}(\omega, k)$ in the first equation given by the objective of minimizing the filtered output power in the MSE sense, i.e.,

$$E_k \left[\left| 0 - \mathbf{x}^H(\omega, k) \mathbf{w}(\omega, k) \right|^2 \right].$$

Considering the measurement update in the Kalman filter,

$$\hat{\mathbf{w}}(\omega, k) = \hat{\mathbf{w}}(\omega, k-1) + \mathbf{K}(\omega, k) (\mathbf{z} - \mathbf{f}(\mathbf{w}(\omega, k-1))) \quad (2-30)$$

It can be shown that $\hat{\mathbf{w}}(\omega, k)$ is the solution to the optimization problem [59]

$$\begin{aligned} \hat{\mathbf{w}}_{\text{SCPO}}(\omega, k) = \arg \min_{\mathbf{w}} & \left\{ (\mathbf{w} - \hat{\mathbf{w}}(\omega, k-1))^H [\mathbf{P}^-(\omega, k-1)]^{-1} (\mathbf{w} - \hat{\mathbf{w}}(\omega, k-1)) \right. \\ & + \frac{1}{\sigma_1^2(\omega)} \left\| 0 - \mathbf{x}^H(\omega, k) \mathbf{w} \right\|^2 \\ & + \frac{1}{\sigma_2^2(\omega)} \left\| 1 - \mathbf{w}^H \tilde{\Phi}_s(\omega) \mathbf{w} \right\|^2 \\ & \left. + \frac{1}{\sigma_3^2(\omega)} \left\| T(\omega) - \mathbf{w}^H \mathbf{w} \right\|^2 \right\} \end{aligned} \quad (2-31)$$

where $\mathbf{K}(\omega, k)$ is the Kalman gain and $\mathbf{P}^-(\omega, k)$ is the *a priori* state error covariance matrix. Now the constraint parameters $\sigma_2^2(\omega)$ and $\sigma_3^2(\omega)$ act as penalty terms. When the constraint parameters approach to zeros, the constraint costs are increasingly weighted, and the solutions that do not satisfy the constraint are increasingly penalized. The solution of the SCPO should approach to the solution of the NCCB problem in (2-17) if the constraint parameters $\sigma_2^2(\omega)$ and $\sigma_3^2(\omega)$ are much smaller than $\sigma_1^2(\omega)$ and the approximation of the nonlinear functions are adequate. To avoid numerical problems, typically the constraint parameters will not be set as zeros. Therefore, the SCPO method does not strictly satisfy the constraints, but provides a flexible approach to incorporate different equality constraints.

2.4 Solutions Using Nonlinear Kalman Filters

In the multi-rank MVDR problems, the multi-rank distortionless constraint and the norm constraint are quadratic. Hence, nonlinear approximation on the measurement equations is needed. Consider the nonlinear measurement equation describe in (2-27), it can be approximated by the Taylor expansion in the second-order around an estimate $\hat{\mathbf{w}}(\omega, k)$ as

$$\begin{aligned} \mathbf{f}(\mathbf{w}(\omega, k)) &= \mathbf{f}(\hat{\mathbf{w}}(\omega, k)) + \mathbf{F}_{\mathbf{w}}(\mathbf{w}(\omega, k) - \hat{\mathbf{w}}(\omega, k)) \\ &\quad + \frac{1}{2} \sum_{i=1}^P \phi_i (\mathbf{w}(\omega, k) - \hat{\mathbf{w}}(\omega, k))^H \mathbf{F}_{\mathbf{w}\mathbf{w}}^{(i)} (\mathbf{w}(\omega, k) - \hat{\mathbf{w}}(\omega, k)) \end{aligned} \quad (2-32)$$

where $\phi_i = [0, \dots, 1, \dots, 0]^T$ and P is the number of measurement equations.

$\mathbf{F}_{\mathbf{w}}$ denotes the Jacobian matrix of the nonlinear function $\mathbf{f}(\mathbf{w}(\omega, k))$, and $\mathbf{F}_{\mathbf{w}\mathbf{w}}^{(i)}$ denotes the Hessian matrix of the i -th measurement equation in $\mathbf{f}(\mathbf{w}(\omega, k))$. The major two categories of the nonlinear Kalman filtering are the extended Kalman filter (EKF) and the unscented Kalman filter (UKF). An overview of the above algorithms has been given in [66]:

- 1) The first-order EKF (FOEKF) [60][62] approximates the Jacobian matrix $\mathbf{F}_{\mathbf{w}}$. This works fine as long as the Hessian matrix (i.e., second-order term) is small, which can depend on the state estimation error or the degree of nonlinearity of $\mathbf{f}(\mathbf{w}(\omega, k))$.
- 2) The second-order EKF (SOEKF) [61][62] approximates both the Jacobian matrix and the Hessian matrices.
- 3) The UKF [62–65] implicitly estimates the first- and second-order terms in the nonlinear transformation in (2-32) instead of estimating the Jacobian and

Hessian matrices. In other words, the UKF approximates the probability distribution using *sigma points* rather than approximating an arbitrary nonlinear function or transformation.

In the following, the solutions using the EKF and the UKF for our problems will be listed. Discussions among the algorithms are investigated in the sequel.

2.4.1 Solutions Using Extended Kalman Filters

The EKF has been widely used in nonlinear filtering [60–62]. It approximates the nonlinear function around the *a priori* estimate of the Kalman filter. In our problems, only parts of the measurement equations are nonlinear. Therefore, only Jacobian and Hessian matrices of the nonlinear measurement equations are needed to be estimated.

Given the state space model in Section 2.3.3, the Jacobian matrix $\mathbf{F}_w(\omega, k)$ and the Hessian matrices $\mathbf{F}_{ww}^{(1)}(\omega, k)$ and $\mathbf{F}_{ww}^{(2)}(\omega, k)$ of the nonlinear functions can be computed as

$$\begin{aligned} \mathbf{F}_w(\omega, k) &= \left[\nabla_w \mathbf{f}^T(\mathbf{w}(\omega, k)) \right]^T \\ &= \begin{bmatrix} \mathbf{x}^H(\omega, k) \\ \mathbf{w}^H(\omega, k) \tilde{\Phi}_s(\omega) \\ \mathbf{w}^H(\omega, k) \end{bmatrix} \end{aligned} \quad (2-33a)$$

$$\mathbf{F}_{ww}^{(1)}(\omega, k) = \nabla_w \nabla_w^H \{ \mathbf{x}^H(\omega, k) \mathbf{w}(\omega, k) \} = \mathbf{0} \quad (2-33b)$$

$$\mathbf{F}_{ww}^{(2)}(\omega, k) = \nabla_w \nabla_w^H \{ f_1(\mathbf{w}(\omega, k)) \} = \tilde{\Phi}_s(\omega) \quad (2-33c)$$

$$\mathbf{F}_{ww}^{(3)}(\omega, k) = \nabla_w \nabla_w^H \{ f_2(\mathbf{w}(\omega, k)) \} = \mathbf{I} \quad (2-33d)$$

For the SOEKF, the Hessian matrices in (2-32) leads to the additional terms in the innovation $\boldsymbol{\pi}(\omega, k)$ and its covariance matrix $\boldsymbol{\Lambda}(\omega, k)$ under the MSE sense (see Appendix I). The bias terms $\boldsymbol{\pi}(\omega, k)$ and $\boldsymbol{\Lambda}(\omega, k)$ in our problem can be expressed as

$$\boldsymbol{\pi}(\omega, k) = \frac{1}{2} \begin{bmatrix} 0 \\ \text{tr}(\tilde{\boldsymbol{\Phi}}_s(\omega) \mathbf{P}^-(\omega, k)) \\ \text{tr}(\mathbf{P}^-(\omega, k)) \end{bmatrix} \quad (2-34a)$$

$$\boldsymbol{\Lambda}(\omega, k) = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \text{tr}\{\tilde{\boldsymbol{\Phi}}_s(\omega) \mathbf{P}^-(\omega, k) \tilde{\boldsymbol{\Phi}}_s(\omega) \mathbf{P}^-(\omega, k)\} & \text{tr}\{\tilde{\boldsymbol{\Phi}}_s(\omega) \mathbf{P}^-(\omega, k) \mathbf{P}^-(\omega, k)\} \\ 0 & \text{tr}\{\tilde{\boldsymbol{\Phi}}_s(\omega) \mathbf{P}^-(\omega, k) \mathbf{P}^-(\omega, k)\} & \text{tr}\{\mathbf{P}^-(\omega, k) \mathbf{P}^-(\omega, k)\} \end{bmatrix} \quad (2-34b)$$

Finally, the EKFs using the first- and second-order Taylor expansion can be summarized as follows [62]:

Multi-Rank MVDR Beamformer Using the FOEKF and SOEKF

The SOEKF, using the second-order Taylor expansion, for the state space model in (2-27) is given by the following recursions initialized with $\hat{\mathbf{w}}(\omega, 0)$ and $\mathbf{P}^+(\omega, 0)$:

$$\mathbf{P}^-(\omega, k) = \mathbf{P}^+(\omega, k-1) + \mathbf{Q}(\omega) \quad (2-35a)$$

$$\mathbf{e}(\omega, k) = \mathbf{z} - \mathbf{f}(\hat{\mathbf{w}}(\omega, k-1)) - \boldsymbol{\pi}(\omega, k) \quad (2-35b)$$

$$\mathbf{S}(\omega, k) = \mathbf{F}_w(\omega, k) \mathbf{P}^-(\omega, k) \mathbf{F}_w^H(\omega, k) + \mathbf{R}(\omega) + \boldsymbol{\Lambda}(\omega, k) \quad (2-35c)$$

$$\mathbf{K}(\omega, k) = \mathbf{P}^-(\omega, k) \mathbf{F}_w^H(\omega, k) \mathbf{S}^{-1}(\omega, k) \quad (2-35d)$$

$$\hat{\mathbf{w}}(\omega, k) = \hat{\mathbf{w}}(\omega, k-1) + \mathbf{K}(\omega, k) \mathbf{e}(\omega, k) \quad (2-35e)$$

$$\mathbf{P}^+(\omega, k) = [\mathbf{I} - \mathbf{K}(\omega, k) \mathbf{F}_w(\omega, k)] \mathbf{P}^-(\omega, k) \quad (2-35f)$$

The FOEKF is obtained by letting both $\boldsymbol{\pi}(\omega, k)$ and $\boldsymbol{\Lambda}(\omega, k)$ be zero.

In (2-35), $\mathbf{P}^+(\omega, k)$ is the *a posteriori* state error covariance matrix; $\mathbf{e}(\omega, k)$ and $\mathbf{S}(\omega, k)$ are

the innovation vector and its covariance matrix. For detailed derivation of the SOE Kalman filter, please refer to the Appendix I.

2.4.2 Solution Using the Unscented Kalman Filter

The UKF uses *sigma points* to approximate the first- and second-order moments of the nonlinear transformation. There are different ways to set the *sigma points* and the weightings [62–66]. In this dissertation, we choose the method given by [62] since it gives positive weightings. The $(2M + 1)$ *sigma points* for the approximation of the nonlinear measurement equations are generated by

$$\hat{\mathbf{W}}_{\sigma}(\omega, k) = \left[\underbrace{\hat{\mathbf{w}}(\omega, k-1)}_{M \times 1}, \underbrace{\hat{\mathbf{w}}(\omega, k-1) \cdot \mathbf{1}^T + \sqrt{M\mathbf{P}^{-}(\omega, k)}}_{M \times M}, \underbrace{\hat{\mathbf{w}}(\omega, k-1) \cdot \mathbf{1}^T - \sqrt{M\mathbf{P}^{-}(\omega, k)}}_{M \times M} \right] \quad (2-36a)$$

And the transformed *sigma points* are given by

$$\left[\hat{\mathbf{Z}}_{\sigma}(\omega, k) \right]_i = \mathbf{f} \left(\left[\hat{\mathbf{W}}_{\sigma}(\omega, k) \right]_i \right), \quad i = 0, \dots, 2M \quad (2-36b)$$

where $\mathbf{1}$ denotes the M -by-1 all-one vector.

The UKF is summarized as follows [62]:

Multi-Rank MVDR Beamformer Using the UKF

By using the *sigma points* given in (2-36), the UKF is given by the following recursions initialized with $\hat{\mathbf{w}}(\omega, 0)$ and $\mathbf{P}^{+}(\omega, 0)$:

$$\mathbf{P}^{-}(\omega, k) = \mathbf{P}^{+}(\omega, k-1) + \mathbf{Q}(\omega) \quad (2-37a)$$

$$\bar{\mathbf{z}}(\omega, k) = \frac{1}{2M+1} \sum_{i=0}^{2M} \left[\hat{\mathbf{Z}}_{\sigma}(\omega, k) \right]_i \quad (2-37b)$$

$$\mathbf{P}_{\mathbf{wz}}(\omega, k) = \frac{1}{2M+1} \sum_{i=0}^{2M} \left(\left[\hat{\mathbf{W}}_{\sigma}(\omega, k) \right]_i - \hat{\mathbf{w}}(\omega, k-1) \right) \left(\left[\hat{\mathbf{Z}}_{\sigma}(\omega, k) \right]_i - \bar{\mathbf{z}}(\omega, k) \right)^H \quad (2-37c)$$

$$\mathbf{P}_{\mathbf{zz}}(\omega, k) = \frac{1}{2M+1} \sum_{i=0}^{2M} \left(\left[\hat{\mathbf{Z}}_{\sigma}(\omega, k) \right]_i - \bar{\mathbf{z}}(\omega, k) \right) \left(\left[\hat{\mathbf{Z}}_{\sigma}(\omega, k) \right]_i - \bar{\mathbf{z}}(\omega, k) \right)^H \quad (2-37d)$$

$$\mathbf{e}(\omega, k) = \mathbf{z} - \bar{\mathbf{z}}(\omega, k) \quad (2-37e)$$

$$\mathbf{S}(\omega, k) = \mathbf{P}_{\mathbf{zz}}(\omega, k) + \mathbf{R}(\omega) \quad (2-37f)$$

$$\mathbf{K}(\omega, k) = \mathbf{P}_{\mathbf{wz}}(\omega, k) \mathbf{S}^{-1}(\omega, k) \quad (2-37g)$$

$$\hat{\mathbf{w}}(\omega, k) = \hat{\mathbf{w}}(\omega, k-1) + \mathbf{K}(\omega, k) \mathbf{e}(\omega, k) \quad (2-37h)$$

$$\mathbf{P}^+(\omega, k) = \mathbf{P}^-(\omega, k) - \mathbf{K}(\omega, k) \mathbf{S}(\omega, k) \mathbf{K}^H(\omega, k) \quad (2-37i)$$

2.4.3 Initial Conditions

The setting of initial conditions is important for constrained Kalman filtering problems. The initial conditions should satisfy the constraints or at least close to the feasible sets of the constraint in the order of the chosen variance parameter in matrix $\mathbf{R}(\omega)$. An improper setting can dramatically degrade the performance with nonlinear constraints.

First, consider the rank-1 MVDR beamforming problem in (2-1). Based on the projection method in the constrained Kalman filtering [55–59], the state error covariance matrix $\mathbf{P}^+(\omega, k)$ converges to the null space of the presumed steering vector $\mathbf{a}_s(\omega)$. Therefore, the initial values of $\hat{\mathbf{w}}(\omega, 0)$ and $\mathbf{P}^+(\omega, 0)$ are chosen as

$$\hat{\mathbf{w}}(\omega, 0) = \frac{\mathbf{a}_s(\omega)}{\|\mathbf{a}_s(\omega)\|^2} \quad (2-38a)$$

$$\begin{aligned}
\mathbf{P}^+(\omega, 0) &= \mathbf{I} - \mathbf{a}_s(\omega) [\mathbf{a}_s^H(\omega) \mathbf{a}_s(\omega)]^{-1} \mathbf{a}_s^H(\omega) \\
&= \mathbf{I} - \frac{\mathbf{a}_s(\omega) \mathbf{a}_s^H(\omega)}{\|\mathbf{a}_s(\omega)\|^2}
\end{aligned} \tag{2-38b}$$

The normalization term in (2-38a) is to satisfy the distortionless constraint $\mathbf{w}^H(\omega) \mathbf{a}_s(\omega) = 1$.

Likewise, for multi-rank signal models, the weight vector $\mathbf{w}(\omega)$ is initialized to satisfy the distortionless constraint, and the state error covariance matrix $\mathbf{P}^+(\omega, k)$ can be set as the null space of $\mathbf{w}(\omega)$. Thus, the initial conditions of $\hat{\mathbf{w}}(\omega, 0)$ and $\mathbf{P}^+(\omega, 0)$ for multi-rank case are chosen as

$$\hat{\mathbf{w}}(\omega, 0) = \frac{\mathbf{w}_{\text{matched}}(\omega)}{\sqrt{\mathbf{w}_{\text{matched}}^H(\omega) \tilde{\Phi}_s(\omega) \mathbf{w}_{\text{matched}}(\omega)}} \tag{2-39a}$$

$$\mathbf{P}^+(\omega, 0) = \mathbf{I} - \frac{\mathbf{w}_{\text{matched}}(\omega) \mathbf{w}_{\text{matched}}^H(\omega)}{\|\mathbf{w}_{\text{matched}}(\omega)\|^2} \tag{2-39b}$$

where $\mathbf{w}_{\text{matched}}(\omega)$ is the matched filter given in (2-7).

2.4.4 Estimation of the Parameter Matrices

In the update equations of the proposed Kalman filters, there are two parameter matrices to be determined: $\mathbf{Q}(\omega)$ and $\mathbf{R}(\omega)$. In general, $\mathbf{Q}(\omega)$ stands for the *random walk* during the state update, which is typically assumed as stochastically white. For stationary environments, $\sigma_0^2(\omega) = 0$ can be chosen. The larger the parameter $\sigma_0^2(\omega)$ is chosen, the larger *random walk* of the state is allowed. That is, the state variation can be large to track the nonstationary environmental changes. Second, $\mathbf{R}(\omega)$ corresponds to the error variances of each measurement. In (2-29), $\sigma_1^2(\omega)$ corresponds to the average output power. It is suggested to be the same order of the optimal output power of the

array (which can be roughly approximated as $\|\mathbf{w}(\omega)\|^2 (M\phi_s(\omega) + \phi_v(\omega))$ [68], where $\phi_s(\omega)$ and $\phi_v(\omega)$ are the PSDs of the desired signal and sensor noise, respectively); however, the PSD of the desired signal is not known *a priori* and it is related to the norm of the state $\mathbf{w}(\omega)$. $\sigma_2^2(\omega)$ and $\sigma_3^2(\omega)$ correspond to the augmented distortionless and norm constraints. $\sigma_2^2(\omega)$ controls the fitness of the distortionless constraint in (2-12). When $\sigma_2^2(\omega)$ approaches to zero, the beamformer approaches to the matched filter in (2-39a), which is distortionless to the presumed signal model but fails to reject the interferences. And $\sigma_3^2(\omega)$ controls the fitness of the norm constraint. The norm constraint controls the sidelobe level of the beamformer. When the norm constraint value $\gamma(\omega)$ is small and $\sigma_3^2(\omega)$ approaches to zero, the beamformer emphasizes on reducing the sidelobe level instead of rejecting interferences.

Since there is only one “true measurement” in the measurement equations, we propose only to estimate the parameter $\sigma_1^2(\omega)$ and consider $\sigma_2^2(\omega)$ and $\sigma_3^2(\omega)$ as adjustable parameters to control the tradeoff between signal distortion and interference rejection. Contrary to the adaptive beamforming problems [67][68], who suggested to set the variance parameters corresponding to the constraints as a very small value 10^{-10} , it is suggested in this dissertation to set an “properly” small value if there are some tradeoffs between the constraint sets. For such case, if both $\sigma_2^2(\omega)$ and $\sigma_3^2(\omega)$ are set very small, the problem may not be feasible. For the uniform linear array (ULA) with 10 elements and half-wavelength spacing, 10^{-5} is a good choice for the tradeoff between the distortionless and norm constraints. Further, to avoid the ill-conditioning problem, the condition number of the parameter matrix $\mathbf{R}(\omega)$ should be controlled. In this case, we propose to use the following form

$$\hat{\mathbf{R}}(\omega) = \hat{\sigma}_1^2(\omega) \begin{bmatrix} 1 & 0 & 0 \\ 0 & \kappa & 0 \\ 0 & 0 & \kappa \end{bmatrix} \quad (2-40)$$

where $10^{-10} < \kappa \leq 1$ is the small variance set for the constraints. In this case, the condition number of $\hat{\mathbf{R}}(\omega)$ is guaranteed to be smaller than 10^{10} .

The parameter $\sigma_1^2(\omega)$ is estimated in a recursive way as

$$\hat{\sigma}_1^2(\omega, k) = \alpha \cdot \hat{\sigma}_1^2(\omega, k-1) + (1-\alpha) \cdot e_1(\omega, k)e_1^*(\omega, k) \quad (2-41)$$

where α is the forgetting factor close to unity and $e_1(\omega, k)$ is the first element of the innovation vector $\mathbf{e}(\omega, k)$. The recursive approach is similar to the work [72]. Compared to the variance estimation which used the *a priori* state error covariance matrix $\mathbf{P}^-(\omega, k)$ [73][74], the recursive estimator in (2-41) depends on only the innovation and is not affected by the initial conditions. This guarantees the stability of the whole system. Typically, the parameter $\sigma_1^2(\omega)$ is initialized as zero. For the simulations, $\alpha = 0.9$ is chosen.

2.5 Summary

In this chapter, robust beamformers with multi-rank signal models based on the Kalman filter are proposed. The original multi-rank MVDR beamforming in [20] is modified with normalized signal models and the norm constraint and transformed into state spaces using the SCPO method. The relationship between the diagonal loading level and the norm constraint value for multi-rank signal models is analyzed, and the problem formulation of the SCPO is given to point out the difference from the original NCCB problem. The settings of initial conditions and parameter matrices are also

studied for our problem. Since the modified problems are quadratic, nonlinear Kalman filters including the first- and second-order extended Kalman filters and the unscented Kalman filter are conducted for on-line implementations.



Chapter 3

Multi-Channel Post-Filtering Based on Spatial Coherence Measure

3.1 Introduction

As discussed in Section 1.1.2, most multi-channel post-filtering algorithms use the spectral densities or phase errors between microphone pairs rather than the whole array elements at one time. For the category of spectral density estimators, the state-of-the-art post-filtering are Zelinski [46] and McCowan [47] post-filters. Under the assumptions of homogeneous sound fields and point source model (rank-1 model), the Zelinski post-filter is equal to the optimal Wiener filter in the incoherent noise field (i.e., spatially white noise). McCowan *et al* generalized the Zelinski post-filter by considering the noise field coherence between microphone pairs. This makes their post-filter to be equal to the Wiener filter for any noise fields if the true noise field coherence is given. However, the assumptions of homogeneous sound fields and rank-1 model are not satisfied in the real-world according to the existence of the local scattering, wavefront fluctuation, or reverberation.

In this dissertation, a new spatial measure is defined on a microphone array which leads to a novel post-filtering algorithm (named *spatial coherence based post-filter*, SCPF). The post-filter belongs to the class of spectral densities estimators (which is inherent in the estimation of the input PSD matrix), while it is guaranteed to lie in the range of $[0, 1]$. Further, the proposed spatial coherence measure can be easily extended to multi-rank signal models encompassing incoherently scattered source etc. It is more convenient to consider various design requirements than previous methods using microphone array.

However, a bias term due to the similarity of the desired signal field and the noise field deteriorates the noise reduction performance. As a result, a bias compensated method is proposed (called *bias compensated spatial coherence based post-filter*, BC-SCPF). It can be shown that the BC-SCPF is equivalent to the optimal Wiener filter if the bias or the noise PSD matrix is perfectly measured.

The remainder of this chapter is organized as follows. In Section 3.2, we briefly review the Zelinski and McCowan post-filters for subsequent comparisons. Section 3.3 introduces a trace inequality and its induced spatial coherence measure. Section 3.4 presents the formulation and mean square error (MSE) analysis of the proposed SCPF. The bias compensated solution is proposed in Section 3.5. Finally, a summary is drawn in Section 3.6.

3.2 Brief Review of Zelinski and McCowan Post-Filters

The Zelinski post-filter [46] and the McCowan post-filter [47] are state-of-the-art multi-channel post-filters. In this dissertation, comparisons of the proposed post-filters with the Zelinski and McCowan post-filters are carried out theoretically and empirically. Thus, a brief review of the post-filters is introduced here.

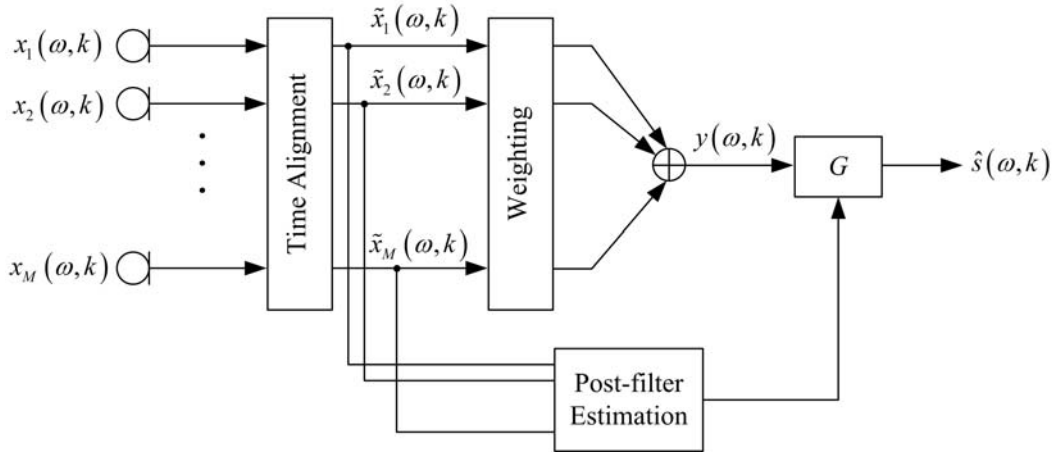


Figure 3-1 System architecture of the Zelinski and McCowan post-filters.

McCowan *et al.* [47] proposed a multi-channel post-filter as a modification of the Zelinski post-filter [46]. The system architecture of the Zelinski and McCowan post-filters is depicted in Figure 3-1. In their systems, the microphones have to pass a time alignment module to adjust the propagation of the desired source between microphones before the post-filter estimation, which is equivalent to the information in the presumed steering vector \mathbf{a}_s . Let us denote the pre-processed input vector after the time alignment module as

$$\tilde{\mathbf{x}}(\omega, k) = \mathbf{x}(\omega, k) \circ \mathbf{a}_s(\omega) \quad (3-1)$$

where $\mathbf{x}(\omega, k)$ and $\tilde{\mathbf{x}}(\omega, k)$ are the microphone input vector before and after the time alignment module, as shown in Figure 3-1; \circ denotes the Schur-Hadamard (elementwise) matrix product.

Compared to the Zelinski post-filter, the work in [47] considered a generalized coherence function to describe the characteristics of the noise field on the aligned inputs. Noises between sensors can be coherent (or correlated). The noise coherence function of the time aligned inputs is defined as

$$\tilde{\Gamma}_{n_i n_j}(\omega) = \tilde{\phi}_{n_i n_j}(\omega) / \sqrt{\tilde{\phi}_{n_i n_i}(\omega) \cdot \tilde{\phi}_{n_j n_j}(\omega)} \quad (3-2)$$

where $\tilde{\phi}_{n_i n_j}(\omega)$ is the cross-spectral density between the noises at the i -th and j -th microphones. Note that the diagonal terms of $\tilde{\Gamma}_n(\omega)$ are 1 and its trace equals to M . In their works, the homogeneous sound fields are assumed. That is, the sources have the same power spectrum at each sensor. Based on this assumption, the spectral densities of the aligned inputs are expressed as [47]

$$\tilde{\phi}_{x_i x_i}(\omega) = \tilde{\phi}_s(\omega) + \tilde{\phi}_n(\omega) \quad (3-3)$$

$$\tilde{\phi}_{x_j x_j}(\omega) = \tilde{\phi}_s(\omega) + \tilde{\phi}_n(\omega) \quad (3-4)$$

$$\tilde{\phi}_{x_i x_j}(\omega) = \tilde{\phi}_s(\omega) + \tilde{\Gamma}_{n_i n_j}(\omega) \tilde{\phi}_n(\omega) \quad (3-5)$$

where $\tilde{\phi}_s(\omega)$, $\tilde{\phi}_n(\omega)$ are the aligned power spectral densities of the desired signal and noise. According to (3-3)–(3-5), the signal power spectral density can be estimated as

$$\tilde{\phi}_s^{(i,j)}(\omega) = \frac{\Re\{\tilde{\phi}_{x_i x_j}(\omega)\} - \frac{1}{2} \Re\{\tilde{\Gamma}_{n_i n_j}(\omega)\}(\tilde{\phi}_{x_i x_i}(\omega) + \tilde{\phi}_{x_j x_j}(\omega))}{(1 - \Re\{\tilde{\Gamma}_{n_i n_j}(\omega)\})} \quad (3-6)$$

where $\tilde{\phi}_{x_i x_j}(\omega)$ is the cross-spectral density between the i -th and j -th aligned inputs and $\Re\{\cdot\}$ is the real operator. The spectral densities can be estimated using a first-order recursive filter. Equation (3-6) can be explained as removing the highly coherent part of the cross-spectral density and then compensating the residual.

The estimation can be improved by averaging the solutions over all sensor combinations, resulting in the post-filter

$$G_{\text{McCowan}}(\omega) = \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M \tilde{\phi}_s^{(i,j)}(\omega)}{\frac{1}{M} \sum_{i=1}^M \tilde{\phi}_{x_i x_i}(\omega)} \quad (3-7)$$

This technique significantly improves the noise reduction in the diffuse noise field, and can be applied to any noise field by modeling the complex coherence function. When the noise field is incoherent, it reduces to the Zelinski post-filter as

$$G_{\text{Zelinski}}(\omega) = \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M \Re\{\tilde{\phi}_{x_i x_j}(\omega)\}}{\frac{1}{M} \sum_{i=1}^M \tilde{\phi}_{x_i x_i}(\omega)} \quad (3-8)$$

3.3 Proposed Spatial Coherence Measure

3.3.1 Definition of the Spatial Coherence Measure

It is known that the trace of the power spectral density (PSD) matrix, obtained from a sensor array, is the summation of the signal powers. This motivates us to use the trace operation to design a coherence measure between two PSD matrices. Let matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{M \times M}$ be positive semi-definite (which also ensures Hermitian), the trace inequality is established as [75]

$$\text{tr}(\mathbf{AB})^p \leq \left\{ \text{tr}(\mathbf{A})^{2p} \text{tr}(\mathbf{B})^{2p} \right\}^{1/2} \quad (3-9)$$

where $\text{tr}(\mathbf{A})$ denotes the trace of matrix \mathbf{A} , and p is an integer. Considering the special case when $p = 1$, we have

$$\text{tr}(\mathbf{AB}) \leq \text{tr}(\mathbf{A}) \text{tr}(\mathbf{B}) \quad (3-10)$$

Based on (3-10), the *spatial coherence measure* between PSD matrices \mathbf{A} and \mathbf{B} is

defined as

$$\mathcal{F}(\mathbf{A}, \mathbf{B}) \equiv \frac{\text{tr}(\mathbf{A}\mathbf{B})}{\text{tr}(\mathbf{A})\text{tr}(\mathbf{B})} \quad (3-11)$$

According to the matrix version of inner product and Kronecker product, (3-11) can be written as

$$\mathcal{F}(\mathbf{A}, \mathbf{B}) = \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\text{tr}(\mathbf{A} \otimes \mathbf{B})} \quad (3-12)$$

where $\langle \mathbf{A}, \mathbf{B} \rangle$ denotes the inner product of PSD matrices \mathbf{A} and \mathbf{B} , and \otimes denotes the Kronecker product. The inner product measures the similarity among the bases in the matrices, and the trace of the Kronecker product gives the normalization. From the positive semi-definite property of the matrices and the inequality given by (3-10), the *spatial coherence measure* $\mathcal{F}(\mathbf{A}, \mathbf{B})$ is guaranteed to be mapped in the interval $[0, 1]$. Since the PSD matrix represents the signal field measured by the sensor array (in the second-order statistics), the proposed *spatial coherence measure* in (3-11) gives the “closeness” between two *measured signal fields* (named MSF hereafter).

3.3.2 Properties of Proposed Spatial Coherence Measure

The PSD matrices can be decomposed as,

$$\mathbf{A} = \sum_{i=1}^M \sigma_i^2(\mathbf{A}) \mathbf{u}_i(\mathbf{A}) \mathbf{u}_i^H(\mathbf{A}) \quad \text{and} \quad \mathbf{B} = \sum_{j=1}^M \sigma_j^2(\mathbf{B}) \mathbf{u}_j(\mathbf{B}) \mathbf{u}_j^H(\mathbf{B}) \quad (3-13)$$

where $\sigma_i^2(\mathbf{A})$ and $\mathbf{u}_i(\mathbf{A})$ denote the i -th eigenvalue and eigenvector of the PSD matrix \mathbf{A} , respectively. By (3-13), the *spatial coherence measure* can be rewritten as

$$\mathcal{F}(\mathbf{A}, \mathbf{B}) = \frac{\sum_{i=1}^M \sum_{j=1}^M \sigma_i^2(\mathbf{A}) \sigma_j^2(\mathbf{B}) |\mathbf{u}_i^H(\mathbf{A}) \mathbf{u}_j(\mathbf{B})|^2}{\sum_{i=1}^M \sigma_i^2(\mathbf{A}) \cdot \sum_{j=1}^M \sigma_j^2(\mathbf{B})} \quad (3-14)$$

It can be seen that the coherence measure is the weighted similarity of the bases, and the eigenvalues give the weighting on each basis. When two MSFs belong to the same 1-dimensional subspace, the *spatial coherence measure* gives a measure of unity. As one of the MSF's dimension increases, the *spatial coherence measure* decreases according to the normalization of eigenvalues. Therefore, given PSD matrices \mathbf{A} and \mathbf{B} , several properties of the proposed *spatial coherence measure* can be listed below:

Property 1: If \mathbf{A} belongs to the null-space of \mathbf{B} , then $\mathcal{F}(\mathbf{A}, \mathbf{B}) = 0$.

Property 2: If \mathbf{A} is rank-1, the self-coherence measure $\mathcal{F}(\mathbf{A}, \mathbf{A}) = 1$. As the eigenvalue spread of \mathbf{A} increases, the $\mathcal{F}(\mathbf{A}, \mathbf{A})$ decreases to $1/M$ until the eigenvalue spread is uniform (i.e., incoherent field, $\mathbf{A} = \sigma^2 \mathbf{I}$ where σ^2 is the signal power).

Property 3: If \mathbf{A} or \mathbf{B} is an incoherent field, then the *spatial coherence measure* equals to a constant value of $\mathcal{F}(\mathbf{A}, \mathbf{B}) = 1/M$ (It can be easily observed from (3-11)).

From *Property 1*, consider \mathbf{A} as the PSD matrix of the desired MSF and \mathbf{B} as the PSD matrix measured by the microphone array. Then $\mathcal{F}(\mathbf{A}, \mathbf{B}) = 0$ could be interpreted as the signals of the microphones do not contain the target source information. Thus, if a multiplicative gain of a post-filter is designed, the gain should be zero. For *Property 2*, the self-coherence measure $\mathcal{F}(\mathbf{A}, \mathbf{A})$ is derived from (3-14) as

$$\mathcal{F}(\mathbf{A}, \mathbf{A}) = \frac{\sum_{i=1}^M \sigma_i^4(\mathbf{A})}{\left(\sum_{i=1}^M \sigma_i^2(\mathbf{A})\right)^2} \quad (3-15)$$

where $\mathcal{F}(\mathbf{A}, \mathbf{A})$ is purely determined by the eigenvalues of \mathbf{A} . According to the natural of coherent speech sources, the eigenvalue spread of the desired MSF typically condenses on some low-dimensional subspace. Therefore, if $\mathcal{F}(\mathbf{A}, \mathbf{B})$ is used as a multiplicative post-filtering gain, the gain approaches to unity when there is only a desired signal.

3.4 The Proposed Spatial Coherence Based Post-Filter (SCPF)

3.4.1 A Description of the Proposed SCPF

The proposed post-filter is designed by comparing the input PSD matrix $\Phi_x(\omega)$ with a desired one $\Phi_s(\omega)$ as,

$$G_{\text{SCPF}}(\omega) = \frac{\text{tr}(\Phi_s(\omega)\Phi_x(\omega))}{\text{tr}(\Phi_s(\omega)) \cdot \text{tr}(\Phi_x(\omega))} \quad (3-16a)$$

where

$$\Phi_x(\omega) = E_k [\mathbf{x}(\omega, k) \mathbf{x}^H(\omega, k)] = \begin{bmatrix} \phi_{x_0 x_0}(\omega) & \phi_{x_0 x_1}(\omega) & \cdots & \phi_{x_0 x_{M-1}}(\omega) \\ \phi_{x_1 x_0}(\omega) & \phi_{x_1 x_1}(\omega) & \cdots & \phi_{x_1 x_{M-1}}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{x_{M-1} x_0}(\omega) & \phi_{x_{M-1} x_1}(\omega) & \cdots & \phi_{x_{M-1} x_{M-1}}(\omega) \end{bmatrix} \quad (3-16b)$$

and $\phi_{x_i x_j}(\omega)$ is the cross-spectral density between the inputs at the i -th and j -th microphones. The post-filter uses the measure directly as the gain function and is called

spatial coherence based post-filter (SCPF). In practice, the PSD matrix $\Phi_x(\omega)$ can be estimated using a first-order recursive update formula

$$\hat{\Phi}_x(\omega, k) = \alpha \hat{\Phi}_x(\omega, k-1) + (1-\alpha) \mathbf{x}(\omega, k) \mathbf{x}^H(\omega, k) \quad (3-17)$$

where α is the forgetting factor close to unity. In practice, the desired MSF $\Phi_s(\omega)$ can be estimated empirically from the clean signal recordings of the microphone array. It is worth to note that the usage of $\phi_s(\omega)$ in (1-5)–(1-7) is not crucial since it is cancelled during the normalization of the *spatial coherence measure*.

In order to compare with the previous algorithms, we consider the special case as the following:

- 1) The desired MSF is assumed to be a point source.
- 2) The sound fields are assumed to be homogeneous [33].

According to these conditions, the theoretical PSD matrix can be expressed as

$$\Phi_x(\omega) = \phi_s(\omega) \mathbf{a}_s(\omega) \mathbf{a}_s^H(\omega) + \phi_n(\omega) \Gamma_n(\omega) \quad (3-18)$$

where $\phi_s(\omega)$, $\phi_n(\omega)$ are the power spectral densities of the desired signal and noise; and $\Gamma_n(\omega)$ denotes the coherence matrix of the noise field. The manifold vector is usually selected such that $\|\mathbf{a}_s(\omega)\|^2 = M$. Note that $\text{tr}(\Gamma_n(\omega)) = M$. With the desired MSF and the theoretical PSD matrix, the SCPF can be expressed by

$$\begin{aligned} G_{\text{SCPF}}(\omega) &= \frac{\phi_s(\omega) \mathbf{a}_s^H(\omega) \Phi_x(\omega) \mathbf{a}_s(\omega)}{\phi_s(\omega) \|\mathbf{a}_s(\omega)\|^2 \cdot \text{tr}(\Phi_x(\omega))} \\ &= \frac{\phi_s^2(\omega) \|\mathbf{a}_s(\omega)\|^4 + \phi_s(\omega) \phi_n(\omega) \mathbf{a}_s^H(\omega) \Gamma_n(\omega) \mathbf{a}_s(\omega)}{\phi_s(\omega) \|\mathbf{a}_s(\omega)\|^2 \cdot (\phi_s(\omega) \|\mathbf{a}_s(\omega)\|^2 + \phi_n(\omega) \text{tr}(\Gamma_n(\omega)))} \\ &= G_{\text{Wiener}}(\omega) + \frac{\mathcal{B}(\omega)}{M^2 (\xi(\omega) + 1)} \end{aligned} \quad (3-19a)$$

where $G_{\text{Wiener}}(\omega)$ is the optimal Wiener filter as,

$$G_{\text{Wiener}}(\omega) = \frac{\xi(\omega)}{\xi(\omega) + 1} \quad (3-19b)$$

and $\xi(\omega) = \phi_s(\omega) / \phi_n(\omega)$ denotes the SNR. The term $\mathcal{B}(\omega)$ denotes the inner product of the coherence matrices of the desired signal field and the noise field.

$$\begin{aligned} \mathcal{B}(\omega) &= \mathbf{a}_s^H(\omega) \mathbf{\Gamma}_n(\omega) \mathbf{a}_s(\omega) \\ &= \text{tr}(\mathbf{\Gamma}_n(\omega) \mathbf{\Gamma}_s(\omega)) \\ &= \langle \mathbf{\Gamma}_s(\omega), \mathbf{\Gamma}_n(\omega) \rangle \end{aligned} \quad (3-20)$$

This can be treated as a bias term to the optimal Wiener filter as shown in (3-19b). Note that $\mathcal{B}(\omega)$ lies in the following range for all kinds of noise fields when the desired MSF is chosen as (1-5):

$$0 \leq \mathcal{B}(\omega) < M^2 \quad (3-21)$$

The lower bound happens when the noise subspace lies in the null-space of the desired MSF, while the supremum happens when the noise MSF is identical to the desired MSF under the rank-1 signal model. Obviously, the SCPF is a function of the SNR and it reduces to the Wiener filter when $\mathcal{B}(\omega) = 0$.

3.4.2 Mean Square Error Analysis of Proposed SCPF

The mean square error (MSE) corresponding to the desired signal in the reference channel can be defined as

$$\text{MSE}(\omega) = E_k \left[\left| \hat{s}(\omega, k) - s(\omega, k) \right|^2 \right] \quad (3-22)$$

where $\hat{s}(\omega, k)$ is the enhanced signal given by a beamformer or a post-filter. Applying

the SCPF on the reference microphone (microphone 1) results in the following MSE,

$$\begin{aligned}
\text{MSE}_{\text{SCPF}}(\omega) &= E_k \left[\left| G_{\text{SCPF}}(\omega) x_1(\omega, k) - s(\omega, k) \right|^2 \right] \\
&= E_k \left[\left| (G_{\text{SCPF}}(\omega) - 1) s(\omega, k) + G_{\text{SCPF}}(\omega) n_1(\omega, k) \right|^2 \right] \\
&= \left| G_{\text{SCPF}}(\omega) - 1 \right|^2 \phi_s(\omega) + G_{\text{SCPF}}^2(\omega) \phi_n(\omega)
\end{aligned} \tag{3-23}$$

By substituting (3-19) into (3-23), we have

$$\text{MSE}_{\text{SCPF}}(\omega) = G_{\text{Wiener}}(\omega) \phi_n(\omega) + \frac{\mathcal{B}^2(\omega)}{M^4 (\xi(\omega) + 1)} \phi_n(\omega) \tag{3-24}$$

It can be shown that the Zelinski post-filter [46] is related to SCPF as: (see Appendix II)

$$G_{\text{Zelinski}}(\omega) = \frac{G_{\text{SCPF}}(\omega) - 1/M}{1 - 1/M} \tag{3-25}$$

By substituting the SCPF in (3-19) into (3-25), we have

$$\begin{aligned}
G_{\text{Zelinski}}(\omega) &= \frac{M \left[(M-1) \xi(\omega) - 1 \right] + \mathcal{B}(\omega)}{M(M-1)(\xi(\omega) + 1)} \\
&= G_{\text{Wiener}}(\omega) - \frac{M - \mathcal{B}(\omega)}{M(M-1)(\xi(\omega) + 1)}
\end{aligned} \tag{3-26}$$

In (3-26), it reveals that the Zelinski post-filter gives a negative gain $-1/(M-1)$ when $\mathcal{B}(\omega) = 0$ and $\xi(\omega) = 0$. The negative gain will introduce unwanted phase flips and leaves some noisy time-frequency blocks in the post-filter output. Similarly, the MSE of the Zelinski post-filter can be derived by following the derivation in (3-23) as

$$\begin{aligned}
\text{MSE}_{\text{Zelinski}}(\omega) &= \left| G_{\text{Zelinski}}(\omega) - 1 \right|^2 \phi_s(\omega) + G_{\text{Zelinski}}^2(\omega) \phi_n(\omega) \\
&= G_{\text{Wiener}}(\omega) \phi_n(\omega) + \frac{(M - \mathcal{B}(\omega))^2}{M^2 (M-1)^2 (\xi(\omega) + 1)} \phi_n(\omega)
\end{aligned} \tag{3-27}$$

The bias terms in (3-24) and (3-27) reveals interesting differences between the proposed SCPF and the Zelinski post-filter for different noise fields:

- 1) $\mathcal{B}(\omega) = M$: the noise field is incoherent, i.e., $\mathbf{\Gamma}_n(\omega) = \mathbf{I}$. In this case, the Zelinski post-filter reduces to the optimal Wiener filter and the proposed SCPF has additional term as $\phi_n(\omega) / [M^2(\xi(\omega)+1)]$.
- 2) $\mathcal{B}(\omega) = 0$: the noise field belongs to the null-space of $\mathbf{a}_s(\omega)$. In this case, on the contrary, the proposed SCPF reduces to the optimal Wiener filter and the Zelinski post-filter has additional term as $\phi_n(\omega) / [(M-1)^2(\xi(\omega)+1)]$.

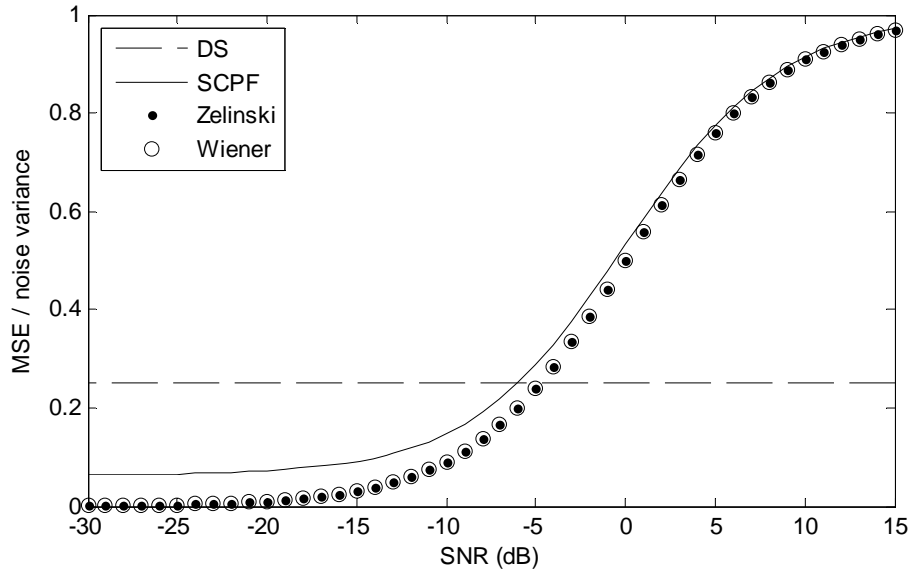
It is worth to note that when the rank-1 desired MSF is chosen, the proposed SCPF is a special case of the post-filter algorithm [5]. In this case, the proposed SCPF can be explained as the ratio of the output power of the delay-and-sum (DS) beamformer to the sum of the input power. Note that the Zelinski and McCowan post-filters also belong to the same family.

It is also interesting to analyze the MSE of the DS beamformer. Given the DS beamformer as $\mathbf{w}_{\text{DS}}(\omega) = \mathbf{a}_s(\omega) / \|\mathbf{a}_s(\omega)\|$, which introduces no distortion on the desired signal and the MSE thereof is derived as

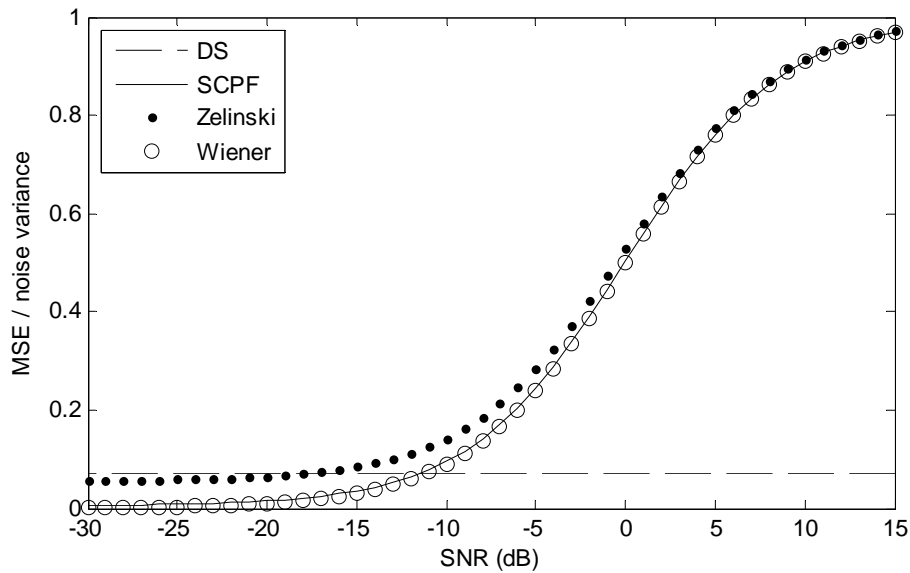
$$\begin{aligned}
\text{MSE}_{\text{DS}}(\omega) &= E_k \left[\left| \mathbf{w}_{\text{DS}}^H(\omega) \mathbf{x}(\omega, k) - s(\omega, k) \right|^2 \right] \\
&= E_k \left[\left| \frac{\mathbf{a}_s^H(\omega) \mathbf{n}(\omega, k)}{M} \right|^2 \right] \\
&= \phi_n \cdot \frac{\mathcal{B}(\omega)}{M^2}
\end{aligned} \tag{3-28}$$

It can be seen that the MSE of the DS beamformer is independent of the desired signal.

Before ending this section, we give an illustration to show the difference between the DS beamformer, the Zelinski post-filter, and the SCPF in the MSE sense. Consider a uniformly distributed linear array (ULA) with $M = 4$ sensors spaced at half-wavelength distance. The $\mathbf{\Gamma}_s(\omega)$ described in (1-5) was steered at $\theta = 0^\circ$.



(a) incoherent noise field



(b) a coherent noise impinged from $\theta = 45^\circ$

Figure 3-2 MSE comparison among the post-filters and the DS beamformer.

Two different noise fields were analyzed: a coherent interference impinging into the array at $\theta = 45^\circ$ and the incoherent white sensor noise. In Figure 3-2, compared to the DS beamformer, it can be seen that both the post-filters attenuate more noise component at low SNRs and preserves more noise at high SNRs. Since speeches are highly nonstationary signal, the post-filters are able to give aggressive noise reduction at low SNRs, especially in the case of incoherent noise fields.

3.5 The Proposed Bias Compensated SCPF (BC-SCPF)

3.5.1 Derivation of Proposed BC-SCPF

Recall from (3-19), the noise reduction ability of the proposed SCPF is limited due to the additional term $\mathcal{B}(\omega) / [M^2(\xi(\omega)+1)]$. Since $\mathcal{B}(\omega)$ is the inner product of the coherence matrices of the desired signal field and the noise field, its effect becomes significant at low frequencies where the similarity between coherence matrices is high due to the insufficient spatial sampling. This can happen both in beamforming and multi-channel post-filtering techniques. When the desired signal is absent in the data, the optimal Wiener filter gives a zero gain, which completely removes the noise. However, the SCPF gives a gain of

$$G_{\text{SCPF}}(\omega) \Big|_{\xi(\omega)=0} = \frac{\text{tr}(\Phi_s(\omega)\Phi_n(\omega))}{\text{tr}(\Phi_s(\omega)) \cdot \text{tr}(\Phi_n(\omega))} \equiv \beta_\omega \quad (3-29)$$

where $\Phi_n(\omega)$ is the noise PSD matrix. Under the assumptions of homogeneous sound fields and point source model, the bias β_ω can be expressed as

$$\beta_\omega = \frac{\mathbf{a}_s^H(\omega)\Gamma_n(\omega)\mathbf{a}_s(\omega)}{M^2} = \frac{\mathcal{B}(\omega)}{M^2} \quad (3-30)$$

Since $\Phi_s(\omega)$ is designed *a priori*, the bias term β_ω only depends on the noise PSD matrix $\Phi_n(\omega)$.

To decrease the effect of the bias β_ω , an intuitive way is to remove the bias and compensate the gain to map the value in the range of [0,1]. The result is called biased-compensated SCPF (BC-SCPF) as the following,

$$G_{\text{BC-SCPF}}(\omega) = \frac{G_{\text{SCPF}}(\omega) - \beta_\omega}{1 - \beta_\omega} \quad (3-31)$$

Note that the bias β_ω lies in the following range for all kinds of noise fields according to

the range of $\mathcal{B}(\omega)$ given in (3-21)

$$1/M \leq \beta_\omega < 1 \quad (3-32)$$

By substituting (3-19) and (3-30) into (3-31), we have

$$\begin{aligned} G_{\text{BC-SCPF}}(\omega) &= \left[\frac{M^2 \xi(\omega) + \mathcal{B}(\omega)}{M^2 (\xi(\omega) + 1)} - \frac{\mathcal{B}(\omega)}{M^2} \right] \bigg/ \left(1 - \frac{\mathcal{B}(\omega)}{M^2} \right) \\ &= \left[\frac{\xi(\omega)(M^2 - \mathcal{B}(\omega))}{M^2 (\xi(\omega) + 1)} \right] \bigg/ \left(\frac{M^2 - \mathcal{B}(\omega)}{M^2} \right) \\ &= G_{\text{Wiener}}(\omega) \end{aligned} \quad (3-33)$$

This gives the optimal Wiener filter if the noise field coherence is perfectly measured. In essence, the BC-SCPF amplifies the small spatial deviation at low frequencies. It is also worth to note that the Zelinski post-filter is a special case of the proposed BC-SCPF with $\beta_\omega = 1/M$ according to (3-25).

For the bias estimation, (3-29) can be used if the PSD matrices of the desired signal and the noise, $\Phi_s(\omega)$ and $\Phi_n(\omega)$, can be obtained in the training process. For the special case of the homogeneous sound fields, the information given by the noise PSD matrix $\Phi_n(\omega)$ equals to that of the noise coherence matrix $\Gamma_n(\omega)$. Furthermore, on-line implementation of the bias estimation can be achieved since the bias is the smallest gain of proposed SCPF at each discrete frequency if the noise field does not change. Thus, the minimum tracking skills [29–31] can be conducted and implemented on-line.

3.5.2 Comparison between BC-SCPF and McCowan Post-Filter

Under the assumption of homogeneous sound field and rank-1 signal model, the McCowan post-filter has been derived as [47]

$$G_{\text{McCowan}}(\omega) = G_{\text{Wiener}}(\omega) + \underbrace{\frac{\phi_s(\omega)}{\phi_s(\omega) + \phi_n(\omega)} \left[\frac{2}{M(M-1)} \Re \left\{ \sum_{i=1}^{M-1} \sum_{j=i+1}^M \frac{\tilde{\Gamma}_{n,n_j}(\omega) - \hat{\Gamma}_{n,n_j}(\omega)}{1 - \hat{\Gamma}_{n,n_j}(\omega)} \right\} \right]}_{e_1} \quad (3-34)$$

where $\tilde{\Gamma}_{n,n_j}(\omega)$ and $\hat{\Gamma}_{n,n_j}(\omega)$ are the actual and estimated noise coherence matrices of the aligned inputs. From (3-34) it can be easily seen that the McCowan post-filter reduces to the Wiener filter when the noise coherence matrix is perfectly measured.

Similarly, the proposed BC-SCPF can be expressed with $\tilde{\Gamma}_{n,n_j}(\omega)$ and $\hat{\Gamma}_{n,n_j}(\omega)$ as following equation (see Appendix III)

$$G_{\text{BC-SCPF}}(\omega) = G_{\text{Wiener}}(\omega) + \underbrace{\frac{\phi_s(\omega)}{\phi_s(\omega) + \phi_n(\omega)} \left[\frac{1}{M^2} \frac{\sum_{i=1}^M \sum_{j=1}^M (\tilde{\Gamma}_{n,n_j}(\omega) - \hat{\Gamma}_{n,n_j}(\omega))}{\sum_{i=1}^M \sum_{j=1}^M (1 - \hat{\Gamma}_{n,n_j}(\omega))} \right]}_{e_2} \quad (3-35)$$

By comparing (3-34) and (3-35), it can be observed that the error term e_1 in the McCowan post-filter is the average of the ratios $(\tilde{\Gamma}_{n,n_j}(\omega) - \hat{\Gamma}_{n,n_j}(\omega)) / (1 - \hat{\Gamma}_{n,n_j}(\omega))$ for each microphone pair, while the error term e_2 in the proposed BC-SCPF is the ratio of averaged $\tilde{\Gamma}_{n,n_j}(\omega) - \hat{\Gamma}_{n,n_j}(\omega)$ and $1 - \hat{\Gamma}_{n,n_j}(\omega)$. It is known that the averaging before division may be robust to the estimation errors. In other words, the error term e_1 is sensitive to the cases such as one of the estimated $\hat{\Gamma}_{n,n_j}(\omega)$ approaches to unity or is significantly different from the true noise coherence matrix. The effects are alleviated after the averaging in the proposed BC-SCPF. Listed below are potential advantages of the proposed BC-SCPF comparing with Zelinski and McCowan post-filters.

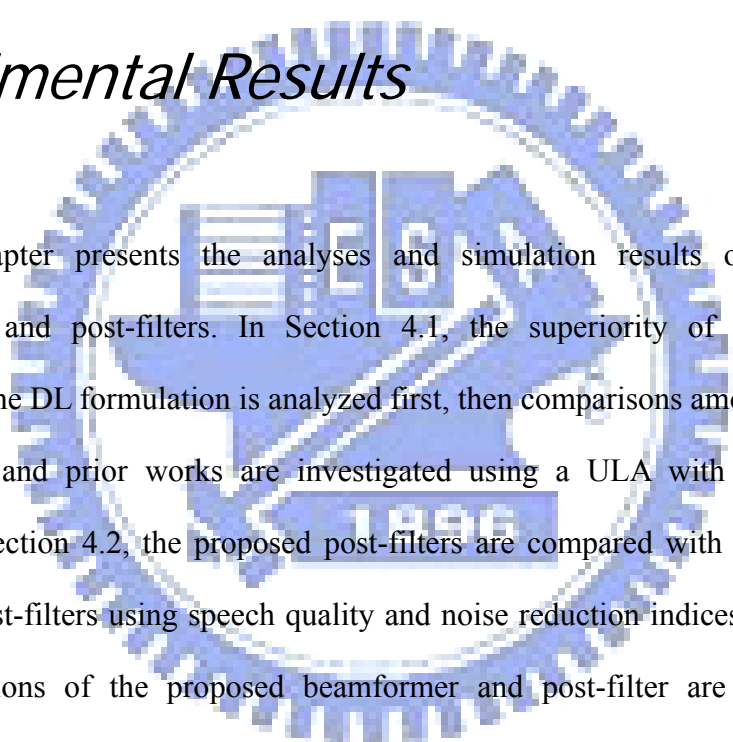
- 1) Multi-rank signal models can be directly adopted in the proposed method.
- 2) The assumption of homogeneous sound fields used in Zelinski and McCowan post-filters can be relaxed in the proposed method.
- 3) Compared to the estimation of the noise field coherence under each microphone pairs, the proposed method merged those as one bias term, which can be designed in many ways.
- 4) The proposed BC-SCPF is less sensitive to the individual estimation error of the noise coherence function (or the noise PSD function).

3.6 Summary

In this chapter, a spatial coherence measure is introduced and analyzed. The spatial coherence measure evaluates the similarity between the measured signal fields using power spectral density matrices. Based on this measure, the SCPF is derived. MSE analysis among the SCPF, the Zelinski post-filter, and the DS beamformer is discussed. Due to the similarity of the desired signal field and the noise field, a bias compensated solution, BC-SCPF, is proposed. It can be shown that the BC-SCPF is equivalent to the optimal Wiener filter if the bias or the noise power spectral density matrix is perfectly measured. Besides, the Zelinski post-filter is a special case of the proposed BC-SCPF. Theoretical comparison between the BC-SCPF and the McCowan post-filter is investigated, and advantages of the proposed BC-SCPF over the McCowan post-filter are listed clearly.

Chapter 4

Experimental Results



This chapter presents the analyses and simulation results of the proposed beamformers and post-filters. In Section 4.1, the superiority of using the norm constraint to the DL formulation is analyzed first, then comparisons among the proposed beamformers and prior works are investigated using a ULA with half-wavelength spacing. In Section 4.2, the proposed post-filters are compared with the Zelinski and McCowan post-filters using speech quality and noise reduction indices. In Section 4.3, the combinations of the proposed beamformer and post-filter are studied for the application of multi-channel speech enhancement with a steering uncertainty.

4.1 Narrow-Band Simulations of the Proposed Robust Beamformer

In this section, narrow-band simulations using a ULA with half-wavelength spacing are investigated, including the analyses of the norm constraint and comparisons among different beamformers. In narrow-band array processing, the ULA with

half-wavelength spacing is called *standard linear array* [1]. For multi-rank sources, we used the following model:

$$\tilde{\Phi}_s(\omega) = \int_{-\pi/2}^{\pi/2} \rho(\theta, \omega) \mathbf{a}(\theta, \omega) \mathbf{a}^H(\theta, \omega) d\theta \quad (4-1a)$$

with

$$\rho(\theta, \omega) = \frac{1}{\sqrt{2\pi\eta}} \exp\left(-\frac{(\theta - \theta_s)^2}{2\eta^2}\right) \quad (4-1b)$$

which is a Gaussian model with incoherently scattered sources, and the parameter η controls the angular spreading of the signal model. In the simulations, $\eta = 5$ is chosen. For all the scenarios, the desired signal is always present in the training data cell.

In Section 4.1.1, the norm constraint value $\gamma(\omega)$ is analyzed with a *standard linear array*. The results were computed by the theoretical covariance matrices, and the relationship between $\lambda(\omega)$ and $\gamma(\omega)$ in (2-20) was used to find the equivalent diagonal loaded beamformer. In Section 4.1.2, the proposed beamformers were compared with the priori works. The comparison between the proposed nonlinear Kalman filters was also investigated. The sources and noises were produced using mutually uncorrelated white Gaussian noise.

4.1.1 Analyses of the Norm Constraint Value for Half-Wavelength Spacing

In this sub-section, we show that an appropriate selection of norm constraint value $\gamma(\omega) = 0.035$ is less sensitive to the input signal powers, central angle mismatches, and number of sensors than the selection of the diagonal loading level $\lambda(\omega)$ with a *standard linear array*. For all the analyses, the theoretical covariance matrices of the signals were used. The equivalent $\lambda(\omega)$ of a chosen $\gamma(\omega)$ can be derived from the relationship in

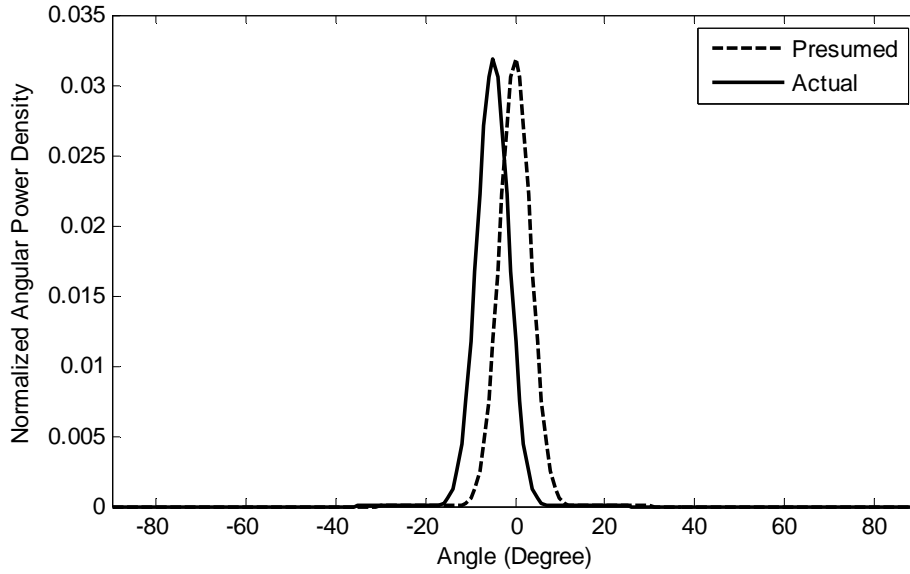
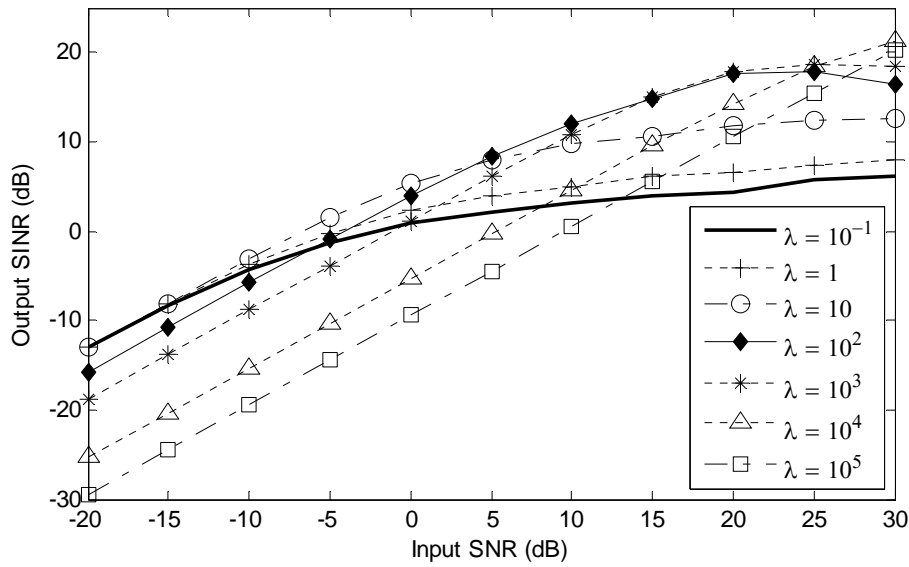


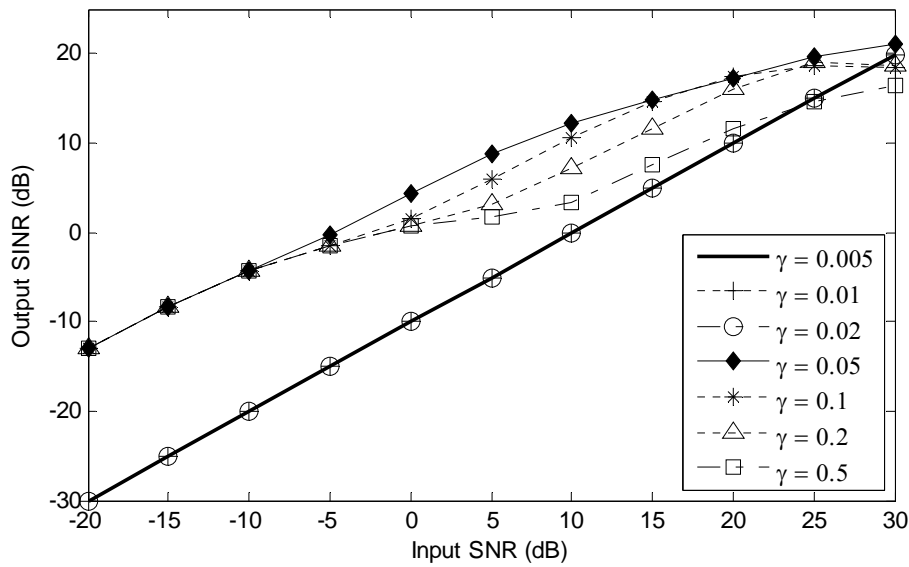
Figure 4-1 Presumed and actual angular power density functions for $\eta = 5$ and 5° central angle mismatch

(2-20).

For the Simulation 1, the sensitivity of $\lambda(\omega)$ and $\gamma(\omega)$ to the input signal powers are studied. A ULA of $M = 10$ sensors with half-wavelength spacing was used. The multi-rank signal model given in (4-1) was utilized with presumed central angle equal to 0° and angular spread $\eta = 5$. The scattered desired source and interference with angular spread $\eta = 5$ impinged into the array from the central angles -5° and 45° , respectively. Thus, 5° central angle mismatch was considered. The presumed and actual normalized angular power density functions are shown in Figure 4-1. The sensor noise power was set to 1, and the interference-to-noise ratio (INR) was 30 dB. The signal-to-noise ratio (SNR) varies from -20 dB to 30 dB. Figure 4-2 shows the output SINRs versus input SNRs for different selections of $\lambda(\omega)$ and $\gamma(\omega)$. It can be seen that $\lambda(\omega) = 10^2$ (note that the sensor noise power = 1) and $\gamma(\omega) = 0.035$ (i.e., $T(\omega) = 1/M + 0.035 = 0.135$) are good choices considering all input SNR conditions. These values will be used as the best choices for the rest simulations. For the selection of $\lambda(\omega)$, high values give more penalties on the spatially white noise (or incoherent noise), which leads to the matched



(a) different diagonal loading levels $\lambda(\omega)$



(b) different norm constraint values $\gamma(\omega)$

Figure 4-2 Output SINR vs. input SNR

filter as $\lambda(\omega) \rightarrow \infty$. The matched filter does not form nulls at the directions of interferences, hence it may have poor output SINR performance. As $\lambda(\omega) \rightarrow 0$, it turns into the MVDR solution without norm constraint, which is sensitive to the array mismatches and has severe *self-cancellation* at high input SNRs. For the selection of $\gamma(\omega)$, it can be expected that the variation of output SINR with different $\gamma(\omega)$ is relatively smaller than that with different $\lambda(\omega)$. This is because in (2-16) and (2-20), the

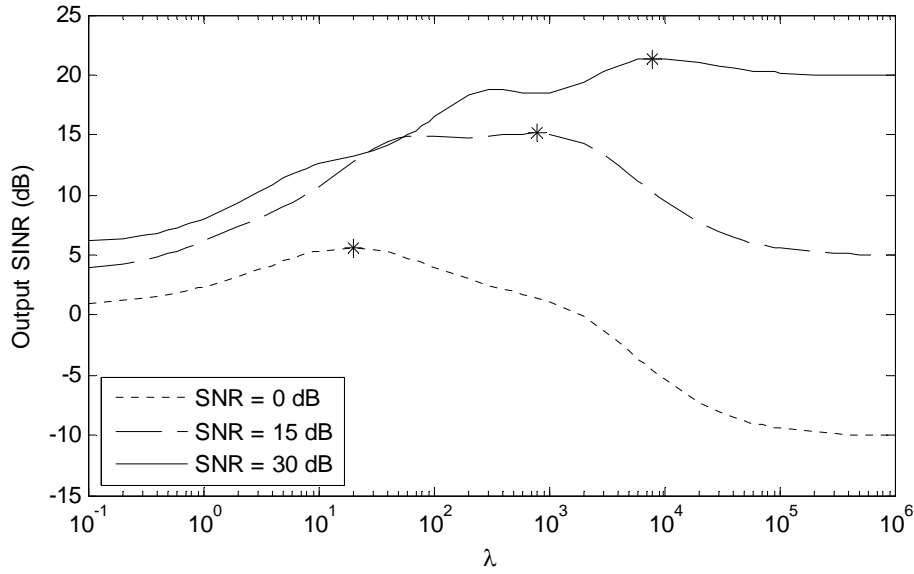


Figure 4-3 Output SINR vs. $\lambda(\omega)$ for different input SNRs.

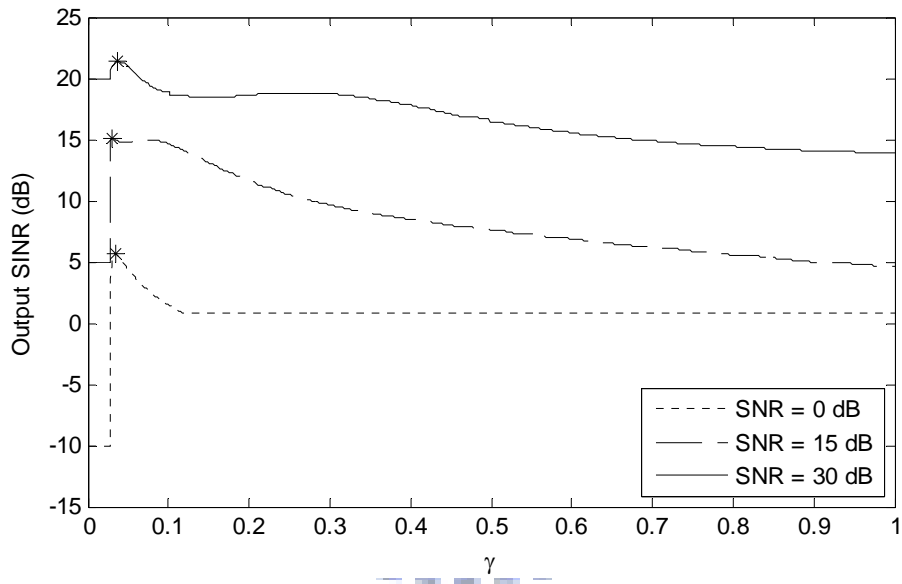


Figure 4-4 Output SINR vs. $\gamma(\omega)$ for different input SNRs.

factor of averaged input power has been removed by the division. When the norm constraint value $\gamma(\omega) \rightarrow 0$, the weight vector turns into the matched filter. When $\gamma(\omega)$ approaches to the upper bound given in (2-23), it gives the MVDR solution without norm constraint. From Figure 4-3 and Figure 4-4, the optimal selections of $\lambda(\omega)$ and $\gamma(\omega)$ for different input SNR conditions are illustrated. In the figures, the stars point out the optimal selections for each case. It is obvious that the optimal selection of $\lambda(\omega)$ is

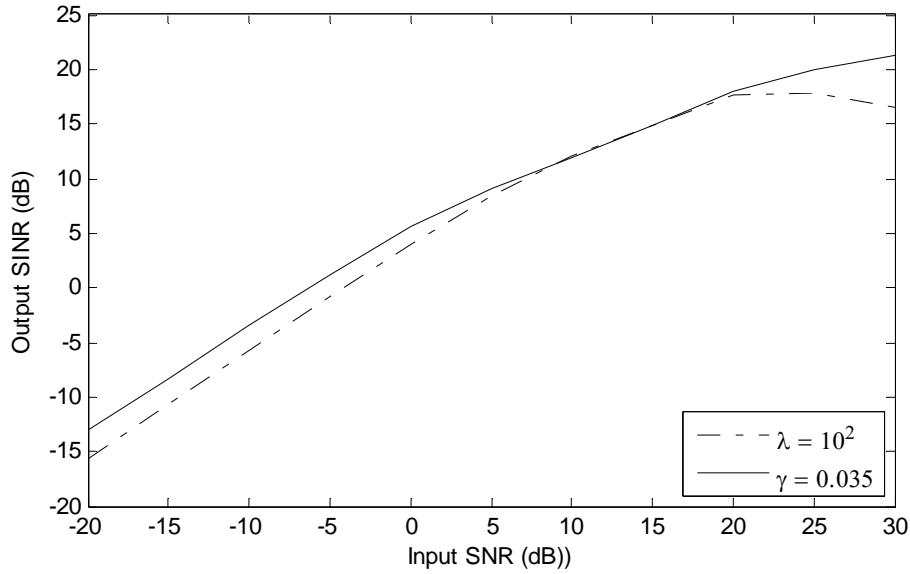


Figure 4-5 Comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different input SNR conditions

related to the input signal power, while the optimal selection of $\gamma(\omega)$ is less sensitive to the input signal power. Since the powers of speech signals are unknown *a priori* and even time varying, the proposed robust beamformer provide consistent SINR performances due to the insensitivity of $\gamma(\omega)$ to the signal powers. It is interesting to note that there is a jump at $\gamma(\omega) \approx 0.027$ in Figure 4-4. A possible explanation of this phenomenon is that the feasible set that satisfies the norm constraint and the distortionless constraint is too small in this case. Thus the solution fails to reject the interference in order to satisfy the constraints, which gives the solution of the matched filter. Figure 4-5 gives the comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different input SNR conditions. Since the optimal choice of $\lambda(\omega)$ depends on the desired signal powers, the proposed norm-constrained robust beamformer with the optimal $\gamma(\omega)$ gives better output SINRs for most SNR conditions.

For the Simulation 2, the comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different central angle mismatches is analyzed. It can be seen from Figure 4-6 that for small angular mismatches, the proposed norm-constrained robust beamformer with the

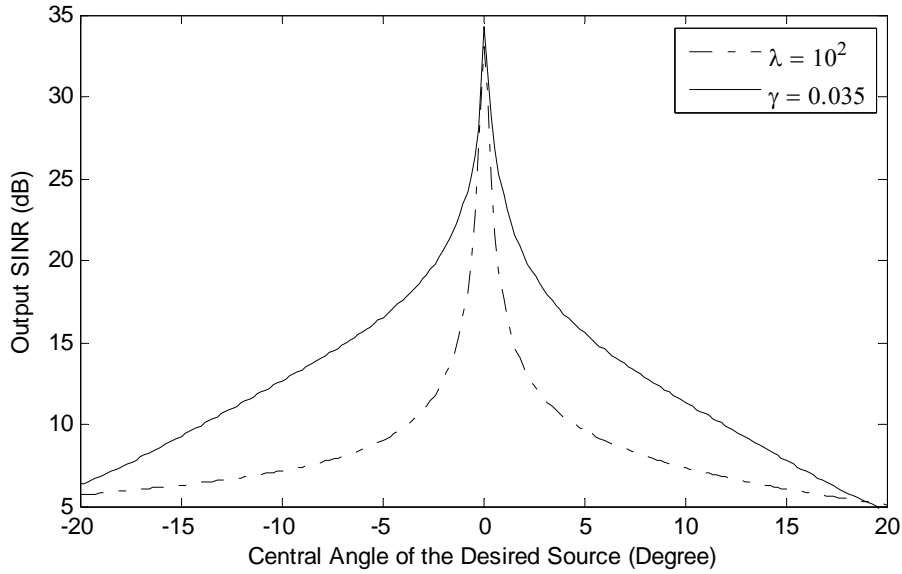


Figure 4-6 Comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different central angle mismatches

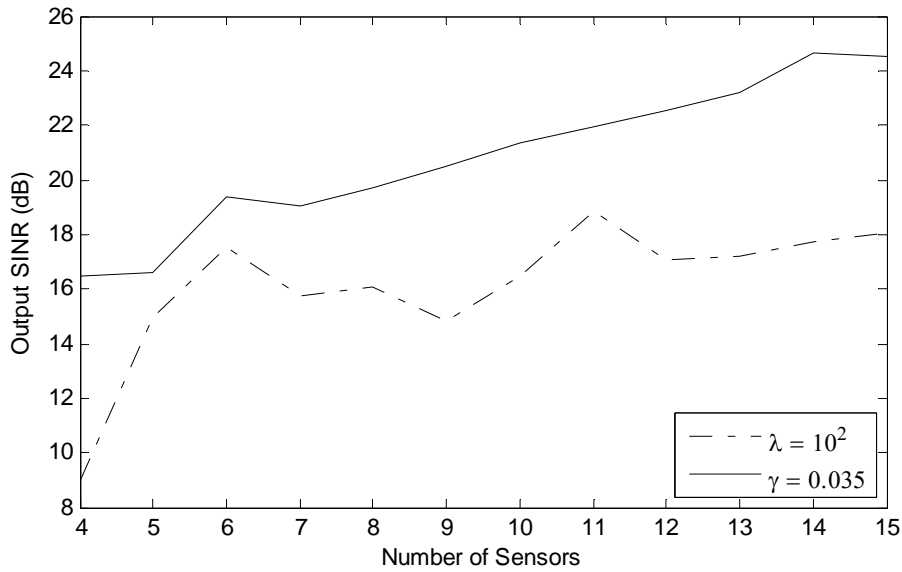


Figure 4-7 Comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different number of sensors

best selected $\gamma(\omega)$ performs better than using the best selected diagonal loading.

For the Simulation 3, the comparison between the best selected $\lambda(\omega)$ and $\gamma(\omega)$ for different number of sensors is investigated. Again, the superiority of using the norm-constrained robust beamformer is demonstrated in Figure 4-7.

In this section, the superiorities of using the norm constraint value $\gamma(\omega)$ over the diagonal loading level $\lambda(\omega)$ were demonstrated in the narrow-band simulations. However, for wide-band applications, the properties differ from frequency bands according to different directivities [1]. The parameter selection strategy for wide-band applications will be discussed in Section 4.3.1.

4.1.2 Narrow-Band Comparisons

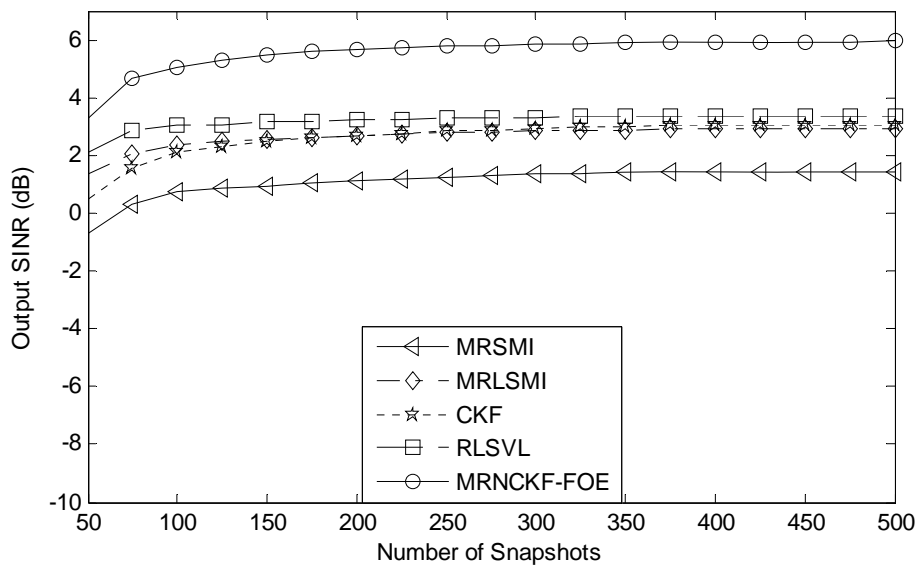
In this section, the comparisons of the proposed beamformers and other adaptive beamformers are studied. The sensor noise power was set to 1. The simulation condition is the same as Simulation 1, except the generated simulated data is used. For each scenario, the average of 100 simulation runs is used to obtain each simulated point. The detailed parameter settings and abbreviations of the algorithms are listed below:

- 1) MRSMI: Multi-rank sample matrix inverse [20]. The algorithm was implemented by (2-8).
- 2) MRLSMI: Multi-rank loaded sample matrix inverse [20]. The algorithm was implemented by (2-10), where the diagonal loading level was chosen as $\lambda(\omega) = 10^2$ (note that the sensor noise power = 1).
- 3) CKF: Constrained Kalman filter [67]. The CKF uses the rank-1 model without the norm constraint and formulates the state space using the SCPO method. The parameter matrices $\mathbf{Q}(\omega)$ was set to a zero matrix, and the diagonal terms of $\mathbf{R}(\omega)$ were estimated as $\hat{\sigma}_1^2(\omega)$ and $10^{-5} \cdot \hat{\sigma}_1^2(\omega)$, where $\hat{\sigma}_1^2(\omega)$ was estimated using (2-41) with the forgetting factor $\alpha = 0.9$.
- 4) RLSVL: Recursive least square with variable loading [71]. The RLSVL uses the rank-1 model and the GSC structure. A norm constraint was imposed on the nulled vector for improving the robustness. The forgetting factor and the

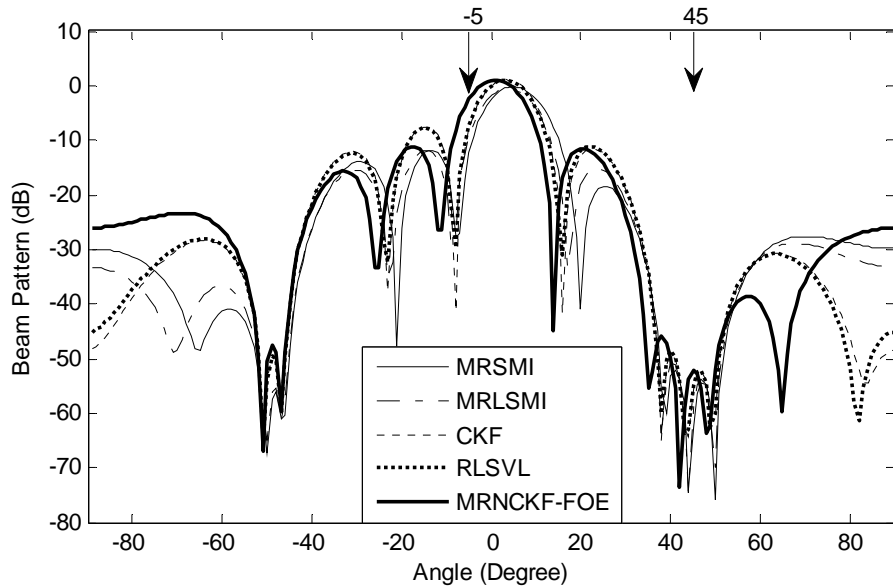
norm constraint were set as 0.999 and 0.2 as suggested in the paper [71].

- 5) MRNCKF-FOE: Multi-rank norm-constrained first-order extended Kalman filter. The proposed beamformer was implemented using (2-35) by letting both $\boldsymbol{\pi}(\omega, k)$ and $\boldsymbol{\Lambda}(\omega, k)$ be zero. The parameter matrix $\mathbf{Q}(\omega)$ was set to a zero matrix, and $\mathbf{R}(\omega)$ was set as in (2-40) with $\kappa = 10^{-5}$ for all the proposed Kalman filters, where $\hat{\sigma}_1^2(\omega)$ was estimated using (2-41) with the forgetting factor $\alpha = 0.9$. The norm constraint value $\gamma(\omega) = 0.035$ is chosen, which corresponds to the constraint $T(\omega) = 1/M + \gamma(\omega) = 0.135$.
- 6) MRNCKF-SOE: Multi-rank norm-constrained second-order extended Kalman filter. The proposed beamformer was implemented using (2-35) with the second-order terms $\boldsymbol{\pi}(\omega, k)$ and $\boldsymbol{\Lambda}(\omega, k)$.
- 7) MRNCKF-U: Multi-rank norm-constrained unscented Kalman filter. The proposed beamformer was implemented using (2-36) and (2-37).

For the first case, the convergences and beam patterns at SNR = 0 dB are studied. Figure 4-8a shows the output SINR performance versus the training size. The presence of the desired signal deteriorates the SINR performance due to the *self-cancellation* phenomenon, which can be observed in Figure 4-8b around the central angle of the desired source -5° . Considering the performance of the pairs (MRSMI, MRLSMI) and (CKF, RLSVL), it can be observed that the norm constraint improves the SINR performance. It is also worth to note that the Kalman filter solutions seem to be more robust to the steering mismatches than the beamformers using the estimation of sample matrix. The Kalman filter is a close-loop system who constrains the weight vector to the desired array response at each iteration; on the other hand, the sample matrix inverse method is an open-loop system who constrains the weight vector after the sample matrix is estimated. Therefore, the latter one can be easily affected by the contaminated sample



(a) Output SINR vs. training size

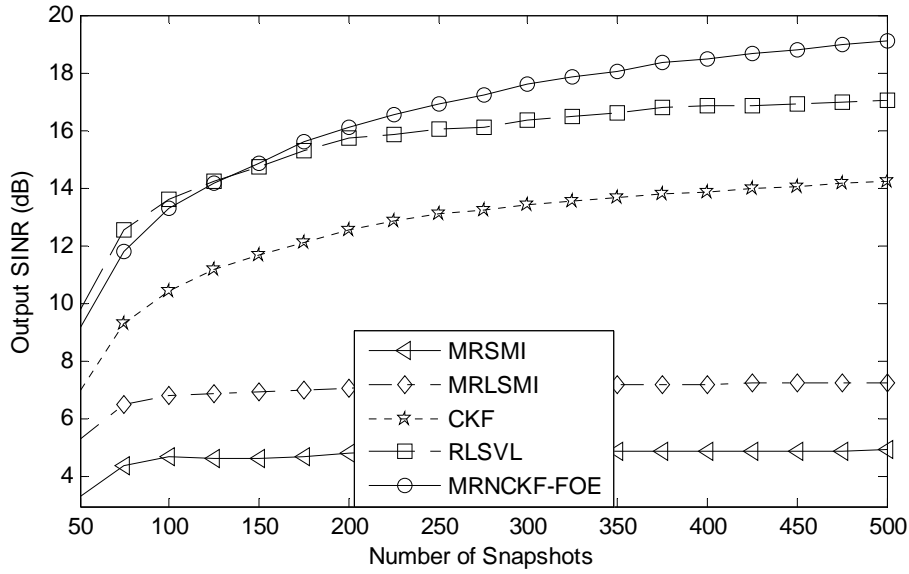


(b) Beam pattern

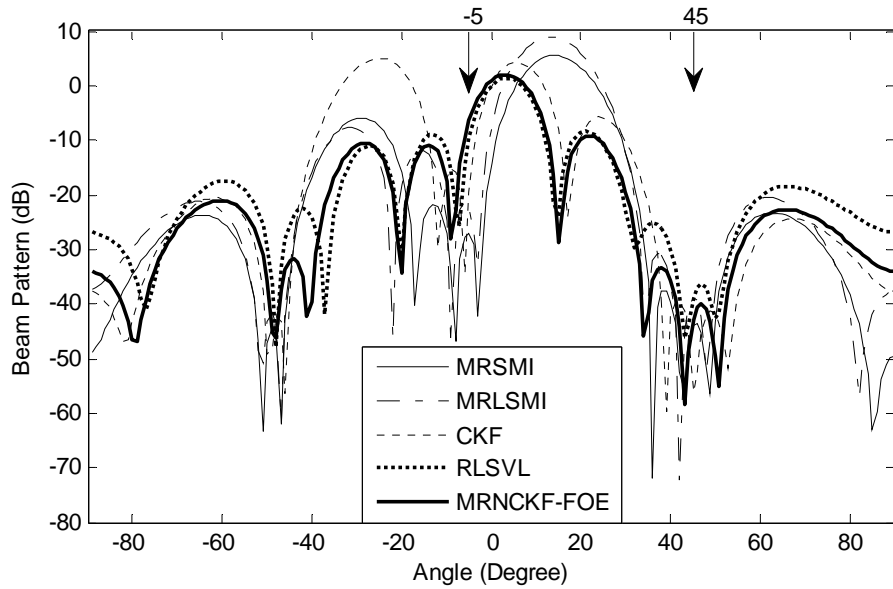
Figure 4-8 Comparisons of the beamformers at 0 dB input SNR.

matrix since the importance of the training data is the same when estimating the sample matrix. From the beam pattern, it is shown that the proposed beamformer, MRNCKF-FOE, gives the best output SINR since it has the smallest signal distortion at -5° while keeping the same order of noise rejection at 45° .

For the second case, the convergences and beam patterns at SNR = 20 dB are studied. In Figure 4-9a, the large signal power slows down the convergence of the



(a) Output SINR vs. training size



(b) Beam pattern

Figure 4-9 Comparisons of the beamformers at 20 dB input SNR.

algorithms. The strong signal power leads to larger *self-cancellation* for the MRSMI and MRLSMI beamformers. Despite of the difference between using the Kalman filter and the sample matrix, this case also reveals that the chosen diagonal loading level of the MRLSMI beamformer is not appropriate under this SNR condition (see Figure 4-3). This demonstrates the advantage of using the norm constraint with a more robust selection of the norm constraint value.

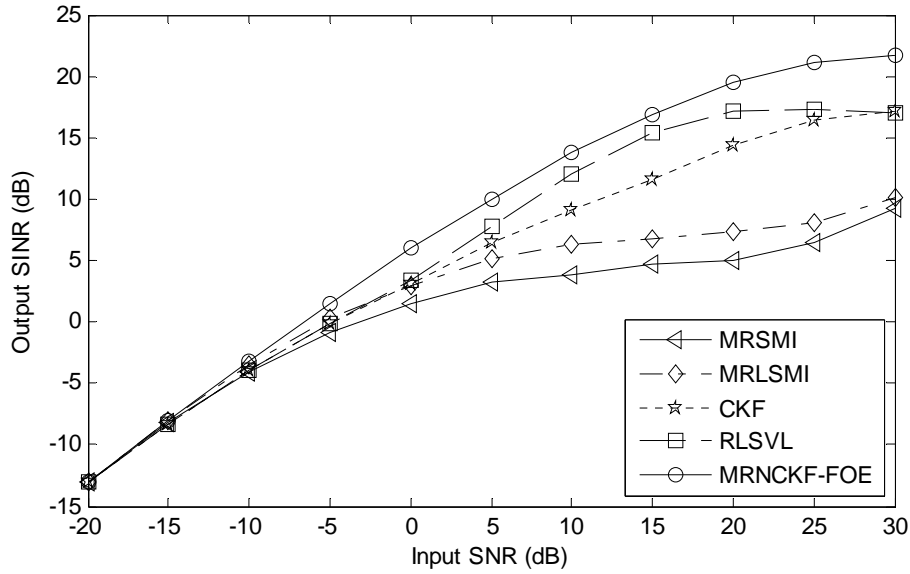


Figure 4-10 Output SINR vs. input SNR for different beamformers

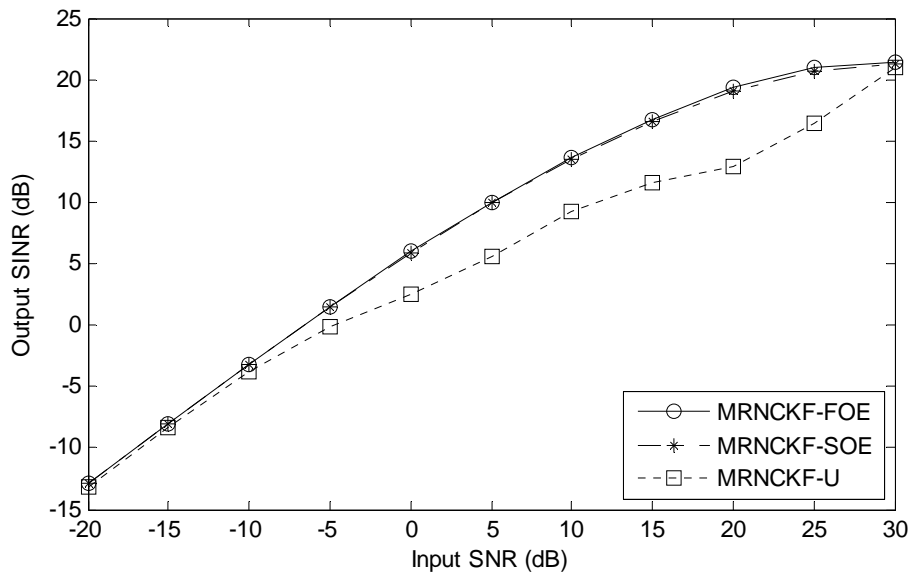


Figure 4-11 Output SINR vs. input SNR for proposed Kalman filters

For the third case, the output SINR of the beamformers versus the input SNR is illustrated in Figure 4-10. The training size of this simulation is $N = 500$. When the input SNR is small, all the beamformers converges to the optimal MVDR solution. As the SNR increases, the differences between algorithms become obvious. It is shown that the proposed beamformer has the best performance through different SNR conditions.

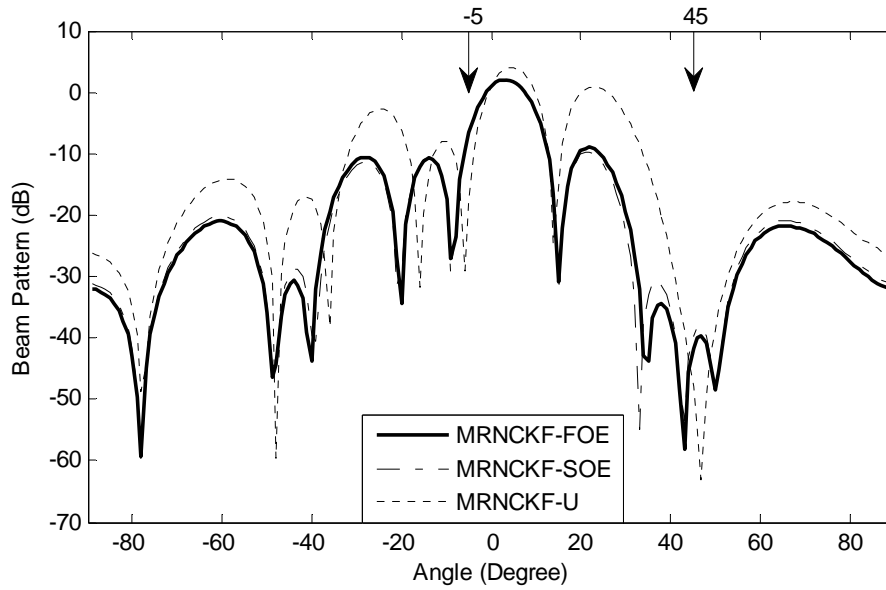


Figure 4-12 Beam patterns of the proposed Kalman filters at 20 dB input SNR.

For the last case, the output SINR of proposed Kalman filter solutions versus the input SNR is illustrated in Figure 4-11. It can be seen that the performances of the first- and second-order extended Kalman filters are almost the same. This indicates that the first-order approximation is good enough for our problem. Compared to the extended solutions, the unscented Kalman filter (UKF) has worse SINR performance. The UKF implicitly estimates the first- and second-order approximation terms of the Taylor expansion using *sigma points*. The *sigma points* were spread based on \sqrt{M} times eigenvectors of the error covariance $\mathbf{P}^-(\omega, k)$. An issue for the spreading of *sigma points* is invoked when some error dominates the covariance $\mathbf{P}^-(\omega, k)$. In this case, some *sigma points* are spread far away from the constraint sets and the neighborhood of the current state estimate, which can induce improper nonlinear transformations that degrade the performance of the UKF. In Figure 4-12, it can be seen that the large error of interference rejection enforces the noise reduction at both 45° and -5° , which results in the *self-cancellation*.

4.2 Speech Enhancement Results of the Proposed SCPF and BC-SCPF

In this section, we use three different noise fields with a local scattered desired source and several SNR conditions to evaluate the proposed post-filters. All the post-filters are processed on the output of the DS beamformer. The detailed parameter settings and abbreviations of the algorithms are listed below:

- 1) DS: Delay-and-sum beamformer.
- 2) SD: Super-directive beamformer [76]. The ratio of the PSDs of the sensor noise and the diffuse noise was chosen as -20 dB.
- 3) Zelinski: Zelinski post-filter [46]. The post-filter was implemented using (3-8), where the spectral densities were estimated using a first-order recursive filter with the forgetting factor $\alpha = 0.9$.
- 4) McCowan: McCowan post-filter based on noise field coherence [47]. The post-filter was implemented using (3-2), (3-6) and (3-7). The coherence matrices were trained with 200 noise-only frames (3.2 s) for each case.
- 5) SCPF: The proposed method was implemented using (3-16) and (3-17). The forgetting factor $\alpha = 0.9$ for estimating the PSD matrices was used for the proposed methods, which is the same as the factor used in Zelinski and McCowan post-filters.
- 6) BC-SCPF: The proposed method was implemented using (3-31) with the same training noise data as the McCowan *et als*. The biases were then computed using (3-29).

4.2.1 The Experimental Setup

The simulations were generated by the room impulse response generator [77] with reverberation corresponding to the reverberation time $RT_{60} = 503$ ms using Sabine-Franklin's formula [78]. There are three noise field conditions: 1) stochastically white noise where noises between microphones are uncorrelated (i.e., incoherent noise field); 2) babble noises which were uttered from four corners of the room to simulate a diffuse noise field; 3) speech interference which is a coherent source impinged into the array at the direction of 45° . The desired source impinged into the array at the direction 0° , which is locally scattered with the angular distribution described by a Gaussian function with $\eta = 10$ in Figure 4-13. A ULA with four omni-directional microphones and 5 cm spacing was used. The simulation environment is illustrated in Figure 4-14. The sampling rate and the fast Fourier transform (FFT) size were 8 kHz and 256, respectively. A female voice and a male voice were used as the desired source and the interference respectively. The white and babble noise signals were taken from the NOISEX-92 database [79]. All the recordings were 60 seconds in duration and combined into different SNR conditions.

4.2.2 Speech Quality and Noise Reduction Evaluations

Three criteria were used to investigate the performance. The speech quality was evaluated by ITU-T P.862 PESQ (Perceptual Evaluation of Speech Quality) [80]. For noise reduction performance, SNRI (Signal-to-Noise Ratio Improvement) and TNLN (Total Noise Reduction Level) from ITU-T G.160 [81] were computed. The simulation results of PESQ score improvement, SNRI, and TNLN are listed in Table 4-1, Table 4-2, and Table 4-3, respectively. For all the indices, the higher score indicates the better performance.

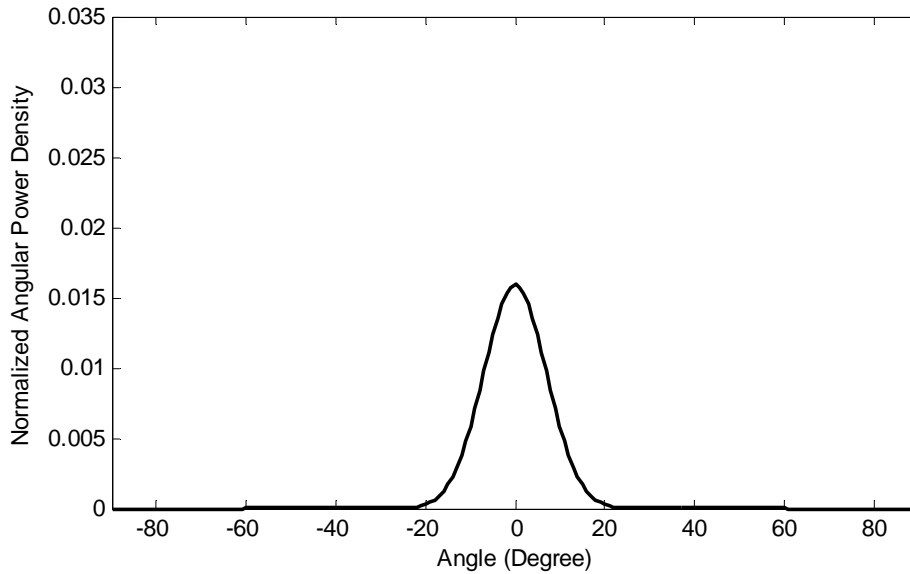


Figure 4-13 Angular power density function for $\eta = 10$

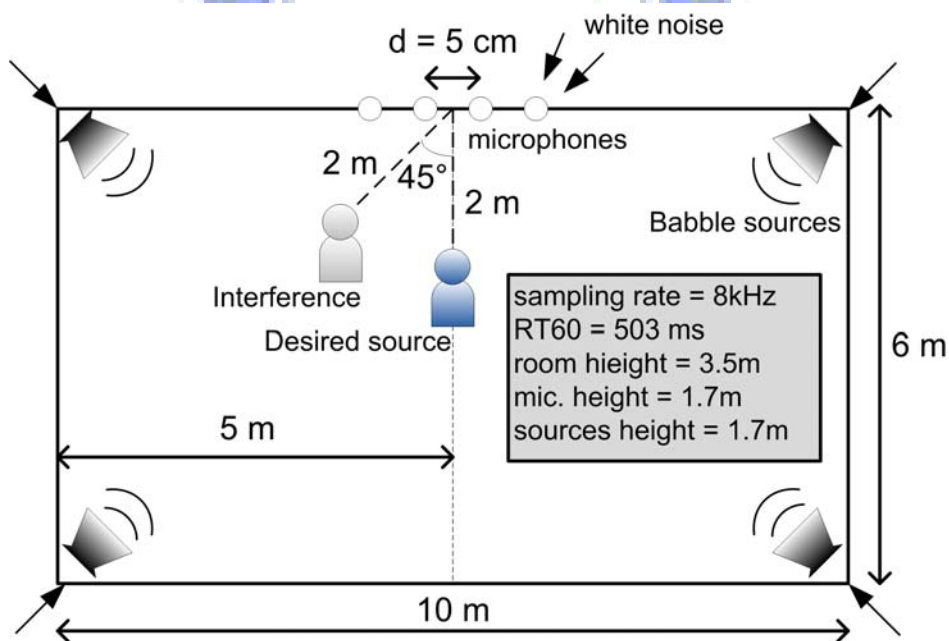


Figure 4-14 Simulation environment with locally scattered sources.

The following discussions are based on the results given by Table 4-1–Table 4-3. First, consider the DS and SD beamformers. It is known that the delay-and-sum (DS) beamformer is optimal for the MVDR design in the incoherent noise field. Therefore, the DS beamformer is ensured to perform better than the SD beamformer in this case. The SD beamformer based on a modified coherence function increases the directivity at

Table 4-1 PESQ score improvement obtained by different input SNRs

Input SNR → Algorithm ↓	5 dB	10 dB	15 dB	20 dB
White noise (Incoherent)				
Original noisy PESQ	1.74	2.02	2.34	2.67
DS	0.35	0.40	0.41	0.42
SD	0.09	0.05	0.00	-0.03
DS + Zelinski	0.87	0.78	0.66	0.57
DS + SCPF	0.72	0.70	0.61	0.55
DS + McCowan	0.87	0.78	0.66	0.57
DS + BC-SCPF	0.89	0.80	0.68	0.59
Babble noise (Diffuse)				
Original noisy PESQ	2.04	2.30	2.58	2.87
DS	0.15	0.14	0.13	0.13
SD	0.32	0.34	0.34	0.35
DS + Zelinski	0.16	0.16	0.15	0.14
DS + SCPF	0.19	0.17	0.16	0.15
DS + McCowan	0.35	0.33	0.28	0.23
DS + BC-SCPF	0.36	0.34	0.29	0.24
Speech interference (Coherent)				
Original noisy PESQ	2.23	2.50	2.79	3.08
DS	0.16	0.16	0.15	0.15
SD	0.28	0.30	0.29	0.29
DS + Zelinski	0.24	0.25	0.23	0.20
DS + SCPF	0.25	0.24	0.22	0.20
DS + McCowan	0.52	0.49	0.40	0.34
DS + BC-SCPF	0.53	0.50	0.41	0.34

low frequencies, or in other words, it amplifies the small deviations between microphones to obtain more noise reduction. As a result, the SD beamformer has better performance than the DS beamformer in the diffuse and coherent noise fields, where the insufficient spatial sampling has to be taken into account in these cases. However, it has some artifacts in the incoherent noise field due to the increased white noise gain [1]. Compared to the DS beamformer, the usages of post-filters give better performances.

Table 4-2 SNRI score obtained by different input SNRs

Input SNR →	SNRI (dB)			
Algorithm ↓	5 dB	10 dB	15 dB	20 dB
White noise (Incoherent)				
DS	6.06	5.79	5.47	4.96
SD	- 1.25	- 1.58	- 1.93	- 2.25
DS + Zelinski	19.78	17.75	15.31	12.50
DS + SCPF	13.75	12.73	11.45	9.73
DS + McCowan	19.65	17.66	15.26	12.48
DS + BC-SCPF	19.95	17.47	14.96	12.14
Babble noise (Diffuse)				
DS	2.12	1.71	1.36	1.01
SD	4.15	3.86	3.57	3.18
DS + Zelinski	2.88	2.52	2.10	1.63
DS + SCPF	2.61	2.21	1.83	1.51
DS + McCowan	9.46	8.60	7.23	5.50
DS + BC-SCPF	8.79	8.42	7.04	5.45
Speech interference (Coherent)				
DS	0.71	0.52	0.32	0.08
SD	1.29	1.18	1.00	0.75
DS + Zelinski	1.80	1.73	1.47	1.10
DS + SCPF	1.38	1.36	1.11	0.67
DS + McCowan	9.91	8.87	6.91	4.73
DS + BC-SCPF	9.97	9.13	7.29	5.11

That means the post-filters followed by a DS beamformer have contributions to both the speech quality and noise reduction.

Second, consider the Zelinski and the McCowan post-filters. The Zelinski post-filter is a special case of the McCowan post-filter when the noise field is incoherent. Hence, it can be seen that in the incoherent noise field, the performances of the Zelinski and the McCowan post-filters are almost the same. While in other noise fields, the consideration of noise field coherence provides evident performance improvements. Likewise, the proposed BC-SCPF after the bias compensation has evident performance improvements compared to the proposed SCPF. However, unlike

Table 4-3 TNLr score obtained by different input SNRs

Input SNR → Algorithm ↓	TNLR (dB)			
	5 dB	10 dB	15 dB	20 dB
White noise (Incoherent)				
DS	5.55	5.42	5.14	4.63
SD	- 2.93	- 2.94	- 2.97	- 3.03
DS + Zelinski	21.10	18.48	15.60	12.49
DS + SCPF	14.62	13.26	11.69	9.78
DS + McCowan	20.96	18.39	15.55	12.47
DS + BC-SCPF	21.38	18.34	15.42	12.31
Babble noise (Diffuse)				
DS	0.64	0.60	0.49	0.20
SD	3.44	3.36	3.18	2.77
DS + Zelinski	1.95	1.77	1.49	1.05
DS + SCPF	1.66	1.50	1.28	0.89
DS + McCowan	9.70	8.44	6.91	5.13
DS + BC-SCPF	9.38	8.53	6.98	5.19
Speech interference (Coherent)				
DS	0.66	0.67	0.71	0.50
SD	1.85	1.86	1.89	1.64
DS + Zelinski	1.94	1.82	1.73	1.38
DS + SCPF	1.70	1.61	1.56	1.27
DS + McCowan	13.26	11.53	9.18	6.60
DS + BC-SCPF	13.70	11.98	9.63	7.07

the relationship between the Zelinski and the McCowan post-filters in the incoherent noise field, the bias compensation still improves the performance in this scenario. For comparison between the Zelinski post-filter and the proposed SCPF, one can refer to the theoretical analysis in Section 3.4.2 to see the effect of different noise fields on $\mathcal{B}(\omega)$. In general, the Zelinski post-filter has better performance than the SCPF in the incoherent noise field. For other noise fields, it depends on the overall effect of $\mathcal{B}(\omega)$. Besides, it can be seen from the SNRI and TNLr scores that the SCPF has less noise reduction than the Zelinski post-filter. However, the noise reduction performance does not guarantee a better speech quality since the point source model used in the Zelinski

post-filter may introduce the signal distortion due to the *self-cancellation* (refer to the PESQ improvements in diffuse and coherent noise fields).

Finally, the performance of the McCowan post-filter and the proposed BC-SCPF is discussed. It can be seen that the proposed BC-SCPF has better speech qualities than the McCowan post-filter among all the noise fields. The results show that the small increment of the noise reduction of the McCowan post-filter does not give a better speech quality since the signal distortion has small impact in the noise reduction indices. Also, the simulation results demonstrate the superiority of the proposed BC-SCPF with a more accurate signal model.

4.3 Speech Enhancement Results of the Proposed Speech Enhancement System (Beamformer + Post-Filter)

In this section, according to the insufficient spatial sampling at low frequency bands, the relaxation of the constraints for the wide-band beamformers based on the Kalman filter is analyzed in Section 4.3.1. Next, the combinations of the proposed beamformer MRNCKF and post-filter BC-SCPF are evaluated using the speech quality and the noise reduction indices in Section 4.3.2.

4.3.1 Relaxation of the Constraints at Low Frequency Bands

For narrow-band applications, we showed the superiorities of the proposed multi-rank norm-constrained beamforming based on the Kalman filter (MRNCKF) with a *standard linear array* (i.e., a ULA with half-wavelength spacing) in Section 4.1. The half-wavelength spacing provides the best spatial information among different wavelengths [1]. However, for wide-band applications, the optimal weight vector norm

and the array response of the multi-rank signal model vary from wavelengths or frequencies. Thus, the previous superiorities will not be persistent for all frequency bands.

In this dissertation, we propose to relax the strength of the distortionless and norm constraints at low frequency bands for wide-band applications. The main objective is to allow more interference rejection for the Kalman filters. At low frequency bands, the wavelengths of the observed signals can be greater than the spacing between the sensors. In this case, the sensors are insufficient to describe the spatial differences between sources. In the following, the impacts of insufficient spatial sampling on the distortionless and norm constraints are investigated.

First, consider the norm constraint $\|\mathbf{w}(\omega, k)\|^2 = T(\omega)$. It is known that the weight vector norm is getting large with smaller steering mismatch due to the contradiction of objectives of the distortionless and minimizing output power [26]. The high spatial coherence of the sources at low frequency bands (which corresponds to long wavelengths) accelerates this phenomenon. As a result, the weight vector norm of the MVDR beamforming is getting higher for lower frequency with coherent noises. Figure 4-15 illustrates the squared vector norm of the optimal weight $\|\mathbf{w}_{\text{opt}}(\omega)\|^2$ versus the ratio between the array spacing and the wavelength for different noise fields. For the simulation condition, a ULA of $M = 4$ sensors was used. The multi-rank signal model given in (4-1) was utilized with presumed central angle equals to 0° and angular spread $\eta = 5$. The theoretical covariance matrices of the sources are used for the computation of the optimal weight and output SINRs. For coherent interference case, the scattered desired interference with angular spread $\eta = 5$ impinged into the array from the central angle 45° . From Figure 4-15, it can be seen that there is no need to add norm constraint for the incoherent noise field (i.e., spatially white). For coherent noise fields, the vector norms at low frequency bands are getting larger with the increasing interference power.

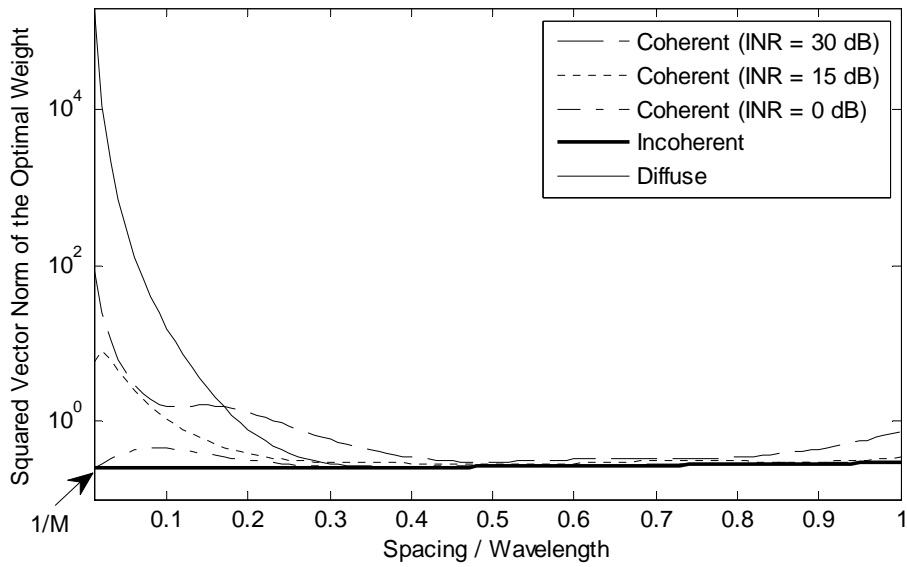


Figure 4-15 Squared norm of the optimal weight vs. the ratio between the array spacing and the wavelength for different noise fields.

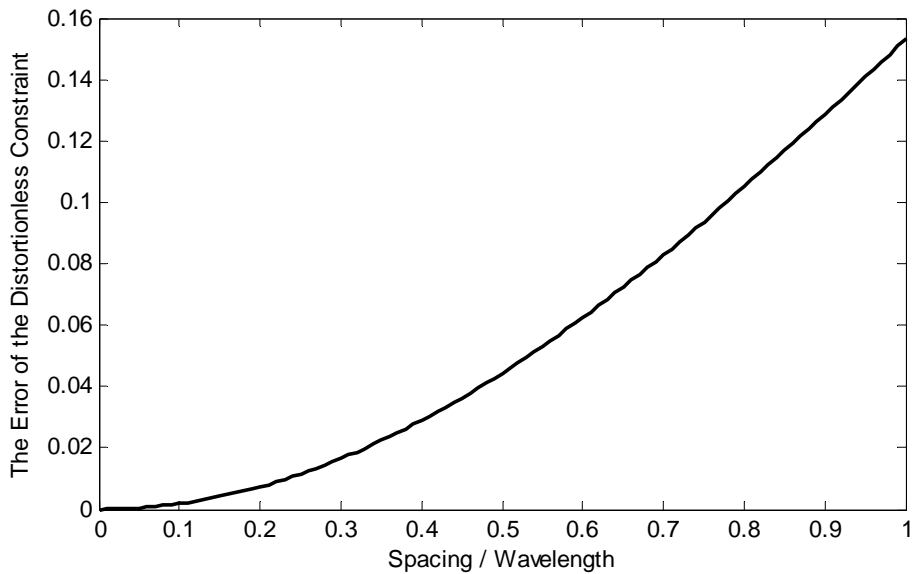


Figure 4-16 The error of the distortionless constraint vs. the ratio between the array spacing and the wavelength.

It becomes the worst in the diffuse noise field since the noise is also present in the desired directions. According to the observations, we propose to keep the norm constraint value $T(\omega)$ but relax the strength of the norm constraint at low frequency bands by increasing the variance parameter $\sigma_3^2(\omega)$. The reason to keep $T(\omega)$ is that the vector norm correlates to the white noise gain [1]. A large white noise gain can amplify

the unwanted noise, thus we would like to keep the value $T(\omega)$ close to $1/M$. According to the SCPO method introduced in Section 2.3.3, the larger $\sigma_3^2(\omega)$ gives less penalty on the error of norm constraint. Thus, the system can put more emphasis on the objective of interference rejection.

Next, consider the distortionless constraint $\mathbf{w}^H(\omega, k) \check{\Phi}_s(\omega) \mathbf{w}(\omega, k) = 1$. For norm-constrained vectors, the error of the distortionless constraint becomes smaller at low frequency bands due to limited spatial information. This can be illustrated by using the steering vector $\mathbf{a}_s(\omega)/M$ for all frequency bands. The error of the distortionless constraint is given by

$$\text{Err}(\omega) = 1 - \frac{\mathbf{a}_s^H(\omega) \check{\Phi}_s(\omega) \mathbf{a}_s(\omega)}{M^2} \quad (4-2)$$

It can be observed from Figure 4-16 that the error increases with the frequency. Here, we also suggest relaxing the distortionless constraint at low frequencies since the unconstrained solution is acceptable with small distortion. The relaxation can be done by increasing the variance parameter $\sigma_2^2(\omega)$. With the proposed estimation of the parameter matrix $\hat{\mathbf{R}}(\omega)$, we only need to increase the parameter κ . The modified κ should be a function of frequency and number of sensors. Here, we propose to design the frequency dependent parameter κ based on the modified squared beam pattern with the largest angle deviation as ([1], Eq. 2.96)

$$\kappa(\omega) = \begin{cases} 1, & f = 0 \\ \varepsilon, & f \geq f_T \\ \max \left(\left[\frac{1}{M} \frac{\sin(\pi M d f / c)}{\sin(\pi d f / c)} \right]^2, \varepsilon \right), & \text{otherwise} \end{cases} \quad (4-3)$$

where d is the spacing between sensors, f is the continuous frequency, c is the sound velocity. $f_T = c/Md$ is the first null of the beam pattern. Since we only tend to relax

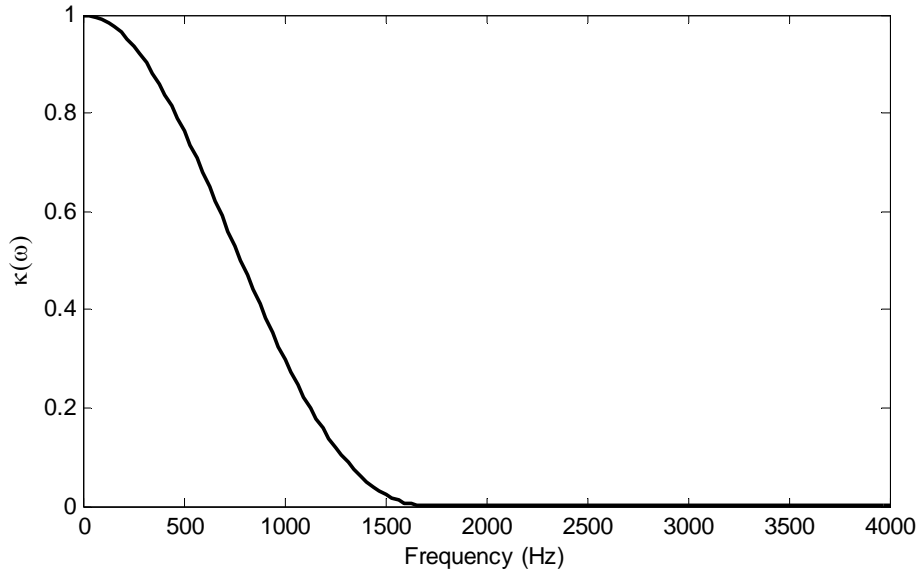


Figure 4-17 κ vs. frequency.

the constraints at low frequencies, the parameters are cut to small value ε (e.g., 10^{-5}). Figure 4-17 gives an illustration with $M = 4$ and $d = 0.05$ m. With the modified parameter $\kappa(\omega)$, we can strengthen the interference rejection at low frequencies. This parameter will be applied to all the Kalman filters in the next section.

4.3.2 Speech Quality and Noise Reduction Evaluations

In this section, we continue the Simulation 1 for the wide-band application. A 5° angle mismatch of the desired source with angular spread $\eta = 5$ was considered, as shown in Figure 4-1. The testing database is the same as in Section 4.2, except for the angle mismatch of the desired source and the angular spread. The beamformers were trained using 60 second recordings for each case, where the desired signal was present in the training data. The trained beamformers were applied as fixed spatial filters, and the proposed post-filter BC-SCPF was imposed afterward for further improvement. The speech indices introduced in Section 4.2.2 were used to evaluate the speech quality and noise reduction performances. The results are shown in Table 4-4–Table 4-6.

Table 4-4 PESQ score improvement with 5° angle mismatch

Input SNR → Algorithm ↓	0 dB	5 dB	10 dB	15 dB
Original noisy PESQ	1.96	2.23	2.51	2.80
DS	0.16	0.14	0.15	0.13
CKF	0.68	0.49	0.28	0.00
MRNCKF	0.81	0.68	0.46	0.14
DS + BC-SCPF	0.53	0.51	0.46	0.37
CKF + BC-SCPF	0.83	0.70	0.49	0.16
MRNCKF + BC-SCPF	0.89	0.86	0.67	0.31

Table 4-5 SNRI score with 5° angle mismatch

Input SNR → Algorithm ↓	SNRI (dB)			
	0 dB	5 dB	10 dB	15 dB
DS	0.85	0.73	0.60	0.45
CKF	6.98	6.30	4.19	-0.95
MRNCKF	7.57	6.94	4.48	-0.62
DS + BC-SCPF	10.55	10.43	9.10	6.94
CKF + BC-SCPF	14.53	13.92	10.81	5.36
MRNCKF + BC-SCPF	14.85	14.29	11.01	5.45

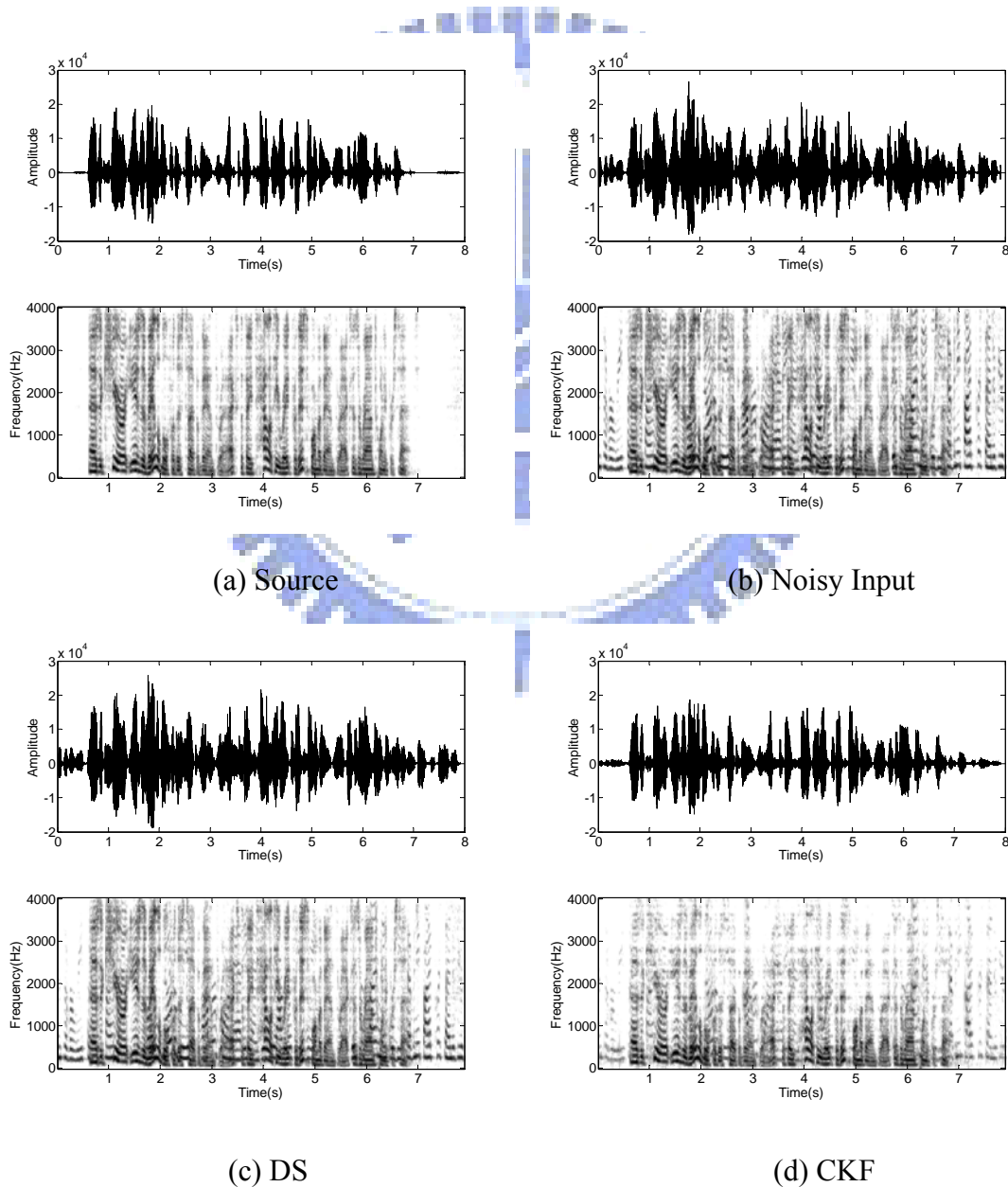
Table 4-6 TNLr score with 5° angle mismatch

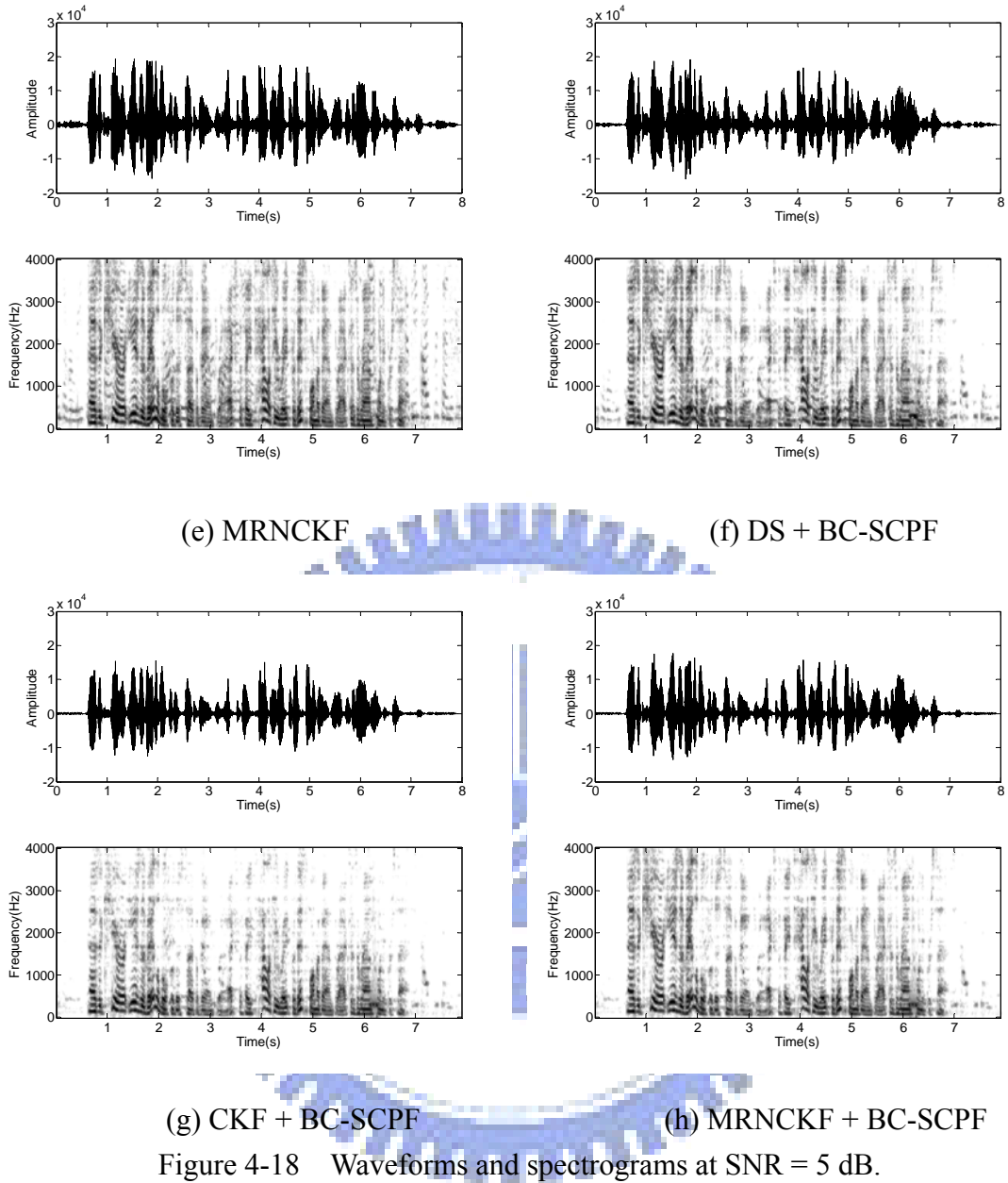
Input SNR → Algorithm ↓	TNLR (dB)			
	0 dB	5 dB	10 dB	15 dB
DS	0.72	0.81	0.87	0.93
CKF	11.29	10.27	7.47	0.98
MRNCKF	11.42	10.46	7.48	0.92
DS + BC-SCPF	16.16	14.75	12.49	9.72
CKF + BC-SCPF	22.72	20.89	16.75	9.47
MRNCKF + BC-SCPF	22.73	20.90	16.61	9.26

First, considering the DS, CKF, and proposed MRNCKF (with FOE) beamformers. The DS beamformer has consistent performance since it is data independent. The CKF and proposed MRNCKF beamformers work well at low SNRs. According to the narrow-band simulation, both the beamformers converge to the optimal MVDR

beamformer when the input SNR approaches to zero. Compared to the CKF, it can be observed the proposed MRNCKF is more robust to the mismatched desired signal since it provides better speech quality (i.e., PESQ) and noise reduction (i.e., SNRI and TNL). However, both beamformers degrade as the SNR increases. At 15 dB SNR, both the data dependent beamformers are worse than the DS beamformer due to the *self-cancellation* (or signal distortion here). Second, the results show that the proposed post-filter, BC-SCPF, always gives better speech quality and noise reduction.

In the end, we give an illustration of the waveforms and spectrograms at 5 dB SNR





as in Figure 4-18. It can be observed that the DS beamformer provides bad noise reduction at low frequency bands. Further, compared to the proposed MRNCKF, the CKF has severe signal distortion at high frequency bands due to the *self-cancellation*.

4.4 Summary

In Section 4.1.1, the superiority of using the norm constraint to the diagonal

loading formulation is analyzed in the narrow-band. It is shown that an proper choice of the norm constraint value $\gamma(\omega)$ (or $T(\omega) = \gamma(\omega) + 1/M$) can be found, which is robust to the unknown signal powers, angle mismatches, and number of sensors. In Section 4.1.2, the proposed beamformer, MRNCKF, is compared with several methods in the narrow-band. The simulation results show that Kalman filter structure is more robust than using the estimation of sample matrix. Besides, the norm constraint and diagonal loading are demonstrated to improve the output SINR when the desired signal is present in the training data and angle mismatch exists. It is also shown that the performances of the proposed beamformer are better than the rank-1 beamformers (such as CKF and RLSVL) with a more similar multi-rank signal model. The comparison of the nonlinear Kalman filters is also studied. For our problem, the performances of the first- and second-order extended solutions are almost the same. This indicates that first-order approximation is enough in our problem. The UKF has worse performance, which can be caused by the spreading of the *sigma points*.

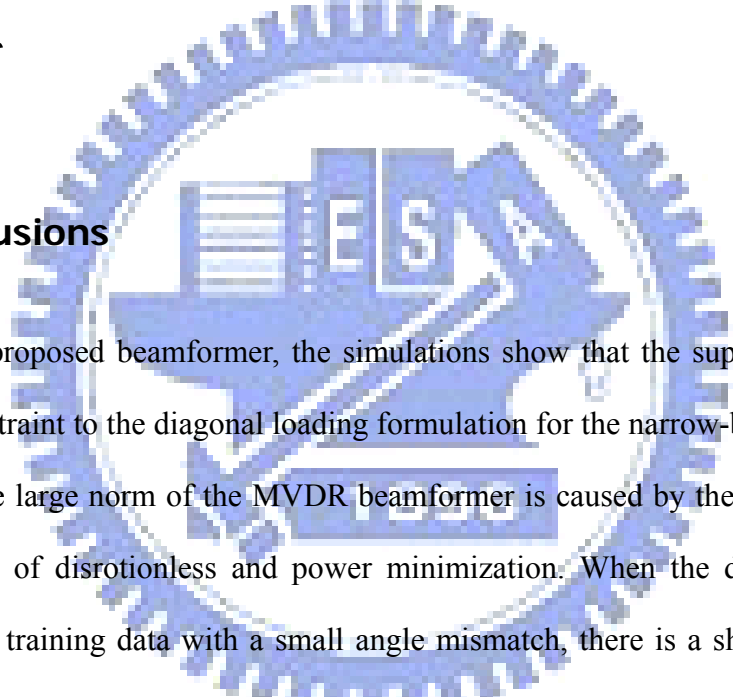
In Section 4.2, the proposed post-filters are evaluated using speech quality and noise reduction indices under three types of noise fields. The results show that the proposed BC-SCPF provides the best speech quality by using a more accurate multi-rank signal model.

In Section 4.3, the wide-band application of the proposed beamformer is investigated. To deal with the insufficient spatial sampling problem at low frequencies, the relaxation of the constraints is proposed to tune the parameters of the Kalman filters in Section 4.3.1. Next, the combinations of the beamformers and the proposed BC-SCPF are evaluated through speech quality and noise reduction indices in Section 4.3.2. The results show the superiority of the proposed MRNCKF for most SNR conditions. It can be observed from the spectrograms that the DS beamformer has small noise reduction and the CKF has signal distortion at high frequencies.

Chapter 5

Conclusions and Potential Research Topics

5.1 Conclusions



For the proposed beamformer, the simulations show that the superiority of using the norm constraint to the diagonal loading formulation for the narrow-band application. Recall that the large norm of the MVDR beamformer is caused by the contradiction of the objectives of distortionless and power minimization. When the desired source is present in the training data with a small angle mismatch, there is a sharp transition to meet the distortionless constraint while minimizing the signal. This transition results in the large norm. Thus, it makes sense that the norm constraint does not work for large angle mismatches. The other thing is the spreading of the scattered source. If the angular spreading of the scattered source is too large, the advantages of the multi-rank MVDR beamformers will be limited to the number of sensors. Similar situation happens when the spatial sampling is insufficient.

Besides, the comparison among the nonlinear Kalman filters for our quadratic problems was discussed in Section 4.1.2. It is shown that the first- and second-order

extended solutions almost have the same performances when the initial conditions are properly set. However, we observed that the second-order approximation is more sensitive to the improper initial conditions, which can even converge to other worse results.

For the proposed multi-channel post-filter, the bias compensated solution gives the optimal Wiener filter theoretically, as the McCowan post-filter did. The results show that with a more accurate multi-rank signal model, the proposed BC-SCPF has better speech quality than the McCowan post-filter. Furthermore, the similarity between the coherence matrices of the desired sound field and the noise field can be merged into a single real-valued bias. Several noise level estimation skills can be adopted to estimate the bias. Compared to the estimation of the noise coherence function, the bias estimation has fewer variables to be estimated. Besides, the noise level estimation can be carried out in the presence of the desired signal, while the estimation of noise coherence function is carried out during noise-only period. Finally, it is relatively easy to describe the similarity between the multi-rank signal models and the noise field using the proposed post-filters. This provides a more flexible design for the real-world environments.

5.2 Potential Research Topics

The future researches can be summarized as follows.

- 1) This dissertation proposed the multi-rank norm-constrained Kalman filter using the SCPO method. Restricted to the SCPO method, only equality constraints can be applied. However, inequality constraints can be applied using the projection method in the constrained Kalman filtering [55–59]. The

benefit of using inequality norm constraint occurs when signal is not always present in the training data. But the drawback is that the formulation of the error covariance is more complicated for implementation. This is left as the future study.

- 2) The beamformer and post-filter proposed in this dissertation is not closely coupled. In [5], the design of the post-filter based on a known beamformer is investigated. However, the best design of the post-filter with an adaptive beamformer is still an opening problem, and it is interesting to impose the multi-rank signal models into the beamformer and post-filter designs.
- 3) For wide-band applications, the best way to design the parameters for the Kalman filter has not been analytically solved. In Section 4.3.1, we proposed to relax the constraints at low frequencies based on a modified model. This seems to work and it provides better noise reduction at low frequencies. However, the improvement to the speech quality is not so clear. For the future study, the database that measuring the speech quality (such as PESQ) can be used to train the best parameters for the Kalman filter.

References

- [1] H. L. Van Trees, *Optimum array processing*. Wiley-Interscience, 2002.
- [2] H. Cox, R.M. Zeskind, and M.M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 35, pp. 1365–1375, Oct. 1987.
- [3] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [4] Y.-H. Chen and C.-T. Chiang, "Adaptive beamforming using the constrained Kalman filter," *IEEE Trans. Antennas and Propagation*, vol. 41, no. 11, pp. 1576–1580, Nov. 1993.
- [5] K.U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," *Microphone Arrays: Signal Process. Techniques and Applications*, vol. 3, pp. 39–60, 2001.
- [6] C. Winter, "Using continuous apertures discretely," *IEEE Trans. Antennas and Propagation*, vol. AP-25, pp. 695–700, Sep. 1977.
- [7] G.W. Elko, T.C. Chou, R.J. Lustberg, and M.M. Goodwin, "A constant-directivity beamforming microphone array," *Journal of Acoust. Soc. Amer.*, vol. 96, no. 5, pp. 3244, Nov. 1994.
- [8] J.H. Doles III and F.D. Benedict, "Broad-band array design using the asymptotic theory of unequally spaced arrays," *IEEE Trans. Antennas and Propagation*, vol. 36, no. 1, pp. 27–33, Jan. 1988.
- [9] B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Speech and Audio Process. Magazine*, pp. 4–24, Apr. 1988.
- [10] M. Doerbecker, "Speech enhancement using small microphone array with optimized directivity," *Int. Workshop Acoust. Echo Noise Control*, pp. 100–103, Sep. 1997.
- [11] J. Bitzer, K.U. Simmer, and K.D. Kammeyer, "An alternative implementation of the super-directive beamformer," in *Proc. IEEE Workshop Application Signal Process. to Audio Acoustics*, New Paltz, NY, USA, pp. 7–10, Oct. 1997.
- [12] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

- [13] N.K. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 8, pp. 996–1012, Aug. 1986.
- [14] K.M. Buckley and L.J. Griffiths, "An adaptive generalized sidelobe canceller with derivative constraints," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 3, pp. 311–319, Mar. 1986.
- [15] J. Bitzer, K.U. Simmer, and K.D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *Proc. IEEE Conf. Acoustics, Speech, and Signal Process.*, vol. 5, pp. 2965–2968, 1999.
- [16] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 19, no. 1, pp. 206–219, Jan. 2011.
- [17] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
- [18] J.E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller," *IEEE Trans. Speech and Audio Process.*, vol. 6, no. 4, pp. 338–351, July 1998.
- [19] J.S. Hu and M.T. Lee, "Spatially pre-processed target-to-jammer ratio weighted Wiener filter using two microphones," *IEEE Int. Conf. on Green Circuits and Systems*, pp. 180–185, Shanghai, China, June 2010.
- [20] S. Shahbazpanahi, A.B. Gershman, Z.Q. Luo, and K.M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.
- [21] A. Pezeshki, B.D. Van Veen, L.L. Scharf, H. Cox, and M.L. Nordenvaad, "Eigenvalue beamforming using a multirank MVDR beamformer and subspace selection," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 1954–1967, May 2008.
- [22] L. Zhang and W. Liu, "Robust forward backward based beamformer for a general-rank signal model with real-valued implementation," *Signal Processing*, vol. 92, no. 1, pp. 163–169, Jan. 2012.
- [23] J.E. Hudson, *Adaptive Array Principles*. London, U.K.: Peter Peregrinus, 1981.
- [24] B.D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, July 1988.
- [25] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1702–1715, July 2003.

- [26] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2407–2423, Sept. 2004.
- [27] S.A. Vorobyov, A.B. Gershman, and Z.Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
- [28] J.S. Hu, M.T. Lee, and C.H. Yang, "Robust adaptive beamformer for speech enhancement using the second-order extended H_∞ filter," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 21, no. 1, pp. 39-50, Jan. 2013.
- [29] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Process.*, vol. 9, no. 5, pp. 504–512, July 2001.
- [30] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.
- [31] S. Rangachari and P.C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 48, no. 2, pp. 220–231, Feb. 2006.
- [32] I. Cohen, "Analysis of two-channel generalized sidelobe canceller (GSC) with post-filtering," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 6, pp. 684–699, Nov. 2003.
- [33] I. Cohen, "Multichannel post-filtering in nonstationary noise environments," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1149–1160, May 2004.
- [34] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
- [35] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.
- [36] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403–2418, Nov. 2001.
- [37] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech and Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.
- [38] M.R. Schroeder, B.S. Atal, and J.L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *Journal of Acoust. Soc. Amer.*, vol. 66, no. 6, pp. 1647–1652, 1979.
- [39] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech and Audio Process.*, vol. 7, no. 2, pp. 126–137, Mar. 1999.

- [40] P.C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech and Audio Process.*, vol. 13, no. 5, pp. 857–869, Sep. 2005.
- [41] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.
- [42] N. Roman, D. Wang, and G.J. Brown, "Speech segregation based on sound localization," in *Proc. IEEE Conf. Neural Networks*, vol. 4, pp. 2861–2866, 2001.
- [43] P. Aarabi and S. Guangji, "Phase-based dual-microphone robust speech enhancement," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 4, pp. 1763–1773, Aug. 2004.
- [44] S. Guangji, P. Aarabi, and H. Jiang, "Phase-based dual-microphone speech enhancement using a prior speech model," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 15, no. 1, pp. 109–118, Jan. 2007.
- [45] P. Aarabi, *Phase-based speech processing*. World Scientific Publishing Company Incorporated, 2006.
- [46] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. IEEE Conf. Acoustics, Speech, and Signal Process.*, vol. 5, pp. 2578–2581, 11–14, Apr. 1988.
- [47] I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 6, pp. 709–716, Nov. 2003.
- [48] R.L. Bouquin-Jeannes, A.A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech and Audio Process.*, vol. 5, no. 5, pp. 484–487, Sep. 1997.
- [49] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," *IEEE Trans. Signal Processing*, vol. 43, no. 9, pp. 2144–2153, Sep. 1995.
- [50] A. Zoubir, Y. Wang, and P. Charge, "Efficient subspace-based estimator for localization of multiple incoherently distributed sources," *IEEE Trans. Signal Processing*, vol. 56, no. 2, pp. 532–542, Feb. 2008.
- [51] A. Zoubir and Y. Wang, "Robust generalised Capon algorithm for estimating the angular parameters of multiple incoherently distributed sources," *IET Signal Processing*, vol. 2, no. 2, pp. 163–168, June 2008.
- [52] H. Ge and I.P. Kirsteins, "Multi-rank processing for passive ranging in underwater acoustic environments subject to spatial coherence loss," in *Proc. IEEE Conf. Acoustics, Speech, and Signal Process.*, pp. 2692–2695, 22–27 May 2011.

- [53] A.B. Gershman, N.D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, May 2010.
- [54] Z.Q. Luo, W.K. Ma, A.M.C. So, Y. Ye, S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.
- [55] D. Simon and T.L. Chia, "Kalman filtering with state equality constraints," *IEEE Trans. Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 128–136, Jan. 2002.
- [56] S.J. Julier and J.J. LaViola, "On Kalman filtering with nonlinear equality constraints," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2774–2784, June 2007.
- [57] C. Yang and E. Blasch, "Kalman filtering with nonlinear state constraints," *IEEE Trans. Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 70–84, Jan. 2009.
- [58] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, Aug. 2010.
- [59] R.J. Hewett, M.T. Heath, M.D. Butala, and F. Kamalabadi, "A robust null space method for linear equality constrained state estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 3961–3971, Aug. 2010.
- [60] S. Schmidt, "Application of state-space methods to navigation problems," *Adv. Contr. Syst.*, pp. 293–340, 1966.
- [61] M. Athans, R.P. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Trans. Automatic Control*, vol. 13, no. 5, pp. 504–514, Oct. 1968.
- [62] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. Wiley.com, 2006.
- [63] S.J. Julier and J.K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*, pp. 182–193, 1997.
- [64] E.A. Wan and R. Van der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Symposium Adaptive Systems for Signal Process., Communications, and Control*, Lake Louise, Alberta, Canada, pp. 153–158, 2000.
- [65] S.J. Julier and J.K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
- [66] F. Gustafsson and G. Hendeby, "Some relations between extended and unscented Kalman filters," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 545–555, Feb. 2012.

- [67] Y.H. Chen and C.T. Chiang, "Adaptive beamforming using the constrained Kalman filter," *IEEE Trans. Antennas and Propagation*, vol. 41, no. 11, pp. 1576–1580, Nov. 1993.
- [68] A. El-Keyi, T. Kirubarajan, and A.B. Gershman, "Robust adaptive beamforming based on the Kalman filter," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3032–3041, Aug. 2005.
- [69] S.P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [70] S. Kwon, I. Oh, and S. Choi, "Adaptive beamforming from the generalized eigenvalue problem with a linear complexity for a wideband CDMA channel," in *Proc. Veh. Technol. Conf.*, vol. 3, Amsterdam, The Netherlands, pp.1890–1894, Sep. 1999.
- [71] Z. Tian, K.L. Bell, and H.L. Van Trees, "A recursive least squares implementation for LCMP beamforming under quadratic constraint," *IEEE Trans. Signal Processing*, vol. 49, no. 6, pp. 1138–1145, June 2001.
- [72] P.K. Dash, S. Hasan, and B.K. Panigrahi, "Adaptive complex unscented Kalman filter for frequency estimation of time-varying signals," *IET Science, Measurement & Technology*, vol. 4, no. 2, pp. 93–103, 2010.
- [73] R.K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automatic Control*, vol. 15, no. 2, pp. 175–184, Apr. 1970.
- [74] K. Myers and B.D. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Trans. Automatic Control*, vol. 21, no. 4, pp. 520–523, Aug. 1976.
- [75] X.M. Yang, X.Q. Yang, and K.L. Teo, "A matrix trace inequality," *Journal of Mathematical Analysis and Applications*, vol. 263, no. 1, pp. 327–331, 2001.
- [76] J. Bitzer, K.U. Simmer, and K.-D. Kammeyer, "Multimicrophone noise reduction techniques for hands-free speech recognition – a comparative study," *Robust Methods for Speech Recognition in Adverse Conditions (ROBUST-99)*, pp. 171–174, 1999.
- [77] E. Habets, "Room impulse response (RIR) generator," Jul. 2006 [Online]. Available: http://home.tiscali.nl/ehabets/rir_generator.html.
- [78] C.F. Eyring, "Reverberation time in "dead" rooms," *Journal of Acoust. Soc. Amer.*, vol. 1, no. 2A, pp. 217–241, 1930.
- [79] 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 Communication*, vol. 12, no. 3, pp. 247–251, July 1993.
- [80] "Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU, ITU-T Rec. P. 862, 2000.

- [81] T. Rohdenburg, V. Hohmann, and B. Kollmeir, "Objective perceptual quality measures for the evaluation of noise reduction schemes," in *Proc. 9th Int. Workshop Acoust. Echo Noise Control*, 2005, pp. 169–172. ITU-T Rec. G. 160, 2008.



Appendix I

In this appendix, the derivation of the SOE Kalman filter is given in detail. This appendix clearly points out the assumptions and approximations in the derivation of the SOE Kalman filter. For more reference, the reader can refer to the paper [61]. In our case, we use the Kalman filter as an adaptive filter. Thus, the state transition is linear and only the output matrix is considered as nonlinear. For brevity, the frequency index ω is omitted in the derivation, and the state-space representation is recalled as follows:

State equation

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mathbf{v}_k^s \quad (\text{A-1a})$$

Measurement equation

$$\mathbf{z} = \mathbf{f}(\mathbf{w}_k) + \mathbf{v}_k^m \quad (\text{A-1b})$$

where \mathbf{w}_k is the state at time index k and $\mathbf{f}(\mathbf{w}_k)$ is the nonlinear output function.

Assumption 1: The process noise \mathbf{v}_k^s and the measurement noise \mathbf{v}_k^m are zero-mean and mutually uncorrelated.

The objective is to minimize errors between the true state and the estimated states in the state equation and the measurement equation in the MSE sense. Let us denote the *a posteriori* and *a posteriori* errors as $\mathbf{e}_k^- = \mathbf{w}_k - \hat{\mathbf{w}}_k^-$ and $\mathbf{e}_k^+ = \mathbf{w}_k - \hat{\mathbf{w}}_k^+$, where $\hat{\mathbf{w}}_k^-$

and $\hat{\mathbf{w}}_k^+$ are the *a priori* and *a posteriori* estimate of the state using the state and measurement equations, respectively. Using the state equation (A-1), the *a priori* state estimate at time k is estimated by the *a posteriori* state estimate at time $k-1$ as

$$\hat{\mathbf{w}}_k^- = \hat{\mathbf{w}}_{k-1}^+ \quad (\text{A-2})$$

It can be seen that the estimate in (A-2) is unbiased since the process noise \mathbf{v}_k^s is assumed to be zero-mean (one can take expectation on (A-1) to find this). Next, the *a priori* state error covariance $\mathbf{P}_k^- = E[\mathbf{e}_k^-(\mathbf{e}_k^-)^H]$ can be expressed as

$$\begin{aligned} \mathbf{P}_k^- &= E\left[(\mathbf{w}_k - \hat{\mathbf{w}}_k^-)(\mathbf{w}_k - \hat{\mathbf{w}}_k^-)^H\right] \\ &= E\left[(\mathbf{w}_{k-1} + \mathbf{v}_{k-1}^s - \hat{\mathbf{w}}_{k-1}^+)(\mathbf{w}_{k-1} + \mathbf{v}_{k-1}^s - \hat{\mathbf{w}}_{k-1}^+)^H\right] \\ &= E\left[(\mathbf{e}_{k-1}^+ + \mathbf{v}_{k-1}^s)(\mathbf{e}_{k-1}^+ + \mathbf{v}_{k-1}^s)^H\right] \end{aligned} \quad (\text{A-3})$$

Assumption 2: The *a posteriori* error \mathbf{e}_k^+ and the process noise \mathbf{v}_k^s are assumed to be uncorrelated, i.e., $E[(\mathbf{e}_k^+)^H \mathbf{v}_k^s] = \mathbf{0}$.

By the Assumption 2, (A-3) can be expressed as

$$\begin{aligned} \mathbf{P}_k^- &= E\left[\mathbf{e}_{k-1}^+(\mathbf{e}_{k-1}^+)^H\right] + E\left[\mathbf{v}_{k-1}^s(\mathbf{v}_{k-1}^s)^H\right] \\ &= \mathbf{P}_{k-1}^+ + \mathbf{Q} \end{aligned} \quad (\text{A-4})$$

Now, consider the update from the measurement equation. Since the output function $\mathbf{f}(\mathbf{w}_k)$ is nonlinear, the Taylor series expansion is used to approximate the function. In our problem, the function is quadratic. Thus, the second-order approximation is sufficient to describe the nonlinear function.

Approximation 1: Assume the complex-valued nonlinear function $\mathbf{f}(\mathbf{w}_k)$ is second-order differentiable around the nominal point $\hat{\mathbf{w}}_k^-$, then the nonlinear function can be expanded as [61]

$$\begin{aligned}\mathbf{f}(\mathbf{w}_k) &\approx \mathbf{f}(\hat{\mathbf{w}}_k^-) + \mathbf{F}_w(\mathbf{w}_k - \hat{\mathbf{w}}_k^-) + \frac{1}{2} \sum_{i=1}^P \phi_i (\mathbf{w}_k - \hat{\mathbf{w}}_k^-)^H \mathbf{F}_{ww}^{(i)} (\mathbf{w}_k - \hat{\mathbf{w}}_k^-) \\ &= \mathbf{f}(\hat{\mathbf{w}}_k^-) + \mathbf{F}_w \mathbf{e}_k^- + \frac{1}{2} \sum_{i=1}^P \phi_i \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \right]\end{aligned}\quad (\text{A-5})$$

where $\phi_i = \left[0, \dots, 1, \dots, 0 \right]^T$ and P is the number of measurement equations.

Given Approximation 1, the *a posteriori* state estimate can be updated in a linear form

$$\hat{\mathbf{w}}_k^+ = \hat{\mathbf{w}}_k^- + \mathbf{K}_k \left[\mathbf{z} - \mathbf{f}(\hat{\mathbf{w}}_k^-) - \boldsymbol{\pi}_k \right] \quad (\text{A-6})$$

where \mathbf{K}_k is the Kalman gain to be estimated and $\boldsymbol{\pi}_k$ is the correction term to ensure that $\hat{\mathbf{w}}_k^+$ is unbiased, i.e., $E[\mathbf{e}_k^+] = E[\mathbf{w}_k - \hat{\mathbf{w}}_k^+] = \mathbf{0}$. Given (A-6), the *a posteriori* error can be rewritten as

$$\begin{aligned}\mathbf{e}_k^+ &= \mathbf{w}_k - \hat{\mathbf{w}}_k^+ \\ &= \mathbf{e}_k^- - \mathbf{K}_k \left[\mathbf{z} - \mathbf{f}(\hat{\mathbf{w}}_k^-) \right] + \boldsymbol{\pi}_k \\ &= \mathbf{e}_k^- - \mathbf{K}_k \left[\mathbf{f}(\mathbf{w}_k) + \mathbf{v}_k^m - \mathbf{f}(\hat{\mathbf{w}}_k^-) - \boldsymbol{\pi}_k \right] \\ &= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{e}_k^- - \frac{1}{2} \mathbf{K}_k \cdot \left\{ \sum_{i=1}^P \phi_i \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \right] \right\} - \mathbf{K}_k \mathbf{v}_k^m + \mathbf{K}_k \boldsymbol{\pi}_k\end{aligned}\quad (\text{A-7})$$

Assumption 3: Assume the *a priori* state estimate is unbiased, i.e., $E[\mathbf{e}_k^-] = \mathbf{0}$.

According to Assumption 1 and Assumption 3, the bias $\boldsymbol{\pi}_k$ can be obtained by taking expectation on (A-7) and equating to zero as

$$\boldsymbol{\pi}_k = \frac{1}{2} \left\{ \sum_{i=1}^P \boldsymbol{\phi}_i \cdot \text{tr} \left[\mathbf{F}_{\text{ww}}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \right] \right\} \quad (\text{A-8})$$

Note that the Kalman gain \mathbf{K}_k is a random variable and it should not be taken out the expectation operator. However, in the derivations of *a posteriori* error \mathbf{e}_k^+ and the *a posteriori* state error covariance $\mathbf{P}_k^+ = E \left[\mathbf{e}_k^+ (\mathbf{e}_k^+)^H \right]$, the Kalman gain is regarded as a determinant matrix, as the way used in the traditional Kalman filter.

$$\textbf{Approximation 2: } E \left(\text{tr} \left[\mathbf{F}_{\text{ww}}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \right] \right) \approx \text{tr} \left[\mathbf{F}_{\text{ww}}^{(i)} E \left[\mathbf{e}_k^- (\mathbf{e}_k^-)^H \right] \right] = \text{tr} \left[\mathbf{F}_{\text{ww}}^{(i)} \mathbf{P}_k^- \right]$$

By Approximation 2, $\boldsymbol{\pi}_k$ can be expressed by

$$\boldsymbol{\pi}_k \approx \frac{1}{2} \left\{ \sum_{i=1}^P \boldsymbol{\phi}_i \cdot \text{tr} \left[\mathbf{F}_{\text{ww}}^{(i)} \mathbf{P}_k^- \right] \right\} \quad (\text{A-9})$$

Now, based on (A-7), we try to find the *a posteriori* state error covariance \mathbf{P}_k^+ .

Before the derivation of \mathbf{P}_k^+ , we give the following assumption.

Assumption 4: The *a priori* error \mathbf{e}_k^- and the measurement noise \mathbf{v}_k^m are assumed to

be mutually uncorrelated, i.e., $E \left[\left[\mathbf{e}_k^+ \right]_i^* \cdot \left[\mathbf{v}_k^m \right]_j \right] = 0$ for all i and j .

According to Assumption 1, Assumption 3, Assumption 4, and (A-7), \mathbf{P}_k^+ can be expressed as (here, $\boldsymbol{\pi}_k$ and \mathbf{K}_k are considered as determinant values)

$$\begin{aligned}
\mathbf{P}_k^+ &= E \left[\left((\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{e}_k^- - \frac{1}{2} \mathbf{K}_k \left\{ \sum_{i=1}^P \boldsymbol{\phi}_i \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \right] \right\} - \mathbf{K}_k \mathbf{v}_k^m + \mathbf{K}_k \boldsymbol{\pi}_k \right) (\bullet)^H \right] \\
&= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^H - \mathbf{K}_k \boldsymbol{\pi}_k \boldsymbol{\pi}_k^H \mathbf{K}_k^H \\
&\quad - \frac{1}{2} (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) E \left[\sum_{i=1}^P \mathbf{e}_k^- (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- \cdot \boldsymbol{\phi}_i^T \right] \mathbf{K}_k^H \\
&\quad - \frac{1}{2} \mathbf{K}_k E \left[\sum_{i=1}^P \boldsymbol{\phi}_i \cdot (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- \cdot (\mathbf{e}_k^-)^H \right] (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H \\
&\quad + \frac{1}{4} \mathbf{K}_k E \left[\sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi}_i \cdot (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(j)} \mathbf{e}_k^- \cdot \boldsymbol{\phi}_j^T \right] \mathbf{K}_k^H
\end{aligned} \tag{A-10}$$

Assumption 5: Assume \mathbf{e}_k^- is complex Gaussian (s.t. the joint distributions are also

Gaussian), then its 3rd-order moment $E \left[e_{k,l}^- (e_{k,m}^-)^* e_{k,n}^- \right] = 0, \quad \forall l, m, n$

Since

$$\begin{aligned}
&E \left[\mathbf{e}_k^- (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- \right] \\
&= E \left[\mathbf{e}_k^- \left(\sum_{m=1}^M \sum_{n=1}^M (e_{k,m}^-)^* \left[\mathbf{F}_{ww}^{(i)} \right]_{m,n} e_{k,n}^- \right) \right] \\
&= E \left[\left(\sum_{m=1}^M \sum_{n=1}^M e_{k,1}^- (e_{k,m}^-)^* \left[\mathbf{F}_{ww}^{(i)} \right]_{m,n} e_{k,n}^- \right), \dots, \left(\sum_{m=1}^M \sum_{n=1}^M e_{k,M}^- (e_{k,m}^-)^* \left[\mathbf{F}_{ww}^{(i)} \right]_{m,n} e_{k,n}^- \right) \right]^T \\
&= \left[\sum_{m=1}^M \sum_{n=1}^M \left[\mathbf{F}_{ww}^{(i)} \right]_{m,n} E \left[e_{k,1}^- (e_{k,m}^-)^* e_{k,n}^- \right], \dots, \sum_{m=1}^M \sum_{n=1}^M \left[\mathbf{F}_{ww}^{(i)} \right]_{m,n} E \left[e_{k,M}^- (e_{k,m}^-)^* e_{k,n}^- \right] \right]^T
\end{aligned} \tag{A-11}$$

By Assumption 5 and (A-11), (A-10) can be rewritten as

$$\begin{aligned}
\mathbf{P}_k^+ &= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^H \\
&\quad + \frac{1}{4} \mathbf{K}_k \left\{ \sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi}_i \cdot E \left[(\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(j)} \mathbf{e}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right\} \mathbf{K}_k^H - \mathbf{K}_k \boldsymbol{\pi}_k \boldsymbol{\pi}_k^H \mathbf{K}_k^H
\end{aligned} \tag{A-12}$$

Substituting $\boldsymbol{\pi}_k$ into (A-12), we have

$$\begin{aligned}
\mathbf{P}_k^+ &= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^H \\
&+ \frac{1}{4} \mathbf{K}_k \left\{ \sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi} \cdot E \left[(\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(j)} \mathbf{e}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right\} \mathbf{K}_k^H \\
&- \frac{1}{4} \mathbf{K}_k \cdot \left[\sum_{i=1}^P \boldsymbol{\phi} \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{P}_k^- \right] \right] \cdot \left[\sum_{j=1}^P \text{tr} \left[\mathbf{F}_{ww}^{(j)} \mathbf{P}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right] \mathbf{K}_k^H \\
&= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^H \\
&+ \mathbf{K}_k \left\{ \frac{1}{4} \sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi} \cdot E \left[(\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(i)} \mathbf{e}_k^- (\mathbf{e}_k^-)^H \mathbf{F}_{ww}^{(j)} \mathbf{e}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right. \\
&\left. - \frac{1}{4} \left[\sum_{i=1}^P \boldsymbol{\phi} \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{P}_k^- \right] \right] \cdot \left[\sum_{j=1}^P \text{tr} \left[\mathbf{F}_{ww}^{(j)} \mathbf{P}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right] \right\} \mathbf{K}_k^H
\end{aligned} \tag{A-13}$$

Lemma 1: If \mathbf{x} is a zero-mean, Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}$, then

$$\begin{aligned}
E \left\{ \mathbf{x} \cdot \text{Tr} \left[\mathbf{x} \mathbf{x}^H \right] \right\} &= \mathbf{0} \\
E \left\{ \text{Tr} \left[\mathbf{A} \mathbf{x} \mathbf{x}^H \mathbf{B} \mathbf{x} \mathbf{x}^H \right] \right\} &= E \left\{ \text{Tr} \left[\mathbf{A} \mathbf{x} \mathbf{x}^H \right] \text{Tr} \left[\mathbf{B} \mathbf{x} \mathbf{x}^H \right] \right\} \\
&= 2 \text{Tr} \left[\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} \boldsymbol{\Sigma} \right] + \text{Tr} \left[\mathbf{A} \boldsymbol{\Sigma} \right] \text{Tr} \left[\mathbf{B} \boldsymbol{\Sigma} \right]
\end{aligned}$$

where \mathbf{A} and \mathbf{B} are symmetric matrices.

Using Lemma 1, (A-13) can be rearranged as

$$\begin{aligned}
\mathbf{P}_k^+ &= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^H \\
&+ \mathbf{K}_k \left\{ \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi} \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{P}_k^- \mathbf{F}_{ww}^{(j)} \mathbf{P}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right. \\
&+ \frac{1}{4} \sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi} \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{P}_k^- \right] \cdot \text{tr} \left[\mathbf{F}_{ww}^{(j)} \mathbf{P}_k^- \right] \cdot \boldsymbol{\phi}_j^T \\
&\left. - \frac{1}{4} \left[\sum_{i=1}^P \boldsymbol{\phi} \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{P}_k^- \right] \right] \cdot \left[\sum_{j=1}^P \text{tr} \left[\mathbf{F}_{ww}^{(j)} \mathbf{P}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right] \right\} \mathbf{K}_k^H \\
&= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^H \\
&+ \frac{1}{2} \mathbf{K}_k \left[\sum_{i=1}^P \sum_{j=1}^P \boldsymbol{\phi} \cdot \text{tr} \left[\mathbf{F}_{ww}^{(i)} \mathbf{P}_k^- \mathbf{F}_{ww}^{(j)} \mathbf{P}_k^- \right] \cdot \boldsymbol{\phi}_j^T \right] \mathbf{K}_k^H \\
&= (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w) \mathbf{P}_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{F}_w)^H + \mathbf{K}_k (\mathbf{R} + \boldsymbol{\Lambda}_k) \mathbf{K}_k^H
\end{aligned} \tag{A-14a}$$

where

$$\Lambda_k = \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \phi_i \cdot \text{tr} \left[\mathbf{F}_{\text{ww}}^{(i)} \mathbf{P}_k^- \mathbf{F}_{\text{ww}}^{(j)} \mathbf{P}_k^- \right] \cdot \phi_j^T \quad (\text{A-14b})$$

Finally, the Kalman gain \mathbf{K}_k can be derived by minimizing the trace of the *a posteriori* state error covariance \mathbf{P}_k^+ . By taking trace operation on (A-14a), we have

$$\begin{aligned} \text{Tr}[\mathbf{P}_k^+] &= \text{Tr}[\mathbf{P}_k^-] - \text{Tr}[\mathbf{P}_k^- \mathbf{F}_w^H \mathbf{K}_k^H] - \text{Tr}[\mathbf{K}_k \mathbf{F}_w \mathbf{P}_k^-] \\ &\quad + \text{Tr}[\mathbf{K}_k (\mathbf{F}_w \mathbf{P}_k^- \mathbf{F}_w^H + \mathbf{R} + \Lambda_k) \mathbf{K}_k^H] \end{aligned} \quad (\text{A-15})$$

For taking derivatives on the traces, the following lemma can be used.

Lemma 2: $\frac{\partial}{\partial \mathbf{X}^*} \text{Tr}[\mathbf{A} \mathbf{X}^H] = \mathbf{A}$, $\frac{\partial}{\partial \mathbf{X}^*} \text{Tr}[\mathbf{X} \mathbf{A}] = \mathbf{0}$, $\frac{\partial}{\partial \mathbf{X}^*} \text{Tr}[\mathbf{X} \mathbf{A} \mathbf{X}^H] = \mathbf{X} \mathbf{A}$

Taking derivative on (A-15) and using Lemma 2, we have

$$\frac{\partial \text{Tr}[\mathbf{P}_k^+]}{\partial \mathbf{K}_k^*} = -\mathbf{P}_k^- \mathbf{F}_w^H + \mathbf{K}_k (\mathbf{F}_w \mathbf{P}_k^- \mathbf{F}_w^H + \mathbf{R} + \Lambda_k) = \mathbf{0} \quad (\text{A-16})$$

The Kalman gain is obtained by

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{F}_w^H (\mathbf{F}_w \mathbf{P}_k^- \mathbf{F}_w^H + \mathbf{R} + \Lambda_k)^{-1} \quad (\text{A-17})$$

Appendix II

This appendix is to find the relationship between the proposed SCPF and the Zelinski post-filter. To compare with the Zelinski post-filter, the special case of the proposed SCPF in (3-19) is used. Recall in Section 3.2, before the post-processing of the Zelinski's and McCowan *et als* methods, the microphones have to pass a time alignment module to adjust the mismatch between microphones, as described in (3-1). Thus, the PSD matrix of the aligned input vector can be denoted as $\tilde{\Phi}_x(\omega)$ and the SCPF can be rewritten as

$$\begin{aligned}
 G_{\text{SCPF}}(\omega) &= \frac{\phi_s(\omega) \mathbf{a}_s^H(\omega) \Phi_x(\omega) \mathbf{a}_s(\omega)}{\phi_s(\omega) \|\mathbf{a}_s(\omega)\|^2 \cdot \text{tr}(\Phi_x(\omega))} \\
 &= \frac{\mathbf{1}^T \tilde{\Phi}_x(\omega) \mathbf{1}}{M \cdot \text{tr}(\tilde{\Phi}_x(\omega))} \\
 &= \frac{\sum_{i=1}^M \sum_{j=1}^M \tilde{\phi}_{x_i x_j}(\omega)}{\left(M \cdot \sum_{i=1}^M \tilde{\phi}_{x_i x_i}(\omega) \right)} \\
 &= \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \tilde{\phi}_{x_i x_j}(\omega) / \left(\frac{1}{M} \cdot \sum_{i=1}^M \tilde{\phi}_{x_i x_i}(\omega) \right)
 \end{aligned} \tag{A-18}$$

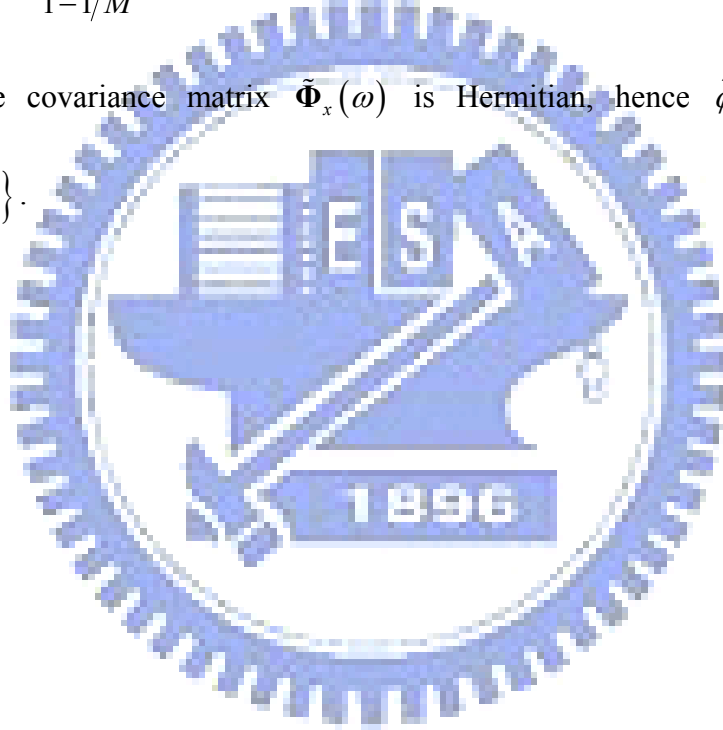
where $\mathbf{1}$ is the all-one vector and $\text{tr}(\tilde{\Phi}_x(\omega)) = \text{tr}(\Phi_x(\omega))$ if the all the magnitudes of the elements in \mathbf{a}_s are equal to unity. $\tilde{\phi}_{x_i x_j}(\omega)$ is the cross-spectral density between the aligned inputs at the i -th and j -th microphones. In this case, the SCPF can be interpreted as the ratio between the average of total spectral densities and the average of auto-spectral densities. Since the Zelinski post-filter is the ratio between the average of cross-spectral densities and the average of auto-spectral densities (as in (3-8)), the relationship between the SCPF and the Zelinski post-filter can be easily derived by

$$\begin{aligned}
G_{\text{SCPF}}(\omega) &= \frac{\frac{1}{M^2} \left(\underbrace{2 \sum_{i=1}^M \sum_{j=1}^M \Re \{ \tilde{\phi}_{x_i x_j}(\omega) \}}_{\text{off-diagonal}} + \underbrace{\sum_{i=1}^M \tilde{\phi}_{x_i x_i}(\omega)}_{\text{diagonal}} \right)}{\frac{1}{M} \cdot \sum_{i=1}^M \tilde{\phi}_{x_i x_i}(\omega)} \\
&= G_{\text{Zelinski}}(\omega) + \frac{1}{M} (1 - G_{\text{Zelinski}}(\omega))
\end{aligned} \tag{A-19}$$

Or it can be written as

$$G_{\text{Zelinski}}(\omega) = \frac{G_{\text{SCPF}}(\omega) - 1/M}{1 - 1/M} \tag{A-20}$$

Note that the covariance matrix $\tilde{\Phi}_x(\omega)$ is Hermitian, hence $\tilde{\phi}_{x_i x_j}(\omega) + \tilde{\phi}_{x_j x_i}(\omega) = 2\Re \{ \tilde{\phi}_{x_i x_j}(\omega) \}$.



Appendix III

This appendix analyzes the proposed BC-SCPF. To compare with the McCowan post-filter, the assumptions of homogeneous sound fields and point source model are used. Assume the actual and estimated noise coherence matrices from the microphones are $\Gamma_n(\omega)$ and $\hat{\Gamma}_n(\omega)$. Then according to (3-30) and (3-33), we have

$$\begin{aligned}
 G_{\text{BC-SCPF}}(\omega) &= \left[\frac{M^2 \xi(\omega) + \mathbf{a}_s^H(\omega) \Gamma_n(\omega) \mathbf{a}_s(\omega)}{M^2 (\xi(\omega) + 1)} - \frac{\mathbf{a}_s^H(\omega) \hat{\Gamma}_n(\omega) \mathbf{a}_s(\omega)}{M^2} \right] / \\
 &\quad \left(1 - \frac{\mathbf{a}_s^H(\omega) \hat{\Gamma}_n(\omega) \mathbf{a}_s(\omega)}{M^2} \right) \\
 &= G_{\text{Wiener}}(\omega) \\
 &\quad + \frac{\phi_s(\omega)}{\phi_s(\omega) + \phi_n(\omega)} \cdot \frac{\mathbf{a}_s^H(\omega) \Gamma_n(\omega) \mathbf{a}_s(\omega) - \mathbf{a}_s^H(\omega) \hat{\Gamma}_n(\omega) \mathbf{a}_s(\omega)}{\left(M^2 - \mathbf{a}_s^H(\omega) \hat{\Gamma}_n(\omega) \mathbf{a}_s(\omega) \right)}
 \end{aligned} \tag{A-21}$$

Using the time alignment expression similar to (A-18) in Appendix II, (A-21) can be rewritten as

$$\begin{aligned}
 G_{\text{BC-SCPF}}(\omega) &= G_{\text{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega) + \phi_n(\omega)} \left[\frac{\mathbf{1}^T \tilde{\Gamma}_n(\omega) \mathbf{1} - \mathbf{1}^T \hat{\Gamma}_n(\omega) \mathbf{1}}{\left(M^2 - \mathbf{1}^T \hat{\Gamma}_n(\omega) \mathbf{1} \right)} \right] \\
 &= G_{\text{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega) + \phi_n(\omega)} \left[\frac{\sum_{i=1}^M \sum_{j=1}^M \tilde{\Gamma}_{n_i n_j}(\omega) - \sum_{i=1}^M \sum_{j=1}^M \hat{\Gamma}_{n_i n_j}(\omega)}{\left(M^2 - \sum_{i=1}^M \sum_{j=1}^M \hat{\Gamma}_{n_i n_j}(\omega) \right)} \right] \\
 &= G_{\text{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega) + \phi_n(\omega)} \left[\frac{1}{M^2} \frac{\sum_{i=1}^M \sum_{j=1}^M \left(\tilde{\Gamma}_{n_i n_j}(\omega) - \hat{\Gamma}_{n_i n_j}(\omega) \right)}{\sum_{i=1}^M \sum_{j=1}^M \left(1 - \hat{\Gamma}_{n_i n_j}(\omega) \right)} \right]
 \end{aligned} \tag{A-22}$$

where $\tilde{\Gamma}_{n,n_j}(\omega)$ and $\hat{\Gamma}_{n,n_j}(\omega)$ are the actual and estimated noise coherence matrices of the aligned inputs.

