KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

MULTI-MICROPHONE NOISE REDUCTION AND DEREVERBERATION TECHNIQUES FOR SPEECH APPLICATIONS

Promotor:
Prof. dr. ir. M. Moonen

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Simon DOCLO

Mei 2003

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

MULTI-MICROPHONE NOISE REDUCTION AND DEREVERBERATION TECHNIQUES FOR SPEECH APPLICATIONS

Jury:
Prof. dr. ir. J. Berlamont, voorzitter
Prof. dr. ir. M. Moonen, promotor
Dr. I. Dologlou (ILSP, Greece)
Prof. dr. ir. P. Sommen (TU Eindhoven)
Prof. dr. ir. D. Van Compernolle
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. J. Vandewalle
Prof. dr. J. Wouters

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Simon DOCLO

U.D.C. 621.395

Mei 2003

D/2003/7515/19

ISBN 90-5682-409-0

Voorwoord

Bij het einde van mijn doctoraat wil ik met plezier een aantal mensen bedanken. In de eerste plaats gaat mijn oprechte dank uit naar mijn promotor Prof. Marc Moonen, die mij overtuigde om een doctoraat in dit boeiend onderzoeksdomein te beginnen. Ik dank Marc omdat ik in alle vrijheid onderzoek heb kunnen uitvoeren, terwijl zijn talrijke nieuwe ideeën en constructieve opmerkingen een grote steun voor mij hebben betekend. Verder wil ik Marc bedanken voor zijn nooit aflatend enthousiasme en de waardevolle 'blauwe bic'-verbeteringen van mijn teksten. Ik had me echt geen betere promotor kunnen wensen.

Verder zou ik alle leden van het leescomité willen bedanken voor hun kostbare dagen – en nachten – die ze besteed hebben aan het nalezen van dit proefschrift. Ik zou Prof. Dirk Van Compernolle graag willen bedanken voor zijn kritische opmerkingen en onverwachte invalshoeken die tot nieuwe inzichten hebben geleid. Prof. Sabine Van Huffel en Prof. Piet Sommen zou ik willen bedanken voor hun waardevolle suggesties en voor de constante interesse die ze in mijn werk hebben betoond. I would also like to thank Dr. Yannis Dologlou for proofreading this manuscript and for the many valuable suggestions and interesting discussions we had during the first year of my research.

Ik dank ook de andere leden van de jury voor hun onmiddellijke bereidheid in de jury te zetelen, ondanks hun drukke agenda. Ik zou Prof. Jan Wouters graag willen bedanken omdat hij mij vele jaren geleden (als thesisstudent) warm heeft gemaakt voor onderzoek naar hoorapparaten en cochleaire implantaten. Nadien is hij mijn onderzoek steeds blijven volgen en ik zou hem willen bedanken voor de zeer goede samenwerking met zijn onderzoeksgroep. Prof. Joos Vandewalle zou ik willen bedanken voor zijn grenzeloze inzet om van SISTA een florerende onderzoeksgroep te maken en voor alle kansen die ik daardoor gekregen heb. Verder dank ik Prof. Jean Berlamont voor het waarnemen van het voorzitterschap van de jury.

Ik ben het I.W.T. erkentelijk voor de financiële ondersteuning gedurende de eerste jaren van mijn doctoraat.

Verder wil ik de talrijke SISTA-leden bedanken voor de aangename sfeer binnen en buiten de werkuren. De discussies op het werk werden gerust op een terrasje voortgezet en de vele Alma-uren waren zeer geschikt om de laatste roddels te weten te komen. Toch zijn er een aantal mensen die ik speciaal wil bedanken: Ann, Geert, Koen, Ruben, bedankt voor[1] de goede samenwerking, de – niet altijd wetenschappelijke – muzikale momenten, het (luide) lachen, het schitterend implementatiewerk, jullie steun. I would also like to thank Sharon for the many valuable discussions and for sharing the intricate details about the Middle-East conflict. Verder wil ik de mensen van het eerste uur Bart, Geert, Katleen, Katrien, Leen, Piet, Tony en Wouter bedanken omdat ze mij direct deden thuisvoelen in SISTA. Ik wil zeker ook de andere mensen in de DSP-groep Benoit, Geert, Geert, Gert, Hilde, Imad, Koen, Olivier, Raphael en Toon niet vergeten die ervoor zorgen dat dit zowel een stimulerende als een aangename werkomgeving is. Ook wil ik onze secretaresse Ida bedanken voor haar administratieve bijstand, de systeemgroep voor hun onderhoudsweekends en de boekhouding om mijn conferentiekosten op tijd te betalen.

Iets verder van het werk zijn er zeker mijn vrienden en huisgenoten die mij op tijd en stond toelieten om achtergrondruis en microfoonroosters uit mijn gedachten te bannen. Bedankt Pim, Erik, Vinkenlaan 25-ers, LUK-vrienden, Fliet Zorro en alle anderen!

En natuurlijk – last but not least – wil ik mijn ouders bedanken voor de mogelijkheden die ze mij geboden hebben en voor hun onvoorwaardelijke steun en vertrouwen.

Simon Doclo
mei 2003

[1] schrappen wat niet past

Abstract

In typical speech communication applications, such as hands-free mobile telephony, voice-controlled systems and hearing aids, the recorded microphone signals are corrupted by background noise, room reverberation and far-end echo signals. This signal degradation can lead to total unintelligibility of the speech signal and decreases the performance of automatic speech recognition systems. In this thesis several multi-microphone noise reduction and dereverberation techniques are developed.

In Part I we present a Generalised Singular Value Decomposition (GSVD) based optimal filtering technique for enhancing multi-microphone speech signals which are degraded by additive coloured noise. Several techniques are presented for reducing the computational complexity and we show that the GSVD-based optimal filtering technique can be integrated into a 'Generalised Sidelobe Canceller' type structure. Simulations show that the GSVD-based optimal filtering technique achieves a larger signal-to-noise ratio improvement than standard fixed and adaptive beamforming techniques and that it is more robust against several deviations from the assumed signal model.

In Part II multi-microphone algorithms for time-delay estimation, dereverberation, and combined noise reduction and dereverberation are discussed. Since these algorithms require an estimate of the acoustic impulse responses, we also present batch and adaptive techniques for estimating the acoustic impulse responses, both in the time-domain and in the frequency-domain. We derive a stochastic gradient algorithm which iteratively estimates the generalised eigenvector corresponding to the smallest generalised eigenvalue and which can be used for time-delay estimation. We show that by integrating the normalised matched filter with the multi-channel Wiener filter, a combined noise reduction and dereverberation technique is obtained.

In Part III several design procedures and cost functions are discussed for designing fixed broadband beamformers with an arbitrary desired spatial directivity pattern for a given arbitrary microphone array configuration, using an FIR filter-and-sum structure. We present two novel cost functions, which are based on eigenfilters. We discuss far-field, near-field and mixed near-field far-field broadband beamformer design, and we present two design procedures for designing broadband beamformers that are robust against gain and phase errors in the microphone characteristics.

Korte Inhoud

In veel spraakcommunicatietoepassingen, zoals handenvrije mobiele telefonie, spraakgestuurde systemen en hoorapparaten, zijn de opgenomen microfoonsignalen vaak van lage kwaliteit ten gevolge van achtergrondlawaai, reverberatie en 'far-end'-echosignalen. Deze slechte signaalkwaliteit kan ertoe leiden dat het gewenste spraaksignaal totaal onverstaanbaar wordt en dat de performantie van systemen voor automatische spraakherkenning aanzienlijk vermindert. In deze doctoraatsthesis worden verschillende technieken ontwikkeld voor ruisonderdrukking en dereverberatie met behulp van meerdere microfoons.

In Deel I stellen we een optimaal-filtertechniek, gebaseerd op de Veralgemeende-Singuliere-Waarde-Ontbinding (GSVD), voor om de signaalkwaliteit van meerkanaals spraaksignalen te verbeteren wanneer additieve gekleurde ruis aanwezig is. Verschillende technieken worden besproken om de berekeningscomplexiteit te verminderen en we tonen aan dat deze GSVD-gebaseerde optimaal-filtertechniek geïntegreerd kan worden in een 'Generalised Sidelobe Canceller'-structuur. Simulaties tonen aan dat de GSVD-gebaseerde optimaal-filtertechniek een grotere verbetering in signaal-ruisverhouding oplevert dan standaard vaste en adaptieve bundelvorming en dat deze techniek robuuster is wanneer afwijkingen in het verondersteld signaalmodel optreden.

In Deel II worden meer-kanaals algoritmes besproken voor het schatten van tijdsvertraging, voor dereverberatie en voor gecombineerde ruisonderdrukking en dereverberatie. Aangezien deze algoritmes een schatting vereisen van de akoestische impulsresponsies, bespreken we ook adaptieve en niet-adaptieve technieken om akoestische impulsresponsies te schatten, zowel in het tijdsdomein als in het frequentiedomein. We leiden een stochastisch-gradiënt-algoritme af dat iteratief de veralgemeende eigenvector berekent behorend bij de kleinste veralgemeende eigenwaarde en dat gebruikt kan worden voor het schatten van tijdsvertraging. We tonen aan dat een gecombineerde techniek voor ruisonderdrukking en dereverberatie kan bekomen worden door het genormaliseerd 'matched' filter te integreren met het meer-kanaals Wiener-filter.

In Deel III worden verschillende ontwerpprocedures besproken voor vaste breedband bundelvormers met een willekeurig spatiaal directiviteitspatroon voor een gegeven willekeurig microfoonrooster, met behulp van een FIR 'filter-and-sum'-structuur. We stellen 2 nieuwe kostfuncties voor die gebaseerd zijn op eigenfilters. We bespreken het ontwerp van 'far-field', 'near-field' en 'mixed near-field far-field' breedband bundelvormers en we ontwikkelen 2 ontwerpprocedures voor breedband bundelvormers die robuust zijn tegen afwijkingen in de versterking en de fase van de microfoons.

Glossary

Mathematical Notation

a	scalar a
$\mathbf{a}$	vector $\mathbf{a}$
$\mathbf{A}$	matrix $\mathbf{A}$
a^*	complex conjugate of a
$\mathbf{A}^T$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
$\mathbf{a}^i$	ith element of vector $\mathbf{a}$
$a_{n,i}$	ith element of vector $\mathbf{a}_n$
$\mathbf{A}^{ij}$	(i,j)-th element of matrix $\mathbf{A}$
$[\mathbf{a}]_i$	ith sub-vector of vector $\mathbf{a}$
$[\mathbf{A}]_{ij}$	(i,j)-th sub-matrix of matrix $\mathbf{A}$
$\{\mathbf{A}\}_{i,i+1}$	2×2 sub-matrix of matrix $\mathbf{A}$ on the intersection of rows $\{i, i+1\}$ and columns $\{i, i+1\}$
$a_R, \mathbf{a}_R, \mathbf{A}_R$	real part of scalar a, vector $\mathbf{a}$, matrix $\mathbf{A}$
$a_I, \mathbf{a}_I, \mathbf{A}_I$	imaginary part of scalar a, vector $\mathbf{a}$, matrix $\mathbf{A}$
$x[k]$	discrete time-filter, time-sequence, stochastic process
$X(z)$	z-transform of $x[k]$
$X(\omega)$	Discrete-Time Fourier Transform of $x[k]$
$X(l,m)$	lth component of DFT of mth frame of $x[k]$
$c_x[k]$	complex cepstrum of $x[k]$
$r_x[k]$	autocorrelation function of $x[k]$
$r_{xy}[k]$	cross-correlation function of $x[k]$ and $y[k]$
$P_x(\omega)$	power spectral density of $x[k]$
$P_{xy}(\omega)$	cross-power spectral density of $x[k]$ and $y[k]$
$\Gamma_{xy}(\omega)$	complex coherence between $x[k]$ and $y[k]$
$G_{xy}(\omega)$	power transfer function between $x[k]$ and $y[k]$
$\bar{\mathbf{R}}_{xx} = \mathcal{E}\{\mathbf{x}\mathbf{x}^T\}$	autocorrelation matrix of vector $\mathbf{x}$
$\bar{\mathbf{R}}_{xy} = \mathcal{E}\{\mathbf{x}\mathbf{y}^T\}$	cross-correlation matrix of vectors $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{R}_{xx}$	empirical autocorrelation matrix of vector $\mathbf{x}$
$\mathbf{R}_{xy}$	empirical cross-correlation matrix of vectors $\mathbf{x}$ and $\mathbf{y}$

$f^{(i)}(a)$	ith derivative of function $f(a)$
$f_\alpha(a)$	probability density function of stochastic variable a
μ_a	mean of probability density function $f_\alpha(a)$
σ_a^2	variance of probability density function $f_\alpha(a)$
$\otimes$	convolution
$\odot$	element-wise multiplication
$\mathcal{O}(M)$	order M
$\mathcal{E}\{\cdot\}$	expectation operator
$\mathcal{F}\{\cdot\}$	Discrete-Time Fourier Transform operator
$\mathcal{F}^{-1}\{\cdot\}$	Inverse Discrete-Time Fourier Transform operator
$\Re\{\cdot\}$	real part
$\Im\{\cdot\}$	imaginary part
$\mathrm{tr}\{\mathbf{A}\}$	trace of matrix $\mathbf{A}$ (sum of diagonal elements)
$\mathrm{diag}\{\mathbf{a}\}$	square diagonal matrix with vector $\mathbf{a}$ as diagonal
$\lvert\cdot\rvert$	absolute value
$\lVert\cdot\rVert_2$	L_2-norm
$\lVert\cdot\rVert_\infty$	L_∞-norm
$\lVert\cdot\rVert_F$	Frobenius-norm
$\hat{a}, \hat{\mathbf{a}}, \hat{\mathbf{A}}$	estimate of scalar a, vector $\mathbf{a}$, matrix $\mathbf{A}$
$\lfloor a \rfloor$	largest integer smaller or equal than a
$\lceil a \rceil$	smallest integer larger or equal than a
$\mathrm{div}(a, b)$	integer division of a and b
$\mathrm{mod}(a, b)$	remainder of integer division of a and b
$a \ll b$	a is much smaller than b
$a \gg b$	a is much larger than b
$a \approx b$	a is approximately equal to b

Fixed Symbols

$A_n, A_n(\omega, \theta)$	microphone characteristic of nth microphone
$D(\omega, \theta), D(\omega, \theta, r)$	desired spatial directivity pattern of beamformer
$F(\omega, \theta), F(\omega, \theta, r)$	weighting function
$H(\omega, \theta), H(\omega, \theta, r)$	spatial directivity pattern of beamformer
I	number of images in acoustic impulse response $h_n[k]$
J	number of linear constraints
J_{MSE}	MSE cost function
K	filter length of acoustic room impulse response $h_n[k]$
L	filter length of FIR filters on microphones
L_{ANC}	filter length of FIR adaptive filter in ANC postprocessing stage
L_f	filter length of FIR filter on far-end echo signal
M	number of microphones × filter length ($M = LN$)
N	number of microphones
P	size of speech data matrix $\mathbf{Y}[k]$

Q	size of noise data matrix $\mathbf{V}[k]$
P_k	number of rows in speech data matrix $\mathbf{Y}[k]$ at time k
Q_k	number of rows in noise data matrix $\mathbf{V}[k]$ at time k
S	surface of room
T	threshold value
T_{60}	reverberation time
V	volume of room
$a_n, a_n(\omega, \theta)$	gain of microphone characteristic of nth microphone
c	speed of sound propagation ($c = 340 \frac{m}{s}$)
d	constant inter-microphone distance
d_n	distance between nth microphone and centre of microphone array
f	frequency-domain variable
f_s	sampling frequency
$f[k]$	total transfer function for speech signal $s[k]$
$f_0[k]$	far-end echo signal
$h_n[k]$	acoustic impulse response between source and nth microphone
$k,\ k'$	discrete-time index
m, n	microphone index
r	distance between source and centre microphone array
$r_n(\theta, r)$	distance between source and nth microphone
s_f, s_g	sub-sampling factors
$s[k]$	clean speech signal at time k
$v_n[k]$	noise component of nth microphone signal at time k
$w_n[k]$	filter on the nth microphone signal
$x_n[k]$	speech component of nth microphone signal at time k
$y_n[k]$	nth microphone signal at time k
$z[k]$	output signal
$z_x[k]$	speech component in the output signal $z[k]$
$z_v[k]$	noise component in the output signal $z[k]$
$\mathbf{b}$	constraint vector
$\mathbf{d}(\omega, \theta)$	steering vector
$\mathbf{e}[k]$	error vector
$\mathbf{e}_i$	vector with ith element equal to 1 and all other elements equal to 0
$\mathbf{e}_v[k]$	residual noise
$\mathbf{e}_y[k]$	signal distortion
$\mathbf{e}(\omega)$	filter delay vector
$\mathbf{g}(\omega, \theta), \mathbf{g}(\omega, \theta, r)$	steering vector broadband beamforming
$\mathbf{s}[k]$	data vector of $s[k]$
$\mathbf{v}[k]$	stacked noise data vector
$\mathbf{v}_n[k]$	L-dimensional data vector of $v_n[k]$

$\mathbf{w}[k]$	stacked filter vector
$\mathbf{w}_n[k]$	L-dimensional FIR filter on nth microphone signal
$\mathbf{w}_{min}$	local/global minimum of cost function
$\mathbf{w}_s$	stationary point
$\mathbf{x}[k]$	stacked speech data vector
$\mathbf{x}_n[k]$	L-dimensional data vector of $x_n[k]$
$\mathbf{y}[k]$	stacked data vector
$\mathbf{y}_n[k]$	L-dimensional data vector of $y_n[k]$

$\mathbf{0}$	zero vector, zero matrix
$\mathbf{1}_M$	$M \times M$-dimensional matrix, all elements equal to 1
$\mathbf{A}(\omega, \theta)$	diagonal matrix containing microphone characteristics
$\mathbf{C}, \hat{\mathbf{C}}$	constraint matrices
$\mathbf{C}_a, \hat{\mathbf{C}}_a$	null space of $\mathbf{C}, \hat{\mathbf{C}}$
$\mathbf{D}$	$L \times L$-dimensional diagonal matrix $0 \ldots L-1$
$\mathbf{F}(\mathbf{w})$	minimax matrix
$\mathbf{G}(\omega, \theta), \mathbf{G}(\omega, \theta, r)$	steering matrix
$\mathbf{H}_{NL}$	Hessian matrix for non-linear cost function
$\mathbf{I}_M$	$M \times M$-dimensional identity matrix
$\mathbf{J}_M$	$M \times M$-dimensional reverse identity matrix
$\mathbf{Q}[k]$	matrix containing generalised singular vectors of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$
$\bar{\mathbf{Q}}[k]$	matrix containing generalised eigenvectors of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$
$\mathbf{Q}_Y[k], \mathbf{R}_Y[k]$	QR-decomposition of $\mathbf{Y}[k]$
$\mathbf{S}_{NL}$	$NL \times NL$-dimensional block-reversal matrix
$\mathbf{U}_Y[k], \mathbf{U}_V[k]$	orthogonal matrices containing generalised singular vectors of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$
$\mathbf{V}[k]$	noise data matrix at time k
$\bar{\mathbf{V}}_y[k]$	orthogonal matrix containing eigenvectors of $\bar{\mathbf{R}}_{yy}[k]$
$\mathbf{W}$	filter matrix
$\mathbf{W}_{WF}[k]$	empirical Wiener filter matrix at time k
$\bar{\mathbf{W}}_{WF}[k]$	Wiener filter matrix at time k
$\mathbf{Y}[k]$	speech data matrix at time k
$\bar{\mathbf{\Delta}}_y[k]$	diagonal matrix containing eigenvalues of $\bar{\mathbf{R}}_{yy}[k]$
$\mathbf{\Delta}_\theta, \mathbf{\Delta}_\omega(\theta)$	diagonal constraint matrices for broadband beamforming
$\mathbf{\Sigma}_Y[k], \mathbf{\Sigma}_Y[k]$	diagonal matrices containing generalised singular values of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$
$\bar{\mathbf{\Lambda}}_y[k]$, $\bar{\mathbf{\Lambda}}_v[k]$	diagonal matrices containing generalised eigenvalues of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$

α_r	weighting factor
$\delta[k]$	Dirac impulse at $k = 0$
δ	speech detection error rate

δ_n	delay of DS beamformer for nth microphone
$\epsilon_y^2[k]$	signal distortion energy
$\epsilon_v^2[k]$	residual noise energy
$\zeta[k]$	VAD output at time k
$\zeta_c[k]$	zero-crossing rate
θ	angle
θ_x	direction of speech source
λ	Lagrange multiplier
λ_y	exponential weighting factor for speech
λ_v	exponential weighting factor for noise
μ	adaptive filter step size
$\sigma_i[k], \eta_i[k]$	generalised singular values of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$
$\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k]$	generalised eigenvalues of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$
τ_n	delay of nth microphone
ϕ_n	angle of nth microphone in planar array
$\psi_n, \psi_n(\omega, \theta)$	phase of microphone characteristic of nth microphone
ω	pulsation
$\Theta, \Theta_p, \Theta_s$	angle region
$\Phi(\omega)$	noise sensitivity
Ψ	phase pdf of microphone characteristics
$\Omega, \Omega_p, \Omega_s$	frequency region

Acronyms and Abbreviations

AEC	Acoustic Echo Cancellation
AG	array gain
ANC	Adaptive Noise Cancellation
APA	Affine Projection Algorithm
ASIC	Application Specific Integrated Circuit
ASR	Automatic Speech Recognition
AR	autoregressive
BSS	blind source separation
BTE	behind-the-ear
cf.	`confer` : see also
CSP	Cross-power Spectrum Phase
DFT	Discrete Fourier Transform
DI	Dereverberation Index
DR	direct-to-reverberant energy ratio
DS	delay-and-sum
DSP	Digital Signal Processor
DTFT	Discrete-Time Fourier Transform
e.g.	`exempli gratia` : for example
EVD	Eigenvalue Decomposition

FIR	finite impulse response
FFT	Fast Fourier Transform
GCC	Generalised Cross-Correlation
GEVD	Generalised Eigenvalue Decomposition
GSC	Generalised Sidelobe Canceller
GSVD	Generalised Singular Value Decomposition
i.e.	`id est` : that is
IDFT	Inverse Discrete Fourier Transform
IDTFT	Inverse Discrete-Time Fourier Transform
iff	if and only if
IFFT	Inverse Fast Fourier Transform
IIR	infinite impulse response
KLT	Karhunen-Loève Transform
LCMV	linearly constrained minimum variance
LMS	Least Mean Squares
LS	Least Squares
MAPD	modified amplitude pdf
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MV	minimum variance
MVDR	minimum variance distortionless response
NLMS	Normalised Least Mean Squares
PAST	Projection Approximation Subspace Tracking
pdf	probability density function
PSD	Power Spectral Density
PTF	Power Transfer Function
QSVD	Quotient Singular Value Decomposition
RLS	Recursive Least Squares
SD	Speech Distortion
SII	Speech Intelligibility Index
SNR	Signal-to-Noise Ratio
STFT	short-time Fourier Transform
SVD	Singular Value Decomposition
SQP	Sequential Quadratic Programming
TDE	Time-Delay Estimation
TLS	Total Least Squares
vs.	versus
VAD	Voice Activity Detection
WNG	white noise gain
WSS	wide-sense stationary
w.r.t.	with respect to

Contents

I GSVD-Based Optimal Filtering for Multi-Microphone Noise Reduction

II Multi-Microphone Dereverberation and Source Localisation

Technieken voor ruisonderdrukking en dereverberatie in spraaktoepassingen met behulp van meerdere microfoons

Hoofdstuk 1: Inleiding

De *motivatie* voor het werk in deze doctoraatsthesis is de snel groeiende markt van spraak- en audiotoepassingen. Handenvrije mobiele telefonie, spraakgestuurde systemen en video-conferencing zijn belangrijke toepassingen in de telecommunicatiesector, terwijl hoorapparaten en cochleaire implantaten belangrijke toepassingen vormen in de biomedische sector. Het gemeenschappelijk probleem voor al deze toepassingen is *de opname van spraaksignalen in een ongunstige akoestische omgeving.* In een typisch handenvrij systeem worden immers microfoons gebruikt op een zekere afstand van de spreker, zodat de opgenomen signalen van lage kwaliteit zijn ten gevolge van achtergrondlawaai, reverberatie (nagalm) en 'far-end'-echosignalen. Deze slechte signaalkwaliteit kan ertoe leiden dat het gewenste spraaksignaal onverstaanbaar wordt en dat de performantie van systemen voor spraakherkenning of spraakcodering aanzienlijk vermindert. Het oplossen van dit probleem vereist performante technieken voor signaalverbetering (ruisonderdrukking, dereverberatie en echo-onderdrukking).

Figuur 1.1 geeft een typische handenvrije-communicatie-omgeving weer, waar een spreker zich vrij kan bewegen zonder een microfoon vast te houden. Het microfoonrooster heeft als doel het (zuivere) signaal van de spreker zo goed mogelijk op te nemen. Door de afstand tussen de spreker en de microfoons zal echter ook *achtergrondlawaai* (bv. radio, andere sprekers, 'far-end' echo) opgenomen worden, en zal niet enkel het direct pad van de spreker opgevangen worden, maar ook de weerkaatsingen van het spraaksignaal tegen muren, vloer en andere objecten (d.i. *reverberatie* of nagalm).

In deze doctoraatsthesis worden verschillende technieken ontwikkeld voor ruisonderdrukking en dereverberatie met behulp van meerdere microfoons. Deze technieken moeten in principe aan meerdere doelstellingen voldoen. We behandelen voornamelijk *meer-kanaals* signaalverbeteringstechnieken, aangezien meer-kanaals technieken zowel de spectrale als de spatiale karakteristieken in de microfoonsignalen kunnen uitbuiten, in tegenstelling tot één-kanaals technieken die enkel de spectrale karakteristieken benutten. Aangezien de signalen en de akoestische omgeving meestal tijdsvariant zijn, dienen de ontwikkelde algoritmes *adaptief* te zijn, zodat ze verschillende ruissituaties en veranderende akoestische omgevingen aankunnen. In het algemeen veronderstellen we dat de ruisbronnen *niet gekend* zijn, dit wil zeggen dat er geen referentiesignaal voor de ruisbronnen beschikbaar is. We zullen ook de *integratie* van verschillende signaalverbeteringstechnieken bespreken, zoals gecombineerde ruis- en echo-onderdrukking en gecombineerde ruisonderdrukking en dereverberatie. Aangezien de meeste meer-kanaals signaalverbeteringstechnieken gevoelig zijn aan afwijkingen in de karakteristieken van het microfoonrooster (versterking, fase, microfoonpositie) en andere afwijkingen (bv. foute schatting van de positie van de spreker, spraakdetectiefouten), zullen we de *robuustheid* van de ontwikkelde algoritmes onderzoeken met betrekking tot deze afwijkingen en, waar mogelijk, zullen we robuustheid tegen deze afwijkingen mee in rekening brengen in het algoritmisch ontwerp. Uiteindelijk zullen we ook rekening houden met de *berekeningscomplexiteit* van de ontwikkelde algoritmes. Nochtans is het voornamelijk de bedoeling in deze thesis om algoritmes te ontwikkelen die een betere performantie en/of robuustheid hebben dan bestaande technieken, waarbij complexiteit slechts op de tweede plaats komt.

Deel I behandelt een GSVD-gebaseerde optimaal-filtertechniek, die gebruikt kan worden voor één-kanaals en meer-kanaals ruisonderdrukking, maar die geen dereverberatie uitvoert. In Deel II wordt een gecombineerde techniek voor ruisonderdrukking en dereverberatie besproken en een techniek voor akoestische bronlokalisatie die robuust is tegen achtergrondlawaai en reverberatie. Deel III behandelt ontwerpprocedures voor robuuste breedband bundelvormers, die zowel voor ruisonderdrukking als voor dereverberatie gebruikt kunnen worden.

In **paragraaf 1.2** worden de algemene voor- en nadelen van handenvrije systemen besproken en wordt dieper ingegaan op de specifieke problemen, economisch belang en bestaande producten voor enkele belangrijke toepassingen:

- *Handenvrije mobiele telefonie*: vanuit een economisch standpunt is mobiele telefonie zeker de voornaamste toepassing, met wereldwijd meer dan één miljard gebruikers. In veel landen is het tegenwoordig verboden om mobiel te telefoneren in de wagen, tenzij een handenvrije kit gebruikt wordt. De voornaamste problemen bij handenvrije mobiele telefonie in de wagen zijn 'far-end'-echosignalen en meerdere ruisbronnen (motor, banden, radio, andere passagiers). De meeste huidige handenvrije kits gebruiken één enkele directionele microfoon, die nog steeds vrij veel achtergrondlawaai opvangt. Daarom wordt verwacht dat in de nabije toekomst meer geavanceerde meer-kanaals systemen aangewend zullen worden. Het feit dat deze systemen vrij goedkoop moeten blijven beperkt echter het aantal microfoons en de benodigde hardware voor signaalverwerking.

- *Video-conferencing*: in plaats dat elke deelnemer aan een video-conferentie zijn eigen microfoon heeft, is het mogelijk om een microfoonrooster te gebruiken dat het geluid van de actieve spreker zo goed mogelijk opvangt. De voornaamste problemen bij video-conferencingsystemen zijn 'far-end'-echosignalen en akoestische bronlokalisatie in omgevingen met veel achtergrondlawaai en reverberatie. Bronlokalisatie kan gebruikt worden om een camera te richten of om het microfoonrooster elektronisch te sturen in de richting van de spreker met behulp van een bundelvormer.

- *Spraakgestuurde systemen*: tegenwoordig kunnen steeds meer apparaten met behulp van spraakcommando's bediend worden (bv. HiFi systemen, PC software, domotica, telematica in de wagen). Opdat spraakgestuurde bediening een toegevoegde waarde zou bieden, moet de spraakherkenning echter betrouwbaar werken in alle omstandigheden. Aangezien de performantie van spraakherkenningssystemen drastisch vermindert in akoestische omgevingen met veel achtergrondlawaai en reverberatie, kunnen signaalverbeteringstechnieken er voor zorgen dat de performantie en betrouwbaarheid terug verbetert in deze omgevingen.

- *Hoorapparaten en cochleaire implantaten*: slechthorendheid is een probleem waaraan wereldwijd meer dan 300 miljoen mensen lijden. De meeste slechthorenden hebben een perceptueel gehoorverlies, waarbij niet alleen alle geluiden verzwakt worden, maar vooral verschillende geluiden niet meer van elkaar onderscheiden kunnen worden. Dit probleem kan dus niet opgelost worden door alle geluiden te versterken, maar enkel door het ongewenst lawaai te verzwakken ten opzichte van het gewenst geluid. Door de recente evolutie in de productie van microfoons en microelektronica is het mogelijk om meerdere microfoons en een DSP in te bouwen in een hoorapparaat. Bestaande meer-kanaals hoorapparaten gebruiken vrij eenvoudige algoritmes voor spraakverbetering, voornamelijk wegens de beperkte rekenkracht van de DSP. In de toekomst zal het echter mogelijk worden om meer geavanceerde algoritmes te implementeren, die zorgen voor een betere performantie en robuustheid. Robuustheid is belangrijk in hoorapparaten wegens de kleine afstand tussen de microfoons (typisch 1 à 2 cm). Voor cochleaire implantaten kunnen natuurlijk ook gelijkaardige signaalverbeteringstechnieken toegepast worden.

In **paragraaf 1.3** worden de belangrijkste karakteristieken van spraak- en ruissignalen en van de akoestische omgeving besproken. *Spraak* is een breedbandig signaal met frequentiecomponenten tussen 100 en 8000 Hz, waarbij voor spraakverstaanbaarheid voornamelijk de frequenties tussen 300 en 3400 Hz belangrijk zijn. Wegens de vraag naar hoge spraakkwaliteit zullen we in deze thesis meestal werken met een bemonsteringsfrequentie van 16 kHz. Aangezien in een typische conversatie gemiddeld slechts 50% spraak aanwezig is, kan van deze aan/af-karakteristiek gebruik gemaakt worden door middel van een spraakdetectie-algoritme (VAD) dat het signaal classificeert in spraak- en ruisperiodes. Spraaksignalen kunnen ook beschreven worden door middel van een lineair lage-rangmodel, waarbij verondersteld wordt dat elke vector van het spraaksignaal voorgesteld kan worden als een lineaire combinatie van een eindig aantal basisvectoren (bv. complexe exponentiëlen). In het algemeen is er minder gekend over de *achtergrondruis*. Achtergrondruis kan komen van een gelokaliseerde ruisbron (bv. radio) of kan diffuse ruis zijn die uit alle richtingen komt (bv. 'cocktail party'). Sommige ruisbronnen hebben een traag-variërend karakter, terwijl andere ruisbronnen zeer niet-stationair zijn of zelfs andere spraaksignalen zijn. De *akoestische omgeving* kan globaal gekarakteriseerd worden door de reverberatietijd T_{60}, die aangeeft hoeveel tijd geluid nodig heeft om te zakken tot -60 dB van het origineel niveau. De akoestische filtering tussen twee punten in een kamer kan goed beschreven worden door middel van een lineair FIR filter, dat *akoestische impulsresponsie* genoemd wordt. Akoestische impulsresponsies kunnen gesimuleerd worden met behulp van de 'image'-methode. Aangezien akoestische impulsresponsies meestal niet-minimum-fasesystemen zijn, kunnen deze impulsresponsies niet eenvoudig geïnverteerd worden. Van de *microfoons* wordt meestal verondersteld dat ze puntsensoren zijn met een ideale omnidirectionele karakteristiek. In een echte opstelling kunnen echter verschillende soorten afwijkingen voorkomen: afwijkingen in de veronderstelde microfoonkarakteristieken (versterking, fase, directiviteit), de plaatsing van de microfoons, en een mogelijk schaduweffect van het hoofd. Het is belangrijk dat signaalverbeteringstechnieken rekening houden met deze afwijkingen. Afhankelijk van de afstand tussen de spreker en de microfoons, bevindt de spreker zich in het zogenaamde 'far-field' of 'near-field' van het microfoonrooster. Formule (1.4) geeft de grens aan waar de 'far-field'-veronderstellingen nog gelden.

Paragraaf 1.4 geeft een kort overzicht van verschillende technieken voor signaalverbetering (ruisonderdrukking, echo-onderdrukking, dereverberatie). Eénkanaals technieken voor *ruisonderdrukking* kunnen ingedeeld worden in enerzijds parametrische technieken zoals Wiener- of Kalman-filtering en anderzijds niet-parametrische technieken zoals spectrale subtractie en deelruimtegebaseerde technieken. Meer-kanaals technieken kunnen ingedeeld worden in enerzijds vaste en adaptieve bundelvorming en anderzijds meer-kanaals Wienerfiltering, een techniek die in Deel I in meer detail zal besproken worden. *Dereverberatie* komt neer op het schatten van het zuivere spraaksignaal uit de microfoonsignalen, zonder enige kennis over de akoestische impulsresponsies.

Standaard één-kanaals technieken zijn cepstrum-technieken of inverse filtering, maar deze technieken hebben een zeer beperkte performantie. Meer-kanaals technieken daarentegen kunnen een spatiale verwerking uitvoeren, zodat het reverberante gedeelte spatiaal gescheiden kan worden van het direct pad. Standaard meer-kanaals technieken zoals inverse filtering of 'matched' filtering vereisen een schatting van de akoestische impulsresponsies, terwijl vaste bundelvormers deze kennis niet vereisen.

In **paragraaf 1.5** wordt een overzicht gegeven van de verschillende hoofdstukken en worden onze bijdragen toegelicht. Figuur 1.4 geeft een schematisch overzicht van de thesis en van de verbanden tussen de verschillende hoofdstukken.

Hoofdstuk 2: Technieken voor signaalverbetering

Dit hoofdstuk beschrijft enkele één-kanaals en meer-kanaals technieken voor ruisonderdrukking en dereverberatie die belangrijk zijn voor het vervolg van de thesis.

Paragraaf 2.1 behandelt enkele basisdefinities van signaalverwerking, zoals Discrete Fourier-Transformatie (DFT), autocorrelatie, kruiscorrelatie, 'Power Spectral Density' (PSD), coherentie en 'Power Transfer Function' (PTF).

In **paragraaf 2.2** wordt het algemeen model beschreven voor de opname van spraaksignalen in een akoestische omgeving met achtergrondlawaai. Elk microfoonsignaal $y_n[k]$ bestaat uit een gefilterde versie van het zuivere spraaksignaal $s[k]$ en additieve ruis. Figuur 2.1 toont een algemene opstelling voor meer-kanaals signaalverbetering, waar de microfoonsignalen (adaptief) gefilterd worden met de filters $w_n[k]$ en gecombineerd worden tot het uitgangssignaal. Alle signaalverbeteringstechnieken in deze thesis verschillen in feite louter in de manier waarop de filters $w_n[k]$ berekend worden. Deze filters kunnen ontworpen worden voor verschillende doelstellingen:

- Het doel van *ruisonderdrukking* is de energie van de residuele ruiscomponent in het uitgangssignaal te minimaliseren, terwijl ook spraakvervorming mee in rekening gebracht wordt.
- Het doel van *dereverberatie* is de filters $w_n[k]$ te berekenen zodat de totale transferfunctie voor het spraaksignaal gelijk is aan een vertraging.
- Het doel van *gecombineerde ruisonderdrukking en dereverberatie* is het schatten van het zuivere spraaksignaal $s[k]$, dit wil zeggen dat gelijktijdig de transferfunctie voor het spraaksignaal een vertraging benadert en de energie van de residuele ruiscomponent geminimaliseerd wordt.

Alle uitdrukkingen kunnen ook voorgesteld worden in het frequentiedomein. In paragraaf 2.2.4 worden verschillende performantiecriteria gedefinieerd. *Ruisonderdrukking* wordt beschreven door de verbetering in signaal-ruisverhouding

(SNR). *Spraakvervorming* kan beschreven worden door de PTF tussen de spraakcomponent in het ingangs- en het uitgangssignaal. *Dereverberatie* kan beschreven worden door de PTF tussen het zuivere spraaksignaal en de spraakcomponent in het uitgangssignaal.

In **paragraaf 2.3** worden twee één-kanaals technieken voor ruisonderdrukking besproken: spectrale subtractie en deelruimte-gebaseerde technieken. Beide technieken benutten enkel de temporele en de spectrale informatie van de spraak- en de ruissignalen. In de meeste spectrale-subtractietechnieken worden de DFT-coëfficiënten vermenigvuldigd met een ruisafhankelijke versterkingsfactor, terwijl in de deelruimte-gebaseerde technieken de KLT-coëfficiënten (Karhunen-Loève-Transformatie) gewijzigd worden. Aangezien beide technieken een schatting nodig hebben van de ruiskarakteristieken, is er een spraakdetectie-algoritme vereist. Deelruimte-gebaseerde technieken veronderstellen dat het zuivere spraaksignaal beschreven kan worden door middel van een lagerangmodel en voeren signaalverbetering uit door de ruisdeelruimte te verwijderen en het zuivere spraaksignaal te schatten in de overblijvende signaaldeelruimte, gebruik makend van een kleinste-kwadraten (LS) of een minimumvariantie (MV) schatter. Beide schatters kunnen voorgesteld worden door middel van een eigenfilterbank, zowel wanneer witte ruis als wanneer gekleurde ruis aanwezig is. Het kan bewezen worden dat de signaalonafhankelijke spectrale-subtractietechnieken en de signaalafhankelijke deelruimte-gebaseerde technieken asymptotisch hetzelfde resultaat produceren wanneer de framelengte oneindig lang wordt en wanneer verondersteld wordt dat de spraak- en de ruissignalen stationair zijn. In Deel I van de thesis zullen we de beschreven deelruimte-gebaseerde technieken uitbreiden naar meerdere microfoons.

In **paragraaf 2.4** worden twee één-kanaals technieken voor dereverberatie besproken: inverse filtering, waarbij de akoestische impulsresponsies gekend verondersteld zijn, en cepstrum-gebaseerde technieken, die geen kennis over de akoestische impulsresponsies vereisen. In de praktijk kunnen één-kanaals inverse-filteringtechnieken maar met beperkt succes toegepast worden, aangezien akoestische impulsresponsies meestal niet-minimum-fasesystemen zijn, terwijl één-kanaals cepstrum-gebaseerde technieken ook meestal een beperkte performantie hebben, omdat het cepstrum van het zuivere spraaksignaal en de akoestische impulsresponsie in grote mate met elkaar overlappen.

Paragraaf 2.5 behandelt vaste en adaptieve bundelvormingstechnieken voor meer-kanaals ruisonderdrukking. *Vaste bundelvormers* zijn data-onafhankelijk en proberen ruimtelijk in te zoomen op de spraakbron. Hierdoor kan reverberatie en achtergrondruis die niet uit de richting van de spraakbron komt onderdrukt worden. Verschillende soorten vaste bundelvormers worden besproken: de eenvoudige – maar vaak gebruikte – 'delay-and-sum' (DS) bundelvormer; eerste-orde differentiële microfoons, gebruik makend van 2 microfoons op een korte afstand van elkaar die vertraagd worden ten opzichte van elkaar; superdirectieve bundelvormers die de directiviteitsindex maximaliseren voor een gekend ruisveld; en de meest algemene 'filter-and-sum' bundelvormers, die in meer detail in Deel III van de thesis bestudeerd worden. *Adaptieve bundelvormers* combineren het ruimtelijk inzoomen van vaste bundelvormers met adaptieve ruisonderdrukking, zodat adaptieve bundelvormers zich kunnen aanpassen aan veranderende akoestische omgevingen en in het algemeen in een betere ruisonderdrukking resulteren dan vaste bundelvormers. In deze paragraaf wordt de 'linearly-constrained minimum-variance' (LCMV) bundelvormer besproken, die de energie van het uitgangssignaal minimaliseert met de beperking dat signalen uit de richting van de spraakbron niet vervormd worden. Dit LCMV-optimalisatieprobleem met beperkingen kan geherformuleerd worden als een optimalisatieprobleem zonder beperkingen, resulterend in de 'Generalised Sidelobe Canceller' (GSC) structuur. Deze GSC-structuur is opgebouwd uit een vaste bundelvormer die een spraakreferentie genereert, een 'blocking'-matrix die ruisreferenties genereert en een 'adaptive noise cancellation' (ANC) trap die gebruik maakt van een meer-kanaals adaptief filter. In de praktijk zal echter door signaalreflecties (reverberatie) en door afwijkingen in de veronderstelde microfoonkarakteristieken signaallek optreden in de ruisreferenties, wat leidt tot signaalvervorming. Verschillende varianten van de standaard GSC-structuur worden besproken die de hoeveelheid signaallek verminderen (bv. door middel van een spatiale 'blocking'-matrix) of het effect van de signaallek op de adaptieve filters beperken (bv. spraakgestuurd adaptatie-algoritme). In Deel I van de thesis zal de performantie van de GSVD-gebaseerde optimaalfiltertechniek vergeleken worden met de performantie van deze vaste en adaptieve bundelvormers.

In **paragraaf 2.6** worden twee meer-kanaals technieken voor dereverberatie besproken: inverse filtering en 'matched' filtering. Beide technieken vereisen dat de akoestische impulsresponsies (gedeeltelijk) gekend zijn. Met behulp van de inverse-filteringtechniek is het mogelijk om perfecte dereverberatie uit te voeren. Deze techniek is echter vrij gevoelig aan de nauwkeurigheid van de opgemeten/geschatte impulsresponsies. In de 'matched'-filteringtechniek worden de microfoonsignalen gefilterd met de tijdsomgekeerde van de (gedeeltelijke) akoestische impulsresponsies. Deze techniek is minder gevoelig aan de nauwkeurigheid van de impulsresponsies, maar perfecte dereverberatie is niet mogelijk. Bovendien treedt er een 'pre-echo'-probleem op, dat verminderd kan worden door de 'matched' filters tot een zekere filterlengte af te kappen. Deze 'matched'-filteringtechniek vormt de basis voor de frequentiedomeintechnieken voor dereverberatie en gecombineerde ruisonderdrukking en dereverberatie, die ontwikkeld worden in Deel II van de thesis.

Deel I : GSVD-gebaseerde optimale filtering voor meer-kanaals ruisonderdrukking

In dit deel stellen we een optimaal-filtertechniek, gebaseerd op de Veralgemeende-Singuliere-Waarde-Ontbinding (GSVD), voor om de signaalkwaliteit van meer-kanaals spraaksignalen te verbeteren wanneer additieve gekleurde ruis aanwezig is. Verschillende technieken worden besproken om de berekeningscomplexiteit te verminderen en we tonen aan dat deze GSVD-gebaseerde optimaalfiltertechniek geïntegreerd kan worden in een GSC-structuur. Simulaties tonen aan dat de GSVD-gebaseerde optimaal-filtertechniek een grotere verbetering in signaal-ruisverhouding oplevert dan standaard vaste en adaptieve bundelvorming en dat deze techniek robuuster is wanneer afwijkingen in het veronderstelde signaalmodel optreden.

Hoofdstuk 3: GSVD-gebaseerde optimale filtering voor één-kanaals en meer-kanaals spraakverbetering

In paragraaf 2.3 zijn één-kanaals deelruimte-gebaseerde technieken voor signaalverbetering besproken. Het kernidee is om het microfoonsignaal voor te stellen in een vectorruimte en deze vectorruimte op te splitsen in 2 orthogonale deelruimtes: de signaaldeelruimte en de ruisdeelruimte. Signaalverbetering kan dan toegepast worden door de ruisdeelruimte te verwijderen en het zuivere spraaksignaal te schatten uit de overblijvende signaaldeelruimte. Eén-kanaals deelruimte-gebaseerde technieken kunnen beschouwd worden als een (signaalafhankelijke) frequentie-filtering, die adaptief de meest energetische formanten uit het spraaksignaal overhoudt en zo achtergrondruis onderdrukt. In dit hoofdstuk stellen we een meer-kanaals uitbreiding voor, die zo de spatio-temporele informatie van de spraak- en de ruisbronnen combineert. Wanneer een MV-schatter gebruikt wordt, leidt dit tot een GSVD-gebaseerde implementatie van het meer-kanaals Wiener-filter, waarbij het lage-rangmodel van het spraaksignaal mee in rekening wordt gebracht.

Paragraaf 3.2 behandelt optimale filtering voor meer-kanaals spraakverbetering. Het optimaal filter in de 'mean square error' (MSE) zin is het meer-kanaals Wiener-filter, dat een 'minimum mean square error' (MMSE) schatting produceert voor de spraakcomponenten in de microfoonsignalen maar dus geen dereverberatie uitvoert. In tegenstelling tot de GSC, dat als een optimaalfilterprobleem met beperkingen beschouwd kan worden, is meer-kanaals Wienerfiltering een optimaal-filterprobleem zonder beperkingen. Door gebruik te maken van de Veralgemeende-Eigenwaarde-Ontbinding (GEVD) van de spraaken de ruiscorrelatiematrices, kan het lage-rangmodel van het spraaksignaal gemakkelijk in rekening gebracht worden. Voor het meer-kanaals Wiener-filter kan spraakvervorming nooit vermeden worden, aangezien de schattingsfout de som is van een term die de residuele ruis voorstelt en een term die spraakvervorming voorstelt. In deze paragraaf stellen we ook een algemene klasse schatters voor, waarbij het mogelijk is om spraakvervorming en ruisonderdrukking tegenover elkaar af te wegen en waarvan de filterparameters ook verkregen kunnen worden uit de GEVD van de correlatiematrices.

In **paragraaf 3.3** tonen we aan dat in de praktijk de Veralgemeende-Singuliere-Waarde-Ontbinding (GSVD) van een spraak- en een ruisdatamatrix gebruikt kan worden om een empirische schatting van de optimaal-filtermatrix te bekomen. Deze datamatrices worden geconstrueerd met behulp van een spraakdetectie-algoritme (VAD), dat bepaalt of een vector tot de spraak- of tot de ruisdatamatrix behoort. Dit spraakdetectie-algoritme is de enige a-priori informatie waarop de GSVD-gebaseerde optimaal-filtertechniek steunt. We tonen aan dat verschillende schattingen voor dezelfde spraakcomponent bekomen worden, en we beschrijven een procedure om te bepalen welke schatting uiteindelijk gebruikt moet worden (in de praktijk wordt meestal de vertraagde spraakcomponent $x_0[k-\frac{L}{2}+1]$ in het eerste microfoonsignaal gekozen). In de 'batch'-versie van het algoritme worden de datamatrices geconstrueerd met behulp van alle beschikbare spraak- en ruisdatavectoren in het beschouwde signaalframe. Deze 'batch'-versie is echter niet geschikt voor een implementatie in reële tijd wegens de grote vertraging veroorzaakt door de frame-gebaseerde verwerking. In de recursieve versie van het algoritme worden de datamatrices voor elke tijdsstap bijgewerkt met de nieuw beschikbare spraak- of ruisdatavector (afhankelijk van de uitgang van het VAD-algoritme), gebruik makend van een venster met exponentiële weging. Aangezien in de recursieve versie voor elke tijdsstap de GSVD en het optimaal filter herberekend moeten worden, is de berekeningscomplexiteit vrij hoog. Daarom worden in hoofdstuk 4 verscheidene technieken beschreven om de berekeningscomplexiteit te verminderen. In deze paragraaf worden ook kort enkele andere implementatietechnieken voor het meer-kanaals Wiener-filter vermeld, zoals een implementatie gebaseerd op de QR-ontbinding, een LMS-gebaseerde implementatie en een subband-gebaseerde implementatie. De subband-gebaseerde implementatie leidt tot een lagere complexiteit en een betere performantie dan de fullband-implementatie, aangezien de MSE in elke subband geminimaliseerd kan worden, wat perceptueel relevanter is.

In **paragraaf 3.4** leiden we een aantal symmetrie-eigenschappen af voor de optimaal-filtermatrix, zowel in het één-kanaals als in het meer-kanaals geval. Deze eigenschappen zijn zowel geldig voor witte ruis als voor gekleurde ruis en voor elke wegingsfunctie van de veralgemeende eigenwaarden. Ook wordt de uitmiddelingsoperatie die toegepast wordt in sommige één-kanaals deelruimtegebaseerde technieken onderzocht. Dit leidt tot het besluit dat deze uitmiddelingsoperatie onnodig en vaak zelfs suboptimaal is.

In **paragraaf 3.5** analyseren we het meer-kanaals Wiener-filter in het frequentiedomein. We tonen aan dat – onder zwakke voorwaarden – het meer-kanaals Wiener-filter gesplitst kan worden in een spatiale filterterm, die afhangt van de spatiale karakteristieken (coherentie) van de spraak- en de ruisbronnen, en een één-kanaals spectraal Wiener-filter, dat afhangt van de spectrale karakteristieken (PSD) van de spraak- en de ruisbronnen. We berekenen de transferfuncties voor de spraak- en de ruiscomponenten en we vereenvoudigen alle uitdrukkingen in het geval van één enkele spraakbron. We tonen aan dat meer spraakvervorming optreedt voor frequenties met een lage SNR en wanneer de spatiale scheiding tussen de spraak- en de ruisbronnen slecht is en dat meer ruisonderdrukking bekomen wordt voor frequenties met een lage SNR en wanneer de spraak- en de ruisbronnen spatiaal goed gescheiden zijn. Bovendien tonen we aan dat de ruisgevoeligheid van de GSC en het meer-kanaals Wiener-filter aan elkaar gelijk zijn in het geval van één enkele spraakbron en wanneer de vaste bundelvormer in de GSC een 'matched' filter is.

In **paragraaf 3.6** tonen we aan dat de meer-kanaals optimaal-filtertechniek ook gebruikt kan worden voor gecombineerde ruis- en echo-onderdrukking door het 'far-end'-echosignaal als extra ingangssignaal te beschouwen. Voor oneindig lange filters bewijzen we dat de 'far-end'-echobron geen invloed heeft op de filters voor de microfoonsignalen, zodat dezelfde performantie bekomen wordt als in het geval waar geen echobron aanwezig is, en dat de 'far-end'-echocomponenten in de microfoonsignalen volledig onderdrukt kunnen worden.

Hoofdstuk 4: Vermindering van berekeningscomplexiteit met behulp van recursieve GSVD en 'ANC-postprocessing'-trap

In dit hoofdstuk worden verschillende technieken besproken om de berekeningscomplexiteit van de GSVD-gebaseerde optimaal-filtertechniek te verminderen.

Zoals reeds gezegd, is de berekeningscomplexiteit van de recursieve versie vrij hoog, aangezien voor elke tijdsstap de GSVD en het optimaal filter herberekend moeten worden. **Paragraaf 4.2** beschrijft technieken om de complexiteit te verminderen door gebruik te maken van recursieve algoritmes van het Jacobi-type om de GSVD te herberekenen en door gebruik te maken van subbemonstering. In plaats van de volledige GSVD opnieuw te berekenen voor elke tijdsstap, berekenen recursieve algoritmes de GSVD op tijdstip k door gebruik te maken van de ontbinding op tijdstip $k-1$. De complexiteit kan verder verlaagd worden door een implementatie te gebruiken die geen wortels nodig heeft om de rotatiehoeken in de Givens-transformaties te berekenen. Voor stationaire akoestische omgevingen kan de complexiteit zonder enig verlies in performantie verder verlaagd worden door sub-bemonsteringstechnieken, waar sub-bemonstering in deze context betekent dat de GSVD en het optimaal filter niet voor elke tijdsstap herberekend worden. Voor niet-stationaire akoestische omgevingen moet sub-bemonstering echter beperkt worden. Voor realistische waarden van de parameters (4 microfoons, 20 filtertaps, 16 kHz) vat Tabel 4.4 de berekeningscomplexiteit samen voor de verschillende implementaties. Deze tabel toont aan dat de berekeningscomplexiteit van de recursieve GSVD-gebaseerde optimaal-filtertechniek significant kan verminderd worden, zodat deze techniek geschikt wordt voor een implementatie in reële tijd.

In **paragraaf 4.3** tonen we aan dat de GSVD-gebaseerde optimaal-filtertechniek geïntegreerd kan worden in een GSC-structuur met een 'ANC-postprocessing'-trap. De uitgang van de GSVD-gebaseerde optimaal-filtertechniek wordt gebruikt als spraakreferentiesignaal, terwijl er verschillende mogelijkheden bestaan om een ruisreferentie te creëren. We maken hiervoor gebruik van het optimaal filter om de ruiscomponenten te schatten, en we tonen aan dat dit filter eenvoudig kan afgeleid worden uit het optimaal filter om de spraakcomponenten te schatten. In hoofdstuk 5 zal door middel van simulaties aangetoond worden dat de 'ANC-postprocessing'-trap ofwel gebruikt kan worden om de performantie te verbeteren, ofwel om de berekeningscomplexiteit te verminderen zonder de performantie te verlagen. Aangezien er net zoals bij een GSC meestal ook signaallek optreedt in de ruisreferenties (signaallek kan verminderd worden door grotere filterlengtes voor het optimaal filter te gebruiken), zullen we het effect van deze signaallek op de ANC adaptieve filters verminderen door gebruik te maken van een spraakgestuurd adaptatie-algoritme, dit wil zeggen dat de adaptieve filters enkel mogen adapteren wanneer er geen spraak aanwezig is.

Hoofdstuk 5: Simulatieresultaten en controle-algoritme

Voor verschillende gesimuleerde akoestische omgevingen en voor een realistische opname bespreekt dit hoofdstuk de performantie (ruisonderdrukking, spraakvervorming, robuustheid) van de GSVD-gebaseerde implementatie van de meer-kanaals optimaal-filtertechniek, waarbij het lage-rangmodel van het spraaksignaal mee in rekening wordt gebracht. De performantie van de GSVD-gebaseerde optimaal-filtertechniek wordt vergeleken met standaard vaste en adaptieve bundelvormingstechnieken, en de robuustheid tegen spraakdetectiefouten en afwijkingen in het veronderstelde signaalmodel wordt onderzocht.

De gebruikte simulatie-omgeving is weergegeven in Figuur 5.1 in **paragraaf 5.1** en bevat een microfoonrooster met 4 microfoons op een afstand van 5 cm van elkaar, een spraakbron op 1.3 m van het microfoonrooster en 3 ruisbronnen. Voor het spraaksignaal gebruiken we Engelse zinnen uit de 'Hearing in Noise Test', terwijl we 3 verschillende ruissignalen gebruiken: stationaire witte ruis, stationaire spraakruis met hetzelfde lange-termijnspectrum als spraak en een niet-stationair muzieksignaal. Deze paragraaf bespreekt ook enkele implementatie-aspecten voor de GSVD-gebaseerde optimaal-filtertechniek en voor de bundelvormingstechnieken (filterlengte, stapgrootte, exponentiële weging).

In **paragraaf 5.2** wordt de performantie (SNR-verbetering en spraakvervorming) van de GSVD-gebaseerde optimaal-filtertechniek met en zonder 'ANC-postprocessing'-trap besproken. Voor eenvoudige akoestische scenario's, wanneer er geen signaalreflecties optreden (reverberatietijd $T_{60} = 0$), toont Figuur 5.4 aan dat de GSVD-gebaseerde optimaal-filtertechniek het gewenste bundelvormingsgedrag vertoont voor spatio-temporele witte ruis en voor gelokaliseerde ruisbronnen. Wanneer er wel reverberatie aanwezig is, tonen simulaties (cf. Figuur 5.5) aan dat de SNR-verbetering verhoogt en de spraakvervorming vermindert voor grotere filterlengtes en voor lagere reverberatietijden. Figuur 5.5 toont ook aan dat de 'batch' en de recursieve versie van de GSVD-gebaseerde optimaal-filtertechniek quasi dezelfde performantie hebben. Figuur

5.7 toont aan dat voor stationaire akoestische omgevingen een hogere subbemonsteringsfactor gebruikt kan worden zonder de performantie te verlagen. In paragraaf 5.2.4 wordt de performantie voor een spectraal niet-stationaire ruisbron onderzocht, dit wil zeggen een ruisbron op een vaste positie maar met een veranderend spectrum. Aangezien we meestal vrij lange datablokken beschouwen in de meer-kanaals GSVD-gebaseerde optimaal-filtertechniek – exponentiële weging dicht bij 1 – zal de performantie voornamelijk afhankelijk zijn van de gemiddelde (lange-termijn) spectrale en spatiale karakteristieken van de ruisbron, zodat de GSVD-gebaseerde optimaal-filtertechniek ook gebruikt kan worden om niet-stationaire ruisbronnen te onderdrukken (cf. Figuur 5.8). In paragraaf 5.2.5 wordt het effect van de 'ANC-postprocessing'-trap bestudeerd, en wordt aangetoond dat de 'ANC-postprocessing'-trap ofwel kan gebruikt worden om de performantie te verbeteren ofwel om de complexiteit te verminderen zonder de performantie te verlagen. Deze 'ANC-postprocessing'-trap zal echter wel leiden tot een verhoogde spraakvervorming, die echter beperkt kan worden door langere filters te gebruiken (cf. Figuur 5.9).

In **paragraaf 5.3** wordt het effect van spraakdetectiefouten op de performantie onderzocht. Eerst wordt een overzicht gegeven van verschillende één-kanaals spraakdetectie-algoritmes ('log-likelihood', log-energie, 'zero crossing rate', spectrale entropie, geometrische VAD), waarvan de performantie bestudeerd wordt voor verschillende ruistypes en signaal-ruisverhoudingen. Daarna wordt het gemiddeld effect van (manueel ingevoerde) spraakdetectiefouten geanalyseerd op de performantie van de GSVD-gebaseerde optimaal-filtertechniek, zowel theoretisch als experimenteel. Aangezien het spraakdetectie-algoritme de enige a-priori informatie is waarop de GSVD-gebaseerde optimaal-filtertechniek steunt, wordt verwacht dat deze techniek vrij gevoelig is voor spraakdetectiefouten. Nochtans kan er theoretisch aangetoond worden dat de SNR-verbetering van het meer-kanaals Wiener-filter niet verminderd wordt door spraakdetectiefouten, noch wanneer spraak foutief als ruis gedetecteerd wordt, noch wanneer ruis foutief als spraak gedetecteerd wordt. Wanneer spraak foutief als ruis gedetecteerd wordt, zal de spraakvervorming echter wel sterk toenemen met het percentage foutief geclassificeerde samples (wanneer dit percentage lager is dan 20%, blijft de spraakvervorming echter beperkt). Wanneer ruis foutief als spraak gedetecteerd wordt, zal de spraakvervorming slechts in geringe mate toenemen. Deze vaststellingen worden ook experimenteel bevestigd. Wanneer we de performantie evalueren van de GSVD-gebaseerde optimaal-filtertechniek in combinatie met de verschillende spraakdetectie-algoritmes, dan blijkt dat de beste performantie voor verschillende ruistypes verkregen wordt door de spraakdetectie-algoritmes gebaseerd op 'log-likelihood' en log-energie.

In **paragraaf 5.4** wordt de performantie van de GSVD-gebaseerde optimaal-filtertechniek vergeleken met standaard bundelvormingstechnieken voor verschillende akoestische scenario's (één en meerdere gesimuleerde ruisbronnen, realistische opname). De figuren 5.12, 5.13 en 5.14 tonen de *SNR-verbetering* en de *spraakvervorming* van verschillende algoritmes (DS-bundelvormer, GSC, spatiale 'blocking'-matrix, GSVD-gebaseerde optimaal-filtertechniek met en zonder 'ANC-postprocessing'-trap) voor verschillende reverberatietijden en ruisscenario's (witte ruis, spraakruis, 3 ruisbronnen). De SNR-verbetering van de GSVD-gebaseerde optimaal-filtertechniek met 'ANC-postprocessing'-trap is steeds beter dan de SNR-verbetering van de GSC voor alle reverberatietijden en voor alle beschouwde akoestische scenario's. Voor de GSVD-gebaseerde optimaal-filtertechniek treedt een grotere spraakvervorming op voor hogere reverberatietijden en wanneer de 'ANC-postprocessing'-trap met meerdere ruisreferenties toegevoegd wordt. Uit deze figuren blijkt ook dat de performantie voor witte ruis beter is dan voor spraakruis en dat de performantie voor één enkele ruisbron beter is dan voor meerdere ruisbronnen, wat volledig in overeenstemming is met de frequentiedomeinanalyse uit paragraaf 3.5. In deze paragraaf wordt ook de *robuustheid* van de GSC en de GSVD-gebaseerde optimaal-filtertechniek geanalyseerd voor verschillende afwijkingen in het veronderstelde signaalmodel: (a) afwijking in de versterking en de fase van de microfoons, (b) afwijking in de microfoonpositie, (c) foutieve veronderstelling over de richting van de spreker. De GSC is zeer gevoelig voor een afwijking in de versterking en de fase (en in mindere mate voor de andere afwijkingen) wanneer de ruisgevoeligheid groot is. Aangezien de GSVD-gebaseerde optimaal-filtertechniek geen a-priori veronderstellingen maakt over de positie van de spreker of over de microfoonkarakteristieken, tonen simulaties aan dat deze techniek robuuster is dan de GSC voor de 3 beschouwde afwijkingen. We kunnen zelfs bewijzen dat de performantie van de GSVD-gebaseerde optimaal-filtertechniek onafhankelijk is van de versterking en de fase van de microfoons.

Deel II : Meer-kanaals dereverberatie en Bronlokalisatie

In dit deel worden meer-kanaals algoritmes besproken voor het schatten van tijdsvertraging, voor dereverberatie en voor gecombineerde ruisonderdrukking en dereverberatie. Aangezien deze algoritmes een schatting vereisen van de akoestische impulsresponsies, bespreken we ook adaptieve en niet-adaptieve technieken om akoestische impulsresponsies te schatten, zowel in het tijdsdomein als in het frequentiedomein. We leiden een stochastisch-gradiëntalgoritme af dat iteratief de veralgemeende eigenvector berekent behorend bij de kleinste veralgemeende eigenwaarde en dat gebruikt kan worden voor het schatten van tijdsvertraging. We tonen aan dat een gecombineerde techniek voor ruisonderdrukking en dereverberatie kan bekomen worden door het genormaliseerd 'matched' filter te integreren met het meer-kanaals Wiener-filter.

Hoofdstuk 6: Robuuste schatting van tijdsvertraging voor akoestische bronlokalisatie

In veel toepassingen, zoals video-conferencing, spraakgestuurde systemen en hoorapparaten, is het wenselijk om de actieve spreker te lokaliseren. Met behulp van een microfoonrooster is het mogelijk om de *positie* van deze spreker te bepalen, zodat het microfoonrooster elektronisch kan gestuurd worden door middel van vaste (en adaptieve) bundelvormers of zodat de videocamera automatisch op de spreker gericht kan worden. In de literatuur is reeds aangetoond dat de positie berekend kan worden uit de *tijdsvertragingen* tussen de verschillende microfoonsignalen. Een nauwkeurige schatting van deze tijdsvertragingen is echter geen eenvoudige taak wegens reverberatie, achtergrondlawaai en het niet-stationaire karakter en het lage-rangmodel van spraaksignalen. Aangezien de meeste standaard technieken (bv. gebaseerd op de veralgemeende kruiscorrelatie) een ideaal kamermodel zonder reverberatie veronderstellen, is hun performantie vrij laag in reverberante omgevingen. Recent is een adaptief Eigenwaarde-Ontbinding (EVD) algoritme voorgesteld voor een (gedeeltelijke) schatting van 2 akoestische impulsresponsies met behulp van een stochastisch-gradiëntalgoritme dat iteratief de eigenvector behorend bij de kleinste eigenwaarde schat. Uit de geschatte akoestische impulsresponsies kan de tijdsvertraging berekend worden als het tijdsverschil tussen de eerste pieken (overeenkomend met het direct pad) of als de piek van de correlatiefunctie tussen de 2 impulsresponsies. De performantie van het adaptief EVD-algoritme is veel beter dan standaard technieken in een reverberante omgeving.

Strikt gesproken is het adaptief EVD-algoritme enkel geldig wanneer er geen ruis aanwezig is of wanneer er spatio-temporele witte ruis aanwezig is. In dit hoofdstuk breiden we daarom het adaptief EVD-algoritme uit voor het geval waar spatio-temporele gekleurde ruis aanwezig is, door een adaptief stochastisch-gradiëntalgoritme af te leiden voor de Veralgemeende-Eigenwaarde-Ontbinding (GEVD) of door een 'prewhitening'-operatie uit te voeren op de microfoonsignalen. Bovendien breiden we alle beschouwde algoritmes voor het schatten van tijdsvertraging uit naar het geval van meer dan 2 microfoons.

Paragraaf 6.2 bespreekt de niet-adaptieve ('batch') schatting van de volledige akoestische impulsresponsies uit de microfoonsignalen, gebruik makend van deelruimte-gebaseerde technieken. We tonen aan dat als de lengte van de akoestische impulsresponsies ofwel gekend is ofwel overschat kan worden, de volledige akoestische impulsresponsies berekend kunnen worden uit de EVD van de spraakcorrelatiematrix (indien geen of spatio-temporele witte ruis aanwezig is) of uit de GEVD van de spraak- en de ruiscorrelatiematrix (indien spatio-temporele gekleurde ruis aanwezig is). Uit simulaties blijkt dat deze procedures vrij gevoelig zijn voor de onafhankelijkheidsveronderstelling tussen spraak en ruis. Hoe beter aan deze veronderstelling voldaan is (bv. hogere SNR, langere spraak- en ruissegmenten), hoe beter de schatting is. In de praktijk kunnen de akoestische impulsresponsies duizenden filtertaps hebben, afhankelijk van de hoeveelheid reverberatie. Wegens het (benaderend) lage-rangmodel van het spraaksignaal zullen correlatiematrices van het zuiver spraaksignaal met deze dimensies rangdeficiënt of op zijn minst slecht geconditioneerd zijn. Daarom is het in de praktijk vrij moeilijk om met deze deelruimte-gebaseerde technieken in het tijdsdomein de volledige akoestische impulsresponsies te schatten, zeker wanneer er achtergrondruis aanwezig is.

Deze 'batch'-technieken voor het schatten van impulsresponsies vormen de basis voor het afleiden van stochastisch-gradiëntalgoritmes die iteratief de (veralgemeende) eigenvector behorend bij de kleinste (veralgemeende) eigenwaarde berekenen. In **paragraaf 6.3** beschrijven we het adaptief EVD-algoritme en leiden we een adaptief GEVD-algoritme en een adaptief 'prewhitening'-algoritme af. In de literatuur is reeds aangetoond dat het adaptief EVD-algoritme gebruikt kan worden voor het schatten van tijdsvertraging, merkwaardig genoeg zelfs wanneer de lengte van de akoestische impulsresponsies onderschat wordt. Door middel van simulaties tonen we aan dat dit resultaat ook geldig is voor het adaptief GEVD-algoritme wanneer spatio-temporele gekleurde ruis aanwezig is.

In **paragraaf 6.4** beschrijven we hoe de ontwikkelde 'batch' en adaptieve algoritmes voor het schatten van tijdsvertraging eenvoudig uitgebreid kunnen worden voor het geval van meer dan 2 microfoons.

In **paragraaf 6.5** worden de simulatieresultaten beschreven. De performantie van de verschillende adaptieve algoritmes voor het schatten van tijdsvertraging (EVD, GEVD, 'prewhitening') wordt onderzocht voor verschillende reverberatiecondities (ideaal en realistisch), verschillende signaal-ruisverhoudingen en verschillende microfoonconfiguraties (2 en 3 microfoons). De simulaties tonen aan dat voor alle beschouwde scenario's de tijdsvertragingen robuuster geschat worden door het adaptief GEVD-algoritme dan door het adaptief EVD-algoritme en het adaptief 'prewhitening'-algoritme.

Hoofdstuk 7: Gecombineerde ruisonderdrukking en dereverberatie

Zoals reeds aangegeven in paragraaf 2.2, vormt het doel van meer-kanaals signaalverbetering ofwel ruisonderdrukking (zonder aandacht te schenken aan residuele reverberatie), dereverberatie (zonder aandacht te schenken aan residuele ruis) of gecombineerde ruisonderdrukking en dereverberatie (waarbij gelijktijdig de transferfunctie voor het spraaksignaal een vertraging moet benaderen en de residuele ruiscomponent geminimaliseerd wordt).

De meeste meer-kanaals dereverberatie-algoritmes (bv. inverse of 'matched' filtering) vereisen een schatting van de akoestische impulsresponsies, in het tijdsdomein of in het frequentiedomein. Zoals reeds aangegeven in paragraaf 6.2, kan met behulp van deelruimte-gebaseerde technieken een schatting in het *tijdsdomein* bekomen worden, maar is het in de praktijk vrij moeilijk om de volledige akoestische impulsresponsies te schatten wegens de lengte van de impulsresponsies, het lage-rangmodel van het spraaksignaal en achtergrondruis. Bovendien blijken deelruimte-gebaseerde technieken in het tijdsdomein vrij gevoelig te zijn voor een onderschatting van de lengte van de impulsresponsies.

Wegens deze redenen zijn er in de literatuur ook technieken in het *frequentiedomein* voorgesteld om de akoestische transferfuncties te schatten. Alhoewel deze technieken in het frequentiedomein minder gevoelig zijn voor het ordeschattingsprobleem, treedt er een (onbekende) schalingsambiguïteit op in elke frequentiebin. Het wegwerken van deze ambiguïteit vereist voorafgaande kennis over de akoestische transferfuncties, wat duidelijk een nadeel is en wat het praktisch gebruik van deze frequentiedomeintechnieken beperkt. Strikt gesproken zijn de voorgestelde technieken enkel geldig wanneer spatiaal witte ruis aanwezig is. In dit hoofdstuk breiden we deze technieken uit voor het geval wanneer spatiaal gekleurde ruis aanwezig is en tonen we aan dat met behulp van de geschatte akoestische transferfuncties zowel dereverberatie als gecombineerde ruisonderdrukking en dereverberatie uitgevoerd kan worden.

In **paragraaf 7.2** stellen we een deelruimte-gebaseerde techniek in het frequentiedomein voor om de akoestische transferfuncties te schatten wanneer spatiaal gekleurde ruis aanwezig is. Het blijkt dat deze techniek vrij gelijkaardig is aan de deelruimte-gebaseerde techniek in het tijdsdomein, waar nu de akoestische-transferfunctievector berekend kan worden uit de veralgemeende eigenvector behorend bij de *grootste* veralgemeende eigenwaarde (in tegenstelling tot de kleinste veralgemeende eigenwaarde voor de tijdsdomeintechniek). De transferfunctievector kan echter maar geschat worden op een (frequentie-afhankelijke) schaleringsfactor na, wat resulteert in een ambiguïteit die enkel opgelost kan worden als de norm van de transferfunctievector gekend is. Alhoewel aangetoond is dat deze norm minder beïnvloed wordt door kleine bewegingen van de spreker dan de individuele transferfuncties, zal deze norm toch drastisch wijzigen wanneer de spreker in de kamer rondloopt. Dit beperkt het gebruik van deze frequentiedomeintechniek tot bv. desktop- of wagentoepassingen, waar de positie van de spreker vrij vast is en de norm van de transferfunctievector op voorhand opgemeten kan worden. Het wegwerken van deze voorafgaande kennis over de norm van de transferfunctievector is een onderwerp voor verder onderzoek. Wanneer spatiaal witte ruis aanwezig is, kan een deelruimte-trackingprocedure gebruikt worden om de voornaamste eigenvector adaptief te schatten. De uitbreiding van deze deelruimte-trackingprocedure naar spatiaal gekleurde ruis is een onderwerp voor verder onderzoek.

Met behulp van de geschatte akoestische-transferfunctievector kan zowel dereverberatie als gecombineerde ruisonderdrukking en dereverberatie uitgevoerd worden. In **paragraaf 7.3** wordt aangetoond dat perfecte *dereverberatie* bekomen kan worden met behulp van het genormaliseerd 'matched' filter. Aangezien dit filter geen rekening houdt met de residuele ruiscomponent, is het zelfs mogelijk dat de ruiscomponenten in de microfoonsignalen versterkt worden door dit dereverberatiefilter. In deze paragraaf tonen we ook aan dat de MMSE-schatting van het zuivere spraaksignaal $s[k]$ bekomen kan worden door de MMSE-schattingen van de spraakcomponenten in de microfoonsignalen te filteren met het genormaliseerd 'matched' filter. We kunnen dus een techniek voor *gecombineerde ruisonderdrukking en dereverberatie* bekomen door het genormaliseerd 'matched' filter te integreren met het meer-kanaals Wienerfilter voor ruisonderdrukking. Aangezien beide algoritmes gebruik maken van dezelfde ontbinding, namelijk de GSVD van een spraak- en een ruisdatamatrix, kunnen ze eenvoudig gecombineerd worden. Merk op dat zowel voor de voorgestelde dereverberatietechniek als voor de techniek voor gecombineerde ruisonderdrukking en dereverberatie voorafgaande kennis over de norm van de akoestische-transferfunctievector nodig is.

In **paragraaf 7.4** worden enkele praktische implementatie-aspecten besproken. Aangezien in feite een convolutie in het frequentiedomein uitgevoerd wordt, moeten de corresponderende filters in het tijdsdomein beperkt worden om circulaire convoluties te vermijden.

In **paragraaf 7.5** worden de simulatieresultaten beschreven voor een kamer met reverberatietijd $T_{60} = 400\,\text{msec}$, een microfoonrooster met 4 microfoons op een afstand van 2 cm van elkaar en een signaal-ruisverhouding van 0 dB. Voor de verschillende algoritmes geeft Tabel 7.1 een overzicht van de objectieve performantiecriteria voor ruisonderdrukking en dereverberatie. De simulatieresultaten tonen aan dat de GSVD-gebaseerde meer-kanaals Wiener-filtertechniek de beste signaal-ruisverhouding oplevert, dat de GSVD-gebaseerde dereverberatietechniek met behulp van het genormaliseerd 'matched' filter de beste dereverberatieperformantie heeft, en dat de techniek voor gecombineerde ruisonderdrukking en dereverberatie een afweging maakt tussen beide doelstellingen.

Deel III : Ontwerp van breedband bundelvormers

In dit deel worden verschillende ontwerpprocedures besproken voor vaste breedband bundelvormers met een willekeurig spatiaal directiviteitspatroon voor een gegeven willekeurig microfoonrooster, met behulp van een FIR 'filter-and-sum'-structuur. We stellen 2 nieuwe kostfuncties voor die gebaseerd zijn op eigenfilters. We bespreken het ontwerp van 'far-field', 'near-field' en 'mixed near-field far-field' breedband bundelvormers en we ontwikkelen 2 ontwerpprocedures voor breedband bundelvormers die robuust zijn tegen afwijkingen in de versterking en de fase van de microfoons.

Hoofdstuk 8: 'Far-field' breedband bundelvorming

Welgekende meer-kanaals signaalverbeteringstechnieken zijn vaste en adaptieve bundelvormers (cf. paragraaf 2.5). Alhoewel adaptieve bundelvormers in het algemeen een betere performantie hebben dan vaste bundelvormers en zich kunnen aanpassen in een veranderende akoestische omgeving, zijn adaptieve bundelvormers vaak vrij gevoelig voor afwijkingen in het veronderstelde signaalmodel (cf. paragraaf 5.4). Daarom worden vaste bundelvormers (met een vast directiviteitspatroon) soms verkozen omdat ze geen controle-algoritme nodig hebben en wegens hun eenvoudige implementatie en lage berekeningscomplexiteit. Vaste bundelvormers worden vaak gebruikt in zeer reverberante omgevingen, om meerdere bundels te creëren, in toepassingen waar de positie van de spreker ongeveer gekend is en om de spraakreferentie in een GSC te creëren.

In het algemeen proberen vaste bundelvormers ruimtelijk in te zoomen op de spraakbron, om zo reverberatie en achtergrondruis te onderdrukken die niet uit dezelfde richting als de spraakbron komt. Voor de meeste vaste bundelvormingstechnieken uit paragraaf 2.5 (DS-bundelvormer, differentiële microfoons, superdirectieve microfoons, frequentie-invariante bundelvormers) is het echter niet mogelijk om een willekeurig spatiaal directiviteitspatroon te ontwerpen voor een willekeurige microfoonroosterconfiguratie. Dit is echter wel mogelijk met behulp van de meest algemene FIR 'filter-and-sum'-structuur (cf. Figuur 8.1), waarmee een – vooraf gedefinieerd – gewenst spatiaal directiviteitspatroon zo goed mogelijk benaderd kan worden door een bepaalde kostfunctie te optimaliseren. In dit hoofdstuk worden verschillende kostfuncties voorgesteld om breedband bundelvormers te ontwerpen, die bv. gebaseerd zijn op gewogen kleinste-kwadraten (LS), een maximum-energie-rooster of niet-lineaire optimalisatietechnieken. Alhoewel we in het algemeen de voorkeur geven aan de niet-lineaire ontwerpprocedure, leidt deze ontwerpprocedure tot een grote berekeningscomplexiteit omdat een iteratieve optimalisatietechniek vereist is. Daarom stellen we ook 2 nieuwe niet-iteratieve ontwerpprocedures voor die gebaseerd zijn op eigenfilters: de conventionele eigenfiltertechniek die een referentiepunt nodig heeft, en de eigenfiltertechniek gebaseerd op een 'Total Least Squares' (TLS) criterium. Door simulaties zal aangetoond worden dat de TLS-eigenfiltertechniek de beste niet-iteratieve ontwerpprocedure is, dit wil zeggen de niet-iteratieve procedure waarvan de performantie het dichtst de niet-lineaire procedure benadert maar met een veel lagere berekeningscomplexiteit. In dit hoofdstuk veronderstellen we dat de spraakbron zich in het 'far-field' van het microfoonrooster bevindt en dat de microfoons een (perfecte) omni-directionele karakteristiek hebben met een vlakke frequentieresponsie gelijk aan 1. In hoofdstuk 9 worden 'near-field' en 'mixed near-field far-field' breedband bundelvormers besproken en in hoofdstuk 10 worden robuuste breedband bundelvormers besproken, die de microfoonkarakteristieken mee in rekening brengen.

Paragraaf 8.2 bespreekt enkele conventies met betrekking tot de notatie en geeft enkele definities (bv. spatiaal directiviteitspatroon). Aangezien in dit hoofdstuk verondersteld wordt dat de spraakbron zich in het 'far-field' van het microfoonrooster bevindt, kunnen we vlakke golfvoortplanting en een gelijke verzwakking voor alle microfoons veronderstellen.

Het ontwerp van een breedband bundelvormer bestaat uit het berekenen van de filtercoëfficiënten, zodanig dat het werkelijk spatiaal directiviteitspatroon het gewenst spatiaal directiviteitspatroon zo dicht mogelijk benadert. Er bestaan verschillende ontwerpprocedures naargelang de kostfunctie die geoptimaliseerd wordt. In **paragraaf 8.3** worden 3 kostfuncties voorgesteld:

- de welgekende *gewogen-kleinste-kwadraten*kostfunctie (LS), die de gewogen-kleinste-kwadratenfout tussen het werkelijk en het gewenst spatiaal directiviteitspatroon minimaliseert. Deze kostfunctie kan geschreven worden als een kwadratische functie.
- de *maximum-energie-rooster*kostfunctie (ME), die de energieverhouding tussen het passband- en het stopbandgebied maximaliseert. Deze kostfunctie geeft aanleiding tot een veralgemeend eigenwaardeprobleem.
- de *niet-lineaire* kostfunctie (NL), die de fout tussen de amplitudes van het werkelijk en het gewenst spatiaal directiviteitspatroon minimaliseert, zonder rekening te houden met de fase van de directiviteitspatronen. We stellen een kleine wijziging voor aan de standaard niet-lineaire kostfunctie, zodat de dubbele integralen slechts eenmalig moeten berekend worden.

In het algemeen geven we de voorkeur aan de (gewijzigde) niet-lineaire kostfunctie. Aangezien het minimaliseren van deze kostfunctie echter aanleiding geeft tot een niet-lineair optimalisatieprobleem, dat met behulp van iteratieve optimalisatietechnieken dient opgelost te worden en dus aanleiding geeft tot een hoge berekeningscomplexiteit, zullen we ook de niet-iteratieve ontwerpprocedures met een lagere berekeningscomplexiteit in aanmerking nemen. In paragraaf 8.4 worden 2 nieuwe niet-iteratieve kostfuncties, gebaseerd op eigenfilters, gedefinieerd en in paragraaf 8.6 wordt de performantie van alle beschouwde niet-iteratieve ontwerpprocedures vergeleken met de niet-lineaire ontwerpprocedure.

Voor alle kostfuncties zullen we het breedband-bundelvormerontwerp uitvoeren voor het volledige frequentie-hoek-gebied, dit wil zeggen dat we het fullband-probleem niet opsplitsen in afzonderlijke smallband-problemen voor verschillende frequenties. Bovendien zullen we de dubbele integralen over frequenties en hoeken niet benaderen door een eindige Riemann-som over een rooster van frequenties en hoeken. Voor elke kostfunctie bespreken we eerst het algemeen ontwerp voor een willekeurig gewenst spatiaal directiviteitspatroon, en beperken we ons dan tot het ontwerp van een breedband bundelvormer met een passband- en een stopbandgebied. Voor elke kostfunctie tonen we ook aan hoe lineaire beperkingen kunnen opgelegd worden aan de filtercoëfficiënten.

In **paragraaf 8.4** stellen we 2 nieuwe niet-iteratieve kostfuncties voor, die gebaseerd zijn op eigenfilters. Eigenfilters zijn reeds gebruikt voor het ontwerp van één-dimensionale FIR filters met lineaire fase, voor 2-dimensionale FIR filters en voor spatiale filters. In deze paragraaf breiden we het toepassingsgebied van eigenfilters uit naar het ontwerp van breedband bundelvormers. In deze paragraaf worden 2 eigenfilter-kostfuncties beschouwd:

- de *conventionele-eigenfilter*kostfunctie (EIG), die de fout tussen de relatieve werkelijke en gewenste spatiale directiviteitspatronen minimaliseert. Deze kostfunctie vereist een referentiepunt in het frequentie-hoek-gebied. Het minimaliseren van deze kostfunctie met of zonder bijkomende beperkingen leidt tot een (veralgemeend) eigenwaarde-probleem. Meestal wordt een kwadratische beperking gebruikt die de oppervlakte onder het spatiaal directiviteitsspectrum gelijkstelt aan 1.

- de *TLS-eigenfilter*kostfunctie (TLS), die de 'Total Least Squares' (TLS) fout tussen het werkelijk en het gewenst spatiaal directiviteitspatroon minimaliseert. Deze kostfunctie vereist geen referentiepunt en leidt ook tot een veralgemeend eigenwaarde-probleem.

In **paragraaf 8.5** worden verschillende types van lineaire beperkingen besproken die opgelegd kunnen worden aan de filtercoëfficiënten. Puntbeperkingen, lijnbeperkingen en afgeleide-beperkingen worden behandeld.

Paragraaf 8.6 beschrijft de simulatieresultaten voor de verschillende kostfuncties en voor drie verschillende ontwerpspecificaties (verschillende passband- en stopbandgebieden, lineaire beperkingen). Het bundelvormerontwerp wordt uitgevoerd voor een lineair uniform microfoonrooster met 5 microfoons op een afstand van 4 cm van elkaar, een bemonsteringsfrequentie van 8 kHz en een filterlengte van 20 taps. Voor alle ontwerpspecificaties vergelijken we de performantie van de niet-iteratieve ontwerpprocedures (LS, EIG, TLS, ME) met de niet-lineaire ontwerpprocedure (NL) en bepalen we welke niet-iteratieve ontwerpprocedure de beste performantie heeft, gebruik makend van de niet-lineaire kostfunctie als performantiecriterium. Uit deze simulaties blijkt dat de TLS-eigenfiltertechniek de beste niet-iteratieve ontwerpprocedure is.

Hoofdstuk 9: 'Near-field' breedband bundelvorming

Wanneer de spraakbron zich dicht genoeg bij het microfoonrooster bevindt (in het zogenaamde 'near-field'), zijn de 'far-field'-veronderstellingen niet meer geldig en moet sferische golfvoortplanting en signaalverzwakking voor de microfoonsignalen in rekening gebracht worden. Dit hoofdstuk bespreekt het ontwerp van 'near-field' breedband bundelvormers. Het ultieme doel is het ontwerp van een breedband bundelvormer waarvan het spatiaal directiviteitspatroon zo goed mogelijk het gewenst spatiaal directiviteitspatroon benadert *voor alle afstanden* tot het microfoonrooster. In dit hoofdstuk zullen we enkel het ontwerp van 'near-field' breedband bundelvormers bespreken voor één welbepaalde afstand tot het microfoonrooster en voor een beperkt aantal afstanden.

In **paragraaf 9.2** tonen we aan dat het ontwerp van 'near-field' breedband bundelvormers voor één welbepaalde afstand zeer gelijkaardig is aan het ontwerp van 'far-field' breedband bundelvormers ('far-field' bundelvormers zijn in feite een speciaal geval voor een oneindig grote afstand). Dezelfde ontwerpprocedures en kostfuncties kunnen gebruikt worden en het enige verschil ligt in de berekening van de dubbele integralen die voorkomen in het ontwerp. Het spatiaal directiviteitspatroon van een 'near-field' bundelvormer ontworpen voor één welbepaalde afstand tot het microfoonrooster kan echter voor andere afstanden in grote mate afwijken van het gewenst spatiaal directiviteitspatroon (cf. simulaties in paragraaf 9.4). Daarom stellen we in deze paragraaf ook ontwerpprocedures voor om breedband bundelvormers te ontwerpen die voor verschillende afstanden werken. Indien één van deze afstanden oneindig is, wordt dit 'mixed near-field far-field' bundelvorming genoemd. Deze uitbreiding is vanzelfsprekend voor de meeste kostfuncties (gewogen kleinste-kwadraten, conventionele eigenfilter, niet-lineaire procedure). Voor de TLS-eigenfiltertechniek en de maximum-energie-roosterkostfunctie leidt deze uitbreiding echter tot een zeer verschillend optimalisatieprobleem, namelijk een som van veralgemeende Rayleigh-quotiënten, waarvoor geen oplossing in gesloten vorm beschikbaar is en waarvoor iteratieve optimalisatietechnieken gebruikt moeten worden.

In **paragraaf 9.3** worden lineaire beperkingen voor het 'near-field'-geval besproken. Enkel punt- en afgeleide-beperkingen worden beschouwd, aangezien lijnbeperkingen niet gedefinieerd kunnen worden voor het 'near-field'-geval.

Paragraaf 9.4 beschrijft de simulatieresultaten voor 'near-field' breedband bundelvorming voor één welbepaalde afstand en voor 'mixed near-field far-field' bundelvorming. Voor het 'near-field'-ontwerp hebben we dezelfde ontwerpcriteria (microfoonrooster, passband- en stopbandgebieden, filterlengte) gebruikt als voor het 'far-field'-ontwerp uit paragraaf 8.6, maar hebben we de bundelvormer ontworpen voor een afstand $r = 0.2\,\mathrm{m}$ tot het microfoonrooster. Uit deze simulaties blijkt opnieuw dat de TLS-eigenfiltertechniek de beste niet-iteratieve ontwerpprocedure is. Voor het 'mixed near-field far-field'-ontwerp hebben we de bundelvormer ontworpen voor de afstanden $r = 0.2\,\mathrm{m}$ en $r = \infty$, gebruik makend van de gewogen-kleinste-kwadratenkostfunctie, de TLS-eigenfiltertechniek en de niet-lineaire kostfunctie. Uit deze simulaties blijkt dat voor alle kostfuncties het 'mixed near-field far-field'-ontwerp een afweging maakt tussen de performantie in 'near-field' en 'far-field'.

Hoofdstuk 10: Breedband bundelvorming robuust tegen afwijkingen in versterking en fase

In de vorige hoofdstukken hebben we verondersteld dat de microfoons een (perfecte) omni-directionele karakteristiek hebben met een vlakke frequentieresponsie gelijk aan 1. In dit hoofdstuk brengen we de microfoonkarakteristieken mee in rekening en bespreken we het ontwerp van breedband bundelvormers die robuust zijn tegen (onbekende) afwijkingen in de versterking en de fase van de microfoons.

Het is gekend dat vaste (en adaptieve) bundelvormers zeer gevoelig kunnen zijn voor afwijkingen in de microfoonkarakteristieken (versterking, fase, microfoonpositie). Kleine afwijkingen in de veronderstelde microfoonkarakteristieken kunnen leiden tot grote afwijkingen in het spatiaal directiviteitspatroon, zeker voor microfoonroosters met een kleine afmeting, die bv. frequent voorkomen in hoorapparaten en cochleaire implantaten. Aangezien het in de praktijk moeilijk is om microfoons met exact dezelfde karakteristiek te produceren, is het in feite onmogelijk om de exacte microfoonkarakteristiek te kennen zonder een meet- of kalibratieprocedure uit te voeren. Deze meet- of kalibratieprocedure zal echter enkel de foutgevoeligheid voor het beschouwd microfoonrooster verminderen, terwijl de kostprijs van zo'n procedure voor elk individueel microfoonrooster zeer groot is. Na kalibratie is het bovendien nog mogelijk dat de microfoonkarakteristieken veranderen in de tijd.

Een standaard techniek om de robuustheid tegen willekeurige afwijkingen te verbeteren bestaat erin de witte-ruis-versterking (WNG) te beperken door een kwadratische beperking op te leggen aan de filtercoëfficiënten. In dit hoofdstuk beschouwen we specifiek afwijkingen in de versterking en de fase van de microfoons en stellen we ontwerpprocedures voor om breedband bundelvormers met een willekeurig spatiaal directiviteitspatroon te ontwerpen die robuust zijn tegen deze specifieke afwijkingen. Aangezien we in dit hoofdstuk microfoonroosters met een kleine afmeting beschouwen, veronderstellen we dat de 'far-field'-veronderstellingen geldig zijn. Alle uitdrukkingen kunnen echter eenvoudig uitgebreid worden naar het 'near-field-geval.

In **paragraaf 10.2** herdefiniëren we de uitdrukkingen en de kostfuncties voor breedband-bundelvormerontwerp die de microfoonkarakteristieken mee in rekening brengen. In het algemeen bestaan de microfoonkarakteristieken uit een frequentie- en hoekafhankelijke versterking en fase. Door gebruik te maken van de geherdefinieerde uitdrukkingen, is het mogelijk om breedband bundelvormers te ontwerpen wanneer de microfoonkarakteristieken exact gekend zijn. Alle uitdrukkingen kunnen significant vereenvoudigd worden wanneer we veronderstellen dat de microfoonkarakteristieken onafhankelijk zijn van frequentie en hoek (zelfs als in de praktijk aan deze veronderstelling niet volledig voldaan is, kunnen we meestal het volledige beschouwde frequentie-hoek-gebied opsplitsen in kleinere gebieden waar wel aan deze veronderstelling voldaan is).

In veel toepassingen zijn de microfoonkarakteristieken echter niet exact gekend en kunnen ze zelfs veranderen in de tijd. In **paragraaf 10.3** stellen we 2 procedures voor om breedband bundelvormers te ontwerpen die robuust zijn tegen willekeurige afwijkingen in de versterking en de fase van de microfoons. In plaats van elk individueel microfoonrooster te kalibreren of op te meten, is het beter om alle mogelijke microfoonkarakteristieken te beschouwen en één van de volgende kostfuncties te optimaliseren:

- de *gemiddelde performantie*, dit wil zeggen de gewogen som van de kostfuncties voor alle mogelijke microfoonkarakteristieken, waar de waarschijnlijkheid van een bepaalde microfoonkarakteristiek als gewicht gebruikt wordt. Deze procedure vereist dus dat de waarschijnlijkheidsdichtheidsfuncties (pdf) van de versterking en de fase gekend zijn. Voor de gewogen-kleinste-kwadraten en de niet-lineaire kostfuncties blijkt dat dezelfde ontwerpprocedures als voor niet-robuust bundelvormerontwerp toegepast kunnen worden, die slechts enkele bijkomende parameters vereisen die eenvoudig uit de versterkings- en fase-pdf berekend kunnen worden (voor de versterking zijn de hogere-orde-momenten van de pdf nodig, terwijl voor de fase in het algemeen kennis over de volledige pdf noodzakelijk is). Wanneer we de gemiddelde performantie optimaliseren, is het toch nog mogelijk dat voor een specifieke combinatie van versterking en fase (typisch met een lage waarschijnlijkheid) de kostfunctie vrij hoog is. Om dit te vermijden, kunnen we de volgende kostfunctie minimaliseren.

- de *'worst-case'-performantie*, dit wil zeggen de maximale kostfunctie voor alle mogelijke microfoonkarakteristieken. Dit criterium is strenger dan de gemiddelde performantie, aangezien nu de kost voor het 'worst-case'-scenario geoptimaliseerd wordt. Deze procedure geeft aanleiding tot een minimax-optimalisatieprobleem over een eindig rooster van microfoonkarakteristieken (versterking en fase). Hoe meer punten dit rooster bevat, hoe hoger de berekeningscomplexiteit om het minimax-optimalisatieprobleem op te lossen. Wanneer enkel afwijkingen in de versterking van de microfoons beschouwd worden en de gewogen-kleinste-kwadratenkostfunctie gebruikt wordt, kunnen we aantonen dat het aantal roosterpunten drastisch gereduceerd kan worden.

Paragraaf 10.4 beschrijft de simulatieresultaten voor robuust breedbandbundelvormerontwerp. Aangezien het effect van afwijkingen groter is voor een microfoonrooster met een kleine afmeting, hebben we een lineair niet-uniform microfoonrooster gebruikt met 3 microfoons op posities $\begin{bmatrix} -0.01 & 0 & 0.015 \end{bmatrix}$ m. We hebben een 'endfire' breedband bundelvormer ontworpen voor een bemonsteringsfrequentie van 8 kHz en een filterlengte van 20 taps. Met behulp van de gewogen-kleinste-kwadratenkostfunctie, de TLS-eigenfiltertechniek en de niet-lineaire kostfunctie hebben we een niet-robuuste en verschillende robuuste bundelvormers ontworpen. Uit de simulaties blijkt dat robuust bundelvormerontwerp een grote verbetering in performantie oplevert zelfs wanneer er kleine afwijkingen in de versterking en/of de fase van de microfoons optreden.

Hoofdstuk 11: Besluit en suggesties voor verder onderzoek

In **paragraaf 11.1** wordt een algemeen besluit gegeven en **paragraaf 11.2** somt enkele suggesties voor verder onderzoek op:

- Alhoewel in hoofdstuk 4 verschillende technieken zijn voorgesteld om de berekeningscomplexiteit van de GSVD-gebaseerde optimaal-filtertechniek te verminderen, blijft de complexiteit vrij hoog – in feite veel hoger dan voor standaard bundelvormingstechnieken. Daarom zou het interessant zijn om andere technieken te bestuderen om de *complexiteit te* *verminderen*, bv. subband-gebaseerde QR-technieken of stochastische-gradiëntalgoritmes, zonder de performantie en de robuustheid drastisch te verlagen.

- Bovendien is het mogelijk dat het spraakdetectie-algoritme (VAD) volledig faalt, bv. bij zeer lage SNR of voor zeer niet-stationaire ruis. In dit geval wordt de performantie van het meer-kanaals Wiener-filter zeer onbetrouwbaar, wat kan resulteren in een onaanvaardbaar hoge spraakvervorming of trage convergentie. Daarom is het noodzakelijk om voor deze scenario's een grotere robuustheid te bekomen. Voor deze scenario's zijn vaste breedband bundelvormers daarentegen zeer robuust aangezien ze niet afhankelijk zijn van een VAD-algoritme. Daarom zou het interessant zijn om *de combinatie van meer-kanaals Wiener-filtering en vaste breedband bundelvorming* te onderzoeken. We verwachten dat de gecombineerde techniek robuuster is dan het meer-kanaals Wiener-filter in situaties waar het VAD-algoritme faalt, terwijl de performantie beter is dan vaste breedband bundelvormers in andere scenario's.
- In paragraaf 6.3 hebben we een stochastisch-gradiëntalgoritme voorgesteld dat de veralgemeende singuliere vector schat behorend bij de *kleinste* veralgemeende singuliere waarde. In paragraaf 7.2 is in feite een stochastisch-gradiëntalgoritme vereist dat de veralgemeende singuliere vector schat behorend bij de *grootste* veralgemeende singuliere waarde. Alhoewel er een deelruimte-trackingprocedure bestaat voor de SVD, blijft de *uitbreiding van deze deelruimte-trackingprocedure voor de GSVD* een onderwerp voor verder onderzoek.
- We geloven dat nog veel onderzoek vereist is voor *het schatten van akoestische impulsresponsies* en voor *dereverberatie*, aangezien hiervoor nog enkele fundamentele problemen dienen opgelost te worden (cf. hoofdstuk 6 en 7). Deelruimte-gebaseerde technieken in het tijdsdomein blijken zeer gevoelig te zijn aan een onderschatting van de lengte van de impulsresponsies, terwijl de onderliggende reden voor deze gevoeligheid niet goed begrepen is. Deelruimte-gebaseerde technieken in het frequentiedomein hebben voorafgaande kennis nodig over de akoestische transferfuncties om een schalingsprobleem op te lossen dat optreedt in elke frequentiebin. Het oplossen van dit schalingsprobleem blijft een onderwerp voor verder onderzoek. Een gelijkaardig schalingsprobleem treedt echter op in frequentiedomeintechnieken voor blinde-signaalscheiding (BSS), waar recent technieken ontwikkeld zijn om het schalings- en permutatieprobleem (gedeeltelijk) op te lossen. Het zou interessant zijn om te onderzoeken of *deze BSS-algoritmes ook kunnen gebruikt worden om het schalingsprobleem op te lossen dat optreedt bij het schatten van akoestische transferfuncties.* Bovendien moeten andere blinde systeemidentificatietechnieken zoals meer-kanaals lineaire predictie en niet-lineaire Kalman-filtering verder onderzocht worden, aangezien deze technieken reeds hun bruikbaarheid hebben bewezen in andere domeinen (bv. digitale communicatie).

Chapter 1

Introduction

1.1 Motivation

The work presented in this thesis is motivated by the rapidly growing market of speech and audio applications. Typical applications in the telecommunication and consumer equipment market include video-conferencing, hands-free mobile telephony and voice-controlled systems, whereas biomedical applications include hearing aids and cochlear implants. The main user benefit for the consumer equipment applications lies in the hands-free operation, enabling the user to walk around freely without wearing a headset or a microphone and hence providing a natural way of communication. Obviously, the main benefit for hearing aid applications is increased hearing capacity, enabling a hearing aid user to interact better with other people.

The common point between the above-mentioned applications is *speech acquisition in a (possibly) adverse acoustic environment*. In hands-free systems, a microphone or a microphone array is typically employed at a certain distance from the speaker. This causes problems not encountered in ordinary telephony or voice-controlled systems, where the microphones are usually installed or held close to the speaker. Therefore, in a hands-free communication system, the recorded speech signals are corrupted in various ways, i.e. by background noise, by room reverberation and by far-end echo signals (cf. Section 1.2.1). This signal degradation can lead to total unintelligibility of the speech signal and decreases the performance of speech coding and speech recognition systems. Hence high-performance signal enhancement procedures are called for.

In this thesis several *multi-microphone noise reduction and dereverberation techniques for speech applications* are discussed. Part I discusses a GSVD-based unconstrained optimal filtering technique, which can be used for single-microphone and multi-microphone noise reduction, but which does not perform dereverberation. Part II describes a combined noise reduction and dereverberation technique as well as an acoustic source localisation technique, which is robust against background noise and reverberation. In Part III design procedures are discussed for designing robust far-field and near-field broadband beamformers, which can be used both for noise reduction and dereverberation.

1.2 Hands-free speech communication systems

In this section, we first describe the general advantages and problems occurring in hands-free speech communication systems and then focus on some important applications, considering specific advantages, problems, economic importance and existing products for each application. However, the algorithms presented in this thesis are not designed with one specific application in mind and can be used for all considered speech communication applications. Of course, tuning an algorithm towards a specific application or user environment can considerably improve its performance.

1.2.1 General problem formulation

Figure 1.1 depicts a typical hands-free speech communication environment. Contrary to classical communication systems (e.g. hand-held telephony), the speaker is allowed to walk around freely in the room without wearing a headset or holding a microphone. The goal of the microphone array – typically located at a fixed position – is to record the (clean) speech signal uttered by the speaker. It is clear from Fig. 1.1 that in a hands-free system several types of signal degradation occur. Due to the large distance between the speaker and the microphone array, background noise sources are also picked up by the microphone array, and not only the direct path signal of the speaker is recorded, but also the signal reflections against walls, floors and objects present in the recording room (i.e. reverberation).

Background noise

Background noise typically arises from computer fans, traffic, audio equipment, or other speakers present in the room (i.e. cocktail party noise). Background noise can seriously reduce the intelligibility of the recorded speech signals. In particular, this is the case for hearing-impaired people, who are much more sensitive to the noise level or more precisely the signal-to-noise ratio (SNR) [212]. E.g. in a restaurant, with some background music and several groups of people talking to each other, a hearing-impaired person will already have severe difficulties in discriminating what his/her interlocutor is saying. It is also well known that the performance of acoustic source localisation techniques [106][271] and automatic speech recognition (ASR) systems [96][201] rapidly

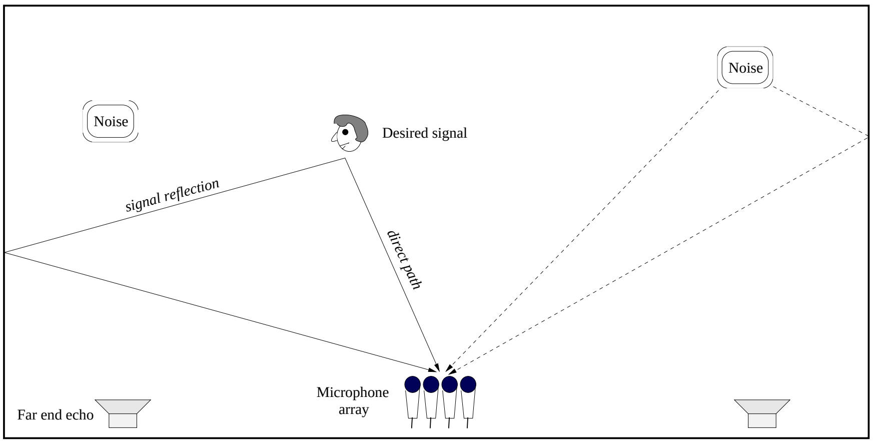


Figure 1.1: Typical hands-free speech communication environment

degrades with decreasing SNR. Signal processing techniques for reducing the background noise level are referred to as *acoustic noise reduction techniques* (cf. Section 1.4.1). Although most of the noise sources are unknown signals, it is sometimes possible to obtain a reference signal for the noise source, e.g. by directly using the emitted noise signal (e.g. other speaker), by recording the noise source with an additional microphone (e.g. car engine) or by using a related signal (e.g. ignition signal of the car engine). In these cases, specific signal enhancement techniques can be used. In this thesis, we will mainly focus on unknown noise sources for which no reference signal is available.

A specific type of noise, also depicted in Fig. 1.1, is *far-end echo*, emitted by a loudspeaker and coming from the remote site in a typical teleconferencing application. Far-end echo signals are also recorded by the microphone array and are sent to the remote site, where usually the same acoustic coupling between the loudspeakers and the microphone array exists. Hence, the local speaker will hear an echo or a delayed version of his/her own speech. In the worst case scenario, the closed loop gain may become too large and the system may become unstable, resulting in a harmful sinusoidal tone (i.e. acoustic feedback). For far-end echo signals, a noise reference can be easily obtained by using the signals emitted by the loudspeakers. Signal processing techniques for cancelling these echo signals are referred to as *acoustic echo cancellation techniques* (cf. Section 1.4.2). In this thesis, we will generally not consider acoustic echo cancellation.

Reverberation

The acoustic environment itself also plays an important role in hands-free speech communication systems. Acoustic waves, coming from the speaker, propagate through the air and are reflected by the walls, the floor, the ceiling, and in principle by any object present in the room before being picked up by the microphones. This propagation results in a signal attenuation and spectral distortion, called reverberation (cf. Section 1.3.3). Although speech and audio sound more pleasant when some reverberation is added [88], in highly reverberant environments the speech intelligibility of the recorded signals drops considerably [129][211]. It is also known that reverberation heavily effects the performance of ASR systems, since the performance of an ASR system trained in one specific environment will drop considerably when used in another acoustic environment [96][201]. In addition, the performance of most acoustic source localisation techniques is seriously degraded by room reverberation [29], even more than by background noise. Signal processing techniques for reducing or removing reverberation are referred to as *dereverberation techniques* or blind deconvolution techniques (cf. Section 1.4.3).

It has to be remarked that the human auditory system is remarkably robust in most adverse situations. We are able to focus on a speech source under severe noise conditions and in extreme reverberant environments. To a large extent this is due to the binaural nature of our hearing and to the (non-linear) adaptive processing in our inner-ear and our brains [20]. On the other hand, speech acquisition systems, speech recognition systems and acoustic source localisation systems do not process the incoming signals as the human auditory system, and their performance seriously degrades with increasing levels of background noise and reverberation.

1.2.2 Adaptive multi-microphone systems

For all above-mentioned problems (background noise, echo, reverberation), *single-microphone* signal enhancement techniques exist (cf. Section 1.4). Generally, the performance of single-microphone techniques is limited, since these techniques can only exploit the temporal and the spectral information present in the microphone signal. Especially for the dereverberation problem, no adequate single-microphone enhancement techniques are presently available. Hence, in many applications (e.g. video-conferencing, hearing aids), a growing tendency exists to move from single-microphone systems to *multi-microphone* systems [18][22][76][104][115][149]. Although multi-microphone systems come at an increased cost (more microphones, D/A converters, memory, signal processing power), they exhibit a huge advantage over single-microphone systems, since multi-microphone techniques are able to additionally exploit the spatial information of the sources. Typically, speech and noise sources are not located at the same position in the room, such that their signals can be spatially separated. Also for the dereverberation problem, it is possible to zoom in on the desired speaker by using multi-microphone beamforming techniques.

It is clear from Fig. 1.1 that the presented situation is far from being static. The speaker is allowed to move around freely in the room, while the noise sources can be non-stationary (both spectrally and spatially) and the acoustic environment can change e.g. by people moving around, doors opening, etc. Especially for hearing aid applications, the acoustic environment can change dramatically since the hearing aid user can be present in many different environments (office, concert hall, outdoors). Therefore there is a need for algorithms which can deal with different noise situations and with changing acoustic environments. *Adaptive multi-microphone algorithms* are suitable candidates for this task. Both for acoustic noise reduction, echo cancellation and dereverberation, separate (multi-microphone) algorithms are available (cf. Section 1.4). However, in order to increase the performance and to reduce the computational complexity of the complete signal enhancement system, there is also a tendency to *integrate the noise reduction, echo cancellation and dereverberation systems.* We will try to address both issues in this thesis.

1.2.3 Typical applications

Hands-free car kits and headsets

From an economic point of view, hands-free mobile telephony certainly is the most important application. The estimated number of worldwide cellular subscribers now exceeds one billion, and it is expected that this number will continue to increase in the near future [77][278]. Hands-free mobile telephony kits can mainly be found in the car, but are also available as small headsets, which are worn around one ear and which communicate with the mobile phone using a wireless (e.g. Bluetooth) protocol.

Recently in many countries – including Belgium – mobile telephony has been forbidden while driving, unless a hands-free car kit is used. This is motivated by the observation that hand-held mobile phone calls distract the driver and increase the number of accidents. During a mobile phone call, the driver misses 4 out of 10 road signs and fails to give way to other vehicles in 25% of the cases. It appears that the accident risk increases with 75%, which reduces to 24% if a hands-free kit is used [269]. Furthermore, it was found that 65% of all mobile phone conversations in North America take place in a car or another form of transport, but that actually less than 15% of the mobile phone users have hands-free accessories [38]. This implies that a large market for hands-free kits can be expected in the near future.

The main problems for *hands-free car kits* are far-end echo signals, emitted through the loudspeakers, and multiple noise sources: engine noise, wind and tire noise, traffic noise, the car radio – which is however generally turned off during calls – and other people talking in the car. Reverberation in a car environment is quite limited, as e.g. measurements in an empty mono-volume have shown small reverberation times in the range of 40 to 70 msec. However, background noise and far-end echo signals are able to cause low speech intelligibility and hence low overall system performance.

Most present-day hands-free car kits still use a single microphone, mounted on the dashboard or on the ceiling of the car. However, this microphone records all signals and is not able to accurately focus on the speaker. Even when using a directional microphone, this microphone has to be installed correctly in order to provide the correct focus. Therefore it is expected that these systems will evolve to multi-microphone systems, which are able to focus on the active speaker and reduce the annoying noise from the recorded signals. For signal enhancement algorithms in the car environment, one can also exploit the fact that the speaker is located close to the microphone array and that the position of the speakers (driver and passengers) is roughly known [115][193]. However, the main impediment for using a large number of microphones in a hands-free car kit is the cost of the microphones and the signal processing hardware.

For mobile telephony *headsets* – which can be used in any environment – background noise is the main problem. Echo signals also arise, however not propagating through the air as is usually the case, but through the headset itself. Reverberation is not an issue since the microphone is located quite close to the mouth of the speaker. However, better focusing and adaptive noise reduction could still be obtained by using multiple microphones.

The most common low-cost hands-free mobile telephony car kits are headsets using a directional microphone and a headphone, which are connected to the mobile phone with a wire (e.g. Panasonic KX-TCA87, ± 25 EUR). Smaller headsets, which are worn on one ear and which communicate with the mobile phone using a wireless (e.g. Bluetooth) protocol, are also available (e.g. Ericsson HBH-30, ± 180 EUR). However, systems without a headset are usually preferred, since these systems provide a more natural way of communication. Complete hands-free car kits, which need to be built in and integrated in the dashboard or the ceiling and which can be connected to the car radio, provide a better sound quality (e.g. Nokia CARK-91US, ± 150 EUR). The most advanced products rely on echo cancellation and noise reduction techniques (e.g. NMS Communications Sonata III), and even some multi-microphone beamforming and noise reduction products for car applications are already available (e.g. Digital Super Directional Array from Andrea Electronics, Mercedes TEMIC StarRec Acoustic Technology). It is expected that in the near future smaller and more advanced solutions for hands-free telephony will be developed, which can be integrated in the mobile phone themselves, and which provide high-quality wideband speech enhancement.

Video-conferencing

Apart from hands-free mobile telephony, audio- and video-conferencing forms another telecommunication application where hands-free speech acquisition is important [136]. In a one-to-one PC video-conferencing setup, a single microphone mounted on the screen of the PC will generally provide acceptable quality. However, for a video-conference with more participants, several switchable microphones would be required and an adaptively steerable microphone array provides a good alternative. Also tele-classing, which enables students to attend classes and lectures from a remote classroom, can be viewed as a special case of video-conferencing. In order to pick up questions of the students in the remote classroom, a microphone array solution can be used.

The main problems for video-conferencing systems are far-end echo signals and acoustic source localisation in noisy and reverberant acoustic environments. In typical video-conferencing locations, background noise is not extremely high (SNR generally larger than 10 dB) and reverberation is quite limited (reverberation time smaller than 500 msec). Acoustic source localisation obviously requires a multi-microphone solution and can be used both for correctly pointing the video camera at the active speaker and for adaptively steering a microphone array which zooms in on the active speaker and which removes the background noise.

Video-conferencing is a rapidly growing hands-free communication application. A market research report states that the market for audio-, video- and web-conferencing systems will reach US$ 9.8 billion by 2006, up from US$ 2.8 billion in 2000 [219]. Powerful teleconferencing systems are already commercially available. Polycom e.g. produces a range of full-duplex audio-conferencing equipment, which range from limited bandwidth solutions, intended for small business meetings, to larger systems, which provide integrated audio- and video-conferencing and which offer better audio quality (e.g. Polycom iPower™).

Voice-controlled systems

Thanks to the significant progress that has been made in the last decades, speech recognition is now sufficiently reliable to be integrated into commercial systems [39]. More and more voice-controlled systems are encountered in daily life, at home as well as at work. Voice-controlled domotic systems can e.g. be used for switching lights on and off, for controlling the central heating, for opening the curtains, etc. Also consumer electronics equipment (e.g. HiFi systems, TV, PC software), where the user is interested in a fast, easy and user-friendly interface, can be controlled using a limited number of voice commands. Another emerging market is telematics for the automotive industry, where speech recognition can e.g. be used to control non-critical cruise functions (ventilation, wipers) or to request navigation information. The global market for telematics equipment is expected to grow to US$ 12 billion in 2007, up from US$ 2.2 billion in 2001 [74].

However, in order that voice-controlled systems provide an added value over existing access techniques, the speech recognition system has to perform reliably and robustly for all users and in all circumstances. It is well known that the performance of speech recognition algorithms rapidly degrades with decreasing SNR and increasing reverberation [96][201]. This is mainly caused by the fact that poor speech recognition models are generally trained on clean speech data. Instead of retraining the models for every conceivable noise situation and acoustic environment, it is easier to apply a general preprocessing operation which suppresses the noise and the reverberation. Depending on the specific application and the user environment, both noise, echo and reverberation have to be taken into account. When using voice control in large rooms (e.g. living rooms), reverberation times can go up to 700 msec and a large distance normally exists between the user and the microphone array. When using voice control for audio equipment, the sound level of the audio equipment is usually higher than the voice level of the user, such that the recorded microphone signals generally have a quite low SNR.

Hearing aids and cochlear implants

Apart from the previously described 'commercial' applications, hearing aids and cochlear implants represent important biomedical applications where multi-microphone signal enhancement techniques can provide a significant performance improvement [149][234][245][266]. Hearing loss is one of the most prevalent chronic conditions affecting more than 300 million people world-wide. Approximately 25% of all people will encounter significant hearing loss during their lifetime. With the constantly aging Western population and more and more people being subject to loud noises (music in disco, noisy work situations) one can only expect this problem to increase. Most hearing impaired people suffer from perceptual hearing loss. This kind of hearing loss is not only caused by the fact that all sounds are decreased in loudness, but mainly because different sounds can not be distinguished any more from each other. Hence, this problem can not be solved by merely amplifying the sounds, but only by reducing the background noise with respect to the useful signal.

The main purpose of traditional hearing aids is pure amplification of all incoming sound. Hearing aids using a single omni-directional microphone amplify all sound, coming from all directions, such that the useful signal and the background noise are equally amplified. Despite their good performance in noise-free environments, they are not able to selectively reduce background noise, such that speech intelligibility considerably drops in noisy situations. In a recent survey with hearing aid users, almost half of them claimed that their hearing aid did not perform adequately in situations with background noise. However, it is mainly this inability to communicate in noisy environments which causes most people to purchase a hearing aid.

Recent developments in microphone manufacturing and micro-electronics have made it possible to integrate two – and even three – microphones and a small digital signal processor (DSP) into a single behind-the-ear (BTE) hearing aid. However, existing multi-microphone hearing aids (e.g. GN Resound, Phonak, Siemens) use rather simple digital signal processing algorithms (e.g. fixed delay-and-sum beamforming, differential microphone array), which have limited noise reduction capabilities. The main reason is the limited processing power of the DSP, partly due to power restrictions of the batteries inside the hearing aid – in some hearing aids, batteries need to be replaced every week. However, it is believed that further advances in battery technology and low-power ASIC design will increase the processing power of the used DSP, such that more advanced multi-microphone signal enhancement techniques - as developed in this thesis - can be implemented, increasing the performance and the robustness.

As already mentioned, mainly background noise is a problem for hearing aid applications. In most cases, one can safely assume that the desired speaker is located in front of the hearing aid user, such that all sounds coming from other directions may certainly be suppressed. Reducing reverberation is only of secondary importance, since even without a hearing aid, the signal received at the ear would sound reverberated. Because of the coupling between the loudspeaker and the microphone(s), which are located very close to each other, acoustic feedback will frequently occur [114][145][148][175][233][243]. In this thesis, we will however not study feedback suppression algorithms. Due to the limited size of a hearing aid, only a small number (2-3) of microphones can be fitted in the hearing aid. Moreover, these microphones will be spaced very close to each other (typically 1-2 cm). Because of the small inter-microphone distance, robustness against errors in the microphone characteristics (gain, phase) and positions becomes very important. Hence, robustness will be an important issue for most of the developed algorithms in this thesis.

A cochlear implant is a device which is implanted into the cochlea of a deaf person (whose auditory nerves are still intact) and which allows this person to perceive sounds again. An externally worn speech processor converts the perceived sounds and speech to electrical stimuli, which are then applied to the auditory nerves through intra-cochlear electrodes. In noise-free environments, the recovered speech intelligibility for a substantial number of users (especially children) is rather good, whereas the performance is considerably reduced – even more than for hearing impaired persons – when background noise is present. Since the same listening situations and noise sources are present as for hearing aids, it is obvious that similar signal enhancement techniques can have a tremendous impact for cochlear implant users.

At this moment, the total hearing aid industry is estimated at US$ 2 billion [19]. Considering the fact that only about 5% of all people who can benefit from a hearing aid actually uses one, there is still a big unexplored market.

1.3 Characterisation of signals and the acoustic environment

The characteristics of the speech and the noise signals and the properties of the acoustic environment have a large influence on the type of signal enhancement algorithm that has to be applied. In this section some properties of the speech and the noise signals and characteristics of the acoustic environment are discussed. Only the properties which are important for the signal processing techniques considered in this thesis are mentioned. More details about speech signal processing and acoustics can be found in [42][88][156][197][215].

1.3.1 Speech signals

Speech is a *wideband signal*, with frequency components ranging from 100 to 8000 Hz. According to its steady-state production model, a speech signal is not inherently band-limited. For voiced sounds, very little energy is present above 4 kHz, and the mean frequency envelope decays with about 6 dB/octave. For unvoiced sounds however, the spectrum is much flatter and does not fall off appreciably even above 8 kHz. For speech understanding mainly the frequencies between 300 and 3400 Hz are of interest, i.e. the classical telephony bandwidth. Hence, a sampling rate of 8 kHz is usually sufficient to obtain an acceptable speech quality. However, because of the demand for higher quality nowadays, higher sampling rates (e.g. 16 kHz) are used for so-called wideband speech systems. In this thesis, we will generally use a sampling rate of 16 kHz.

Speech is also a *non-stationary* signal, with both time envelope and spectrum continuously changing. Sometimes speech can be considered quasi-periodic (e.g. vowels), at other times it resembles coloured noise (e.g. fricatives) or is more impulse-like (e.g. plosives). Short-time stationarity in the order of 20-30 msec can be assumed for speech analysis, but generally this property is not relevant in multi-microphone speech enhancement algorithms. Furthermore, speech is an *intermittent signal*, i.e. silences exist between the words and in a typical conversation more than 50% of the time will consist of pauses. This on/off characteristic of speech signals can be exploited by speech enhancement algorithms e.g. by using a voice activity detection (VAD) algorithm which classifies noise-only periods and speech-and-noise periods.

A *low-rank linear model* has often been attributed to clean speech signals. This model assumes that each vector of the speech signal can be represented as the linear combination of a finite number of basis vectors, i.e. the L-dimensional speech vector $\mathbf{s}[k] = \begin{bmatrix} s[k] & s[k-1] & \dots & s[k-L+1] \end{bmatrix}^T$ can be written as

$$\mathbf{s}[k] = \sum_{i=1}^{R} \mathbf{s}_i a_i[k] \ , \tag{1.1}$$

with $R \leq L$ (if $R = L$, this representation is of course always possible) and $\{\mathbf{s}_1, \ldots, \mathbf{s}_R\}$ the set of L-dimensional linearly independent basis vectors. For the basis vectors, the *complex exponential* model is the best known model. Other related models are the exponential model, the damped sinusoidal model and the sinusoidal model [92][178]. Depending on the specific model and the specific speech frame, typical values for R range from 12 to 20. Note that these models have also been frequently used for speech coding algorithms [42].

1.3.2 Noise signals

In general, less is known about the noise sources. Background noise can originate from a *localised* noise source or can be *diffuse noise*, coming from all directions. E.g. in car applications, noise generated by the engine or the car radio can be considered to be localised, whereas noise from the wind passing around the car cabin or from the contact between the road and the tires can be considered diffuse noise [257]. For hearing aid applications background noise sources have been classified in [147].

Some of the noise sources are *stationary* (e.g. fans) or have a slowly varying spectral content, whereas other noise sources can be highly *non-stationary* (e.g. radio). The most difficult problem arises when the noise sources are also speech signals (e.g. concurrent speakers), which are similar in structure to the desired signal. Furthermore, the noise sources can be smallband (e.g. siren) or wideband, intermittent or persistent, and they may have the same spectral characteristics and/or angle of arrival as the desired speech signal. As already mentioned, if a reference signal can be obtained for the noise sources, the noise reduction problem is greatly simplified.

1.3.3 Acoustic environment

The acoustic environment plays an important role in hands-free communication systems, affecting both speech intelligibility and the performance of speech recognition systems [96][129][201][211]. Also the performance of most speech enhancement and acoustic source localisation algorithms is strongly influenced by the properties of the acoustic environment.

Reverberation is caused by the fact that acoustic waves are reflected by room walls and by other objects present in the room, such that the signals recorded by the microphone array consist of a direct path signal and multiple delayed and attenuated versions. Obviously, the acoustic path is different for each source-microphone pair. Since the positions of the sources are not necessarily fixed and objects can also move around through the room, acoustic paths are generally time-varying. It appears that the acoustic path can be modelled quite well by a linear transfer function. In real-life, additional (non-linear) phenomena occur [88], such as diffraction, diffusion, dissipation in the air, non-linear absorption and temperature-dependent effects. Although the linear transfer function model therefore is only an approximation, it nevertheless represents a useful tool for analysing and simulating many room acoustics problems.

Reverberation time

A global characterisation of a room can be given with a single parameter: the *reverberation time* T_{60}. The reverberation time is defined as the time needed for the sound pressure level to decay to -60 dB of its original value. The reverberation time is a function of the dimensions of the room and the materials used for walls, floor and ceiling. A typical office room has a reverberation time in the order of 200-500 ms, where T_{60} for a car is typically smaller than 200 ms and T_{60} for a church can be several seconds.

W.C. Sabine, the Harvard pioneer in acoustics, who introduced this concept, used a portable wind chest and organ pipes as a sound source, a stopwatch, and a pair of keen ears to measure the time from the interruption of the source to inaudibility [88]. Of course, today we have better techniques for measuring acoustic impulse responses, e.g. using impulse sound sources (electrical spark discharges, pistols, pricked balloons) or steady-state sound sources (frequency-bands of random noise), and we have better techniques for computing the reverberation time from the measured impulse responses [90].

If the dimensions of the room and the absorption (or reflection) coefficients of the used materials are known [156], then the reverberation time of the room can be calculated using different formulas. For a room with volume V [m^3] and absorption coefficient α_i for each room surface S_i [m^2], an approximate expression for T_{60} [sec] is given by Eyring's formula [89], i.e.

$$T_{60} = \frac{0.163\ V}{-S\log(1 - \frac{\sum_i S_i\alpha_i}{S})}\ , \tag{1.2}$$

with $S = \sum_i S_i$, or by Sabine's formula, which is a further approximation for small absorption coefficients,

$$T_{60} = \frac{0.163\ V}{\sum_i S_i\alpha_i}\ . \tag{1.3}$$

More advanced formulas for calculating the reverberation time have been described in [189].

Acoustic impulse response

While a room can be globally characterised by its reverberation time T_{60}, the linear filter incorporating the reverberation effects between two points in the room is described by an *acoustic impulse response.* Typical acoustic impulse responses, for different reverberation times, are shown in Fig. 1.2 and consist of three parts:

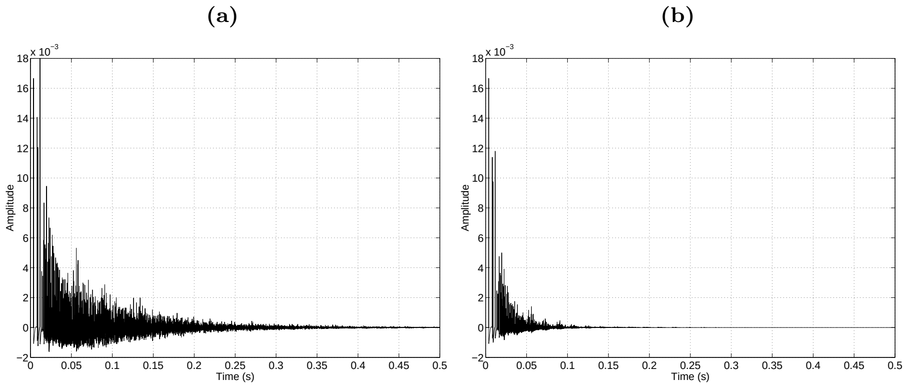


Figure 1.2: Acoustic impulse responses simulated using the image method [4][210] for different reverberation times. The simulated room has dimensions $6 \times 3 \times 2.5\,\mathrm{m}$, the source is located at $[2.4\ \ 2\ \ 1.5]\,\mathrm{m}$ and the microphone at $[3\ \ 1\ \ 1]\,\mathrm{m}$. The reverberation time T_{60} is equal to (**a**) 1000 ms, corresponding to a large auditorium, and (**b**) 300 ms, corresponding to a typical office.

- a *dead time*, i.e. the time needed for the acoustic wave to propagate from the source to the microphone along the shortest, direct acoustic path;
- a set of *early reflections*, whose amplitude and delay is strongly determined by the shape of the room and the positions of the source and the microphone;
- a set of *late reflections*, also called reverberation, which decay exponentially in time.

Acoustic impulse responses are typically modelled using finite impulse response (FIR) filters, having several hundreds or thousands of filter taps, depending on the room reverberation and the sampling frequency. In order to reduce the filter order, infinite impulse response (IIR) models may also be used. Although the filter order can indeed be reduced, it appears that it still remains large, i.e. several hundreds of taps [118]. Moreover, IIR signal enhancement algorithms typically lead to an increased computational load, stability problems or convergence to local minima [176][230]. Hence, IIR models are not really appropriate for modelling acoustic impulse responses.

Unfortunately, inverting an acoustic impulse response is not readily possible. In [188] it has been shown that acoustic impulse responses are generally *non-minimum-phase* systems. Non-minimum-phase systems have zeroes outside the unit circle and do not have a stable causal inverse, but still an approximate delayed inverse can be constructed [213]. When considering multiple microphones (and multiple impulse responses), it has been shown that the inverse system can be computed under certain (mild) conditions, cf. Section 2.6.1 [180].

A program for simulating acoustic impulse responses is of great help for speech enhancement research. It allows to build simulation experiments in a much wider variety of environments and situations than one would ever be able to measure. Acoustic impulse responses in a rectangular room can be easily simulated by the *image method* [4], which has been extended for microphone arrays by applying an additional low-pass filtering [210]. The image method is an elegant ray-tracing method. Instead of tracing all reflections, mirror images of the sound source with respect to the room boundaries are created. From each image source, a direct path is traced towards the receiving microphone. The order of the image is related to the number of reflections that have occurred and hence the ray can be multiplied with the correct attenuation factor. The image method has been widely used and is known to be sufficiently accurate in many applications. In this thesis, we will use the image method (up to order 3) for validating the signal enhancement algorithms in different environments. Of course, we will also perform simulations using real-life recordings.

1.3.4 Microphones for speech recordings

The used microphones can be positioned in different *microphone array configurations*. The term ‘array’ does not necessarily stand for a one-dimensional positioning, but the microphones can be positioned on a line, on an arc, in a planar or even a 3-dimensional ordering. In general, the microphone array configuration will have an influence on the performance of the multi-microphone signal enhancement algorithms. In this thesis, we will however not study the influence of the microphone array configuration on the performance.

Microphones are generally considered to be perfect point sensors with ideal omni-directional properties and a flat frequency response equal to 1. However, this is quite unrealistic, since generally the microphones also perform a spatial and a spectral filtering operation.

In a real microphone array setup, different kinds of *imperfections* occur. A first imperfection arises because the *characteristics* (gain, phase, directivity) of the microphones and the preamplifiers deviate from the assumed nominal characteristics. Moreover, the microphone and the preamplifier characteristics are ideally assumed to be equal for all sensors in the microphone array. Matched microphone pairs are commercially available, but these are difficult to produce and therefore are extremely expensive. One should realise that the performance of most microphone array algorithms is dependent on the individual microphone characteristics. E.g., it is well known that fixed and adaptive beamforming techniques are highly sensitive to errors in the assumed microphone array characteristics, especially for small-size microphone arrays (e.g. hearing aids) [24][37][86][128][192]. In practice, it is however impossible to exactly know the microphone array characteristics without a measurement or a calibration procedure [24][248]. Obviously, the cost of such a measurement or calibration procedure for every individual microphone array is objectionable. Moreover, after calibration the characteristics can still drift over time [137]. Therefore, signal enhancement algorithms should be designed to be robust against these (small) microphone array imperfections. Secondly, the *mounting* of the microphones onto the array may not be perfect, i.e. the distance from one microphone to another may also deviate from its nominal value. Again, fixed and adaptive beamforming techniques are quite sensitive to errors in the array geometry, as the complete filter design relies on it. Therefore, signal enhancement algorithms should be designed to be robust against these imperfections and should not critically rely on the exact mathematical relations derived from the array geometry. Finally, one has to take into account the potential *shadow effect* from each microphone on its adjacent microphones (especially for high frequencies) and the shadow effect of the head for a microphone array mounted on a BTE hearing aid. The average effect of the head shadow can e.g. be obtained through measurements.

Speech and noise sources can either be located close to the microphone array or far away from it. The former case is called a *near-field* situation, the latter a *far-field* situation. In far-field situations, plane wave propagation can be assumed and signal attenuation can be assumed to be equal for all microphones. In a near-field situation, spherical wave propagation and signal attenuation have to be taken into account. The typical rule of thumb is that far-field assumptions are no longer valid [169] when

$$r < \frac{d_{tot}^2 f_s}{c} \,, \tag{1.4}$$

with r the radial distance of the source to the centre of the microphone array, d_{tot} the total length (aperture) of the microphone array, f_s the sampling frequency and c the speed of sound ($c = 340\frac{m}{s}$). E.g. for $d_{tot} = 0.2\,\mathrm{m}$ and $f_s = 8\,\mathrm{kHz}$, the minimum source distance for the far-field assumptions to be valid is $r = 0.94\,\mathrm{m}$. If the far-field assumption holds, this can seriously simplify algorithm design. The speaker position with respect to the microphone array may only be known approximately, but in some cases (e.g hearing aids and car applications) more precise assumptions about the look direction and/or the speaker position hold.

1.4 Overview of speech enhancement techniques

For each form of signal degradation (noise, echo, reverberation), a brief overview of existing single and multi-microphone algorithms is given in this section. Chapter 2 discusses some noise reduction and dereverberation algorithms that are important for the remainder of this thesis in more (mathematical) detail. A good overview of several acoustic signal processing techniques and microphone array algorithms can be found in [12][22][104].

1.4.1 Acoustic noise reduction

Single-channel noise reduction

Single-channel noise reduction techniques have attracted a great deal of interest in the last decades [161]. These techniques are also called speech enhancement techniques, since their goal is to reduce as much noise as possible, without distorting the speech signal. Single-microphone speech enhancement algorithms can be broadly classified in parametric and non-parametric techniques. Parametric techniques model the noisy speech signal as a stochastic autoregressive (AR) model embedded in coloured Gaussian noise [162]. Speech enhancement then roughly consists of estimating the speech AR parameters and applying a (non-causal) Wiener filter [119][244] or Kalman filter [97][99][107] to the noisy signal, where the optimal filters are based on the estimated AR parameters. Non-parametric techniques do not estimate the speech parameters, but require a noise fingerprint in a transform domain (mainly the Discrete Fourier Transform (DFT) or the Karhunen-Loève Transform (KLT) domain). This noise fingerprint is estimated during noise-only periods and used during subsequent speech-and-noise periods in order to obtain an estimate of the clean speech signal. Well-known non-parametric techniques include spectral subtraction [21][45][83][84][163][172][177][268][280] and signal subspace-based techniques [43][49][61][85] [121][130][138][179][220]. These last two techniques will be discussed in more mathematical detail in Section 2.3.

Single-microphone speech enhancement techniques can only exploit the temporal and the spectral information of the speech and the noise signals and can therefore be considered a signal-adaptive frequency filtering of the noisy speech signal. Since the speech and the noise signal usually occupy overlapping frequency bands, single-microphone speech enhancement techniques generally have problems to reduce the background noise without introducing noticeable artifacts (e.g. musical noise [28][109]) or speech distortion. However, when the speech and the noise sources are physically located at different positions, spatial diversity can be exploited by using multi-microphone noise reduction techniques, such that both spectral and spatial characteristics of the signal sources can be used.

Fixed and adaptive beamforming techniques

A first class of multi-microphone noise reduction techniques is fixed and adaptive beamforming. A good overview of beamforming techniques can be found in [258][264]. *Fixed beamforming techniques* filter the microphone signals with fixed filters and hence are data-independent. A fixed beamformer tries to obtain spatial focusing on the speech source, thereby reducing reverberation and suppressing the background noise not coming from the same direction as the speech source. Typical fixed beamforming techniques include delay-and-sum beamforming, differential microphone arrays [24][75], superdirective microphone arrays [16][36][146] and frequency-invariant beamformers [274]. However, using these types of fixed beamformers, it is generally not possible to design arbitrary spatial directivity patterns for arbitrary microphone array configurations. This is however possible using a general filter-and-sum structure, where the filter coefficients of the fixed beamformer are calculated such that the spatial directivity pattern optimally fits the desired spatial directivity pattern with respect to some cost function [65][59][144][155][157][159][192]. Part III discusses several techniques for designing fixed far-field and near-field broadband beamformers, which additionally can be designed to be robust against errors in the microphone array characteristics.

Adaptive beamforming techniques combine the spatial focusing of fixed beamformers with adaptive noise suppression. This typically gives rise to constrained optimisation problems, yielding constrained solutions for the filters. In an LCMV (linearly constrained minimum variance) beamformer, i.e. the Frost beamformer [95], the energy of the output signal is minimised under the constraint that signals arriving from the look direction, i.e. the direction of the speech source, undergo a fixed filtering operation. A well-known alternative implementation of this LCMV beamformer is the Generalised Sidelobe Canceller (GSC), i.e. the Griffiths-Jim broadband beamformer [116], where the constrained optimisation problem is reformulated as an unconstrained optimisation problem. The GSC consists of a fixed beamformer, creating a so-called speech reference signal; a blocking matrix, creating so-called noise reference signals; and a multichannel adaptive filter [123], eliminating the (noise) components in the speech reference signal which are correlated with the noise reference signals. Because of room reverberation, microphone mismatch and look direction error, the speech signal may leak into the noise references, such that signal cancellation and signal distortion often cannot be avoided in the standard GSC. In order to limit signal cancellation and signal distortion, different variants of the standard GSC implementation exist, e.g. using a speech-controlled (VAD) adaptation algorithm [113][128][194][254], a spatial filter designed blocking matrix [191][194], norm-constrained [37] and coefficient-constrained adaptive filters [128] or incorporating a transfer function model [100]. Some of these variants will be discussed in more detail in Section 2.5.3.

Adaptive beamforming techniques generally have a better noise reduction performance than fixed beamforming techniques and are able to adapt to changing acoustic environments. However, they are also quite sensitive to modelling errors, resulting in speech distortion and cancellation. Therefore fixed beamforming techniques are sometimes preferred for their robustness and easy implementation. Fixed beamformers are frequently used in highly reverberant acoustic environments, in applications where the position of the speech source is approximately known (e.g. hearing aids), for creating multiple beams [150][259] and for creating the speech and noise reference signals in a GSC.

Multi-channel Wiener filtering

A second class of multi-microphone noise reduction techniques is multi-channel Wiener filtering [55][56][61][195][196][224][232][242]. These techniques are unconstrained optimal filtering techniques, which compute an optimal, i.e. minimum mean square error (MMSE), estimate of either the speech component in a microphone signal [56][61][224][242], the clean unreverberated speech signal [55][232] or a reference signal, which can e.g. be a linear combination of pre-recorded speech signals [195][196]. In these techniques, inevitably some linear speech distortion will be introduced, but speech distortion can be easily traded off with noise reduction [61]. Multi-channel Wiener filters can be implemented in different ways. Part I discusses a full-band GSVD-based implementation, which can in fact be considered a multi-microphone extension of the single-channel subspace-based speech enhancement algorithms. Other possible implementations include a QR-based implementation [221][224] or subband implementations [242][240]. Multi-channel unconstrained optimal filtering techniques give rise to a higher noise reduction performance than standard fixed and adaptive beamforming techniques and are more robust to deviations from the assumed signal model (e.g. look direction error, microphone mismatch, speech detection errors), but are computationally more demanding.

1.4.2 Acoustic echo cancellation

In order to suppress echo, several conventional acoustic echo cancellation techniques can be applied [122]. E.g. highly directional loudspeakers and microphones and sound absorbing materials can be used in order to avoid reflections. In practice nowadays, acoustic echo cancellers are based on *adaptive filtering techniques* [11][123][214]. Adaptive filtering techniques are a powerful signal processing tool which can be used for system modelling and for signal enhancement thanks to their self-learning capabilities. Adaptive filters can start from zero a-priori knowledge and by their inherent feedback structure and continuous updating they are able to form a model for the unknown system and track possible system variations.

Figure 1.3 depicts a typical adaptive filtering setup. The input signal $x[k]$ passes through an unknown system $h[k]$, leading to the desired signal $d[k] = h[k] \otimes x[k]$, with $\otimes$ denoting convolution between signals. The goal of the adaptive filter $w[k]$ is to model the unknown system $h[k]$. This can be achieved by updating the adaptive filter such that the energy of the error signal $e[k] = d[k] - y[k]$ is minimised. If the adaptive filter perfectly models the unknown system, then the error signal will be equal to 0. Hence, all adaptive filtering techniques consist of two major parts : a filtering operation and a filter adaptation. The adaptation is done continuously through feedback of the error signal.

A large set of adaptive filtering techniques has been developed during the last decades, differing in terms of performance (convergence speed, tracking, delay),

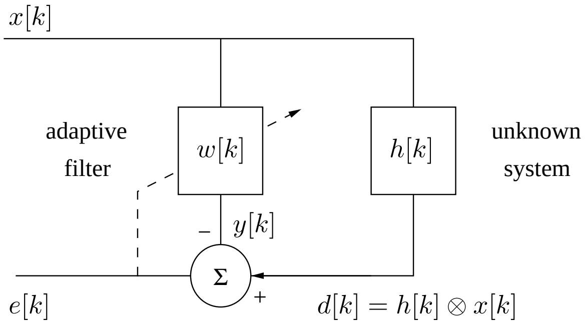


Figure 1.3: Adaptive filtering setup

complexity and stability. In general it is not easy to decide which algorithm is 'optimal' for a certain application, as it strongly depends on the available computational power and the designer's preference. In acoustic echo cancellation, the unknown system $h[k]$ that has to be modelled by the adaptive filter is the acoustic impulse response from the far-end echo loudspeaker to the microphone(s). Since this acoustic impulse response can be quite long and highly time-varying, the adaptive filter will require several hundreds or thousands of taps and high-performance (i.e. fast converging), but low complexity adaptive filtering algorithms are desirable. Moreover, the delay introduced by the algorithm cannot be too large. Advanced adaptive algorithms with a high complexity, such as the recursive least-squares (RLS) algorithm [123], are therefore not often taken into consideration, unless fast versions [10][82][182][214] are used. For acoustic applications, cheap algorithms, such as the least mean squares (LMS) and normalised LMS (NLMS) algorithm [123], are typically used. However, these algorithms exhibit a slow convergence behaviour, especially for coloured signals such as speech. Therefore, also 'in between'-solutions, such as the affine projection algorithm (APA) and its variants [103][186][203][222][249], have been investigated. These algorithms have a better convergence behaviour than LMS-type algorithms but a lower complexity than RLS-type algorithms.

The complexity of adaptive algorithms can still be reduced by using *frequency-domain or subband* algorithms [13][78][79][80][171][231]. Frequency-domain algorithms mainly have two advantages over time-domain algorithms. First, a considerable cost reduction can be obtained thanks to the block-processing and the use of fast signal transforms (FFT). Secondly, a better convergence behaviour is expected due to the decorrelation properties of the signal transforms. The main disadvantage is the quite large algorithmic delay, which can however be reduced by using block partitioning [44][73][81][237].

In fact, we can state that conceptually (mono) echo cancellation is a 'solved' problem, and that nowadays research is mainly focused on the development of less complex algorithms which provide the same (or better) performance.

However, in present-day setups, all consumer equipment produces two (stereo) or even more channels (surround sound), such that multiple loudspeaker signals need to be cancelled. It can be proved that *multi-channel acoustic echo cancellation* inherently suffers from a non-uniqueness problem, since the loudspeaker signals are generally highly correlated [102][238]. In practice, a unique solution for the multi-channel echo cancellation problem does exist, but the underlying optimisation problem appears to be ill-conditioned. In order to improve the conditioning of the problem, uncorrelated components can be introduced into the different loudspeaker signals. These components should however be inaudible and should not affect the stereo perception. Different techniques have been proposed for (partially) decorrelating the loudspeaker signals, e.g. using a half-wave rectifier [15], complementary comb filters [14], time-varying all-pass filters [3], or inserting psycho-acoustically-masked noise [108].

Traditionally, noise reduction and echo cancellation have been addressed independently, either by first cancelling the echo components in all microphone signals and then performing multi-microphone noise reduction, or vice-versa, by first performing multi-microphone noise reduction, followed by a single-channel echo canceller. Both schemes have their own advantages and disadvantages with respect to performance and complexity. Recently, it has been recognised that both problems are better solved using a *combined approach*, certainly when using multiple microphones. Initial results indicate that a combined approach yields a better performance at a lower complexity [47][125][126][151][152][158][173][174].

1.4.3 Dereverberation

Dereverberation consists of extracting the clean speech signal from the (reverberated) microphone signals, without any knowledge about neither the acoustic impulse responses nor the clean speech signal. Therefore dereverberation is also referred to as blind deconvolution.

For the *single-microphone dereverberation* problem, a direct solution is provided by conventional inverse filtering techniques. If the acoustic impulse response is known (from calculations or measurements), reverberation can be removed by using the inverse filter or by MMSE deconvolution [180]. However, since typical acoustic impulse responses are non-minimum-phase and therefore do not have stable causal inverses [188], inverse filtering-based single-microphone techniques have a limited scope in practice [180]. The situation is further complicated by the difficulty of estimating and tracking the acoustic impulse response in real-time applications. An alternative approach is provided by cepstrum-based techniques [8][202][251]. The underlying motivation is the fact that deconvolution in the time domain corresponds to division in the frequency-domain and subtraction in the cepstrum domain. Since the complex cepstrum of a speech signal is usually concentrated around the cepstral origin, while that of its echoes is composed of pulses away from the origin, dereverberation can be achieved by low-quefrency 'liftering' or peak-picking in the cepstrum domain. While cepstrum filtering has been applied successfully for the enhancement of speech degraded by simple echoes, its use for dereverberating single-microphone speech signals poses several practical problems [8]. We can conclude that single-microphone blind deconvolution techniques exhibit a limited performance.

On the other hand, *multi-microphone dereverberation techniques* provide spatial processing, such that the reverberant part can be spatially separated from the direct path signal. A first class of multi-microphone techniques is fixed beamforming, as apart from suppressing background noise, these techniques are also known to partially dereverberate the microphone signals. Fixed beamforming techniques try to capture the sound coming from the direction of the speaker and attenuate sounds coming from other directions, thereby reducing reverberation. In an early paper [5], a two-microphone fixed beamforming technique has been proposed which compensates for the phase and the amplitude differences of the two microphones signals and sums them coherently. Adaptive beamforming techniques have also been considered for the suppression of reverberation. However, the key assumption in adaptive beamforming algorithms is that the desired speech signal is statistically independent from all sources of interference, which means that the interference can not be due to delayed versions of the speech signal, as is the case in a reverberant room. Another class of multi-microphone dereverberation techniques combines beamforming with cepstrum-based processing. In [164] the different microphone signals are factored into a minimum-phase and an all-pass component, which are separately processed in the cepstrum and in the frequency-domain.

Standard delay-and-sum beamforming techniques use a simple time-delay compensation for aligning the different microphone signals, but do not achieve substantial dereverberation. More sophisticated beamforming techniques use matched filtering instead of this simple time-delay compensation [91]. When the different acoustic impulse responses are known, matched filtering simply consists of filtering the microphone signals with the time-reversed acoustic impulse responses. When the acoustic impulse responses are not known, matched filtering is still possible by e.g. estimating the acoustic impulse responses in the frequency-domain using subspace-tracking techniques [2].

In the case of a single speaker and multiple microphones, the total system to be identified and inverted can be considered a single-input multiple-output (SIMO) system. Blind system identification techniques [1][7][101][117][187][252][265] may therefore be used to estimate the different acoustic impulse responses, followed by an inverse multi-channel filtering operation [180][272]. However, most of these blind system identification techniques are not robust to over- and underestimation of the length of the acoustic impulse responses – which cannot be exactly estimated in practice – and furthermore have a high computational complexity. Moreover, because of the low-rank model of speech (cf. Section

1.3.1), these system identification problems become rank-deficient or at least very ill-conditioned. Therefore it is quite difficult to estimate and track the complete impulse responses, certainly when a large amount of background noise is present [101]. Hence, blind channel estimation and dereverberation for speech communication systems remains a topic of further research. In Part II time-domain and frequency-domain techniques are discussed for estimating and tracking the (partial) acoustic impulse responses in adverse acoustic environments, which can either be used for dereverberation or for acoustic source localisation.

1.5 Outline of the thesis and main contributions

In this section a chapter by chapter overview of the thesis is given, summarising the main contributions. We also provide references to the publications that have been produced in the frame of this work.

1.5.1 Objectives of the developed algorithms

In this thesis several multi-microphone signal enhancement techniques for noise reduction and dereverberation are discussed. As already mentioned, background noise and reverberation can seriously degrade the speech intelligibility and the performance of speech recognition systems in hands-free applications. In order to improve the ‘quality’ of the recorded microphone signals, different signal enhancement techniques are called for.

In this thesis we will focus on **multi-microphone** signal enhancement techniques, since multi-microphone techniques can exploit both the spectral and the spatial characteristics present in the microphone signals. Part I discusses a GSVD-based unconstrained optimal filtering technique for multi-microphone noise reduction. Part II describes an acoustic source localisation technique, obviously requiring the use of multiple microphones, and a combined noise reduction and dereverberation technique. In part III design procedures for broadband beamformers are discussed.

Since the signals and the acoustic environment are highly time-varying, most developed algorithms will be **adaptive**, enabling these algorithms to deal with different noise situations and with changing acoustic environments. Generally, we will assume that the background noise sources are **unknown**, i.e. that no reference for these noise sources is available. We will also discuss the **integration** of different signal enhancement techniques, e.g. combined noise reduction and echo cancellation in Part I and combined noise reduction and dereverberation in Part II.

Since most multi-microphone signal enhancement techniques are sensitive to errors in the microphone array characteristics (gain, phase, position) and other deviations (e.g. look direction error, speech detection errors), we will analyse the **robustness** of the developed algorithms against these deviations and where possible, we will take these deviations into account in the algorithm design. In Part I we will analyse the robustness of the GSVD-based optimal filtering technique with respect to deviations from the assumed signal model (microphone characteristics, look direction error, speech detection errors) and in Part III we will discuss design procedures for broadband beamformers that are robust against gain and phase errors in the microphone characteristics.

Finally, we will also consider the **computational complexity** of the developed algorithms. E.g. it will be shown that the complexity of the GSVD-based optimal filtering technique can be drastically reduced such that it becomes suitable for real-time implementation. However, our main concern is to develop algorithms that have a better performance and/or robustness than existing techniques, where computational complexity is only of secondary importance.

1.5.2 Chapter by chapter overview and contributions

This thesis consists of three parts. Each of these parts is divided into two or three chapters. A schematic overview of the thesis is given in Fig. 1.4.

In **Chapter 2** several existing single- and multi-microphone noise reduction and dereverberation techniques are briefly reviewed, i.e. single-microphone signal subspace-based noise reduction, fixed and adaptive beamforming, and inverse filtering and matched filtering techniques for dereverberation.

Part I: GSVD-Based Optimal Filtering for Multi-Microphone Noise Reduction

In this part we present a Generalised Singular Value Decomposition (GSVD) based optimal filtering technique for enhancing multi-microphone speech signals which are degraded by additive coloured noise. Several techniques are discussed for reducing the computational complexity and we show that the GSVD-based optimal filtering technique can be integrated into a Generalised Sidelobe Canceller (GSC) type structure. Simulations show that the GSVD-based optimal filtering technique achieves a larger SNR improvement than standard fixed and adaptive beamforming techniques and that it is more robust against several deviations from the assumed signal model. Publications related to this part of the thesis are [47][48][49][50][51][52][53][54][56][61][67].

In **Chapter 3** multi-channel Wiener filtering is discussed. This optimal filter produces a minimum mean square error (MMSE) estimate of the speech components in the microphone signals, thereby reducing the background noise but not the reverberation. We also discuss a more general class of estimators, making it possible to trade off speech distortion and noise reduction. When incorporating

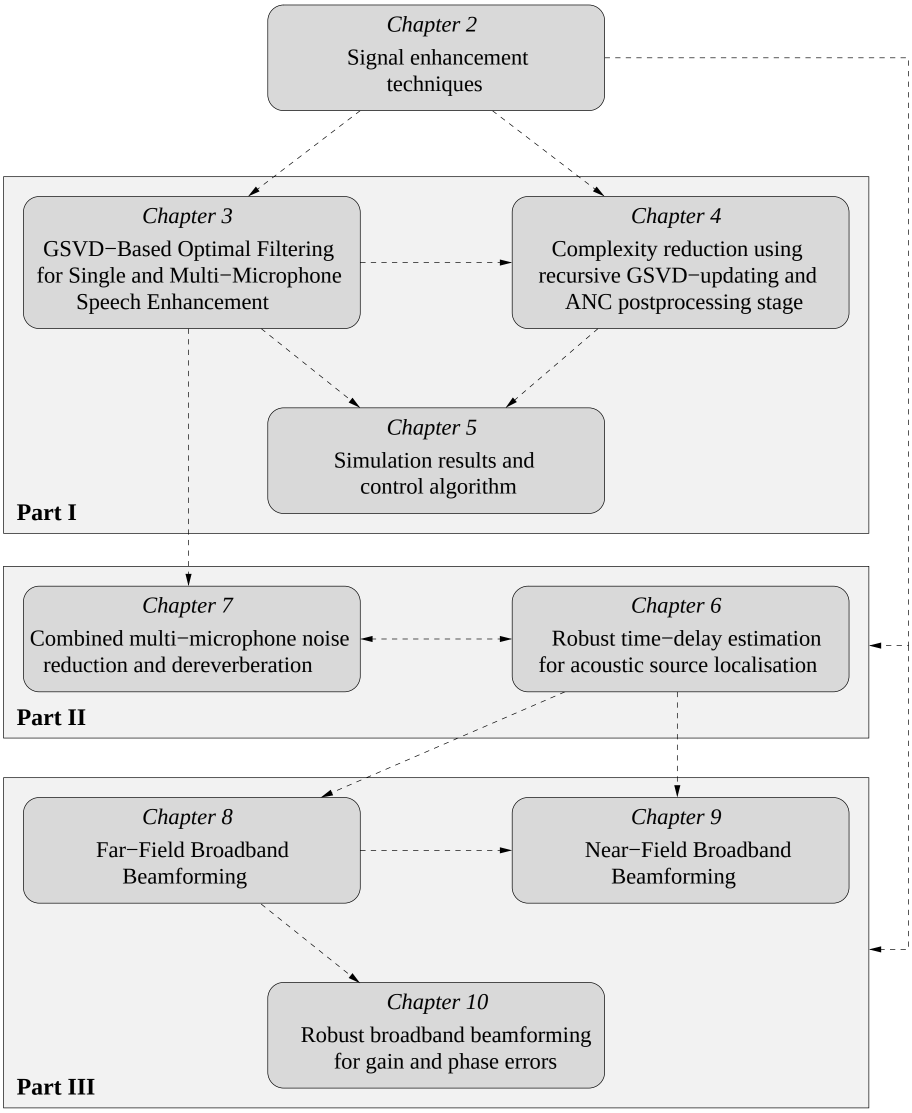

Chapter 2
Signal enhancement techniques
Chapter 3
GSVD−Based Optimal Filtering for Single and Multi−Microphone Speech Enhancement
Chapter 4
Complexity reduction using recursive GSVD−updating and ANC postprocessing stage
Chapter 5
Simulation results and control algorithm
Part I
Chapter 7
Combined multi−microphone noise reduction and dereverberation
Chapter 6
Robust time−delay estimation for acoustic source localisation
Part II
Chapter 8
Far−Field Broadband Beamforming
Chapter 9
Near−Field Broadband Beamforming
Chapter 10
Robust broadband beamforming for gain and phase errors
Part III

Figure 1.4: Schematic overview of the thesis

the low-rank model of the speech signal, this class of optimal filtering techniques can be considered a multi-microphone extension of the single-microphone subspace-based speech enhancement techniques. In practice, the optimal filter matrix can be computed using the GSVD of a speech and a noise data matrix. We derive a number of symmetry properties for this optimal filter matrix, which are valid for the white noise case as well as for the coloured noise case. In addition, the averaging step of some single-microphone subspace-based algorithms is examined, leading to the conclusion that this averaging operation is unnecessary and even suboptimal. When analysing the multi-channel Wiener filter in the frequency-domain, we show that this filter can be decomposed into a spectral and a spatial filtering term and we derive conditions under which the noise sensitivity of the GSC and the multi-channel Wiener filter are equal. Finally, we show that unconstrained optimal filtering can also be used for combined noise and echo reduction. When assuming infinite-length filters, we prove that the far-end echo source has no influence on the noise reduction performance.

In **Chapter 4** we present several techniques for reducing the computational complexity of the GSVD-based optimal filtering technique, such that it becomes suitable for real-time implementation. Instead of recomputing the GSVD from scratch at each time step, recursive Jacobi-type GSVD-updating algorithms can be used. The computational complexity can be further reduced by using a square root-free implementation and by using sub-sampling techniques. In addition, we show how to incorporate the GSVD-based optimal filtering technique in a GSC-type structure with an ANC postprocessing stage. This ANC postprocessing stage can either be used for increasing the noise reduction performance or for reducing the complexity without decreasing the performance.

Chapter 5 analyses the performance of the GSVD-based optimal filtering technique for several simulated acoustic environments and for real-life recordings. For higher filter lengths and for lower reverberation times, the SNR improvement increases and the speech distortion decreases. Simulations show that the SNR improvement of the GSVD-based optimal filtering technique outperforms the SNR improvement of standard fixed and adaptive beamforming techniques for all considered acoustic scenarios. In addition, robustness issues such as the effect of speech detection errors and deviations from the assumed signal model are analysed. It is shown that the SNR improvement is not degraded by speech detection errors but that speech distortion increases with increasing error rate. It is also shown that the GSVD-based optimal filter is more robust than the GSC for microphone mismatch, microphone displacement and look direction error.

Part II: Multi-Microphone Dereverberation and Source Localisation

In this part, multi-microphone algorithms are discussed for time-delay estimation (TDE)[1], dereverberation, and combined noise reduction and dereverberation. Since the presented TDE and dereverberation algorithms both require a (partial) estimate of the acoustic impulse responses, we also present batch and adaptive algorithms for (partially) estimating the acoustic impulse responses, both in the time-domain and in the frequency-domain. Publications related to

[1] The estimated time-delays between the microphone signals can be used for acoustic source localisation.

this part of the thesis are [55][57][66].

In **Chapter 6** two adaptive time-domain algorithms are presented for robust TDE in noisy and reverberant acoustic environments. We first discuss batch, i.e. non-adaptive, procedures for estimating the complete acoustic impulse responses in the time-domain. These batch estimation procedures are based on the generalised eigenvalue decomposition (GEVD) of the speech and the noise correlation matrices and form the basis for deriving adaptive algorithms. We extend a recently developed adaptive EVD algorithm for TDE to noisy environments, by using an adaptive GEVD or by pre-whitening the microphone signals. For the adaptive GEVD, we derive a stochastic gradient algorithm which iteratively estimates the generalised eigenvector corresponding to the smallest generalised eigenvalue. In addition, we extend all TDE algorithms to the case of more than two microphones. Simulations show that the time-delays can be estimated more robustly using the adaptive GEVD algorithm than using the adaptive EVD algorithm and the adaptive pre-whitening algorithm.

In **Chapter 7** frequency-domain algorithms for dereverberation and for combined noise reduction and dereverberation are discussed. We first present a frequency-domain technique for estimating the acoustic transfer functions when the microphone signals are corrupted by spatially coloured noise. Unlike the time-domain techniques presented in Chapter 6, this frequency-domain technique requires some prior knowledge about the acoustic transfer functions. Using the estimated transfer functions, dereverberation can be performed with a normalised matched filtering approach. In addition, we show that the MMSE estimate of the clean dereverberated speech signal can be obtained by dereverberating the MMSE estimates of the speech components in the microphone signals. Hence, by combining the normalised matched filter with the multi-channel Wiener filter, presented in Chapter 3, we obtain a combined noise reduction and dereverberation technique. Simulations show that this combined technique provides a trade-off between the noise reduction and the dereverberation objectives.

Part III: Broadband Beamformer Design

In this part several design procedures are discussed for designing fixed broadband beamformers with an arbitrary desired spatial directivity pattern for a given arbitrary microphone array configuration, using an FIR filter-and-sum structure. We will present far-field, near-field and mixed near-field far-field beamformers, and we will take into account robustness against errors in the microphone characteristics. Publications related to this part of the thesis are [58][59][60][64][65][62].

In **Chapter 8** we assume that the speech source is located in the far-field of the microphone array. Several cost functions are discussed for designing far-field broadband beamformers: a weighted least-squares (LS), a maximum energy array and a non-linear cost function. Although in general we would like to use the non-linear design procedure, this procedure gives rise to a high computational complexity, since it requires an iterative optimisation technique. Hence, we will also consider non-iterative design procedures with a lower computational complexity. We present two novel non-iterative cost functions, which are both based on eigenfilters. Eigenfilters have already been used for designing 1-D and 2-D linear-phase FIR filters and spatial filters; in this chapter we extend their usage to the design of broadband beamformers. In the conventional eigenfilter technique a reference frequency-angle point is required, whereas in the eigenfilter technique based on a Total Least Squares (TLS) error criterion, this reference point is not required. Simulations show that among all considered non-iterative design procedures the TLS eigenfilter technique has the best performance, i.e. best resembling the performance of the non-linear design procedure but having a significantly lower computational complexity.

In **Chapter 9** the design of near-field broadband beamformers is discussed. It is shown that the design of a near-field broadband beamformer operating at one specific distance is very similar to the design of a far-field broadband beamformer. The same design procedures and cost functions can be used, while the only difference lies in the calculation of the double integrals involved in the design. Since the spatial directivity pattern of a near-field broadband beamformer designed for one specific distance can be quite unsatisfactory at other distances, we present design procedures for broadband beamformers which operate at several distances, e.g. mixed near-field far-field design. Simulations show that for near-field broadband beamformer design the TLS eigenfilter technique again is the preferred non-iterative design procedure and that mixed near-field far-field design provides a trade-off between the near-field and the far-field performance.

In Chapters 8 and 9 it is assumed that the microphones are omni-directional microphones with a flat frequency response equal to 1. **Chapter 10** discusses the design of broadband beamformers that are robust against errors in the microphone array characteristics. First, we extend the broadband beamformer design procedures in case the microphone characteristics, i.e. frequency- and angle-dependent gain and phase, are exactly known. Since in many applications these characteristics are not known in practice and can even change over time, we present two procedures for designing broadband beamformers that are robust against (unknown) gain and phase errors. The first design procedure optimises the mean performance for all feasible microphone characteristics, requiring the gain and the phase probability density functions, whereas the second design procedure optimises the worst-case performance, leading to a minimax problem. Simulations show that robust broadband beamformer design gives rise to a significant performance improvement when gain and phase errors occur, especially when using small-size microphone arrays.

In **Chapter 11** we give an overall conclusion and we list some suggestions for further research.

1.6 Conclusions

In general, the speech and audio preprocessing market is an expanding market. People are requesting better sound quality, user-friendliness and interactivity in digital signal processing applications. This implies that there will be an ever growing demand for signal enhancement and signal conditioning techniques.

In Section 1.2 the advantages and the problems occurring in hands-free speech acquisition systems have been described. In hands-free systems people are allowed to walk around freely without wearing a headset or holding a microphone. Hence, several types of signal degradation (noise, echo, reverberation) occur in the microphone recordings. The need for adaptive and integrated multi-microphone speech enhancement techniques has been discussed in order to improve speech intelligibility and speech recognition performance and some important applications (hands-free mobile telephony, voice control, hearing aids) have been presented.

In Section 1.3 several important characteristics of the speech and the noise signals and the acoustic environment have been discussed, since these properties have a large influence on the signal degradation and on the performance of speech enhancement algorithms. It has been shown that acoustic environments can be highly time-varying and that the microphone array characteristics (gain, phase, position) should be taken into account in the algorithm design.

In Section 1.4 a brief overview has been given of existing single- and multi-microphone algorithms for noise reduction, echo cancellation and dereverberation. Some of these algorithms will be discussed in more mathematical detail in Chapter 2.

Section 1.5 presents an outline of the thesis and summarises the main contributions.

Chapter 2

Signal enhancement techniques

This chapter briefly discusses some single- and multi-microphone signal enhancement techniques for noise reduction and dereverberation that are important for the remainder of the text.

Section 2.1 discusses some signal processing basics. Section 2.2 describes the recording model for speech signals in a noisy acoustic environment, gives the general setup for multi-microphone speech enhancement in the time-domain and in the frequency-domain and defines some performance measures. Sections 2.3 and 2.4 discuss single-microphone speech enhancement techniques such as spectral subtraction, signal subspace-based and cepstrum-based techniques, whereas Sections 2.5 and 2.6 discuss multi-microphone speech enhancement techniques, such as beamforming, inverse filtering and matched filtering.

2.1 Signal processing basics

The following basic signal processing definitions and theorems can be found in any signal processing handbook, e.g. [42][213], but are repeated here for the sake of self-containedness. Linear algebra definitions and matrix decompositions are reviewed in Appendix A.

Most of the signals, filters and systems that are used in this thesis are *discrete-time* variables. They can be represented in the time-domain or in the frequency-domain. The *time-domain* representation of a variable x,

$$x[k] = \{\ \ldots\quad x[-1]\quad x[0]\quad x[1]\quad x[2]\quad \ldots\ \}\ , \tag{2.1}$$

depends on the discrete time k, which is related to the actual time $t = k/f_s$ through the sampling frequency f_s. The *frequency-domain* representation of $x[k]$ is obtained by applying the Discrete-Time Fourier Transform (DTFT),

$$X(\omega) = \mathcal{F}\{x[k]\} = \sum_{k=-\infty}^{\infty} x[k]\, e^{-jk\omega} \ . \tag{2.2}$$

The spectrum $X(\omega)$ is periodic in ω with period 2π. For the evaluation of the frequency-domain characteristics, the fundamental interval is usually considered, i.e. $-\pi < \omega \leq \pi$ or $0 \leq \omega < 2\pi$, with $\omega = \pi$ representing the Nyquist-frequency. The inverse transform, the Inverse Discrete-Time Fourier Transform (IDTFT), can be used to transform $X(\omega)$ back to the time-domain,

$$x[k] = \mathcal{F}^{-1}\{X(\omega)\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega)\, e^{jk\omega} \ . \tag{2.3}$$

In practice, the (continuous) spectrum $X(\omega)$ is approximated by considering frames of $x[k]$ with length L, multiplying these frames with a window function (e.g. rectangular, Hanning, Kaiser) of length L, and applying the Discrete Fourier Transform (DFT), also called short-time Fourier Transform (STFT). The lth component of the DFT of the mth frame of $x[k]$ is obtained as

$$X(l,m) = \sum_{k=0}^{L-1} w[k]\, x[mL+k]\, e^{-j2\pi kl/L} \ , \quad l = 0 \ldots L-1 \ , \tag{2.4}$$

with L the frame length and the size of the DFT and $w[k]$ the window function. In fact, this component is an approximation for $X(l\frac{2\pi}{L})$. Fast Fourier Transform (FFT) algorithms can be used for efficiently computing these DFT-components. The inverse transform is the Inverse Discrete Fourier Transform (IDFT),

$$x[mL+k] = \frac{1}{L} \sum_{l=0}^{L-1} X(l,m)\, e^{j2\pi kl/L} \ , \quad k = 0 \ldots L-1 \ , \tag{2.5}$$

which can be efficiently calculated using an Inverse Fast Fourier Transform (IFFT) algorithm.

The *autocorrelation function* $r_x[l]$ of a wide-sense stationary (WSS) stochastic process $x[k]$ is defined as

$$r_x[l] = \mathcal{E}\{x[k+l]x^*[k]\} \ , \tag{2.6}$$

with $\mathcal{E}\{\cdot\}$ the expectation operator. The *cross-correlation function* $r_{xy}[l]$ of two jointly WSS stochastic processes $x[k]$ and $y[k]$ is defined as

$$r_{xy}[l] = \mathcal{E}\{x[k+l]y^*[k]\} \ . \tag{2.7}$$

The *Power Spectral Density* (PSD) $P_x(\omega)$ of $x[k]$ is defined as the DTFT of the autocorrelation function $r_x[l]$ (i.e. the Wiener-Khintchine theorem), and the *Cross-Power Spectral Density* $P_{xy}(\omega)$ of $x[k]$ and $y[k]$ is defined as the DTFT of the cross-correlation function $r_{xy}[l]$. For ergodic processes (a property we will assume to be valid for all considered stochastic processes), $P_x(\omega)$ and $P_{xy}(\omega)$ can be computed as

$$\begin{aligned} P_x(\omega) &= \mathcal{E}\{|X(\omega)|^2\}\ , && (2.8)\\ P_{xy}(\omega) &= \mathcal{E}\{X(\omega)Y^*(\omega)\}\ . && (2.9) \end{aligned}$$

The *complex coherence* $\Gamma_{xy}(\omega)$ between the signals $x[k]$ and $y[k]$ is defined as

$$\Gamma_{xy}(\omega) = \frac{P_{xy}(\omega)}{\sqrt{P_x(\omega)P_y(\omega)}}\ . \tag{2.10}$$

The *Power Transfer Function* (PTF) $G_{xy}(\omega)$ between the (input) signal $x[k]$ and the (output) signal $y[k]$ is defined as the ratio of their PSDs,

$$G_{xy}(\omega) = \frac{P_y(\omega)}{P_x(\omega)}\ . \tag{2.11}$$

2.2 Problem statement

2.2.1 Recording model

The recording of a speech signal in a noisy environment can be described as follows. Consider an acoustic environment consisting of N microphones, one speech source and multiple background noise sources and far-end echo sources (see Fig. 1.1). Each microphone signal $y_n[k], n = 0 \ldots N-1$, at time k, consists of a filtered version of the clean speech signal $s[k]$ and additive noise,

$$\boxed{y_n[k] = h_n[k] \otimes s[k] + v_n[k] = x_n[k] + v_n[k]} \tag{2.12}$$

with $x_n[k]$ and $v_n[k]$ the speech and the noise component received at the nth microphone, $h_n[k]$ the acoustic impulse response (cf. Section 1.3.3) between the speech source and the nth microphone and $\otimes$ denoting convolution. Reverberation can thus be considered a convolutional noise source. Generally the acoustic impulse responses are modelled using FIR-filters $\mathbf{h}_n[k]$ of length K,

$$\mathbf{h}_n[k] = \begin{bmatrix} h_{n,0}[k] & h_{n,1}[k] & \ldots & h_{n,K-1}[k] \end{bmatrix}^T\ , \tag{2.13}$$

such that nth microphone signal at time k can be written as

$$y_n[k] = \sum_{i=0}^{K-1} h_{n,i}[k]\, s[k-i]\ . \tag{2.14}$$

In this model we assume that the microphones are perfectly omni-directional and have a flat frequency response equal to 1. If this is not the case, the microphone characteristics are usually incorporated in the acoustic impulse responses $h_n[k]$.

The additive noise component $v_n[k]$ in the nth microphone signal can be coloured and is assumed to be uncorrelated with the clean speech signal $s[k]$. This noise component actually consists of three parts:

- a contribution $v_n^u[k]$ from the *unknown noise sources* (e.g. ventilator, radio, wind, other people). In the case of localised noise sources this contribution can be written as

  $$v_n^u[k] = \sum_j h_{jn}^u[k] \otimes u_j[k] \ , \tag{2.15}$$

  with $u_j[k]$ the jth localised (unknown) noise source and $h_{jn}^u[k]$ the acoustic impulse response between the jth noise source and the nth microphone. This contribution also includes diffuse noise sources.

- a contribution $v_n^f[k]$ from the *known noise sources* (we will only consider far-end echo sources). This contribution can be written as

  $$v_n^f[k] = \sum_l h_{ln}^f[k] \otimes f_l[k] \ , \tag{2.16}$$

  with $f_l[k]$ the (known) loudspeaker signal of the lth echo source and $h_{ln}^f[k]$ the acoustic impulse response between the lth echo source and the nth microphone. In this model we assume that we can neglect the non-linear characteristics of the loudspeakers.

- independent (uncorrelated) recording noise $v_n^r[k]$. The level of this sensor noise depends on the type and the quality of the used microphones. In general we will not consider sensor noise.

2.2.2 Noise reduction and dereverberation

Figure 2.1 depicts a general setup for multi-microphone speech enhancement, where the microphone signals $y_n[k]$ are (adaptively) filtered with the filters $w_n[k]$ and are combined in order to obtain the enhanced signal $z[k]$,

$$z[k] = \sum_{n=0}^{N-1} w_n[k] \otimes y_n[k] \ . \tag{2.17}$$

The filters $w_n[k]$ generally are FIR filters $\mathbf{w}_n[k]$ of length L, i.e.

$$\mathbf{w}_n[k] = \begin{bmatrix} w_{n,0}[k] & w_{n,1}[k] & \dots & w_{n,L-1}[k] \end{bmatrix}^T \ , \tag{2.18}$$

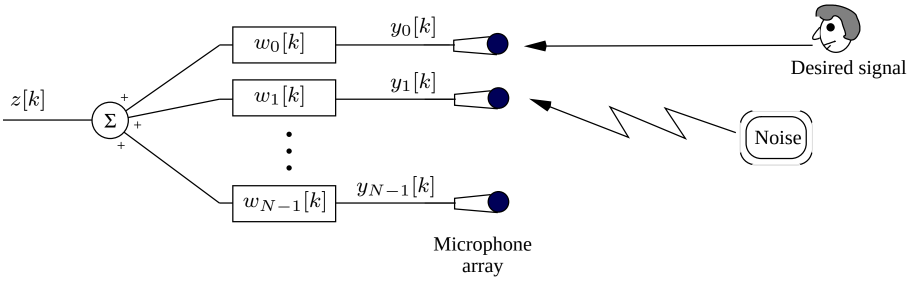


Figure 2.1: Multi-microphone filtering for speech enhancement

and if we consider the L-dimensional data vector $\mathbf{y}_n[k]$ and the M-dimensional stacked filter vector $\mathbf{w}[k]$ and stacked data vector $\mathbf{y}[k]$,

$$\begin{aligned} \mathbf{y}_n[k] &= \left[\begin{array}{cccc} y_n[k] & y_n[k-1] & \dots & y_n[k-L+1] \end{array} \right]^T \ , && (2.19) \\ \mathbf{w}[k] &= \left[\begin{array}{cccc} \mathbf{w}_0^T[k] & \mathbf{w}_1^T[k] & \dots & \mathbf{w}_{N-1}^T[k] \end{array} \right]^T \ , && (2.20) \\ \mathbf{y}[k] &= \left[\begin{array}{cccc} \mathbf{y}_0^T[k] & \mathbf{y}_1^T[k] & \dots & \mathbf{y}_{N-1}^T[k] \end{array} \right]^T \ , && (2.21) \end{aligned}$$

with $M = LN$, then the signal $z[k]$ at time k can be written as

$$\boxed{z[k] = \sum_{n=0}^{N-1} \mathbf{w}_n^T[k]\mathbf{y}_n[k] = \mathbf{w}^T[k]\mathbf{y}[k]} \qquad (2.22)$$

All signal enhancement algorithms discussed in this thesis can be described by this equation and 'merely' differ in the way the filters $w_n[k]$ are computed. By combining (2.12) and (2.17), the enhanced signal $z[k]$ can be written as a function of the speech and noise components,

$$z[k] = z_x[k] + z_v[k] = \underbrace{\sum_{n=0}^{N-1} w_n[k] \otimes h_n[k]}_{f[k]} \otimes s[k] + \sum_{n=0}^{N-1} w_n[k] \otimes v_n[k] \ , \qquad (2.23)$$

with $z_x[k]$ and $z_v[k]$ the speech and the noise component in the output signal $z[k]$ and $f[k]$ the total transfer function for the clean speech signal $s[k]$. The filters $w_n[k], n = 0 \dots N-1$ can be designed with different objectives in mind.

- The goal of *noise reduction* is to minimise the energy in the residual noise component $z_v[k]$, which can e.g. be achieved by putting $w_n[k] = 0, n = 0 \dots N-1$. Clearly, in this case also all speech components are removed, such that constraints should be introduced which take into account speech distortion. E.g. in adaptive beamformers (cf. Section 2.5.3) a constraint is imposed such that all signals coming from the direction of the speech source are not distorted, whereas in multi-channel Wiener filters (cf. Part I) a trade-off can be made between speech distortion and noise reduction.

- The goal of *dereverberation* is to compute the filters $w_n[k]$ such that the total speech transfer function $f[k]$ is equal to 1 (or more realistically a delay). However, since the residual noise component is not constrained in any way, it is even possible that the noise components $v_n[k]$ are amplified.
- The goal of *combined noise reduction and dereverberation* is to extract the clean speech $s[k]$ from the noisy microphone signals $y_n[k]$, i.e. the filters $w_n[k]$ should be designed such that both $f[k]$ approximates a delay and the energy of the residual noise component $z_v[k]$ is minimised.

One should keep in mind that only the noisy microphone signals $y_n[k]$ are available, i.e. neither the acoustic impulse responses $h_n[k]$ nor the noise components $v_n[k]$ are available and hence both should be estimated from the microphone signals. Moreover, both the acoustic impulse responses and the noise sources can be (highly) time-varying, necessitating the use of adaptive filters $w_n[k]$.

When a reference is available for the noise sources (as is the case for the far-end echo signals $f_l[k]$), these reference signals can also be used in the signal enhancement algorithms, and additional terms are added to (2.17),

$$z[k] = \sum_{n=0}^{N-1} w_n[k] \otimes y_n[k] + \sum_l w_l^f[k] \otimes f_l[k] \ , \tag{2.24}$$

with $w_l^f[k]$ the filters applied to the reference signals $f_l[k]$. This model will be used in Section 3.6, where a combined noise reduction and echo cancellation algorithm is discussed.

2.2.3 Frequency-domain representation

All previous expressions can also be represented in the frequency-domain[1]. Using (2.12), the microphone signal $Y_n(\omega), n = 0 \ldots N-1$, can be written as

$$\boxed{Y_n(\omega) = H_n(\omega)S(\omega) + V_n(\omega) = X_n(\omega) + V_n(\omega)} \tag{2.25}$$

with $H_n(\omega)$ the acoustic transfer function between the speech source and the nth microphone. The stacked vector of microphone signals can be written as

$$\mathbf{Y}(\omega) \quad = \quad \begin{bmatrix} Y_0(\omega) \\ Y_1(\omega) \\ \vdots \\ Y_{N-1}(\omega) \end{bmatrix} = \begin{bmatrix} H_0(\omega) \\ H_1(\omega) \\ \vdots \\ H_{N-1}(\omega) \end{bmatrix} S(\omega) + \begin{bmatrix} V_0(\omega) \\ V_1(\omega) \\ \vdots \\ V_{N-1}(\omega) \end{bmatrix} \tag{2.26}$$

[1] For this frequency-domain representation, we assume that all signals are stationary and the Discrete-Time Fourier Transform can be used.

$$= \mathbf{H}(\omega)S(\omega) + \mathbf{V}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega) \ . \tag{2.27}$$

The general expression (2.17) for multi-microphone speech enhancement can be written as

$$Z(\omega) = \sum_{n=0}^{N-1} W_n(\omega) Y_n(\omega) \ , \tag{2.28}$$

with $W_n(\omega) = \sum_{k=0}^{L-1} w_n[k] e^{-jk\omega}$. Using (2.27), the output signal $Z(\omega)$ can also be written as

$$Z(\omega) = \mathbf{W}^T(\omega)\mathbf{Y}(\omega) = \underbrace{\mathbf{W}^T(\omega)\mathbf{H}(\omega)}_{F(\omega)} S(\omega) + \mathbf{W}^T(\omega)\mathbf{V}(\omega) \ , \tag{2.29}$$

with $F(\omega)$ the total transfer function for the speech signal and

$$\mathbf{W}(\omega) = \left[\begin{array}{cccc} W_0(\omega) & W_1(\omega) & \ldots & W_{N-1}(\omega) \end{array} \right]^T \ . \tag{2.30}$$

For convenience, (2.29) is generally written as

$$\boxed{Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega) = \sum_{n=0}^{N-1} W_n^*(\omega) Y_n(\omega)} \tag{2.31}$$

which in the time-domain corresponds to a convolution with the time-reversed filters $w_n[-k]$.

2.2.4 Performance measures

Noise reduction performance

The noise reduction performance will be described by the improvement in unbiased signal-to-noise ratio (SNR) between the output signal $z[k]$ and the input signal, which is generally chosen to be the first microphone signal $y_0[k]$. The *unbiased SNR* of $z[k] = z_x[k] + z_v[k]$ is defined as

$$\boxed{\mathrm{SNR}_z = 10 \log_{10} \frac{\sum z_x^2[k]}{\sum z_v^2[k]}} \tag{2.32}$$

which is computed during speech-and-noise periods. The unbiased SNR *improvement* is then given as the difference between the unbiased SNR of the input and the output signal, i.e. $\mathrm{SNR}_z - \mathrm{SNR}_{y_0}$.

In the frequency-domain, the unbiased SNR is defined as

$$\boxed{\mathrm{SNR}_z(\omega) = 10 \log_{10} \frac{P_{z_x}(\omega)}{P_{z_v}(\omega)}} \tag{2.33}$$

with $P_{z_x}(\omega)$ and $P_{z_v}(\omega)$ the PSD of the speech and the noise components of $z[k]$ (cf. Section 2.1). Using (2.31), the PSD $P_{z_x}(\omega)$ can be written as

$$P_{z_x}(\omega) = \mathcal{E}\{|Z_x(\omega)|^2\} = P_s(\omega)|\mathbf{W}^H(\omega)\mathbf{H}(\omega)|^2 \;, \tag{2.34}$$

with $P_s(\omega) = \mathcal{E}\{|S(\omega)|^2\}$, while the PSD $P_{z_v}(\omega)$ can be written as

$$P_{z_v}(\omega) = \mathcal{E}\{|Z_v(\omega)|^2\} = \mathbf{W}^H(\omega)\bar{\mathbf{R}}_{vv}(\omega)\mathbf{W}(\omega) \;, \tag{2.35}$$

with

$$\bar{\mathbf{R}}_{vv}(\omega) = \mathcal{E}\{\mathbf{V}(\omega)\mathbf{V}^H(\omega)\} = \begin{bmatrix} P_{v_0}(\omega) & P_{v_0v_1}(\omega) & \dots & P_{v_0v_{N-1}}(\omega) \\ P_{v_1v_0}(\omega) & P_{v_1}(\omega) & \dots & P_{v_1v_{N-1}}(\omega) \\ \vdots & \vdots & & \vdots \\ P_{v_{N-1}v_0}(\omega) & P_{v_{N-1}v_1}(\omega) & \dots & P_{v_{N-1}}(\omega) \end{bmatrix} . \tag{2.36}$$

If we assume that the noise-field is homogeneous, i.e. $P_{v_n}(\omega) = P_v(\omega)$, $n = 0 \dots N-1$, then $\bar{\mathbf{R}}_{vv}(\omega)$ can be written as a function of the noise coherence matrix $\mathbf{\Gamma}_v(\omega)$, cf. Section 2.1, i.e.

$$\bar{\mathbf{R}}_{vv}(\omega) = P_v(\omega)\mathbf{\Gamma}_v(\omega) = P_v(\omega) \begin{bmatrix} 1 & \Gamma_{v_0v_1}(\omega) & \dots & \Gamma_{v_0v_{N-1}}(\omega) \\ \Gamma_{v_1v_0}(\omega) & 1 & \dots & \Gamma_{v_1v_{N-1}}(\omega) \\ \vdots & \vdots & & \vdots \\ \Gamma_{v_{N-1}v_0}(\omega) & \Gamma_{v_{N-1}v_1}(\omega) & \dots & 1 \end{bmatrix} , \tag{2.37}$$

such that the unbiased output SNR in (2.33) can be written as

$$\mathrm{SNR}_z(\omega) = 10\log_{10}\frac{P_s(\omega)}{P_v(\omega)}\frac{|\mathbf{W}^H(\omega)\mathbf{H}(\omega)|^2}{\mathbf{W}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}(\omega)} \;. \tag{2.38}$$

Speech distortion

Speech distortion can be analysed by considering the PTF between the speech component of the input signal and the speech component of the output signal $z_x[k]$. For the GSVD-based optimal filtering technique (cf. Section 3), the input signal is the first microphone signal $y_0[k]$, such that speech distortion will be analysed by considering the PTF

$$G_{x_0z_x}(\omega) = \frac{P_{z_x}(\omega)}{P_{x_0}(\omega)} = \frac{|\mathbf{W}^H(\omega)\mathbf{H}(\omega)|^2}{|H_0(\omega)|^2} \;. \tag{2.39}$$

The average *speech distortion* (SD) is computed as the average of the PTF (in dB) over the full frequency band, i.e.

$$\boxed{SD = \frac{1}{2\pi}\int_{-\pi}^{\pi} |10\log_{10} G_{x_0z_x}(\omega)|\, d\omega} \tag{2.40}$$

Noise reduction can also be analysed by considering the PTF between the noise components of the input and the output signal, i.e.

$$G_{v_0 z_v}(\omega) = \frac{P_{z_v}(\omega)}{P_{v_0}(\omega)} = \mathbf{W}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}(\omega) \ . \tag{2.41}$$

Dereverberation

It is quite difficult to define a good performance measure for dereverberation. We will analyse the dereverberation performance by considering the PTF between the clean speech signal $s[k]$ and the speech component of the output signal $z_x[k]$, i.e.

$$G_{sz_x}(\omega) = \frac{P_{z_x}(\omega)}{P_s(\omega)} = |\mathbf{W}^H(\omega)\mathbf{H}(\omega)|^2 \ . \tag{2.42}$$

The *dereverberation index* (DI) is computed as the average of the PTF (in dB) over the full frequency band, i.e.

$$\boxed{DI = \frac{1}{2\pi}\int_{-\pi}^{\pi} |20\log_{10}|\mathbf{W}^H(\omega)\mathbf{H}(\omega)||\, d\omega} \tag{2.43}$$

We will use this performance measure, although it does not give full information about the amount of reverberation present in the output signal (i.e. even if $DI = 0\,\text{dB}$, it is still possible that $z_x[k] \neq s[k]$, since phase information is neglected in computing the dereverberation index).

2.3 Single-microphone noise reduction

In single-microphone speech enhancement the number of microphones $N = 1$, such that the model (2.12) simplifies to

$$y_0[k] = x_0[k] + v_0[k] \ . \tag{2.44}$$

As already mentioned in Section 1.4.1, single-channel noise reduction techniques can only exploit the temporal and the spectral information of the speech and the noise signals. In this section we will briefly describe two non-parametric techniques, which require a noise fingerprint in the Discrete Fourier Transform (DFT) or the Karhunen-Loève Transform (KLT) domain.

2.3.1 Spectral subtraction techniques

Spectral subtraction is a well-known and commonly used single-microphone speech enhancement technique. Transforming (2.44) to the frequency-domain using an L-point short-time Fourier transform, cf. (2.4), gives

$$Y_0(l,m) = X_0(l,m) + V_0(l,m), \quad l = 0 \ldots L-1 \ , \tag{2.45}$$

Magnitude subtraction [21]	$W(l,m) = \left[1 - \frac{\lvert\mu_V(l)\rvert}{\lvert Y_0(l,m)\rvert}\right]$
Power subtraction [163][177]	$W(l,m) = \sqrt{1 - \frac{\lvert\mu_V(l)\rvert^2}{\lvert Y_0(l,m)\rvert^2}}$
Wiener estimation [163][177]	$W(l,m) = 1 - \frac{\lvert\mu_V(l)\rvert^2}{\lvert Y_0(l,m)\rvert^2}$
Generalised Spectral subtraction [163]	$W(l,m) = \left[1 - \alpha\left(\frac{\lvert\mu_V(l)\rvert}{\lvert Y_0(l,m)\rvert}\right)^{\gamma}\right]^{\beta}$

Table 2.1: Common gain functions for spectral subtraction

with m the frame index. Typically frame lengths of $20 - 30\,\text{ms}$ are used and $50 - 66\%$ overlap between the frames is taken [42]. For each frame, the clean speech spectrum $X_0(l,m)$ is estimated from the noisy speech spectrum $Y_0(l,m)$ using an estimate of the noise spectrum $\mu_V(l)$, i.e.

$$\mu_V(l) = \mathcal{E}\{V_0(l,m)\}, \quad l = 0 \ldots L-1 \, . \tag{2.46}$$

This noise fingerprint can be calculated by assuming that the noise characteristics change slowly over time, such that the noise spectrum can be estimated by averaging over the spectra of noise-only frames. This estimation procedure obviously requires a voice activity detection (VAD) algorithm (cf. Section 5.3), which classifies the frames into noise-only and speech-and-noise frames.

The estimate of the clean speech spectrum $\hat{X}_0(l,m)$ is obtained by multiplying the noisy speech spectrum $Y_0(l,m)$ with a gain function $W(l,m)$, i.e.

$$\hat{X}_0(l,m) = W(l,m) Y_0(l,m) \, , \tag{2.47}$$

where the gain function $W(l,m)$ depends on the noisy speech spectrum and on the estimated noise spectrum, i.e.

$$W(l,m) = f(Y_0(l,m), \mu_V(l)) \, . \tag{2.48}$$

Hence, all spectral subtraction techniques can be considered frequency-domain techniques in which the STFT coefficients are multiplied with a noise-dependent gain. Several gain functions have been proposed in the literature, of which some common functions are listed in Table 2.1 [45]. More complicated functions include non-linear gain functions [280] or the Ephraim-Malah gain functions [83][84], which make a minimum mean square error (MMSE) estimate of the amplitude of the clean speech spectrum in the spectral or in the log-spectral domain and which are frequently used in practice. Other spectral subtraction techniques incorporate properties of the human auditory system [268] or try to estimate the noise spectrum even during speech-and-noise periods [172].

In all frames it is however possible that for some frequencies the estimated amplitude of the noise spectrum $|\mu_V(l)|$ is larger than the instantaneous amplitude of the noisy speech spectrum $|Y_0(l,m)|$. Since this could lead to negative estimates for the amplitude of the clean speech spectrum $|\hat{X}_0(l,m)|$, for these frequencies the gain function $W(l,m)$ is usually put to zero (i.e. half-wave rectification [21]) or equal to a small noise floor value [268]. However, because of the non-stationary character of the speech signal, this non-linear rectification mapping leads to a specific kind of residual noise, called musical noise, which consists of short-lived tones with randomly distributed frequencies. Different techniques have been proposed to eliminate this annoying residual noise, e.g. by averaging the (instantaneous) noisy speech spectrum over a number of frames, by augmenting the gain function with a soft-decision VAD [83][177] or by using non-linear spectral subtraction techniques [165][279][280].

2.3.2 Signal subspace-based techniques

Recently, several single-microphone signal subspace-based speech enhancement techniques for additive (coloured) noise have been proposed. These techniques are based on a (generalised) singular value decomposition (SVD) [43][49][121] [138], which is also referred to as Karhunen-Loève transform (KLT) [85][130] [179][220]. The main idea is to consider the noisy signal as a vector in an L-dimensional vector space and to separate this space into 2 orthogonal subspaces: the *signal subspace* (with dimension smaller than L, corresponding to the clean signal) and the *noise subspace*, i.e. the orthogonal complement of the signal subspace. Of course, this separation is only possible if the clean signal can be modelled with a low-rank model (cf. Section 1.3.1). Signal enhancement is performed by removing the noise subspace and by estimating the clean speech signal from the remaining signal subspace. Similar subspace-based signal enhancement techniques have also been used in other applications, e.g. biomedical and image processing applications, where the signals can be modelled as the sum of a finite number of complex exponentials [69][70][247].

Signal subspace-based speech enhancement techniques can be *classified* according to the noise assumptions (white noise vs. coloured noise), type of estimate (least-squares, minimum variance, perceptually relevant criterion), type of processing (block-based vs. adaptive) and on whether an additional averaging step is performed or not.

In this section we will discuss two techniques, described in [85] and in [138], which use a (slightly) different approach, but result in practically the same algorithm. In [85] a statistical approach is followed, where the clean speech data vector $\mathbf{x}_0[k]$ is estimated from the noisy data vector $\mathbf{y}_0[k]$ using the speech and the noise correlation matrices. In [138] a deterministic approach is followed, where the clean speech data matrix $\mathbf{X}_0[k]$ is estimated from the noisy data matrix $\mathbf{Y}_0[k]$. We will discuss two linear signal estimators for both approaches: *least-squares (LS) estimation* and *minimum-variance (MV) estimation*. In [85] another interesting linear estimator is described which trades off speech distortion and noise distortion. This estimator will be discussed in Section 3.2.3, where the multi-microphone case is presented.

Data model: statistical approach

Consider the L-dimensional data vectors $\mathbf{y}_0[k]$, $\mathbf{x}_0[k]$ and $\mathbf{v}_0[k]$, with the speech and the noise data vectors $\mathbf{x}_0[k]$ and $\mathbf{v}_0[k]$ similarly defined as in (2.19). Using (2.44), we can write

$$\mathbf{y}_0[k] = \mathbf{x}_0[k] + \mathbf{v}_0[k] \ . \tag{2.49}$$

We assume that each speech data vector $\mathbf{x}_0[k]$ can be represented as a linear combination of R linearly independent L-dimensional basis vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_R\}$, with $R < L$ (cf. Section 1.3.1), i.e.

$$\mathbf{x}_0[k] = \sum_{i=1}^{R} \mathbf{x}_i a_i[k] \ = \mathbf{X}_R \, \mathbf{a}_R[k] \ , \tag{2.50}$$

with $\mathbf{X}_R$ an $L \times R$-dimensional matrix and $\mathbf{a}_R[k]$ an R-dimensional vector[2]. When $R < L$, the set of all possible signal vectors $\mathbf{x}_0[k]$ lies in a subspace of the Euclidean space $\mathbb{R}^L$ spanned by the columns of $\mathbf{X}_R$. This subspace is referred to as the *signal subspace.* The $L \times L$-dimensional correlation matrix $\bar{\mathbf{R}}_{xx}[k]$ of the speech signal $\mathbf{x}_0[k]$ is equal to

$$\bar{\mathbf{R}}_{xx}[k] = \mathcal{E}\{\mathbf{x}_0[k]\mathbf{x}_0^T[k]\} \ , \tag{2.51}$$

which can be written, using (2.50), as $\bar{\mathbf{R}}_{xx}[k] = \mathbf{X}_R \bar{\mathbf{R}}_a[k] \mathbf{X}_R^T$, with $\bar{\mathbf{R}}_a[k] = \mathcal{E}\{\mathbf{a}_R[k] \, \mathbf{a}_R^T[k]\}$ the $R \times R$-dimensional correlation matrix of the vector $\mathbf{a}_R[k]$. Hence, $L - R$ eigenvalues of $\bar{\mathbf{R}}_{xx}[k]$ are equal to zero. The correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ are similarly defined as in (2.51). The noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$ is assumed to be known, but should not necessarily be equal to $\sigma_v^2 \, \mathbf{I}_L$, i.e. no white Gaussian noise assumption is made here. Since the noise correlation matrix is assumed to be positive definite, noise vectors have a component in the signal subspace as well as in the complement of the signal subspace, which is referred to as the *noise subspace.*

The generalised eigenvalue decomposition (GEVD) of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ is equal to (cf. Appendix A.2)

$$\left\{ \begin{array}{lcl} \bar{\mathbf{R}}_{yy}[k] & = & \bar{\mathbf{Q}} \, \bar{\mathbf{\Lambda}}_y \bar{\mathbf{Q}}^T \\ \bar{\mathbf{R}}_{vv}[k] & = & \bar{\mathbf{Q}} \, \bar{\mathbf{\Lambda}}_v \bar{\mathbf{Q}}^T \ , \end{array} \right. \tag{2.52}$$

with $\bar{\mathbf{Q}}$ an $L \times L$-dimensional invertible, but not necessarily orthogonal, matrix and $\bar{\mathbf{\Lambda}}_y = \text{diag}\{\bar{\sigma}_i^2\}$, $i = 1 \ldots L$, and $\bar{\mathbf{\Lambda}}_v = \text{diag}\{\bar{\eta}_i^2\}$, $i = 1 \ldots L$. Since the speech and the noise components are assumed to be uncorrelated, the correlation matrix $\bar{\mathbf{R}}_{xx}[k]$ can be written using (2.52) as

$$\bar{\mathbf{R}}_{xx}[k] = \bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k] = \bar{\mathbf{Q}} \left(\bar{\mathbf{\Lambda}}_y - \bar{\mathbf{\Lambda}}_v \right) \bar{\mathbf{Q}}^T \ . \tag{2.53}$$

[2] Typical values for R range from 12 to 20 (cf. Section 1.3.1), while typical values for L range from 20 to 80.

Since $\bar{\mathbf{R}}_{xx}[k]$ has rank R and all correlation matrices are assumed to be positive semi-definite, it readily follows that

$$\left\{ \begin{array}{lllr} \bar{\sigma}_i^2 & > & \bar{\eta}_i^2 & i = 1 \ldots R \,, \\ \bar{\sigma}_i^2 & = & \bar{\eta}_i^2 & i = R+1 \ldots L \,, \end{array} \right. \tag{2.54}$$

such that $\bar{\mathbf{R}}_{xx}[k]$ can be decomposed as

$$\bar{\mathbf{R}}_{xx}[k] = \left[\begin{array}{cc} \bar{\mathbf{Q}}_1 & \bar{\mathbf{Q}}_2 \end{array} \right] \left[\begin{array}{cc} \bar{\mathbf{\Lambda}}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] \left[\begin{array}{c} \bar{\mathbf{Q}}_1^T \\ \bar{\mathbf{Q}}_2^T \end{array} \right] = \bar{\mathbf{Q}}_1 \bar{\mathbf{\Lambda}}_x \bar{\mathbf{Q}}_1^T \,, \tag{2.55}$$

with $\bar{\mathbf{Q}}_1$ an $L \times R$-dimensional matrix, whose columns span the signal subspace, and $\bar{\mathbf{Q}}_2$ an $L \times (L-R)$-dimensional matrix. Note that $\bar{\mathbf{Q}}_2$ is not necessarily orthogonal to $\bar{\mathbf{Q}}_1$, such that the columns of $\bar{\mathbf{Q}}_2$ do not necessarily span the noise subspace, which is defined as the orthogonal complement of the signal subspace. Only in the white noise case, i.e. $\bar{\mathbf{R}}_{vv}[k] = \sigma_v^2 \mathbf{I}_L$ with σ_v^2 the noise power, the matrix $\bar{\mathbf{Q}}$ is an orthogonal matrix.

Data model: deterministic approach

Instead of using correlation matrices, we can also consider data matrices. Consider the $P \times L$ Toeplitz data matrix $\mathbf{X}_0[k]$ (with $P > L$)[3],

$$\mathbf{X}_0[k] = \left[\begin{array}{llll} x_0[k-P+1] & x_0[k-P] & \ldots & x_0[k-P-L+2] \\ \vdots & \vdots & & \vdots \\ x_0[k-1] & x_0[k-2] & \ldots & x_0[k-L] \\ x_0[k] & x_0[k-1] & \ldots & x_0[k-L+1] \end{array} \right] , \tag{2.56}$$

$$= \left[\begin{array}{l} \mathbf{x}_0^T[k-P+1] \\ \vdots \\ \mathbf{x}_0^T[k-1] \\ \mathbf{x}_0^T[k] \end{array} \right] = \left[\begin{array}{l} \mathbf{a}_R^T[k-P+1] \\ \vdots \\ \mathbf{a}_R^T[k-1] \\ \mathbf{a}_R^T[k] \end{array} \right] \mathbf{X}_R^T \,, \tag{2.57}$$

which clearly is rank-deficient (rank R), independent of the exact type of linear model that is used in (2.50). One can also choose to work with Hankel matrices instead of Toeplitz matrices [138], but there are no fundamental differences. Using (2.44), we can write

$$\mathbf{Y}_0[k] = \mathbf{X}_0[k] + \mathbf{V}_0[k] \,, \tag{2.58}$$

with $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$ Toeplitz data matrices, similarly defined as in (2.56). We will assume that $\mathbf{V}_0[k]$ and $\mathbf{Y}_0[k]$ are full-rank matrices (rank L). In addition, we assume that the speech and the noise matrices are orthogonal, i.e. $\mathbf{X}_0^T[k]\,\mathbf{V}_0[k] = \mathbf{0}$.

[3] Since short-time stationarity of speech is in the order of $20-30$ msec (cf. Section 1.3.1), typical values for P range from 300 to 500 ($f_s = 16$ kHz).

The generalised singular value decomposition (GSVD) of the noisy speech and the noise data matrices $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$ is equal to (cf. Appendix A.2),

$$\begin{cases} \mathbf{Y}_0[k] &= \mathbf{U}_Y \, \mathbf{\Sigma}_Y \, \mathbf{Q}^T \\ \mathbf{V}_0[k] &= \mathbf{U}_V \, \mathbf{\Sigma}_V \, \mathbf{Q}^T \ , \end{cases} \tag{2.59}$$

with $\mathbf{U}_Y$ and $\mathbf{U}_V$ $P\times L$-dimensional orthogonal matrices, $\mathbf{Q}$ an $L\times L$-dimensional invertible, but not necessarily orthogonal, matrix and $\mathbf{\Sigma}_Y = \text{diag}\{\sigma_i\}$, $i = 1\dots L$, and $\mathbf{\Sigma}_V = \text{diag}\{\eta_i\}$, $i = 1\dots L$. If $P \to \infty$ and assuming stationarity for the considered signals, the generalised singular vectors and singular values of $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$ converge to the generalised eigenvectors and eigenvalues of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$. Using (2.59) and the orthogonality assumption, we can write

$$\mathbf{X}_0^T\mathbf{X}_0 = \mathbf{Y}_0^T\mathbf{Y}_0 - \mathbf{V}_0^T\mathbf{V}_0 = \mathbf{Q}(\mathbf{\Sigma}_Y^2 - \mathbf{\Sigma}_V^2)\mathbf{Q}^T \ . \tag{2.60}$$

Since $\mathbf{X}_0$ has rank R, it again follows that $\sigma_i > \eta_i$, $i = 1\dots R$, and $\sigma_i = \eta_i$, $i = R+1\dots L$. Hence, the GSVD of $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$ can be rewritten as

$$\begin{cases} \mathbf{Y}_0[k] &= \begin{bmatrix} \mathbf{U}_{Y1} & \mathbf{U}_{Y2} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{Y1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{V2} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \\ \mathbf{V}_0[k] &= \begin{bmatrix} \mathbf{U}_{V1} & \mathbf{U}_{V2} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{V1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{V2} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} , \end{cases} \tag{2.61}$$

with $\mathbf{Q}_1$ an $L\times R$-dimensional matrix, whose columns span the signal subspace, and $\mathbf{Q}_2$ an $L\times(L-R)$-dimensional matrix (not necessarily orthogonal to $\mathbf{Q}_1$). Only in the white noise case, the matrix $\mathbf{Q}$ is an orthogonal matrix.

In the deterministic approach, we also require the QR-decomposition of the noise data matrix $\mathbf{V}_0[k]$, which is defined as (cf. Appendix A.2)

$$\mathbf{V}_0[k] = \mathbf{Q}_V \, \mathbf{R}_V \ , \tag{2.62}$$

with $\mathbf{Q}_V$ a $P\times L$-dimensional orthogonal matrix and $\mathbf{R}_V$ an $L\times L$-dimensional upper-triangular matrix.

Least-squares (LS) estimation

Deterministic approach: in [138] the least-squares estimation problem is formulated as estimating the clean speech data matrix $\mathbf{X}_0[k]$ from $\mathbf{Y}_0[k]$ by approximating the pre-whitened matrix $\mathbf{Y}_0[k]\,\mathbf{R}_V^{-1}$ by a pre-whitened matrix $\hat{\mathbf{X}}_0^{LS}[k]\,\mathbf{R}_V^{-1}$ of rank R, i.e.

$$\boxed{\min_{\text{rank}(\hat{\mathbf{X}}_0^{LS}[k])=R} ||\big(\mathbf{Y}_0[k] - \hat{\mathbf{X}}_0^{LS}[k]\big)\cdot\mathbf{R}_V^{-1}||_F^2} \tag{2.63}$$

Using (2.62) and (2.59), the matrix $\mathbf{R}_V^{-1}$ can be written as

$$\mathbf{R}_V^{-1} = \big(\mathbf{Q}_V^T\,\mathbf{V}_0[k]\big)^{-1} = \mathbf{Q}^{-T}\mathbf{\Sigma}_V^{-1}\big(\mathbf{Q}_V^T\,\mathbf{U}_V\big)^{-1} = \mathbf{Q}^{-T}\mathbf{\Sigma}_V^{-1}\big(\mathbf{Q}_V^T\,\mathbf{U}_V\big)^T \ , \tag{2.64}$$

since $\mathbf{Q}_V^T\,\mathbf{U}_V$ is an orthogonal matrix (cf. Appendix B.1). Using (2.59) and (2.64), the pre-whitened matrix $\mathbf{Y}_0[k]\,\mathbf{R}_V^{-1}$ can hence be written as

$$\mathbf{Y}_0[k]\,\mathbf{R}_V^{-1} = \mathbf{U}_Y \cdot \mathbf{\Sigma}_Y \mathbf{\Sigma}_V^{-1} \cdot \left(\mathbf{Q}_V^T\,\mathbf{U}_V\right)^T , \tag{2.65}$$

which is the singular value decomposition (SVD) of $\mathbf{Y}_0[k]\,\mathbf{R}_V^{-1}$. From this equation it can be seen that the singular values of $\mathbf{Y}_0[k]\,\mathbf{R}_V^{-1}$ are equal to the generalised singular values of $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$. The rank-R approximation $\hat{\mathbf{X}}_0^{LS}[k]\,\mathbf{R}_V^{-1}$ of $\mathbf{Y}_0[k]\,\mathbf{R}_V^{-1}$ in the LS sense is obtained by putting the $L - R$ smallest (generalised) singular values in (2.65) to zero [110][228], i.e.

$$\hat{\mathbf{X}}_0^{LS}[k]\,\mathbf{R}_V^{-1} = \mathbf{U}_Y \begin{bmatrix} \mathbf{\Sigma}_{Y1}\mathbf{\Sigma}_{V1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \left(\mathbf{Q}_V^T\,\mathbf{U}_V\right)^T , \tag{2.66}$$

such that using (2.64)

$$\begin{aligned} \hat{\mathbf{X}}_0^{LS}[k] &= \mathbf{U}_Y \begin{bmatrix} \mathbf{\Sigma}_{Y1}\mathbf{\Sigma}_{V1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \left(\mathbf{Q}_V^T\,\mathbf{U}_V\right)^{-1} \left(\mathbf{Q}_V^T\,\mathbf{U}_V\right)\mathbf{\Sigma}_V\mathbf{Q}^T & (2.67) \\ &= \mathbf{U}_Y \begin{bmatrix} \mathbf{\Sigma}_{Y1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^T = \mathbf{U}_{Y1}\mathbf{\Sigma}_{Y1}\mathbf{Q}_1^T\,. & (2.68) \end{aligned}$$

Hence, the LS approximation $\hat{\mathbf{X}}_0^{LS}[k]$ is obtained by truncating the GSVD in (2.61) to order R, and can also be written as

$$\boxed{\hat{\mathbf{X}}_0^{LS}[k] = \mathbf{Y}_0[k] \cdot \mathbf{Q}^{-T} \begin{bmatrix} \mathbf{I}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^T} \tag{2.69}$$

Statistical approach: in [85], the LS estimation problem is similarly formulated using the pre-whitened data vectors $\bar{\mathbf{R}}_{vv}^{-T/2}[k]\,\mathbf{y}_0[k]$, with $\bar{\mathbf{R}}_{vv}^{T/2}[k]$ equal to the square-root factor in the Cholesky-decomposition of $\bar{\mathbf{R}}_{vv}[k]$, i.e.

$$\bar{\mathbf{R}}_{vv}[k] = \bar{\mathbf{R}}_{vv}^{T/2}[k]\,\bar{\mathbf{R}}_{vv}^{1/2}[k]\,. \tag{2.70}$$

The LS estimation problem is then formulated as estimating the vector $\hat{\mathbf{x}}_0^{LS}[k]$ in the signal subspace (i.e. calculating the coefficients $\hat{\mathbf{a}}_R[k]$) from $\mathbf{y}_0[k]$ using the optimisation problem

$$\boxed{\min_{\hat{\mathbf{x}}_0^{LS}[k]=\mathbf{X}_R\hat{\mathbf{a}}_R[k]} ||\bar{\mathbf{R}}_{vv}^{-T/2}[k]\left(\mathbf{y}_0[k] - \hat{\mathbf{x}}_0^{LS}[k]\right)||^2} \tag{2.71}$$

which is equivalent to

$$\min_{\hat{\mathbf{a}}_R[k]} ||\bar{\mathbf{R}}_{vv}^{-T/2}[k]\left(\mathbf{y}_0[k] - \mathbf{X}_R\hat{\mathbf{a}}_R[k]\right)||^2\,, \tag{2.72}$$

where the signal subspace $\mathbf{X}_R$ is assumed to be known. The estimated coefficients $\hat{\mathbf{a}}_R[k]$ are given by

$$\hat{\mathbf{a}}_R[k] = \left(\mathbf{X}_R^T\bar{\mathbf{R}}_{vv}^{-1}[k]\mathbf{X}_R\right)^{-1}\mathbf{X}_R^T\bar{\mathbf{R}}_{vv}^{-1}[k]\,\mathbf{y}_0[k]\,, \tag{2.73}$$

such that

$$\hat{\mathbf{x}}_0^{LS}[k] = \mathbf{X}_R\hat{\mathbf{a}}_R[k] = \mathbf{X}_R\left(\mathbf{X}_R^T\bar{\mathbf{R}}_{vv}^{-1}[k]\mathbf{X}_R\right)^{-1}\mathbf{X}_R^T\bar{\mathbf{R}}_{vv}^{-1}[k]\,\mathbf{y}_0[k]\,. \tag{2.74}$$

The signal subspace is also spanned by the columns of $\bar{\mathbf{Q}}_1$, cf. (2.55), which can therefore be written as $\bar{\mathbf{Q}}_1 = \mathbf{X}_R\,\mathbf{T}$, with $\mathbf{T}$ an $R\times R$-dimensional full-rank matrix. Hence, $\hat{\mathbf{x}}_0^{LS}[k]$ is also equal to

$$\hat{\mathbf{x}}_0^{LS}[k] = \bar{\mathbf{Q}}_1\left(\bar{\mathbf{Q}}_1^T\bar{\mathbf{R}}_{vv}^{-1}[k]\bar{\mathbf{Q}}_1\right)^{-1}\bar{\mathbf{Q}}_1^T\bar{\mathbf{R}}_{vv}^{-1}[k]\,\mathbf{y}_0[k]\,. \tag{2.75}$$

It can be proved that this expression can be simplified to

$$\boxed{\hat{\mathbf{x}}_0^{LS}[k] = \left[\begin{array}{cc}\bar{\mathbf{Q}}_1 & \mathbf{0}\end{array}\right]\bar{\mathbf{Q}}^{-1}\,\mathbf{y}_0[k] = \bar{\mathbf{Q}}\left[\begin{array}{cc}\mathbf{I}_R & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\bar{\mathbf{Q}}^{-1}\,\mathbf{y}_0[k]} \tag{2.76}$$

yielding a similar result as (2.69), with the generalised eigenvectors $\bar{\mathbf{Q}}$ instead of the generalised singular vectors $\mathbf{Q}$. In the white noise case, this operation boils down to projecting the noisy data vectors onto the signal subspace.

When using the LS estimator, no speech distortion is introduced, but the amount of noise reduction achieved may be limited. A practical problem that also arises is the proper determination of the rank R, which can change from frame to frame in speech signals. This rank determination is not required in the minimum-variance estimator (or one of its variants).

Minimum-variance (MV) estimation

Deterministic approach: in [40][138][261] the minimum-variance (MV) or minimum mean square error (MMSE) estimation problem is formulated as calculating the $L \times L$-dimensional matrix $\mathbf{W}$ that minimises

$$\boxed{\min_{\mathbf{W}} ||\mathbf{Y}_0[k]\mathbf{W} - \mathbf{X}_0[k]||_F^2} \tag{2.77}$$

for the time being assuming that $\mathbf{X}_0[k]$ is known. The matrix $\mathbf{W}$, minimising this cost function is equal to (cf. Appendix B.2)

$$\mathbf{W} = \left(\mathbf{Y}_0^T[k]\mathbf{Y}_0[k]\right)^{-1}\mathbf{Y}_0^T[k]\mathbf{X}_0[k]\,. \tag{2.78}$$

Using (2.59), (2.60) and (2.61), this matrix can be rewritten as

$$\begin{aligned}\mathbf{W} &= \left(\mathbf{Q}\mathbf{\Sigma}_Y^2\mathbf{Q}^T\right)^{-1}\mathbf{X}_0^T[k]\mathbf{X}_0[k] = \mathbf{Q}^{-T}\mathbf{\Sigma}_Y^{-2}\mathbf{Q}^{-1}\mathbf{Q}(\mathbf{\Sigma}_Y^2 - \mathbf{\Sigma}_V^2)\mathbf{Q}^T \quad (2.79)\\ &= \mathbf{Q}^{-T}(\mathbf{I}_L - \mathbf{\Sigma}_Y^{-2}\mathbf{\Sigma}_V^2)\mathbf{Q}^T = \mathbf{Q}^{-T}\left[\begin{array}{cc}\mathbf{I}_R - \mathbf{\Sigma}_{Y1}^{-2}\mathbf{\Sigma}_{V1}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\mathbf{Q}^T\,. \quad (2.80)\end{aligned}$$

Using (2.59), the minimum-variance estimate $\hat{\mathbf{X}}_0^{MV}[k]$ can then be obtained as

$$\boxed{\begin{aligned}\hat{\mathbf{X}}_0^{MV}[k] &= \mathbf{Y}_0[k]\mathbf{W} = \mathbf{U}_Y\mathbf{\Sigma}_Y(\mathbf{I}_L - \mathbf{\Sigma}_Y^{-2}\mathbf{\Sigma}_V^2)\mathbf{Q}^T\\ &= \mathbf{U}_{Y1}\mathbf{\Sigma}_{Y1}\big(\mathbf{I}_R - \mathbf{\Sigma}_{Y1}^{-2}\mathbf{\Sigma}_{V1}^2\big)\mathbf{Q}_1^T\end{aligned}} \tag{2.81}$$

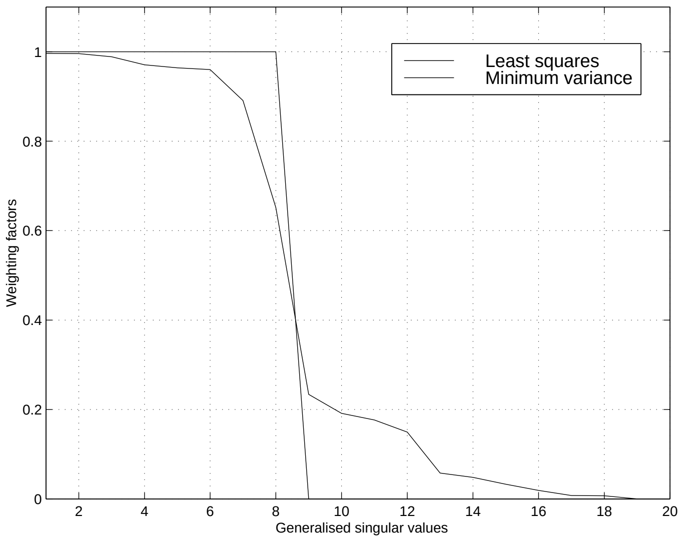


Figure 2.2: LS/MV weighting factors ($L = 20$, $R = 8$)

Comparing (2.68) and (2.81), it can be seen that both the LS and the MV estimate of the data matrix $\hat{\mathbf{X}}_0[k]$ can be written as

$$\hat{\mathbf{X}}_0[k] = \mathbf{U}_Y \left(\mathbf{\Sigma}_Y \mathbf{\Delta} \right) \mathbf{Q}^T \ , \tag{2.82}$$

i.e. the matrix $\hat{\mathbf{X}}_0[k]$ is constructed by multiplying the generalised singular values $\mathbf{\Sigma}_Y$ with an $L \times L$-dimensional diagonal matrix $\mathbf{\Delta} = \text{diag}\{\delta_i\}$, which is equal to

$$\mathbf{\Delta}_{LS} = \begin{bmatrix} \mathbf{I}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \ , \quad \mathbf{\Delta}_{MV} = \mathbf{I}_L - \mathbf{\Sigma}_Y^{-2} \mathbf{\Sigma}_V^2 \ . \tag{2.83}$$

In the MV estimation the rank R does not have to be determined beforehand, since the diagonal elements δ_i^{MV} are automatically zero for $i > R$. Even when the clean speech signal is not (perfectly) rank-deficient, the MV estimator will still reduce noise. Figure 2.2 shows typical LS/MV weighting factors δ_i, $i = 1 \ldots L$, for an example with $L = 20$ and $R = 8$ when the clean signal is not (perfectly) rank-deficient.

Statistical approach: in [85] the MV estimation problem is similarly formulated using data vectors. Suppose that $\hat{\mathbf{x}}_0^{MV}[k] = \mathbf{W}^T \mathbf{y}_0[k]$ is the estimate for the speech data vector $\mathbf{x}_0[k]$. The estimation error vector $\mathbf{e}[k]$ is defined as

$$\mathbf{e}[k] = \mathbf{x}_0[k] - \hat{\mathbf{x}}_0^{MV}[k] = \mathbf{x}_0[k] - \mathbf{W}^T \mathbf{y}_0[k] \ , \tag{2.84}$$

which is in fact the sum of a term $\mathbf{e}_y[k]$ representing signal distortion and a term $\mathbf{e}_v[k]$ representing the residual noise,

$$\mathbf{e}[k] = \mathbf{x}_0[k] - \mathbf{W}^T(\mathbf{x}_0[k] + \mathbf{v}_0[k]) = \underbrace{\big(\mathbf{I}_L - \mathbf{W}^T\big)\mathbf{x}_0[k]}_{\mathbf{e}_y[k]} - \underbrace{\mathbf{W}^T\mathbf{v}_0[k]}_{\mathbf{e}_v[k]} \ . \tag{2.85}$$

If we assign equal importance to signal distortion and noise reduction, and hence minimise the mean square error (MSE)

$$\mathcal{E}\{||\mathbf{e}[k]||_2^2\} = \mathcal{E}\{\mathbf{x}_0^T[k]\mathbf{x}_0[k]\} - 2\mathcal{E}\{\mathbf{y}_0^T[k]\mathbf{W}\mathbf{x}_0[k]\} + \mathcal{E}\{\mathbf{y}_0^T[k]\mathbf{W}\mathbf{W}^T\mathbf{y}_0[k]\} \ , \tag{2.86}$$

the optimal filter matrix $\bar{\mathbf{W}}$ is the well-known Wiener filter [227],

$$\boxed{\bar{\mathbf{W}}_{WF} = \bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{yx}[k]} \tag{2.87}$$

with $\bar{\mathbf{R}}_{yx}[k] = \mathcal{E}\{\mathbf{y}_0[k]\mathbf{x}_0^T[k]\}$ the cross-correlation matrix between $\mathbf{y}_0[k]$ and $\mathbf{x}_0[k]$. Since the speech and the noise signal are uncorrelated and using (2.52) and (2.53), this matrix can be written as

$$\begin{aligned} \bar{\mathbf{W}}_{WF} &= \bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{xx}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k]\,(\bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k]) & (2.88)\\ &= \bar{\mathbf{Q}}^{-T}\bar{\mathbf{\Lambda}}_y^{-1}\bar{\mathbf{Q}}^{-1}\,\bar{\mathbf{Q}}(\bar{\mathbf{\Lambda}}_y - \bar{\mathbf{\Lambda}}_v)\bar{\mathbf{Q}}^T = \bar{\mathbf{Q}}^{-T}(\mathbf{I}_L - \bar{\mathbf{\Lambda}}_y^{-1}\bar{\mathbf{\Lambda}}_v)\bar{\mathbf{Q}}^T \ , & (2.89) \end{aligned}$$

yielding a similar result as (2.80). The MV estimate $\hat{\mathbf{x}}_0^{MV}[k]$ can now be written as

$$\boxed{\hat{\mathbf{x}}_0^{MV}[k] = \bar{\mathbf{W}}_{WF}^T\mathbf{y}_0[k] = \bar{\mathbf{Q}}(\mathbf{I}_L - \bar{\mathbf{\Lambda}}_y^{-1}\bar{\mathbf{\Lambda}}_v)\bar{\mathbf{Q}}^{-1}\,\mathbf{y}_0[k]} \tag{2.90}$$

As already mentioned, it is also possible to trade off the noise reduction term $\mathbf{e}_v[k]$ and the signal distortion term $\mathbf{e}_y[k]$ of (2.85) in the estimation procedure. This estimator will be discussed in Section 3.2.3.

Averaging operation

In order to extract the estimated speech signal $\hat{x}_0[k]$ from the LS and MV matrices $\hat{\mathbf{X}}_0^{LS}[k]$ and $\hat{\mathbf{X}}_0^{MV}[k]$, one can e.g. consider the first row and/or the first column of these matrices. However, in general the matrices $\hat{\mathbf{X}}_0^{LS}[k]$ and $\hat{\mathbf{X}}_0^{MV}[k]$ will have lost their Toeplitz structure. Therefore, some procedures [43][49][121][138] first average along the diagonals in order to restore the Toeplitz matrix structure. After this averaging operation, the matrices will in general not be rank-deficient (rank R) any more. Hence, one could iterate this procedure (rank reduced LS/MV estimation and averaging along diagonals) and it has been shown that this iterative procedure converges to a Toeplitz matrix having rank R [27]. However, in Chapter 3 it will be shown that in fact this averaging operation is unnecessary and often suboptimal, since it typically gives rise to a larger MSE (when using the MV estimation) while it increases the computational complexity.

FIR filterbank interpretation

In [68] it has been shown that for the white noise case the overall procedure of rank reduced LS/MV estimation and averaging along the diagonals, is equivalent to subtracting zero-phase filtered versions from the noisy speech signal $y_0[k]$. The used zero-phase filters are constructed from the singular vectors corresponding to the $L-R$ smallest singular values of the data matrix $\mathbf{Y}_0[k]$.

In [120] a complete FIR filterbank representation is given for the LS/MV estimation algorithms in terms of the SVD of $\mathbf{Y}_0[k]$ (for the white noise case) or the GSVD of $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$ (for the coloured noise case). From (2.82), the matrix $\hat{\mathbf{X}}_0[k]$ of the enhanced signal can be decomposed as

$$\hat{\mathbf{X}}_0[k] = \sum_{i=1}^{R} \delta_i \sigma_i \mathbf{u}_{Yi} \mathbf{q}_i^T \ , \tag{2.91}$$

with $\delta_i, i = 1 \ldots R$, the weighting factors for the LS/MV estimation, cf. (2.83). If we define the matrix $\mathbf{T} = \mathbf{Q}^{-T}$, containing the vectors $\mathbf{t}_i$, then $\mathbf{Y}_0[k]\mathbf{t}_i = \sigma_i \mathbf{u}_{Yi}$, cf. (A.30), such that $\hat{\mathbf{X}}_0[k]$ can be written as

$$\hat{\mathbf{X}}_0[k] = \mathbf{Y}_0[k] \sum_{i=1}^{R} \delta_i \mathbf{t}_i \mathbf{q}_i^T \ . \tag{2.92}$$

When no averaging step is performed, we can e.g. choose the jth column of $\hat{\mathbf{X}}_0[k]$ as the estimate for the speech signal $\hat{x}_0[k]$. This column is equal to

$$\hat{\mathbf{X}}_{0,j}[k] = \mathbf{Y}_0[k] \sum_{i=1}^{R} \delta_i q_{ji} \mathbf{t}_i \ , \tag{2.93}$$

such that the enhanced speech signal $\hat{x}_0[k]$ can be obtained by summing R filtered versions of the noisy speech signal $y_0[k]$ (see Fig. 2.3). Each version is obtained by passing the signal $y[k]$ through an FIR filter $q_{ji}\mathbf{t}_i$ having L taps and multiplying the output with a weight δ_i. Since the filter coefficients are derived from the generalised singular vectors of $\mathbf{Y}_0[k]$ and $\mathbf{V}_0[k]$ (which converge to the generalised eigenvectors of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ when $P \rightarrow \infty$), this filterbank is called an *eigenfilterbank*.

In [120] it is shown that when an additional averaging operation along the diagonals is performed, the total estimation procedure can still be represented using an eigenfilterbank (see Fig. 2.4). After passing the signal $y_0[k]$ through the FIR filter $\mathbf{t}_i$, the result is now filtered with a second FIR filter $\mathbf{J}_L\mathbf{q}_i$ having L taps, with $\mathbf{J}_L$ the $L \times L$ reversal matrix, cf. (A.8). This filtering operation corresponds to a backward filtering with $\mathbf{q}_i$. After summing the R filtered versions, multiplication with a diagonal matrix D is necessary due to the different lengths of the diagonals over which the averaging is performed,

$$D = \mathrm{diag}\Big\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{L}, \frac{1}{L}, \ldots, \frac{1}{L}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\Big\} \ . \tag{2.94}$$

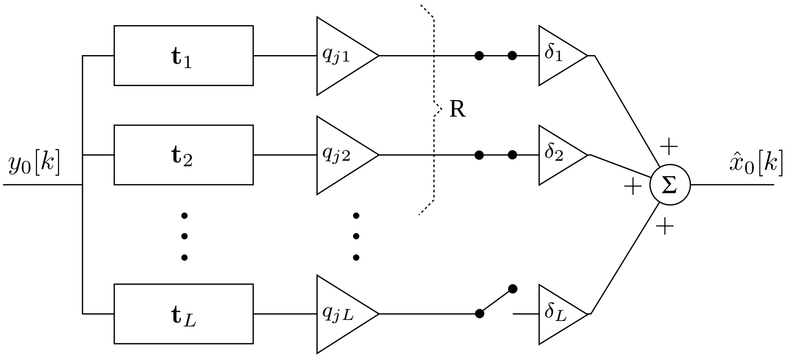


Figure 2.3: FIR filterbank representation (coloured noise case, no averaging)

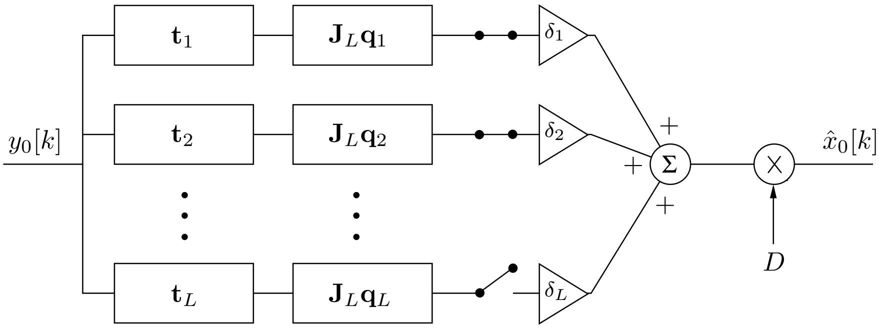


Figure 2.4: FIR filterbank representation (coloured noise case, with averaging)

In the white noise case, $\mathbf{Q}$ is an orthogonal matrix, such that $\mathbf{t}_i = \mathbf{q}_i$. The total forward and backward filtering operation with $\mathbf{q}_i$ corresponds to a zero-phase filtering operation with an FIR filter having $2L-1$ taps. Hence, the enhanced speech signal $\hat{x}_0[k]$ consists of the sum of R zero-phase filtered versions of the signal $y_0[k]$. An interpretation of this algorithm in the frequency-domain can now be given. Because the SVD can be considered as a signal decomposition based on energetic criteria, the zero-phase filtered versions corresponding to the largest singular values correspond to the frequency components with the largest amplitudes. For speech this means that the zero-phase filters corresponding to the largest singular values capture the formants of the speech with large energy, whereas the other zero-phase filtered versions mainly contain noise.

Relation to spectral subtraction

Actually, spectral subtraction and signal subspace-based techniques are quite related since they both transform the noisy signal $y_0[k]$ to a transform domain, multiply the transform domain coefficients with a certain weight, which

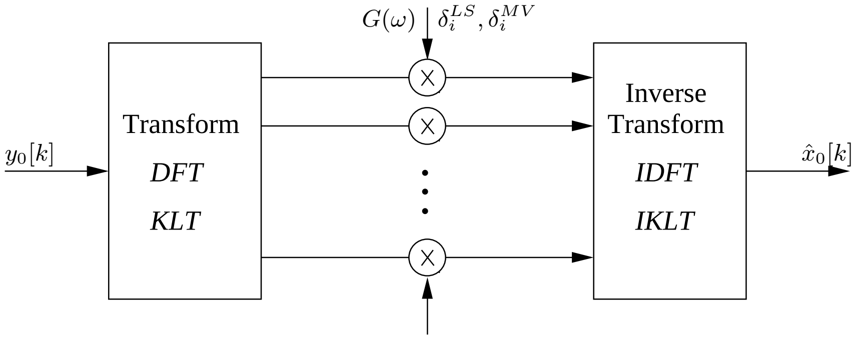


Figure 2.5: Transform domain filter representation for spectral subtraction and signal subspace-based techniques

depends on the speech and the noise characteristics, and finally perform a transformation back to the time-domain (see Fig. 2.5). The main difference between both techniques is the fact that spectral subtraction techniques use a *signal-independent* transform (DFT-type), whereas signal subspace-based techniques use a *signal-dependent* transform (KLT-type). However, since the DFT and the KLT are related[4], spectral subtraction can be considered an approximate signal subspace approach. In [85] it has been proved that if the Wiener gain function is used (cf. Table 2.1), then spectral subtraction is optimal in an asymptotic minimum mean square error (MMSE) sense when the frame length P goes to infinity and the speech and the noise are assumed stationary, i.e.

$$\lim_{P\to\infty} \frac{1}{P}\mathcal{E}\{||\hat{\mathbf{x}}_0^{MV}[k] - \hat{\mathbf{x}}_0^{SS}[k]||_2^2\} = 0 \ , \tag{2.95}$$

with $\hat{\mathbf{x}}_0^{SS}[k]$ the estimated speech signal using spectral subtraction.

2.4 Single-microphone dereverberation

In the single-microphone case (and assuming no background noise is present), the microphone signal $y_0[k]$ consists of the filtered speech signal $s[k]$, cf. (2.12),

$$y_0[k] = x_0[k] = h_0[k] \otimes s[k] \ , \tag{2.96}$$

with $h_0[k]$ the acoustic impulse response between the speech source and the microphone. Dereverberation consists of extracting the clean speech signal $s[k]$ from $y_0[k]$ without any prior knowledge about the acoustic impulse response $h_0[k]$. In this section we will briefly discuss 2 techniques: *inverse filtering* [42][180], which however assumes that $h_0[k]$ is known (from measurements or calculations), and *cepstrum-based techniques* [8][42][202][251].

[4] Both transforms are equivalent when the correlation matrix $\mathbf{R}_{yy}[k]$ is circulant [111].

2.4.1 Inverse filtering

If we assume that the acoustic impulse response $h_0[k]$ is known (e.g. from measurements or an estimation procedure, cf. Section 6.2), reverberation can be removed by filtering $y_0[k]$ with the inverse filter $h_0^{-1}[k]$. However, typical acoustic impulse responses are non-minimum-phase and therefore do not have stable causal inverses [188]. Using (2.13) and (2.14), the L-dimensional data vector $\mathbf{y}_0[k]$ can be written as

$$\mathbf{y}_0[k] = \begin{bmatrix} y_0[k] \\ y_0[k-1] \\ \vdots \\ y_0[k-L+1] \end{bmatrix} = \underbrace{\begin{bmatrix} \boxed{\mathbf{h}_0^T[k]} & 0 & \dots & 0 \\ 0 & \boxed{\mathbf{h}_0^T[k]} & \dots & 0 \\ & & \ddots & \\ 0 \;\; 0 & \dots & & \boxed{\mathbf{h}_0^T[k]} \end{bmatrix}}_{\mathbf{H}_0[k]} \underbrace{\begin{bmatrix} s[k] \\ s[k-1] \\ \vdots \\ s[k-K-L+2] \end{bmatrix}}_{\mathbf{s}[k]} , \tag{2.97}$$

with $\mathbf{H}_0[k]$ an $L\times(K+L-1)$-dimensional Toeplitz matrix and $\mathbf{s}[k]$ a $(K+L-1)$-dimensional vector. Dereverberation consists of computing an L-dimensional vector $\mathbf{w}_0[k]$, such that

$$\mathbf{w}_0^T[k]\mathbf{y}_0[k] = s[k], \; \forall k \; , \tag{2.98}$$

i.e.

$$\mathbf{w}_0^T[k]\mathbf{H}_0[k] = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} = \mathbf{d}^T \; . \tag{2.99}$$

However, in general no exact solution exists, since $\mathbf{H}_0[k]$ has more columns than rows. The LS solution of (2.99) is equal to

$$\boxed{\mathbf{w}_{0,LS}[k] = \left(\mathbf{H}_0^{\dagger}[k]\right)^T \mathbf{d} = \left(\mathbf{H}_0[k]\mathbf{H}_0^T[k]\right)^{-1} \mathbf{H}_0[k]\mathbf{d}} \tag{2.100}$$

with $\mathbf{H}_0^{\dagger}[k]$ the pseudo-inverse of $\mathbf{H}_0[k]$, cf. Appendix A.2. However, in [180] it has been proved that the average energy of the LS estimation error, i.e. $s[k] - \mathbf{w}_{0,LS}^T[k]\mathbf{y}_0[k]$, does not converge to 0 if $h_0[k]$ is non-minimum phase, even when the filter length L goes to infinity. Moreover, in most cases the acoustic impulse response $h_0[k]$ is not known, such that single-microphone inverse filtering techniques have a limited scope in practice.

2.4.2 Cepstrum-based techniques

Cepstrum-based techniques have been proposed for single-microphone dereverberation when the acoustic impulse response $h_0[k]$ is unknown. The *complex cepstrum* $c_{y_0}[k]$ of a signal $y_0[k]$ is defined as the inverse Fourier transform of the (complex) logarithm of the Fourier transform of $y_0[k]$, i.e.

$$c_{y_0}[k] = \mathcal{F}^{-1}\big\{ \log\big(\mathcal{F}\{y_0[k]\}\big)\big\} = \mathcal{F}^{-1}\big\{ \log Y_0(\omega)\big\} \; . \tag{2.101}$$

The complex cepstrum can therefore be considered as a spectral analysis of the log-spectrum, consisting of low and high-'quefrency' components. Because of the logarithm, operations in the cepstrum domain perform non-linear operations on the signal in the time-domain. As can be easily seen from (2.101), convolution in the time-domain is equivalent to addition in the cepstrum domain, i.e.

$$\begin{aligned} c_{y_0}[k] &= \mathcal{F}^{-1}\big\{\log Y_0(\omega)\big\} = \mathcal{F}^{-1}\big\{\log\big(H_0(\omega)S(\omega)\big)\big\} && (2.102)\\ &= \mathcal{F}^{-1}\big\{\log H_0(\omega)\big\} + \mathcal{F}^{-1}\big\{\log S(\omega)\big\} = c_{h_0}[k] + c_s[k]\ , && (2.103)\end{aligned}$$

such that dereverberation corresponds to subtraction in the cepstrum domain [42]. Since the cepstrum of the clean speech signal $c_s[k]$ is usually concentrated around the cepstral origin, while that of the acoustic impulse response $c_{h_0}[k]$ is composed of pulses (far) away from the origin, dereverberation can be achieved by low-quefrency filtering $c_{y_0}[k]$ (called 'liftering') or by peak-picking.

While cepstrum filtering has been successfully applied to the enhancement of speech degraded by simple echoes [202], its use for the enhancement of speech affected by room reverberation poses several practical problems. Typically, frame-based processing is used to calculate the cepstrum of a signal. Since reverberation effects are generally much longer than typical frame lengths, the current frame m does not contain all the reverberation effects of this frame, while it also contains reverberation effects from previous frames. Moreover, the cepstrum of the clean speech signal $c_s[k]$ and the cepstrum of the acoustic impulse response $c_{h_0}[k]$ typically have a large overlap, resulting in signal distortion when using low-quefrency 'liftering'. By using an exponential windowing procedure and cepstral averaging in order to identify the room impulse response $h_0[k]$ before inverse filtering, a significant improvement is possible [8][251]. However, in practice single-microphone cepstrum based techniques for dereverberation have a (very) limited performance, also since these techniques exhibit problems when additive background noise is present. Cepstrum-based processing has also been combined with microphone arrays. In [164] a technique is described, where the different microphone signals are factored into their minimum-phase and all-pass component, which are separately processed in the cepstrum and in the frequency-domain. However, since better multi-microphone techniques are available for dereverberation (e.g. fixed beamforming and dereverberation techniques discussed in Chapter 7), we will not consider cepstrum-based techniques in this thesis.

2.5 Multi-microphone noise reduction

As already mentioned, multi-microphone noise reduction techniques can exploit the spatial information in the microphone signals when the speech and the noise sources are located at different positions, and hence are able to perform both spectral and spatial filtering. In this section we will briefly discuss fixed and adaptive beamforming techniques. *Fixed beamforming techniques* are data-independent and try to obtain spatial focusing on the speech source, thereby reducing reverberation and suppressing background noise not coming from the same direction as the speech source. *Adaptive beamforming techniques* combine the spatial focusing of fixed beamformers with adaptive noise suppression, such that they are able to adapt to changing acoustic environments and generally have a better noise reduction performance. First we discuss some commonly used definitions and performance measures for beamformers.

2.5.1 Beamformer definitions and performance measures

In this section we will assume linear microphone arrays and sources in the far-field of the microphone array (cf. Section 1.3.4), such that planar wave propagation and equal signal attenuation for all microphones can be assumed. However, all expressions can be easily extended to other microphone configurations and to the near-field case (cf. Chapter 9).

Consider the microphone array depicted in Fig. 2.6, which is the frequency-domain representation of Fig. 2.1, with N microphones and with d_n the distance between the nth microphone and the centre of the microphone array. The speech source $S(\omega)$ is located at an angle θ from the microphone array. The direction $\theta = 90°$ is called *broadside*, whereas the directions $\theta = 0°$ and $\theta = 180°$ are called *endfire*. For a *uniform* linear microphone array, the distance

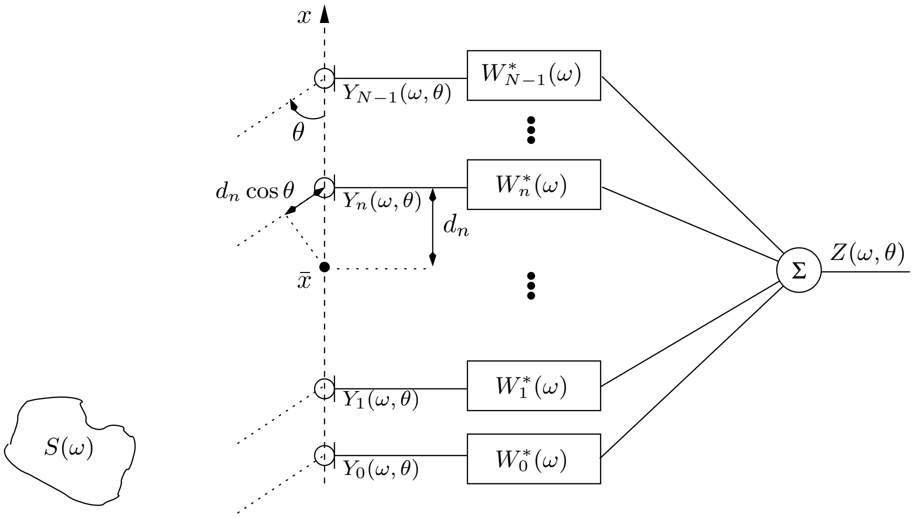


Figure 2.6: Linear microphone array configuration

between adjacent microphones is equal, i.e. $d_n - d_{n-1} = d$, $n = 1 \ldots N-1$. Under far-field conditions, the microphone signals $Y_n(\omega, \theta)$ are delayed versions of the speech source $S(\omega)$, and hence can be considered delayed versions of the signal $\bar{Y}(\omega, \theta)$, received at the centre of the array,

$$Y_n(\omega,\theta) = S(\omega)e^{-j\omega\bar{\tau}(\theta)}e^{-j\omega\tau_n(\theta)} = \bar{Y}(\omega,\theta)e^{-j\omega\tau_n(\theta)}, \; -\pi \leq \theta \leq \pi \; , \quad (2.104)$$

with the delay $\tau_n(\theta)$ in number of samples equal to

$$\tau_n(\theta) = \frac{d_n \cos\theta}{c} f_s \; , \quad (2.105)$$

with c the speed of sound ($c = 340\,\frac{m}{s}$) and f_s the sampling frequency. Using (2.31), the output signal $Z(\omega, \theta)$ can be written as

$$Z(\omega,\theta) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega,\theta) = \bar{Y}(\omega,\theta)\,\mathbf{W}^H(\omega)\mathbf{d}(\omega,\theta) \; , \quad (2.106)$$

with $\mathbf{d}(\omega, \theta)$ the *steering vector*, which is equal to

$$\mathbf{d}(\omega,\theta) = \left[\begin{array}{cccc} e^{-j\omega\tau_0(\theta)} & e^{-j\omega\tau_1(\theta)} & \ldots & e^{-j\omega\tau_{N-1}(\theta)} \end{array} \right]^T \; . \quad (2.107)$$

Spatial directivity pattern

The spatial directivity pattern $H(\omega, \theta)$ of the beamformer $\mathbf{W}(\omega)$ is defined as the transfer function from the source $S(\omega)$ at an angle θ to the output signal $Z(\omega, \theta)$ of the microphone array, i.e.

$$\boxed{H(\omega,\theta) = \frac{Z(\omega,\theta)}{\bar{Y}(\omega,\theta)} = \mathbf{W}^H(\omega)\mathbf{d}(\omega,\theta)} \quad (2.108)$$

Array gain

If background noise is present, the microphone signal $Y_n(\omega, \theta)$ is equal to

$$Y_n(\omega,\theta) = \underbrace{S(\omega)e^{-j\omega\bar{\tau}(\theta)}e^{-j\omega\tau_n(\theta)}}_{X_n(\omega,\theta)} + V_n(\omega) \; , \quad (2.109)$$

with $V_n(\omega)$ the noise received at the nth microphone. Using (2.38) with $H_n(\omega) = e^{-j\omega\bar{\tau}(\theta)}e^{-j\omega\tau_n(\theta)}$ and assuming a homogeneous noise-field (cf. Section 2.2.4), the unbiased output SNR is equal to

$$\mathrm{SNR}_z(\omega,\theta) = 10\log_{10}\frac{P_s(\omega)}{P_v(\omega)}\frac{|\mathbf{W}^H(\omega)\mathbf{d}(\omega,\theta)|^2}{\mathbf{W}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}(\omega)} \; , \quad (2.110)$$

whereas the average unbiased input SNR is equal to

$$\mathrm{SNR}_{in}(\omega,\theta) \; = \; 10\log_{10}\frac{\mathcal{E}\{\mathbf{X}^H(\omega,\theta)\mathbf{X}(\omega,\theta)\}}{\mathcal{E}\{\mathbf{V}^H(\omega)\mathbf{V}(\omega)\}} \quad (2.111)$$

$$= 10\log_{10}\frac{\mathcal{E}\{|S(\omega)|^2\}\,\mathbf{d}^H(\omega,\theta)\mathbf{d}(\omega,\theta)}{\sum_{n=0}^{N-1}\mathcal{E}\{|V_n(\omega)|^2\}} = 10\log_{10}\frac{P_s(\omega)}{P_v(\omega)}\ . \quad (2.112)$$

The *array gain* $AG(\omega,\theta)$ is defined as the SNR improvement achieved by the microphone array. Hence, using (2.110) and (2.112), the array gain is equal to

$$\boxed{AG(\omega,\theta) = \mathrm{SNR}_z(\omega,\theta) - \mathrm{SNR}_{in}(\omega,\theta) = 10\log_{10}\frac{|\mathbf{W}^H(\omega)\mathbf{d}(\omega,\theta)|^2}{\mathbf{W}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}(\omega)}} \quad (2.113)$$

Directivity and white noise gain

When the noise field is known, the array gain can be computed analytically. E.g. for a localised noise source at an angle θ_v, the complex coherence $\Gamma_v^{mn}(\omega) = \mathcal{E}\{V_m(\omega)V_n^*(\omega)\}$ is equal to

$$\Gamma_v^{mn}(\omega) = e^{-j\omega\left(\tau_m(\theta_v)-\tau_n(\theta_v)\right)} = e^{-j\omega(d_m-d_n)\cos\theta_v f_s/c}\ , \quad (2.114)$$

while for a diffuse noise source, i.e. equally distributed uncorrelated white noise coming from all directions, the complex coherence $\Gamma_v^{mn}(\omega)$ is equal to [16]

$$\Gamma_v^{mn}(\omega) = \frac{\sin\left(\omega(d_m-d_n)f_s/c\right)}{\omega(d_m-d_n)f_s/c}\ . \quad (2.115)$$

The *directivity* $DI(\omega,\theta)$ of a beamformer is defined as the array gain for diffuse noise, whereas the *white noise gain* $WNG(\omega,\theta)$ is defined as the array gain for spatially uncorrelated noise (e.g. sensor noise), i.e. $\mathbf{\Gamma}_v(\omega) = \mathbf{I}_N$. The white noise gain can be used as a measure for the robustness of a beamformer.

2.5.2 Fixed beamforming

In a fixed beamformer the filters $w_n[k]$, i.e. $W_n(\omega)$, are designed in a non-adaptive way such that sounds coming from the direction of the speech source (arriving from an angle θ_x) are passed without distortion, whereas sounds coming from other directions are suppressed. If the speaker is moving, the speaker position should be continuously tracked using an acoustic source localisation algorithm (cf. Chapter 6) and for each speaker position different – fixed – beamformer weights may be applied [150]. Alternatively, fixed beamformers can be designed to be robust against (small) speaker movements (cf. Part III). In this section we will discuss the delay-and-sum beamformer, differential microphones, the filter-and-sum beamformer and superdirective beamformers. These fixed beamformers will either be used as a building block in adaptive beamformers, for comparison purposes with the GSVD-based optimal filtering technique in Part I, or as a starting point for beamformer design in Part III.

Delay-and-sum (DS) beamformer

A delay-and-sum (DS) beamformer is the simplest structure to obtain spatial selectivity. The theory of DS beamforming originates from narrowband antenna array processing, where the plane waves at different sensors are delayed appropriately to be added exactly in phase. In this way, the array can be electronically steered towards a specific direction. This principle is also valid for broadband signals, although the directivity will then be frequency-dependent.

A DS beamformer spatially aligns the microphone signals to the direction of the speech source by delaying and summing the microphone signals [258][264], i.e.

$$\boxed{z[k] = \frac{1}{N} \sum_{n=0}^{N-1} y_n[k - \delta_n]} \tag{2.116}$$

where the delays δ_n are computed as

$$\delta_n = -\frac{d_n \cos\theta_x}{c} f_s = -\tau_n(\theta_x) \ . \tag{2.117}$$

Angular selectivity is obtained based on constructive interference ($\theta = \theta_x$) and destructive interference ($\theta \neq \theta_x$). Apart from noise reduction, the DS beamformer will therefore also perform some dereverberation, since the direct path contribution is added in phase, whereas other reflections are typically added randomly. Since the delays δ_n are generally non-integer values, the filtering operation in (2.116) is generally implemented using interpolation filters [140]. In the frequency-domain, the filter $\mathbf{W}(\omega)$ is equal to $\mathbf{d}(\omega, \theta_x)/N$. For a *uniform* DS beamformer with inter-microphone distance d, the spatial directivity pattern $H(\omega, \theta)$ is equal to

$$H(\omega, \theta) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-j\omega n d(\cos\theta - \cos\theta_x) f_s/c} = \frac{1}{N} \frac{e^{-jN\gamma/2 \sin(N\gamma/2)}}{e^{-j\gamma/2 \sin(\gamma/2)}} \ , \tag{2.118}$$

with $\gamma = \omega d(\cos\theta - \cos\theta_x) f_s/c$. Hence, $H(\omega, \theta)$ has a sinc-like shape and is frequency-dependent, i.e. not all frequency components of the speech signal will undergo the same spatial filtering operation. For the parameters $N = 4$, $d = 0.03\,\mathrm{m}$, $\theta_x = 60°$ and $f_s = 16\,\mathrm{kHz}$, Fig. 2.7 plots the spatial directivity pattern $H(\omega, \theta)$ for all frequencies and for the specific frequency $f = 5000\,\mathrm{Hz}$. As can be seen, the beamwidth of a DS beamformer is frequency-dependent and for low frequencies the directivity is quite poor.

Because of the periodicity of $H(\omega, \theta)$ in θ, for frequencies

$$f \geq \frac{c}{d(1 + |\cos\theta_x|)} \tag{2.119}$$

an ambiguity, called *spatial aliasing*, occurs. This is analogous to time-domain aliasing, where now the spatial sampling d is too large. The result is that high

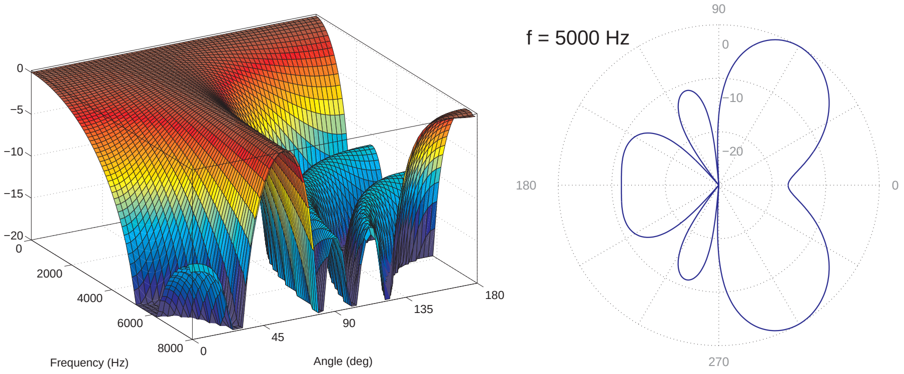


Figure 2.7: Spatial directivity pattern of a uniform DS beamformer ($N = 4$, $d = 0.03\,\mathrm{m}$, $\theta_x = 60°$, $f_s = 16\,\mathrm{kHz}$)

frequency noises (and reverberant components) pass without attenuation at angles different from the steering angle θ_x. In order to avoid spatial aliasing for all steering angles, the maximum inter-microphone distance is $d_{max} = c/f_s$. Hence, for broadband signals like speech, nested logarithmic array configurations are typically used, with a large inter-microphone distance for the lower frequencies and a small distance for the higher frequencies. However, in this thesis we will not study the influence of the microphone array configuration on the performance of the speech enhancement algorithms.

An additional beam shaping is possible when introducing a sensor-dependent complex weight, also called tapering, before summation, i.e.

$$z[k] = \frac{1}{N} \sum_{n=0}^{N-1} w_n \, y_n[k - \delta_n] \; . \tag{2.120}$$

Using these weights w_n it is e.g. possible to design a beam pattern with a uniform sidelobe level, i.e. Dolph-Chebyshev design [71].

Differential microphones

A good overview of first- and higher-order differential microphones can be found in [75]. A *first-order differential microphone* is a directional microphone array which consists of 2 closely spaced microphones at a distance d, where one microphone is delayed – generally in hardware – and whose outputs are then subtracted from each other (see Fig. 2.8). The spatial directivity pattern is equal to

$$\boxed{H(\omega, \theta) = 1 - e^{-j\omega(\tau + d\cos\theta/c)}} \tag{2.121}$$

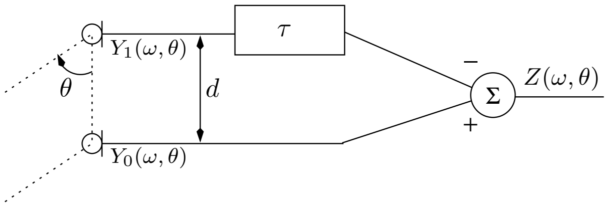


Figure 2.8: First-order differential microphone array configuration

Since the microphones are closely spaced, one can assume that $\omega d/c \ll \pi$ and $\omega\tau \ll \pi$, such that $H(\omega,\theta)$ can be approximated by

$$H(\omega,\theta) = \omega(\tau + d\cos\theta/c) \; . \tag{2.122}$$

As one can see, a first-order differential microphone has a first-order high-pass frequency characteristic. If we compensate for this high-pass frequency-characteristic and introduce $\alpha_\tau = \tau/(\tau + d/c)$, then the normalised directional response $H(\theta)$ can be written as

$$H(\theta) = \alpha_\tau + (1-\alpha_\tau)\cos\theta \; , \tag{2.123}$$

with $0 \leq \alpha_\tau \leq 1$, having a maximum at $\theta = 0°$ and a minimum or a zero between 90° and 180°. The delay τ (or equivalently α_τ) can be computed for different objectives. The most commonly used first-order differential microphones are a dipole ($\alpha_\tau = 0$) having a zero at 90°, a cardioid microphone ($\alpha_\tau = 0.5$) having a zero at 180°, a hypercardioid microphone ($\alpha_\tau = 0.25$) maximising the directivity index ($DI_{max} = 6.0\,\text{dB}$) and a supercardioid microphone ($\alpha_\tau \approx 0.35$) maximising the front-to-back ratio. However, all differential microphones are quite sensitive to microphone imperfections (gain, phase, position), as has e.g. been shown in [24].

Filter-and-sum beamformer

The filter-and-sum structure, depicted in Fig. 2.1 and 2.6, obviously is the most general beamformer structure. Using this structure it is possible to design a fixed beamformer whose spatial directivity pattern optimally fits a (predefined) desired spatial directivity pattern $D(\omega,\theta)$, which can be an arbitrary two-dimensional function in the variables ω and θ. The filter coefficients $\mathbf{w}_n[k]$, $n = 0 \ldots N-1$, of the fixed beamformer are computed such that a specific cost function is minimised, e.g. the weighted LS cost function

$$\min_{\mathbf{w}_n[k]} \int_{\omega_1}^{\omega_2} \int_{\theta_1}^{\theta_2} F(\omega,\theta) |H(\omega,\theta) - D(\omega,\theta)|^2 \, d\omega d\theta \; , \tag{2.124}$$

with $F(\omega,\theta)$ a weighting function, or the non-linear cost function [144]

$$\min_{\mathbf{w}_n[k]} \int_{\omega_1}^{\omega_2} \int_{\theta_1}^{\theta_2} F(\omega,\theta) \left[|H(\omega,\theta)| - |D(\omega,\theta)| \right]^2 \, d\omega d\theta \; , \tag{2.125}$$

where only the error between the amplitudes of the spatial directivity patterns is taken into account, since the phase is typically of less importance. Some procedures use a frequency-domain approach and solve easier (decoupled) optimisation problems for separate frequencies ω_i, $i = 0 \ldots L-1$,

$$\min_{\mathbf{W}_n(\omega_i)} \int_{\theta_1}^{\theta_2} |H(\omega_i, \theta) - D(\omega_i, \theta)|^2 \, d\theta \; . \tag{2.126}$$

However, using this frequency-domain approach it is not possible to control the spatial directivity pattern at intermediate frequencies and to apply a frequency-dependent weighting function $F(\omega, \theta)$. Many other cost functions have been proposed, e.g. based on a maximum energy array [155], an eigenfilter approach [65][59] or a non-linear minimax optimisation problem [157][159][192]. Some of these cost functions will be discussed in more detail in Part III, where the design of robust broadband beamformers in the far-field and the near-field of a microphone array is discussed. A special type of filter-and-sum beamformer is a *frequency-invariant beamformer* [274], which – as the name suggests – has a spatial directivity pattern which is independent of ω, i.e. $H(\omega, \theta) = H(\theta)$. Frequency-invariance is a desirable property for beamformers in speech communication applications, since all frequency components of the speech signal then undergo the same spatial filtering operation. The design procedure in [274] starts from a continuous sensor, for which a necessary condition for frequency-invariance is derived and which is then discretised using a microphone array with microphones at discrete positions.

Superdirective beamformer

The ‘super’-aspect of a superdirective beamformer lies in the fact that it maximises the directivity index $DI(\omega, \theta)$ in the direction of the speech source for a known (diffuse) noise field. A good overview of superdirective beamformers is given in [16]. The filter $\mathbf{W}(\omega)$ that maximises (2.113), with $\theta = \theta_x$, for a known noise field $\mathbf{\Gamma}_v(\omega)$, is given by the largest generalised eigenvector of $\mathbf{d}(\omega, \theta_x)\mathbf{d}^H(\omega, \theta_x)$ and $\mathbf{\Gamma}_v(\omega)$, i.e.

$$\mathbf{W}(\omega) = \alpha \, \mathbf{\Gamma}_v^{-1}(\omega)\mathbf{d}(\omega, \theta_x) \; , \tag{2.127}$$

where α is usually determined such that the spatial directivity pattern $H(\omega, \theta)$ is equal to 1 for the steering angle, i.e. $\mathbf{W}^H(\omega)\mathbf{d}(\omega, \theta_x) = 1$, such that

$$\boxed{\mathbf{W}(\omega) = \frac{\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{d}(\omega, \theta_x)}{\mathbf{d}^H(\omega, \theta_x)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{d}(\omega, \theta_x)}} \tag{2.128}$$

It can be shown that the maximum directivity for a diffuse noise field and for endfire steering, i.e. $\theta_x = 0°$, is equal to $10 \log_{10} N^2$ [75]. The beamformer in (2.128) is also called an MVDR (minimum variance distortionless response) beamformer, since this beamformer minimises the signal output power, subject to a unit response for the steering angle, i.e.

$$\boxed{\min_{\mathbf{W}(\omega)} \mathbf{W}^H(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{W}(\omega), \quad \text{subject to } \mathbf{W}^H(\omega)\mathbf{d}(\omega,\theta_x) = 1} \tag{2.129}$$

with $\bar{\mathbf{R}}_{yy}(\omega)$ similarly defined as in (2.36). The proof is given in Appendix B.3.

The same optimisation problem will be used for the adaptive Frost beamformer (cf. Section 2.5.3), where the noise field then is unknown and has to be adaptively estimated from the microphone data. Superdirective beamformers are known to be very sensitive to microphone mismatch and will boost uncorrelated noise at lower frequencies. Therefore an additional WNG constraint is generally added to the optimisation problem (2.129) in order to enhance the beamformer robustness [36][146], yielding

$$\mathbf{W}(\omega) = \frac{(\mathbf{\Gamma}_v(\omega) + \lambda\mathbf{I}_N)^{-1}\mathbf{d}(\omega,\theta_x)}{\mathbf{d}^H(\omega,\theta_x)(\mathbf{\Gamma}_v(\omega) + \lambda\mathbf{I}_N)^{-1}\mathbf{d}(\omega,\theta_x)} \, . \tag{2.130}$$

For uncorrelated white noise, i.e. $\mathbf{\Gamma}_v(\omega) = \mathbf{I}_N$, the MVDR beamformer in (2.128) is equal to

$$\mathbf{W}(\omega) = \mathbf{d}(\omega,\theta_x)/N \, , \tag{2.131}$$

which is in fact the delay-and-sum beamformer. Therefore the DS beamformer is the beamformer which maximises the WNG. The maximum value for the WNG is $10\log_{10} N$.

2.5.3 Adaptive beamforming

In practice, since the background noise is unknown and can change both spectrally and spatially, information about the noise field has to be adaptively estimated from the microphone data. Adaptive beamformers combine the spatial focusing of fixed beamformers with adaptive noise suppression, such that they are able to adapt to changing acoustic environments and generally exhibit a better noise reduction performance than fixed beamformers. Adaptive beamformers typically give rise to constrained optimisation problems, in order not to distort signals coming from the direction of the speech source. A good overview of adaptive beamforming techniques can be found in [258][264].

Frost beamformer – LCMV-beamformer

Frost [95] was the first to formulate the adaptive beamforming problem as a constrained optimisation problem. He however assumes that no reverberation is present and that the speech source is located at broadside, i.e. $\theta_x = 90°$. If this is not the case, delays can be added to the microphone signals in order to steer the microphone array towards the direction of the speech source (cf. DS beamforming). The (adaptive) filters $\mathbf{w}_n[k]$, $n = 0 \ldots N-1$, in the filter-and-sum structure of Fig. 2.1 are computed such that the variance of the output signal $z[k]$, i.e.

$$\mathcal{E}\{z^2[k]\} = \mathcal{E}\{(\mathbf{w}^T[k]\mathbf{y}[k])^2\} = \mathbf{w}^T[k]\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k] \;, \tag{2.132}$$

with $\bar{\mathbf{R}}_{yy}[k] = \mathcal{E}\{\mathbf{y}[k]\mathbf{y}^T[k]\}$ and $\mathbf{w}[k]$ the stacked filter vector, is minimised. In order to avoid that the speech signal is distorted or cancelled out, J linear constraints are added, i.e.

$$\mathbf{C}\mathbf{w}[k] = \mathbf{b} \;, \tag{2.133}$$

with $\mathbf{C}$ a $J \times M$-dimensional constraint matrix and $\mathbf{b}$ a J-dimensional constraint vector. These J linear constraints restrict the filter $\mathbf{w}[k]$ to lie in an $(M-J)$-dimensional hyperplane, which is orthogonal to the subspace spanned by the rows of $\mathbf{C}$. In general the goal of these constraints is to obtain better control over certain spectral and spatial regions and to obtain a solution that is less sensitive to deviations from the assumed signal model. The optimisation problem

$$\boxed{\min_{\mathbf{w}[k]} \mathbf{w}^T[k]\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k] \quad \text{subject to} \;\; \mathbf{C}\mathbf{w}[k] = \mathbf{b}} \tag{2.134}$$

is in fact a time-domain formulation of (2.129) for a certain choice of $\mathbf{C}$ and $\mathbf{b}$. If the speech and the noise are uncorrelated (and no reverberation is assumed), then constrained output energy minimisation corresponds to constrained noise energy minimisation. The filter $\mathbf{w}[k]$ minimising (2.134) is given by

$$\boxed{\mathbf{w}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k]\mathbf{C}^T(\mathbf{C}\bar{\mathbf{R}}_{yy}^{-1}[k]\mathbf{C}^T)^{-1}\mathbf{b}} \tag{2.135}$$

cf. Appendix B.4. This filter is called an LCMV (linearly-constrained minimum-variance) beamformer, of which the MVDR-beamformer is a special case. A typical constraint is a predefined frequency response $F(\omega) = \sum_{k=0}^{L-1} f[k]e^{-jk\omega}$ in the direction of the speech source, which corresponds to $J = L$ constraints with

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_L & \mathbf{I}_L & \ldots & \mathbf{I}_L \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} f[0] & f[1] & \ldots & f[L-1] \end{bmatrix}^T \;. \tag{2.136}$$

When the filter $F(\omega) = 1$, the LCMV-beamformer corresponds to the MVDR-beamformer. Other linear constraints typically include multiple-point, eigenvector and derivative constraints [87][264], cf. Section 8.5.

In practice, the correlation matrix $\bar{\mathbf{R}}_{yy}[k]$ is unknown and hence has to be estimated using an adaptive technique. To this aim the optimisation problem (2.134) can be solved using a gradient-descent optimisation technique, where in each iteration step the filters are updated in the direction of the constrained gradient, leading to the update equation (cf. Appendix B.5)

$$\mathbf{w}[k+1] = \mathbf{P}_C(\mathbf{w}[k] - \mu\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k]) + \mathbf{b}_C \;, \tag{2.137}$$

with μ the step size of the adaptive algorithm and

$$\begin{aligned} \mathbf{P}_C &= \mathbf{I}_M - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C} \; , & (2.138) \\ \mathbf{b}_C &= \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b} \; . & (2.139) \end{aligned}$$

By using the instantaneous gradient, i.e. using a stochastic approximation for the correlation matrix, $\bar{\mathbf{R}}_{yy}[k] \approx \mathbf{y}\mathbf{y}^T[k]$, the *constrained LMS* algorithm[5] is obtained,

$$\boxed{\mathbf{w}[k+1] = \mathbf{P}_C(\mathbf{w}[k] - \mu z[k]\mathbf{y}[k]) + \mathbf{b}_C} \qquad (2.140)$$

For a geometric interpretation of this equation, we refer to Appendix B.5. Using this adaptive algorithm, it is possible to reduce background noise in unknown noise fields and to adapt to changing acoustic environments.

Frost has developed this algorithm under the assumption that no reverberation is present. However, in reverberant acoustic environments, only a portion of the speech energy impinges on the array from the steering direction θ_x. Hence, multi-path propagation combined with output energy minimisation will result in signal distortion and even signal cancellation, because the filter coefficients are partially adapted to minimise output power corresponding to the speech signal $s[k]$ itself. A possible solution is to switch off adaptation during speech-and-noise periods.

Griffiths-Jim beamformer – Generalised Sidelobe Canceller (GSC)

Griffiths and Jim [116] reformulated the constrained LCMV optimisation problem (2.134) as an *unconstrained* optimisation problem, leading to an easier adaptation scheme. Consider the $M \times M$-dimensional full-rank matrix $\mathbf{C}_t$,

$$\mathbf{C}_t = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}_a \end{bmatrix} , \qquad (2.141)$$

with $\mathbf{C}_a$ the $(M-J) \times M$-dimensional null-space of $\mathbf{C}$. If we define the M-dimensional vector

$$\begin{array}{r} J \updownarrow \\ M-J \updownarrow \end{array} \begin{bmatrix} \mathbf{v}[k] \\ -\mathbf{w}_a[k] \end{bmatrix} = \mathbf{C}_t^{-T}\mathbf{w}[k] \; , \qquad (2.142)$$

such that the filter $\mathbf{w}[k]$ can be parametrised as

$$\mathbf{w}[k] = \mathbf{C}^T\mathbf{v}[k] - \mathbf{C}_a^T\mathbf{w}_a[k] \; , \qquad (2.143)$$

then the linear constraint can be written as

$$\mathbf{C}\mathbf{w}[k] = \mathbf{C}\mathbf{C}^T\mathbf{v}[k] - \underbrace{\mathbf{C}\mathbf{C}_a^T}_{\mathbf{0}}\mathbf{w}_a[k] = \mathbf{b} \; , \qquad (2.144)$$

[5] For a brief introduction on the LMS algorithm, we refer to page 66.

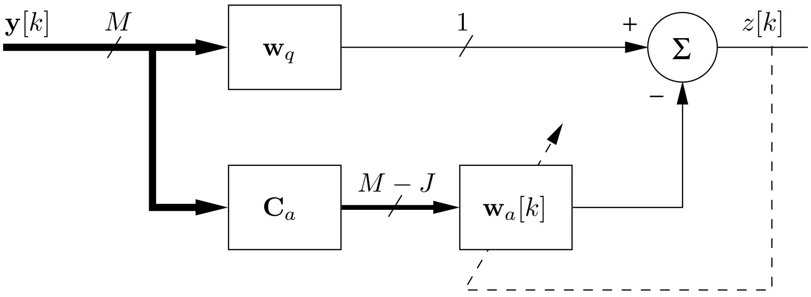


Figure 2.9: Griffiths-Jim beamformer structure

such that

$$\mathbf{v}[k] = (\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b}\,. \tag{2.145}$$

The filter $\mathbf{w}[k]$ can now be decomposed into a fixed (also called quiescent) part $\mathbf{w}_q$ and a variable part $\mathbf{w}_a[k]$,

$$\mathbf{w}[k] = \underbrace{\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b}}_{\mathbf{w}_q} - \mathbf{C}_a^T\mathbf{w}_a[k]\,, \tag{2.146}$$

such that the constraint and the minimisation problem are naturally separated and the constrained minimisation of $\mathbf{w}[k]$ is equivalent to the unconstrained minimisation of $\mathbf{w}_a[k]$ [25][116][123]. This Griffiths-Jim structure is depicted in Fig. 2.9, with $\mathbf{w}_q$ the *fixed beamformer* and $\mathbf{C}_a$ the *blocking matrix*, which is orthogonal to $\mathbf{w}_q$.

When the constraint is a predefined frequency response $F(\omega)$ in the look direction, cf. (2.136), the fixed beamformer $\mathbf{w}_q$ is equal to

$$\mathbf{w}_q = \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b} = \frac{1}{N}\begin{bmatrix} \mathbf{I}_L & \mathbf{I}_L & \dots & \mathbf{I}_L \end{bmatrix}^T\mathbf{b}\,, \tag{2.147}$$

which corresponds to summing the microphone signals $y_n[k]$ and filtering with $f[k]$, while the $(N-1)L \times NL$-dimensional blocking matrix $\mathbf{C}_a$ e.g. is equal to

$$\mathbf{C}_a = \begin{bmatrix} \mathbf{I}_L & -\mathbf{I}_L & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{I}_L & \mathbf{0} & -\mathbf{I}_L & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{I}_L & \mathbf{0} & \mathbf{0} & \dots & -\mathbf{I}_L \end{bmatrix}\,, \tag{2.148}$$

which effectively corresponds to first creating $N-1$ signals by subtracting the microphone signals $y_n[k]$, $n = 1\dots N-1$, from $y_0[k]$ and applying an L-dimensional filter $w_{an}[k]$ on each of these $N-1$ signals. The resulting beamformer structure is also called a *Generalised Sidelobe Canceller* (GSC) and is depicted in Fig. 2.10. The GSC consists of three parts:

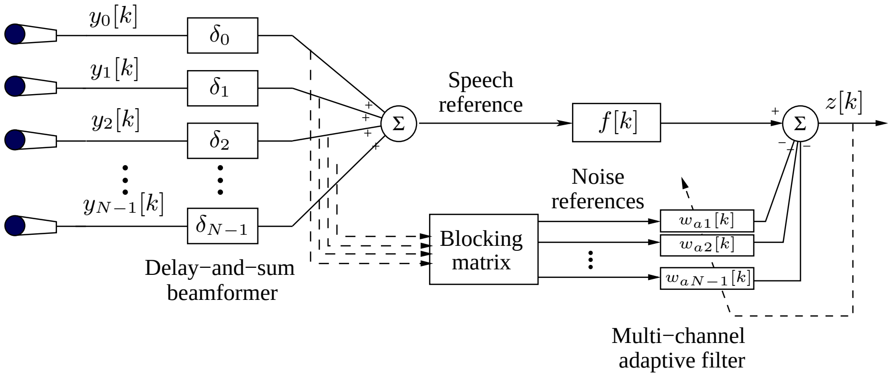


Figure 2.10: Generalised Sidelobe Canceller (GSC) structure

- a fixed DS beamformer, which spatially aligns the microphone signals to the direction of the speech source and which creates a so-called *speech reference* signal. This speech reference signal can be filtered with the constraint filter $f[k]$.

- a blocking matrix, usually orthogonal to the fixed beamformer, creating so-called *noise reference* signals by forming spatial zeros in the direction of the speech source. Maximally $N-1$ independent noise reference signals can be created. Generally we will use the Griffiths-Jim blocking matrix, creating the $N-1$ noise reference signals as

$$\mathbf{r}_{noise}^{GSC}[k] = \begin{bmatrix} y_0[k-\delta_0] - y_1[k-\delta_1] \\ y_0[k-\delta_0] - y_2[k-\delta_2] \\ \vdots \\ y_0[k-\delta_0] - y_{N-1}[k-\delta_{N-1}] \end{bmatrix} . \qquad (2.149)$$

- an *adaptive noise cancellation (ANC)* stage, using a multi-channel adaptive filter, which removes the remaining correlation between the (residual) noise component in the speech reference signal and the noise reference signals. If the noise components in the microphone signals are correlated, then the adaptive filter can reduce a considerable amount of noise from the speech reference signal. A GSC will therefore perform considerably better for highly correlated noise than for uncorrelated noise [17]. Generally we will use a time-domain NLMS adaptive algorithm (cf. page 66) for adapting the filter coefficients and we will take $f[k] = \delta[k - \frac{L_{ANC}}{2}]$, with L_{ANC} the length of the adaptive filters, such that some acausal taps can be modelled by the adaptive filter.

Frequency-domain representation

In the frequency-domain, the fixed beamformer is represented by the N-dimensional vector $\mathbf{W}_q(\omega)$, whereas the blocking matrix is represented by the $J \times N$-dimensional matrix $\mathbf{C}_a(\omega)$, with $J < N$ (usually $J = N-1$). The speech reference signal then is equal to $\mathbf{W}_q^H(\omega)\mathbf{Y}(\omega)$, whereas the noise reference signals are equal to $\mathbf{Y}_a(\omega) = \mathbf{C}_a(\omega)\mathbf{Y}(\omega)$. The J-dimensional filter $\mathbf{W}_a(\omega)$ which minimises the cost function

$$\min_{\mathbf{W}_a(\omega)} \mathcal{E}\{|\mathbf{W}_q^H(\omega)\mathbf{Y}(\omega) - \mathbf{W}_a^H(\omega)\mathbf{Y}_a(\omega)|^2\} \tag{2.150}$$

is equal to

$$\mathbf{W}_a(\omega) = \left[\mathbf{C}_a(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{C}_a^H(\omega)\right]^{-1}\mathbf{C}_a(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{W}_q(\omega)\ , \tag{2.151}$$

with $\bar{\mathbf{R}}_{yy}(\omega) = \mathcal{E}\{\mathbf{Y}(\omega)\mathbf{Y}^H(\omega)\}$. If we assume that no signal leakage is present in the noise references, i.e. $\mathbf{C}_a(\omega)\mathbf{X}(\omega) = \mathbf{0}$ and hence $\mathbf{C}_a(\omega)\bar{\mathbf{R}}_{xx}(\omega) = \mathbf{0}$ (or that the filters $\mathbf{W}_a(\omega)$ are only calculated during noise-only periods) and if we assume a homogeneous noise field, i.e. $\bar{\mathbf{R}}_{vv}(\omega) = P_v(\omega)\mathbf{\Gamma}_v(\omega)$, then

$$\mathbf{W}_a(\omega) = \left[\mathbf{C}_a(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{C}_a^H(\omega)\right]^{-1}\mathbf{C}_a(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}_q(\omega)\ . \tag{2.152}$$

The overall filter $\mathbf{W}(\omega)$ can then be written as

$$\begin{aligned}\mathbf{W}(\omega) &= \mathbf{W}_q(\omega) - \mathbf{C}_a^H(\omega)\mathbf{W}_a(\omega) \qquad (2.153)\\ &= \Big[\mathbf{I}_N - \mathbf{C}_a^H(\omega)\left[\mathbf{C}_a(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{C}_a^H(\omega)\right]^{-1}\mathbf{C}_a(\omega)\mathbf{\Gamma}_v(\omega)\Big]\mathbf{W}_q(\omega)\ ,\end{aligned}$$

which *only depends on the spatial characteristics of the noise field.* If the blocking matrix and the fixed beamformer are orthogonal, i.e. $\mathbf{C}_a(\omega)\mathbf{W}_q(\omega) = \mathbf{0}$, and if $\mathbf{C}_a^H(\omega)$ and $\mathbf{W}_q(\omega)$ span the entire N-dimensional space, i.e. $J = N-1$, then it has been proved in [139] that

$$\begin{aligned}&\mathbf{I}_N - \mathbf{C}_a^H(\omega)\left[\mathbf{C}_a(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{C}_a^H(\omega)\right]^{-1}\mathbf{C}_a(\omega)\mathbf{\Gamma}_v(\omega) = \\ &\qquad \mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)\left[\mathbf{W}_q^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)\right]^{-1}\mathbf{W}_q^H(\omega)\ ,\end{aligned} \tag{2.154}$$

such that the filter $\mathbf{W}(\omega)$ can be written as

$$\boxed{\mathbf{W}(\omega) = \frac{\mathbf{W}_q^H(\omega)\mathbf{W}_q(\omega)}{\mathbf{W}_q^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)}\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)} \tag{2.155}$$

which is equal to the superdirective beamformer (2.128) if $\mathbf{W}_q(\omega) = \mathbf{d}(\omega, \theta_x)/N$, i.e. if the fixed beamformer is a DS beamformer.

If we also assume that the speech field is homogeneous, i.e. that the PSD of the speech components $P_{x_n}(\omega) = P_x(\omega)$, $n = 0 \dots N-1$, then $P_{y_n}(\omega) =$

$P_y(\omega) = P_x(\omega) + P_v(\omega)$, $n = 0 \ldots N-1$, such that $\bar{\mathbf{R}}_{yy}(\omega) = P_y(\omega)\mathbf{\Gamma}_y(\omega)$ and $\bar{\mathbf{R}}_{xx}(\omega) = P_x(\omega)\mathbf{\Gamma}_x(\omega)$, with the coherence matrices $\mathbf{\Gamma}_y(\omega)$ and $\mathbf{\Gamma}_x(\omega)$ similarly defined as in (2.37). Using $Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega)$, the PSD $P_z(\omega)$ of the output signal is equal to

$$P_z(\omega) = \mathbf{W}^H(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{W}(\omega) = P_y(\omega)\mathbf{W}^H(\omega)\mathbf{\Gamma}_y(\omega)\mathbf{W}(\omega) \ , \tag{2.156}$$

such that, using (2.153), the PTF of the speech component is equal to

$$G_{x_0 z_x}(\omega) = \frac{P_{z_x}(\omega)}{P_{x_0}(\omega)} = \mathbf{W}^H(\omega)\mathbf{\Gamma}_x(\omega)\mathbf{W}(\omega) = \mathbf{W}_q^H(\omega)\mathbf{\Gamma}_x(\omega)\mathbf{W}_q(\omega) \ . \tag{2.157}$$

and that, using (2.155), the PTF of the noise component is equal to

$$G_{v_0 z_v}(\omega) = \frac{P_{z_v}(\omega)}{P_{v_0}(\omega)} = \mathbf{W}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}(\omega) = \frac{\left[\mathbf{W}_q^H(\omega)\mathbf{W}_q(\omega)\right]^2}{\mathbf{W}_q^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)} \ . \tag{2.158}$$

Variants of standard GSC-implementation

Instead of using a simple DS beamformer and scalar elements in the blocking matrix, it is also possible to use more advanced (fixed) filter-and-sum beamformers with FIR filters (cf. Section 2.5.2), in order to provide better spatial and spectral control for the speech and the noise references [34][191]. Generally, the blocking matrix and the fixed beamformer are designed to be orthogonal to each other [191].

In the standard GSC-implementation, it is also assumed that no speech components are present in the noise reference signals. However, this is only true when no signal reflections (reverberation) occur, when the direction of the speech source is exactly known (or is correctly estimated) and when the microphone characteristics (gain, phase, position) do not deviate from the assumed characteristics. Clearly, in a practical implementation these conditions will not be satisfied. Therefore, *signal leakage* will occur in the noise references and the adaptive filter will also remove part of the speech component from the speech reference signal, leading to signal distortion or even signal cancellation. In order to limit signal leakage and the resulting signal distortion, different variants of the standard GSC-implementation have been proposed, which are based on:

- *reducing the amount of signal leakage*, e.g. by using a spatial filter designed blocking matrix [191][194], creating better speech and (especially) noise references covering a region around the speaker location. Although the amount of signal leakage into the noise reference will be reduced (certainly when dealing with microphone mismatch, look direction error or spatially distributed sources), it can never be completely avoided (certainly not in highly reverberant acoustic environments).

- *limiting the effect of the signal leakage on the adaptive filters*, e.g. by using a speech-controlled (VAD) adaptation algorithm, where the adaptive filter is only allowed to adapt during noise-only periods [46][113][128][194] [254], by using norm-constrained [37] and coefficient-constrained adaptive filters [128], or by using a leaky-LMS algorithm. In case of using a speech-controlled adaptation algorithm with a perfect VAD, signal leakage will have no effect on the adaptive filters, but of course signal distortion will still occur (since speech components are present in the noise reference).

In the remainder of the thesis, we will use a standard GSC-implementation with a Griffiths-Jim blocking matrix and a GSC-implementation with a spatial filter designed blocking matrix. In both implementations, we will use a speech-controlled adaptation algorithm, switching off the adaptation during speech-and-noise periods, since signal leakage can never be completely avoided.

NLMS-algorithm

The *least-mean-squares (LMS) algorithm*, originally proposed by Widrow and Hoff, is a simple but effective method for adapting the coefficients of an adaptive FIR filter [123][277]. Consider Fig. 1.3, with $d[k]$ the desired signal, $x[k]$ the input signal of the adaptive filter $w[k]$ and $y[k] = w[k] \otimes x[k]$. If we use an FIR filter $\mathbf{w}[k]$ of length L, then the output signal $y[k]$ at time k can be written as

$$y[k] = \mathbf{w}^T[k]\mathbf{x}[k] \ , \tag{2.159}$$

with $\mathbf{x}[k] = \begin{bmatrix} x[k] & x[k-1] & \ldots & x[k-L+1] \end{bmatrix}^T$. The goal of any adaptive filtering algorithm is to minimise the average energy of the error signal $e[k] = d[k] - y[k]$. By using a gradient-descent algorithm and approximating the gradient by its instantaneous value, the LMS algorithm is obtained as

$$\boxed{\mathbf{w}[k+1] = \mathbf{w}[k] + \mu\mathbf{x}[k]e[k]} \tag{2.160}$$

with μ the step size of the adaptive algorithm. This step size controls the convergence speed and the stability of the adaptive filter. It can be shown that the LMS algorithm is stable if $0 \leq \mu < 2/\lambda_{max}$, with λ_{max} the largest eigenvalue of the correlation matrix $\bar{\mathbf{R}}_{xx}[k] = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^T[k]\}$. It can also be shown that the eigenvalue spread $\lambda_{max}/\lambda_{min}$, with λ_{min} the smallest eigenvalue of $\bar{\mathbf{R}}_{xx}[k]$, essentially determines the convergence speed. Therefore the LMS algorithm has a better convergence speed for white noise ($\lambda_{max}/\lambda_{min} = 1$) than for speech-like signals with a large eigenvalue spread. In order to obtain a better convergence speed for signals with a large eigenvalue spread, other adaptive algorithms, such as RLS [123] and APA [103][186][203][222][249] should be used. These algorithms however tend to have a large computational complexity, certainly for large filter lengths L. Because of its small computational complexity and its simplicity, the LMS algorithm therefore still is a commonly used adaptive algorithm in acoustic applications, despite its slower convergence speed.

Since the step size of the LMS algorithm is dependent on the signal statistics, generally the *normalised LMS (NLMS) algorithm* is used, where an implicit normalisation of the step size is performed with the signal energy, i.e.

$$\boxed{\mathbf{w}[k+1] = \mathbf{w}[k] + \frac{\mu}{\mathbf{x}^T[k]\mathbf{x}[k] + \alpha}\mathbf{x}[k]e[k]} \tag{2.161}$$

with α a small constant in order to prevent division by zero in case no input signal is present. Since $\lambda_{max} \leq \sum_{i=1}^{L} \lambda_i = \mathcal{E}\{\mathbf{x}^T[k]\mathbf{x}[k]\}$, the NLMS algorithm is stable if $0 \leq \mu < 2$.

The NLMS adaptive filter can also be used with multiple input signals, as e.g. in the ANC stage of the GSC (see Fig. 2.10). In that case the filter $\mathbf{w}[k]$ is the stacked filter vector of all adaptive filters $\mathbf{w}_{an}[k]$, $n = 1 \ldots N-1$, while the vector $\mathbf{x}[k]$ is the stacked data vector of all data vectors $\mathbf{x}_{an}[k]$, $n = 1 \ldots N-1$. Only one error signal $z[k]$ is used for updating all adaptive filters.

2.6 Multi-microphone dereverberation

As has been indicated in Section 2.4, single-microphone dereverberation techniques have a limited scope in practice, even when the complete acoustic impulse response between the speech source and the microphone is known. Using multi-microphone fixed beamforming techniques, discussed in Section 2.5.2, it is possible to obtain spatial focusing on the speech source, thereby reducing some reverberation. However, since fixed beamforming techniques only take into account the direct path of the acoustic impulse responses, their dereverberation performance is limited, especially in highly reverberant acoustic environments. If the complete acoustic impulse responses are known, more advanced multi-microphone techniques can be used, such as inverse filtering (cf. Section 2.6.1) and matched filtering (cf. Section 2.6.2). The acoustic impulse responses can either be measured or can be estimated by blind system identification techniques in the time-domain and in the frequency-domain. These blind system identification techniques will be discussed in Chapters 6 and 7.

2.6.1 Inverse filtering

In Section 2.4.1 it has been shown that in the single-microphone case generally no perfect dereverberation can be obtained using inverse filtering techniques, even when the filter length L goes to infinity. However, *in the multi-microphone case, perfect dereverberation is always possible if the acoustic impulse responses are known* (even if the impulse responses are non-minimum-phase) [180]. Assuming that no background noise is present[6], the L-dimensional vector $\mathbf{y}_n[k]$

[6] Dereverberation when (coloured) noise is present and combined noise reduction and dereverberation will be discussed in Chapter 7.

can be written, using (2.97), as

$$\mathbf{y}_n[k] = \mathbf{H}_n[k]\,\mathbf{s}[k] \;, \tag{2.162}$$

with $\mathbf{H}_n[k]$ an $L \times (K + L - 1)$-dimensional Toeplitz matrix, consisting of the coefficients of the K-dimensional acoustic impulse response $\mathbf{h}_n[k]$, and $\mathbf{s}[k]$ a $(K + L - 1)$-dimensional vector, consisting of the clean speech samples. The M-dimensional stacked data vector $\mathbf{y}[k]$ can then be written as

$$\mathbf{y}[k] = \begin{bmatrix} \mathbf{H}_0[k] \\ \mathbf{H}_1[k] \\ \vdots \\ \mathbf{H}_{N-1}[k] \end{bmatrix} \mathbf{s}[k] = \mathcal{H}[k]\,\mathbf{s}[k] \;, \tag{2.163}$$

with $\mathcal{H}[k]$ an $M \times (K + L - 1)$-dimensional matrix. Dereverberation consists of computing an M-dimensional vector $\mathbf{w}[k]$, such that

$$\mathbf{w}^T[k]\mathbf{y}[k] = s[k],\, \forall k \;, \tag{2.164}$$

i.e.

$$\mathbf{w}^T[k]\mathcal{H}[k] = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} = \mathbf{d}^T \;. \tag{2.165}$$

If we assume that $\mathcal{H}[k]$ has full column rank, i.e. the acoustic impulse responses do not have common zeros, it is clear that this set of equations can always be solved when $M \geq K + L - 1$, i.e.

$$L \geq \frac{K-1}{N-1} \;, \tag{2.166}$$

and the solution is given by

$$\boxed{\mathbf{w}_{LS}[k] = \left(\mathcal{H}^\dagger[k]\right)^T \mathbf{d} = \mathcal{H}[k]\left(\mathcal{H}^T[k]\mathcal{H}[k]\right)^{-1}\mathbf{d}} \tag{2.167}$$

Obviously, the condition (2.166) can never be fulfilled in the single-microphone case ($N = 1$). Although this condition can always be fulfilled in the multi-microphone case, numerical problems may occur when calculating the pseudo-inverse of $\mathcal{H}[k]$ and since the solution in (2.167) is quite sensitive to the accuracy of the measured/estimated acoustic impulse responses (in Chapter 6 it will be shown that it is quite difficult to accurately identify the complete acoustic impulse responses in practice, especially when a large amount of background noise is present).

2.6.2 Matched filtering

In [91] a multi-microphone matched filtering technique for dereverberation has been presented. This technique requires the acoustic impulse responses to be (partially) known and is less sensitive to the accuracy of the estimated impulse responses than inverse filtering. However, using this matched filtering technique perfect dereverberation can never be obtained and a pre-echo problem occurs.

In [91] it is assumed that we can model the acoustic impulse responses $h_n[k]$ as consisting of a direct path contribution and I images (reflections), all having a comparable amplitude. The DS-beamformer, discussed in Section 2.5.2, produces a beam in the direction of the speech source by spatially aligning the direct path contributions of the microphone signals and hence attenuating signals coming from other directions. The output of the DS beamformer provides N coherent arrivals (direct path contributions) and IN incoherent arrivals (images), which are distributed in time and typically add powerwise. The direct-to-reverberant energy ratio (DR) for the DS beamformer is equal to

$$\mathrm{DR} = \frac{N^2}{NI} = \frac{N}{I}\ , \tag{2.168}$$

which decreases monotonically with the number of images I, i.e. for highly reverberant acoustic environments the dereverberation performance of the DS beamformer is limited.

In [91] a *multiple beamformer* is proposed, which not only produces a beam in the direction of the speech source (direct path), but also in the direction of B major images. Hence, the output of the multiple beamformer provides $(B+1)N$ coherent arrivals and $(B+1)IN$ incoherent arrivals. The direct-to-reverberant energy ratio for the multiple beamformer is equal to

$$\mathrm{DR} = \frac{[(B+1)N]^2}{(B+1)IN} = \frac{(B+1)N}{I}\ . \tag{2.169}$$

In the *matched filtering* technique, the filters $w_n[k]$ on the microphone signals are equal to the time-reversed acoustic impulse responses $h_n[-k]$ (which therefore need to be known), such that the output signal $z[k]$ is equal to

$$z[k] = \sum_{n=0}^{N-1} h_n[-k] \otimes y_n[k] = \sum_{n=0}^{N-1} h_n[-k] \otimes h_n[k] \otimes s[k]\ , \tag{2.170}$$

which in the frequency-domain corresponds to

$$Z(\omega) = \sum_{n=0}^{N-1} H_n^*(\omega) Y_n(\omega) = \|\mathbf{H}(\omega)\|^2 S(\omega)\ . \tag{2.171}$$

As can be clearly seen from these equations, no perfect dereverberation can be obtained using this matched filtering technique. In Section 7.3.1 it will be shown that using the normalised matched filter $\mathbf{H}(\omega)/\|\mathbf{H}(\omega)\|^2$ perfect dereverberation is obtained. The direct-to-reverberant energy ratio is equal to

$$\mathrm{DR} = \frac{[(I+1)N]^2}{(I+1)IN} = \frac{(I+1)N}{I} \approx N\ , \tag{2.172}$$

if the number of images I is large enough.

In Fig. 2.11a typical acoustic impulse responses $h_n[k]$ are depicted, which have been constructed using the image method (cf. Section 1.3.3) with $N = 4$, $T_{60} = 300\,\text{ms}$, $f_s = 8\,\text{kHz}$ and $K = 1000$. Using the inverse filtering technique, perfect dereverberation can be obtained with filters $w_n[k]$ having a filter length $L = 333$, cf. (2.166). These inverse filters are depicted in Fig. 2.11b. Note that the filter coefficients of the inverse filters have quite large amplitudes, implying that this inverse filtering technique is quite sensitive to errors (e.g. estimation errors of the acoustic impulse responses $h_n[k]$).

Figures 2.12a and 2.12b depict the matched filters $w_n[k] = h_n[-k]$ with filter length $L = 1000$ and the total transfer function $f[k]$ for the speech component. As can be clearly seen from Fig. 2.12b, no perfect dereverberation is obtained, and also a pre-echo phenomenon occurs, i.e. a long impulse response tail is present *before* the main peak, as indicated in this figure. In order to reduce this pre-echo effect, the matched filters can be truncated [216]. Figure 2.13a depicts the matched filters which have been truncated to $L = 80$, and Fig. 2.13b depicts the total transfer function $f[k]$ using these truncated matched filters. As can be seen from Fig. 2.13b, the pre-echo is reduced, but perfect dereverberation can never be obtained using matched filtering techniques (with or without truncation).

2.7 Conclusion

In this chapter, we have discussed the noise reduction problem and the dereverberation problem from a mathematical point of view and we have briefly described several existing single- and multi-microphone signal enhancement techniques, both in the time-domain and in the frequency-domain.

In Section 2.2 we have described the recording model for speech signals in noisy acoustic environments, making a distinction between additive and convolutional noise (i.e. reverberation). Additive noise may consist of contributions from both unknown and known noise sources, e.g. far-end echo. The goal of any signal enhancement algorithm is to compute the filters $w_n[k], n = 0 \ldots N-1$, on the microphone signal(s) with either noise reduction, dereverberation or combined noise reduction and dereverberation as the objective. In this section we have also discussed a frequency-domain representation and we have defined performance measures for the signal enhancement algorithms, such as unbiased SNR improvement, average speech distortion and a dereverberation index.

In Section 2.3 two single-microphone noise reduction techniques have been discussed: spectral subtraction and signal subspace-based techniques. Both techniques only exploit the temporal and the spectral information of the speech and the noise signals. In spectral subtraction techniques the DFT-coefficients

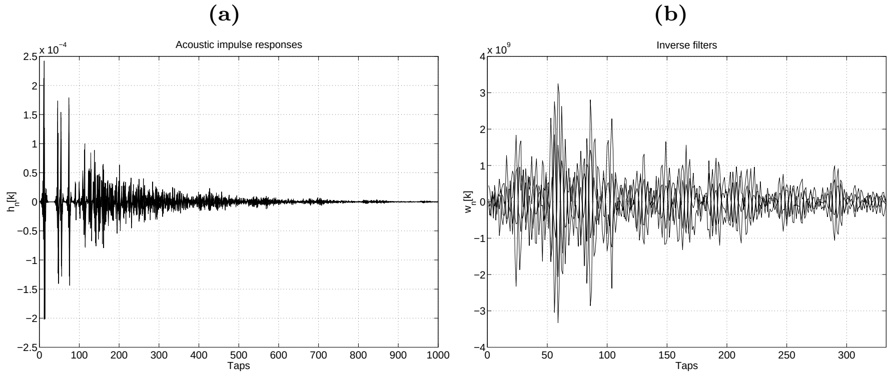


Figure 2.11: (**a**) Acoustic impulse responses $h_n[k]$ ($N = 4$, $T_{60} = 300\,\mathrm{ms}$, $f_s = 8\,\mathrm{kHz}$, $K = 1000$), (**b**) Inverse filters $w_n[k]$ ($L = 333$)

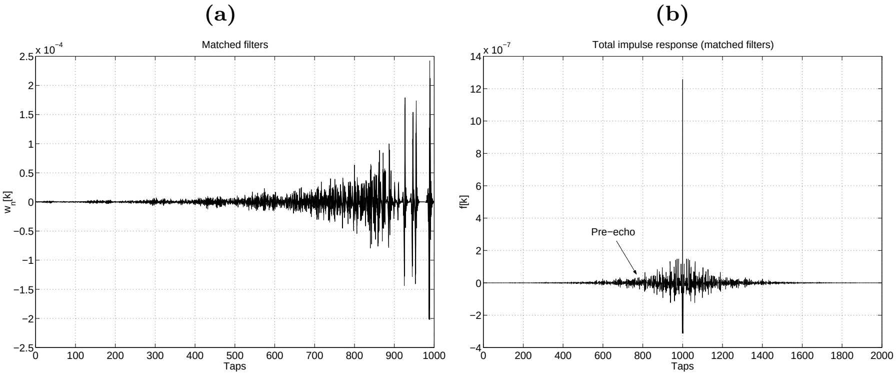


Figure 2.12: (**a**) Matched filters $h_n[-k]$ ($L = 1000$), (**b**) Total impulse response $f[k]$ for the speech component using matched filters

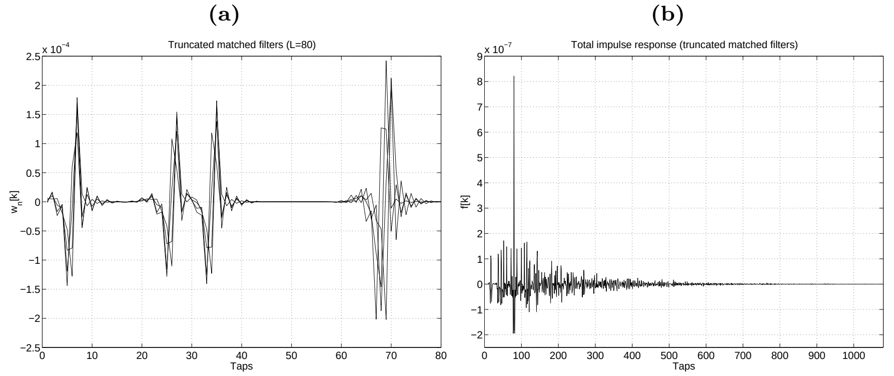


Figure 2.13: (**a**) Truncated matched filters $h_n[-k]$ ($L = 80$), (**b**) Total impulse response $f[k]$ for the speech component using truncated matched filters

are multiplied with a noise-dependent gain, whereas in signal subspace-based techniques the KLT-coefficients are modified. Since both techniques need an estimate of the noise characteristics, a VAD algorithm is required. Signal subspace-based techniques assume that the clean speech signal can be modelled with a low-rank model and perform signal enhancement by removing the noise subspace and by estimating the clean speech signal from the remaining signal subspace, using a LS or a MV estimator. Both estimators can be represented as an FIR eigenfilterbank (in the white noise case and in the coloured noise case). Actually, it can be proved that the signal-independent spectral subtraction techniques and the signal-dependent subspace-based techniques asymptotically produce the same result when the frame length goes to infinity and when the speech and the noise signals are assumed to be stationary. In Part I the presented signal subspace-based techniques will be extended to the multi-microphone case.

In Section 2.4 two single-microphone dereverberation techniques have been discussed: inverse filtering, requiring the acoustic impulse response to be known, and cepstrum-based techniques, not requiring any prior knowledge about the impulse response. In practice, single-microphone inverse filtering has a limited scope since the acoustic impulse response typically is non-minimum-phase, while cepstrum-based techniques also have a limited performance since the cepstrum of the clean speech signal and the acoustic impulse response typically have a large overlap.

In Section 2.5 fixed and adaptive beamforming techniques for multi-microphone noise reduction have been discussed. Fixed beamformers are data-independent and try to obtain spatial focusing on the speech source, thereby reducing reverberation and suppressing noise not coming from the direction of the speech source. We have discussed several types of fixed beamformers: the simple – but still widely used – delay-and-sum beamformer; first-order differential microphones using 2 closely spaced microphones which are delayed and subtracted; superdirective beamformers maximising the directivity index for a known noise field; and filter-and-sum beamformers, which will be discussed in more detail in Part III. Adaptive beamformers combine the spatial focusing of fixed beamformers with adaptive noise suppression, such that they are able to adapt to changing acoustic environments and generally exhibit a better noise reduction performance than fixed beamformers. We have discussed the LCMV-beamformer, which minimises the output power under the constraint that signals from the direction of the speech source are not distorted. This constrained LCMV optimisation problem can be reformulated as an unconstrained optimisation problem, resulting in the Generalised Sidelobe Canceller, which consists of a fixed beamformer, a blocking matrix and a multi-channel adaptive filter (e.g. NLMS). Several variants of the standard GSC have been discussed which reduce the amount of signal leakage and/or limit the effect of the signal leakage on the adaptive filters. The performance of the GSVD-based optimal filtering techniques in Part I will be compared with the performance of these fixed and adaptive beamformers.

In Section 2.6 two multi-microphone dereverberation techniques have been discussed: inverse filtering and matched filtering. Both techniques require the acoustic impulse responses to be (partially) known. Using the inverse filtering technique perfect dereverberation is possible. However, this technique is quite sensitive to the accuracy of the measured/estimated acoustic impulse responses. In the matched filtering technique the microphone signals are filtered with the time-reversed acoustic impulse responses. This technique is less sensitive to the accuracy of the impulse responses, but no perfect dereverberation can be obtained. In addition, a pre-echo problem occurs, which can be limited by truncating the matched filters. This matched filtering technique forms the basis for the frequency-domain dereverberation and combined noise reduction and dereverberation technique discussed in Part II.

Part I

GSVD-Based Optimal Filtering for Multi-Microphone Noise Reduction

Chapter 3

GSVD-Based Optimal Filtering for Single and Multi-Microphone Speech Enhancement

In this chapter a Generalised Singular Value Decomposition (GSVD) based algorithm is discussed for enhancing multi-microphone speech signals degraded by additive coloured noise. This GSVD-based algorithm is a specific implementation of the multi-channel Wiener filter, taking into account the low-rank model of the speech signal, and can be considered an extension of the single-microphone signal subspace-based algorithms (cf. Section 2.3.2), now combining the spatio-temporal information of the speech and the noise sources.

In Section 3.2 unconstrained optimal filtering for enhancing multi-microphone noisy speech signals is described. The MMSE estimator, i.e. the multi-channel Wiener filter, as well as a more general class of estimators is discussed. Section 3.3 discusses the practical implementation using a GSVD and it is shown that the optimal filter matrix can be written as a function of the generalised singular vectors and singular values of a so-called speech and noise data matrix. In Section 3.4 a number of symmetry properties are derived for the single-microphone and the multi-microphone optimal filter, which are valid for the white noise case as well as for the coloured noise case. In addition, the averaging step of some single-microphone signal subspace-based algorithms is re-examined, leading to the conclusion that this averaging operation is unnecessary and typically even suboptimal. In Section 3.5 a frequency-domain analysis is given for the unconstrained optimal filtering technique, showing that the optimal filter can be decomposed into a spectral and a spatial filtering term, and in Section 3.6 it is shown that this multi-microphone signal enhancement technique can also be used for combined noise reduction and echo cancellation.

3.1 Introduction

In Section 2.3.2 single-microphone signal subspace-based speech enhancement techniques have been discussed. The main idea is to consider the noisy signal in a vector space and to separate this vector space into 2 orthogonal subspaces: the signal subspace and the noise subspace. Signal enhancement is then performed by removing the noise subspace and by estimating the clean speech signal from the remaining signal subspace. Depending on the specific optimisation criterion, different estimates for the clean speech signal can be obtained. In Section 2.3.2 the least-squares (LS) and the minimum-variance (MV) estimator have been discussed for the single-microphone case.

Single-microphone subspace-based speech enhancement techniques only use the spectral information present in the microphone signal. These techniques can be viewed as a (signal-dependent) frequency filtering operation on the noisy speech signal [120], which adaptively extracts the most important, i.e. most energetic, formants of the speech signal, thereby reducing the background noise. When multiple microphones are available, both the spectral and the spatial characteristics of the speech and the noise sources can be exploited. In the literature, signal subspace-based algorithms have already been used for processing multi-channel signals. Hansen [121] suggests to use a single-channel subspace-based speech enhancement algorithm on each microphone signal separately, followed by a delay-and-sum beamformer. Jabloun and Champagne [135] exploit the multi-microphone information to design a (single-channel) signal subspace-based post-filter, following a delay-and-sum beamformer. However these techniques cannot be considered integrated multi-microphone subspace-based speech enhancement techniques. Subspace-based techniques have been used for processing (multi-channel) images and biomedical signals in [69][247], but for speech applications these procedures do not allow to exploit the spatial information present in the multi-microphone signals.

In this chapter we present a multi-microphone extension of the single-microphone subspace-based speech enhancement techniques, combining the spatio-temporal information of the speech and the noise sources. For the multi-microphone case, we mainly consider the optimal estimator (in the MSE sense), which produces an MMSE estimate of the speech component in one of the microphone signals, but also a related estimator that trades off speech distortion and noise reduction. Since speech components in the microphone signals are estimated, no dereverberation will be achieved and inevitably some (linear) speech distortion will be introduced. It will be shown that the optimal filter can be written as a function of the generalised singular vectors and values of a speech and a noise data matrix, where the specific function used provides a means to trade off noise reduction and speech distortion. When analysing the multi-channel optimal filter in the frequency domain, it will be shown that this filter can indeed be decomposed into a spatial and a spectral filtering term.

3.2 Unconstrained optimal filtering

In this section we discuss the unconstrained optimal filtering technique for multi-microphone speech enhancement. The optimal filter (in the MSE sense) is the multi-channel Wiener filter, which produces an MMSE estimate for different delayed versions of the speech components in the microphone signals. By using the generalised eigenvalue decomposition (GEVD) of the speech and the noise correlation matrices, the low-rank model of the speech signal can be easily taken into account, such that this signal enhancement technique can be considered a multi-microphone extension of the single-microphone signal subspace-based techniques. We also discuss a more general class of estimators, which trades off noise reduction and speech distortion and for which the filter parameters are also obtained from the GEVD of the correlation matrices.

3.2.1 Multi-channel Wiener filter

Data Model

Consider again Fig. 2.1, which depicts the general setup for multi-microphone speech enhancement using N microphones. Each microphone signal $y_n[k]$, $n = 0 \ldots N-1$, consists of the filtered speech signal $s[k]$ and additive noise,

$$y_n[k] = h_n[k] \otimes s[k] + v_n[k] = x_n[k] + v_n[k] \ , \tag{3.1}$$

with $x_n[k]$ and $v_n[k]$ the speech and the noise component of the nth microphone signal. The additive noise is assumed to be uncorrelated with the speech signal. As shown in (2.22), the output signal $z[k]$ at time k can be written as

$$z[k] = \sum_{n=0}^{N-1} \mathbf{w}_n^T[k]\mathbf{y}_n[k] = \mathbf{w}^T[k]\mathbf{y}[k] \ , \tag{3.2}$$

with the L-dimensional filter vector $\mathbf{w}_n[k]$ and data vector $\mathbf{y}_n[k]$ equal to

$$\begin{aligned} \mathbf{w}_n[k] &= \begin{bmatrix} w_{n,0}[k] & w_{n,1}[k] & \ldots & w_{n,L-1}[k] \end{bmatrix}^T \ , & (3.3) \\ \mathbf{y}_n[k] &= \begin{bmatrix} y_n[k] & y_n[k-1] & \ldots & y_n[k-L+1] \end{bmatrix}^T \ , & (3.4) \end{aligned}$$

and the M-dimensional stacked filter vector $\mathbf{w}[k]$ and stacked data vector $\mathbf{y}[k]$, with $M = LN$, equal to

$$\mathbf{w}[k] = \begin{bmatrix} \mathbf{w}_0^T[k] & \mathbf{w}_1^T[k] & \ldots & \mathbf{w}_{N-1}^T[k] \end{bmatrix}^T \ , \tag{3.5}$$

$$\mathbf{y}[k] \;=\; \left[\begin{array}{cccc} \mathbf{y}_0^T[k] & \mathbf{y}_1^T[k] & \ldots & \mathbf{y}_{N-1}^T[k] \end{array}\right]^T . \qquad (3.6)$$

The goal of multi-microphone speech enhancement is to compute the filter vector $\mathbf{w}[k]$ such that the speech signal $s[k]$ or one of the speech components $x_n[k]$ is recovered. The GSC (LCMV beamformer) is formulated as a *constrained optimal filtering* problem, cf. (2.134), which attempts to recover the speech signal $s[k]$ by minimising the output energy and constraining the array response to unity in the direction of the speech source. In this section we will discuss an *unconstrained optimal filtering* problem, which 'optimally' estimates the speech components $x_n[k]$ from the noisy microphone signals $y_n[k]$.

Optimal filtering – MMSE estimation

Consider the filtering problem depicted in Fig. 3.1: $\mathbf{y}[k]$ is the M-dimensional filter input vector, $\mathbf{z}[k] = \mathbf{W}^T[k]\,\mathbf{y}[k]$ is the filter output vector with $\mathbf{W}[k]$ an $M \times M$ filter matrix. The M-dimensional vector $\mathbf{x}[k]$ is the desired response vector and $\mathbf{e}[k] = \mathbf{x}[k] - \mathbf{z}[k]$ is the estimation error vector. The optimal filter is defined as the filter that minimises the MSE (mean square error) cost function

$$\begin{aligned} J_{MSE}(\mathbf{W}[k]) &= \mathcal{E}\{||\mathbf{e}[k]||_2^2\} = \mathcal{E}\{||\mathbf{x}[k] - \mathbf{W}^T[k]\,\mathbf{y}[k]||_2^2\} \qquad (3.7)\\ &= \mathcal{E}\{\mathbf{x}^T[k]\mathbf{x}[k]\} - 2\mathcal{E}\{\mathbf{y}^T[k]\mathbf{W}[k]\mathbf{x}[k]\} + \mathcal{E}\{\mathbf{y}^T[k]\mathbf{W}[k]\mathbf{W}[k]^T\mathbf{y}[k]\}. \end{aligned}$$

The optimal filter matrix $\bar{\mathbf{W}}_{WF}[k]$ is found by setting the derivative

$$\frac{\partial J_{MSE}(\mathbf{W}[k])}{\partial \mathbf{W}[k]} = -2\mathcal{E}\{\mathbf{y}[k]\,\mathbf{x}[k]^T\} + 2\mathcal{E}\{\mathbf{y}[k]\,\mathbf{y}^T[k]\}\,\mathbf{W}[k] \qquad (3.8)$$

to zero and is equal to the well-known multi-dimensional Wiener filter [227],

$$\boxed{\bar{\mathbf{W}}_{WF}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{yx}[k]} \qquad (3.9)$$

with $\bar{\mathbf{R}}_{yy}[k] = \mathcal{E}\{\mathbf{y}[k]\,\mathbf{y}^T[k]\}$ the $M \times M$ correlation matrix of the input signal and $\bar{\mathbf{R}}_{yx}[k] = \mathcal{E}\{\mathbf{y}[k]\,\mathbf{x}^T[k]\}$ the $M \times M$ cross-correlation matrix of the input and the desired signal. Note that for multiple microphones, both the correlation and the cross-correlation matrix contain spatio-temporal information.

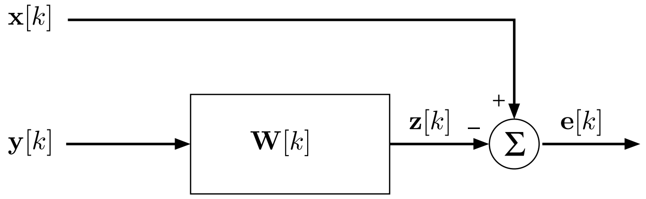


Figure 3.1: Optimal filtering problem with desired response vector $\mathbf{x}[k]$

When considering multi-microphone noisy speech signals, the input vector $\mathbf{y}[k]$ consists of the speech component and the additive noise component,

$$\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k] \; , \tag{3.10}$$

with $\mathbf{y}[k]$ defined in (3.6) and $\mathbf{x}[k]$ and $\mathbf{v}[k]$ similarly defined. Since the desired signal $\mathbf{x}[k]$ is an unobservable signal, this poses a particular problem which may be solved based on the on/off characteristics of the speech signal, cf. Section 1.3.1. If we use a robust voice activity detection (VAD) algorithm [50] [250][260], noise-only observations can be made during speech pauses (denoted here with time index k'), where $\mathbf{y}[k'] = \mathbf{v}[k']$, which allows to estimate the spatio-temporal correlation properties of the noise signal. The output of the VAD-algorithm at time k is represented by $\zeta[k]$, where $\zeta[k] = 1$ represents a speech-and-noise observation and $\zeta[k] = 0$ represents a noise-only observation.

We now make two *assumptions*: we assume that the second-order statistics of the noise signal are sufficiently stationary such that the noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$, which can be estimated during noise-only periods, can also be used during subsequent speech-and-noise periods, i.e.

$$\bar{\mathbf{R}}_{vv}[k] = \mathcal{E}\{\mathbf{v}[k]\,\mathbf{v}^T[k]\} = \mathcal{E}\{\mathbf{v}[k']\,\mathbf{v}^T[k']\} = \bar{\mathbf{R}}_{vv}[k'] \; , \tag{3.11}$$

and secondly, we assume that the speech and the noise signals are statistically independent, implying that

$$\bar{\mathbf{R}}_{xv}[k] = \mathcal{E}\{\mathbf{x}[k]\,\mathbf{v}^T[k]\} = \mathbf{0} \; . \tag{3.12}$$

From the second assumption it is easily verified that

$$\bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{R}}_{xx}[k] + \bar{\mathbf{R}}_{vv}[k], \quad \bar{\mathbf{R}}_{yx}[k] = \bar{\mathbf{R}}_{xx}[k] \; , \tag{3.13}$$

such that the optimal filter matrix in (3.9) can be written as

$$\boxed{\bar{\mathbf{W}}_{WF}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k] \; \left(\bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k]\right)} \tag{3.14}$$

where $\bar{\mathbf{R}}_{yy}[k]$ is estimated during speech-and-noise periods and $\bar{\mathbf{R}}_{vv}[k]$ is estimated during noise-only periods.

By using the joint diagonalisation of the symmetric block-Toeplitz correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$, the low-rank model of the clean speech signal $s[k]$ can be easily taken into account (cf. Section 3.2.2) and one can also easily provide a trade-off between noise reduction and speech distortion (cf. Section 3.2.3)[1]. The joint diagonalisation, i.e. generalised eigenvalue decomposition

[1] Note that it is also possible to implement the multi-dimensional Wiener filter $\bar{\mathbf{W}}_{WF}[k]$ directly using (3.14) or using a QRD-based implementation (cf. Section 3.3.4). Using the GEVD should be considered one possible way to implement the multi-dimensional Wiener filter, enabling to easily incorporate a low-rank signal model, but is certainly not imperative.

(GEVD), of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ is defined as (cf. Appendix A.2)

$$\begin{cases} \bar{\mathbf{R}}_{yy}[k] & = & \bar{\mathbf{Q}}[k]\,\bar{\mathbf{\Lambda}}_y[k]\,\bar{\mathbf{Q}}^T[k] \\ \bar{\mathbf{R}}_{vv}[k] & = & \bar{\mathbf{Q}}[k]\,\bar{\mathbf{\Lambda}}_v[k]\,\bar{\mathbf{Q}}^T[k]\ , \end{cases} \tag{3.15}$$

with $\bar{\mathbf{Q}}[k]$ an $M \times M$-dimensional invertible, but not necessarily orthogonal, matrix and $\bar{\mathbf{\Lambda}}_y[k] = \text{diag}\{\bar{\sigma}_i^2[k]\}$, $i = 1 \ldots M$, and $\bar{\mathbf{\Lambda}}_v[k] = \text{diag}\{\bar{\eta}_i^2[k]\}$, $i = 1 \ldots M$. Substituting (3.15) into (3.14) gives an expression for the optimal filter matrix,

$$\boxed{\bar{\mathbf{W}}_{WF}[k] = \bar{\mathbf{Q}}^{-T}[k]\ \text{diag}\Big\{1 - \frac{\bar{\eta}_i^2[k]}{\bar{\sigma}_i^2[k]}\Big\}\ \bar{\mathbf{Q}}^T[k]} \tag{3.16}$$

In the spatio-temporally white noise case, the noise correlation matrix is equal to $\bar{\mathbf{R}}_{vv}[k] = \bar{\sigma}_v^2\,\mathbf{I}_M$, with $\bar{\sigma}_v^2$ the noise power. The matrix $\bar{\mathbf{Q}}[k]$ then reduces to an orthogonal matrix, such that $\bar{\mathbf{W}}_{WF}[k]$ is a symmetric matrix,

$$\bar{\mathbf{W}}_{WF}[k] = \bar{\mathbf{Q}}[k]\ \text{diag}\Big\{1 - \frac{\bar{\sigma}_v^2}{\bar{\sigma}_i^2[k]}\Big\}\ \bar{\mathbf{Q}}^T[k]\ . \tag{3.17}$$

The enhanced speech vector $\hat{\mathbf{x}}[k] = \mathbf{z}[k]$ is obtained as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{W}}_{WF}^T[k]\,\mathbf{y}[k]\ , \tag{3.18}$$

such that the M-dimensional vector $\hat{\mathbf{x}}[k]$ contains an estimate for all the speech samples $x_n[k-l]$, $n = 0 \ldots N-1$, $l = 0 \ldots L-1$, i.e. for all L delayed versions of the speech components in all N microphone signals. The ith element of $\hat{\mathbf{x}}[k]$, which is obtained by filtering the microphone signals with the ith column $\bar{\mathbf{w}}_{WF,i}[k]$ of $\bar{\mathbf{W}}_{WF}[k]$, represents an optimal estimate for the speech component in the mth microphone signal with delay Δ,

$$\hat{x}_m[k-\Delta] = \mathbf{y}^T[k]\,\bar{\mathbf{w}}_{WF,i}[k]\ , \tag{3.19}$$

with

$$m = \text{div}(i-1, L), \quad \Delta = \text{mod}(i-1, L)\ , \tag{3.20}$$

where $\text{div}(i-1, L)$ denotes the integer part of $\frac{i-1}{L}$ and $\text{mod}(i-1, L)$ denotes the remainder of this division.

Estimation error – speech distortion

The estimation error vector $\mathbf{e}[k] = \mathbf{x}[k] - \hat{\mathbf{x}}[k]$, such that the error covariance matrix $\bar{\mathbf{R}}_{ee}[k] = \mathcal{E}\{\mathbf{e}[k]\,\mathbf{e}[k]^T\}$ can be written using (3.14) as

$$\begin{aligned}\bar{\mathbf{R}}_{ee}[k] &= \mathcal{E}\{(\mathbf{x}[k] - \bar{\mathbf{W}}_{WF}^T[k]\,\mathbf{y}[k])\,(\mathbf{x}[k] - \bar{\mathbf{W}}_{WF}^T[k]\,\mathbf{y}[k])^T\} \\ &= \bar{\mathbf{R}}_{xx}[k] - \bar{\mathbf{R}}_{xy}[k]\bar{\mathbf{W}}_{WF}[k] - \bar{\mathbf{W}}_{WF}^T[k]\bar{\mathbf{R}}_{yx}[k] + \bar{\mathbf{W}}_{WF}^T[k]\bar{\mathbf{R}}_{yy}[k]\bar{\mathbf{W}}_{WF}[k] \\ &= \bar{\mathbf{R}}_{xx}[k] - \bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{W}}_{WF}[k] - \bar{\mathbf{W}}_{WF}^T[k]\,\bar{\mathbf{R}}_{xx}[k] + \bar{\mathbf{W}}_{WF}^T[k]\,\bar{\mathbf{R}}_{xx}[k]\end{aligned}$$

$$\begin{aligned}
&= \bar{\mathbf{R}}_{xx}[k] - \bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{W}}_{WF}[k] = (\bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k])\,(\mathbf{I}_M - \bar{\mathbf{W}}_{WF}[k]) \\
&= \bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k] - \bar{\mathbf{R}}_{yy}[k]\bar{\mathbf{W}}_{WF}[k] + \bar{\mathbf{R}}_{vv}[k]\bar{\mathbf{W}}_{WF}[k] \\
&= \bar{\mathbf{R}}_{vv}[k]\,\bar{\mathbf{W}}_{WF}[k]\ . \qquad (3.21)
\end{aligned}$$

The elements $\bar{\mathbf{R}}_{ee}^{ii}[k], i = 1 \ldots M$, on the diagonal of the error covariance matrix indicate how well the ith component of $\mathbf{x}[k]$ is estimated. The smallest diagonal element therefore corresponds to the 'best' estimator.

As already indicated in (2.85), when using an unconstrained MMSE optimal filtering technique, some (linear) speech distortion cannot be avoided, since the estimation error $\mathbf{e}[k]$ is the sum of a term $\mathbf{e}_y[k]$ representing speech distortion and a term $\mathbf{e}_v[k]$ representing the residual noise, i.e.

$$\mathbf{e}[k] = \mathbf{x}[k] - \bar{\mathbf{W}}_{WF}^T[k]\,\mathbf{y}[k] = \underbrace{\big(\mathbf{I}_M - \bar{\mathbf{W}}_{WF}^T[k]\big)\,\mathbf{x}[k]}_{\mathbf{e}_y[k]} - \underbrace{\bar{\mathbf{W}}_{WF}^T[k]\,\mathbf{v}[k]}_{\mathbf{e}_v[k]}\ , \qquad (3.22)$$

The MMSE estimator attributes equal importance to noise reduction and speech distortion. However, it is also possible to attribute more importance to either speech distortion or noise reduction, cf. Section 3.2.3.

3.2.2 Low-rank modelling of speech signals

As for the single-microphone case, it is possible to simplify expression (3.16) for the optimal filter matrix, when the signal to be estimated can be represented using a low-rank model, which is the case for the speech components $x_n[k]$. Using (2.163), the M-dimensional stacked data vector $\mathbf{x}[k]$ can be written as

$$\mathbf{x}[k] = \mathcal{H}[k]\,\mathbf{s}[k]\ , \qquad (3.23)$$

with $\mathcal{H}[k]$ an $M \times (K+L-1)$-dimensional matrix (with typically $K \gg M$), consisting of the coefficients of the acoustic impulse response $\mathbf{h}_n[k]$, and $\mathbf{s}[k]$ a $(K+L-1)$-dimensional vector, consisting of the clean speech samples. We assume that the clean speech signal $s[k]$ can be modelled with a low-rank model of rank R, with $R \leq K+L-1$, such that the signal vector $\mathbf{s}[k]$ can be written as a linear combination of R linearly independent basis vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_R\}$, cf. Section 1.3.1. Since the correlation matrix $\bar{\mathbf{R}}_{ss}[k] = \mathcal{E}\{\mathbf{s}[k]\,\mathbf{s}^T[k]\}$ then is a rank-R matrix[2], also the correlation matrix $\bar{\mathbf{R}}_{xx}[k]$, which can be written as

$$\bar{\mathbf{R}}_{xx}[k] = \mathcal{H}[k]\,\bar{\mathbf{R}}_{ss}[k]\,\mathcal{H}^T[k]\ , \qquad (3.24)$$

is a rank-R matrix (if $R \leq M$ and $\mathcal{H}[k]$ is assumed to be of full row-rank), such that $M-R$ eigenvalues of $\bar{\mathbf{R}}_{xx}[k]$ are equal to zero, independent of the

[2] In practice, $\bar{\mathbf{R}}_{ss}[k]$ has $K+L-1-R$ eigenvalues which are very small, but which are not exactly equal to zero. In this section, we will however assume that these eigenvalues are exactly equal to zero. Recall from Chapter 2 that typical values for K range from 1000 to 2000, typical values for L range from 20 to 80 and typical values for R range from 12 to 20.

exact type of linear model used. Hence, all possible vectors $\mathbf{x}[k]$ lie in an R-dimensional subspace, which is referred to as the *signal subspace*. Since the noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$ is assumed to be positive definite, noise vectors have a component in the signal subspace as well as in the complement of the signal subspace, which is referred to as the *noise subspace*.

The GEVD of $\bar{\mathbf{R}}_{xx}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ is equal to

$$\left\{ \begin{array}{lcl} \bar{\mathbf{R}}_{xx}[k] & = & \bar{\mathbf{Q}}[k] \left[\begin{array}{cc} \bar{\mathbf{\Lambda}}_x[k] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right] \bar{\mathbf{Q}}^T[k] \\ \bar{\mathbf{R}}_{vv}[k] & = & \bar{\mathbf{Q}}[k] \left[\begin{array}{cc} \bar{\mathbf{\Lambda}}_{v1}[k] & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}}_{v2}[k] \end{array} \right] \bar{\mathbf{Q}}^T[k] , \end{array} \right. \tag{3.25}$$

with $\bar{\mathbf{\Lambda}}_x[k]$ and $\bar{\mathbf{\Lambda}}_{v1}[k]$ $R \times R$-dimensional diagonal matrices and $\bar{\mathbf{\Lambda}}_{v2}[k]$ an $(M-R) \times (M-R)$-dimensional diagonal matrix. Since $\bar{\mathbf{R}}_{xx}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ can be assumed to be positive (semi-)definite, all diagonal elements are positive or equal to zero. The correlation matrix $\bar{\mathbf{R}}_{yy}[k]$ can now be written as

$$\bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{R}}_{xx}[k] + \bar{\mathbf{R}}_{vv}[k] = \bar{\mathbf{Q}}[k] \left[\begin{array}{cc} \bar{\mathbf{\Lambda}}_x[k] + \bar{\mathbf{\Lambda}}_{v1}[k] & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}}_{v2}[k] \end{array} \right] \bar{\mathbf{Q}}^T[k] \; . \tag{3.26}$$

Comparing this equation to (3.15), we see that

$$\left\{ \begin{array}{lcll} \bar{\sigma}_i^2[k] & > & \bar{\eta}_i^2[k] & i = 1 \ldots R \\ \bar{\sigma}_i^2[k] & = & \bar{\eta}_i^2[k] & i = R+1 \ldots M \; . \end{array} \right. \tag{3.27}$$

implying that the diagonal matrix in (3.16) has R positive non-zero elements. Even if the signal cannot be modelled with a low-rank model, i.e. $R = M$, none of the diagonal elements can ever become negative. This fact will be used in the practical computation of the optimal filter matrix (cf. Section 3.3).

In the spatio-temporally white noise case, all $\bar{\eta}_i^2[k]$, $i = 1 \ldots M$, are equal to $\bar{\sigma}_v^2$, such that the noise power $\bar{\sigma}_v^2$ can be estimated from the smallest eigenvalues of $\bar{\mathbf{R}}_{yy}[k]$ if the speech components can be modelled with a low-rank model. This also implies that in this case no VAD is required.

We can conclude that the unconstrained MMSE optimal filtering technique for multi-microphone speech enhancement comes down to removing the noise subspace and estimating the speech components from the remaining signal subspace using a MV estimator. Therefore this signal enhancement technique can be considered a multi-microphone extension of the single-microphone signal subspace-based techniques, now combining the spatio-temporal information of the speech and the noise sources. In Section 3.5 it will be shown that this multi-microphone optimal filtering operation can indeed be decomposed into a spatial and a spectral filtering term.

3.2.3 General class of estimators

The filter matrix $\bar{\mathbf{W}}_{WF}[k]$ in (3.16) in fact belongs to a more general class of estimators, which can be represented as

$$\boxed{\bar{\mathbf{W}}[k] = \bar{\mathbf{Q}}^{-T}[k]\,\mathrm{diag}\Big\{f(\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k])\Big\}\,\bar{\mathbf{Q}}^T[k]} \tag{3.28}$$

with $f(\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k])$ a function of the generalised eigenvalues, depending on the specific cost criterion that is being optimised. This formula can be interpreted as an analysis filterbank $\bar{\mathbf{Q}}^{-T}[k]$ which performs a transformation from the time-domain to a signal-dependent transform domain, a gain function $f(\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k])$ which modifies the transform domain parameters and a synthesis filterbank $\bar{\mathbf{Q}}^T[k]$ which performs a transformation back to the time domain. This is similar to the eigenfilterbank interpretation which has been given for the single-microphone signal subspace-based techniques [120], cf. Section 2.3.2. In Section 3.4 we will prove symmetry properties for this filter matrix.

If the MSE criterion is optimised, then the filter $\bar{\mathbf{W}}[k]$ is equal to (3.16). If the output SNR is maximised, the solution corresponds to a least-squares (LS) estimate of rank 1, where only the principal generalised eigenvector is retained, such that the gain function is $f(\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k]) = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}$. This gain function will however introduce a significant amount of signal distortion.

For the single-microphone and white noise case, two perceptually relevant cost criteria have been presented in [85], which trade off noise reduction and speech distortion. This trade-off can be achieved by minimising the signal distortion while keeping the residual noise energy below some given threshold (or vice versa). We have extended the cost criteria presented in [85] to the multi-microphone and the coloured noise case. Using (3.22), the signal distortion energy is $\epsilon_y^2[k] = \mathcal{E}\{\mathbf{e}_y^T[k]\,\mathbf{e}_y[k]\}$, while the residual noise energy is $\epsilon_v^2[k] = \mathcal{E}\{\mathbf{e}_v^T[k]\,\mathbf{e}_v[k]\}$. The goal is to minimise the signal distortion energy under the constraint that the residual noise energy is smaller than a given threshold T, i.e.

$$\min_{\bar{\mathbf{W}}[k]} \epsilon_y^2[k], \quad \text{subject to } \epsilon_v^2[k] \leq T = \alpha T_{max}\ , \tag{3.29}$$

with $0 \leq \alpha \leq 1$ and the maximum threshold equal to the input noise energy $T_{max} = \mathcal{E}\{\mathbf{v}^T[k]\mathbf{v}[k]\}$, i.e. when $\bar{\mathbf{W}}[k] = \mathbf{I}_M$. Although this may seem trivial, we have mathematically proved in Appendix C.1 that the smaller the signal distortion energy $\epsilon_y^2[k]$, the larger the residual noise energy $\epsilon_v^2[k]$. Hence, for the filter matrix $\bar{\mathbf{W}}[k]$ minimising (3.29), the residual noise energy $\epsilon_v^2[k]$ is exactly equal to T, since otherwise the signal distortion energy $\epsilon_y^2[k]$ has not been minimised. The optimisation problem (3.29) therefore is equivalent to

$$\min_{\bar{\mathbf{W}}[k]} \epsilon_y^2[k], \quad \text{subject to } \epsilon_v^2[k] = T\ , \tag{3.30}$$

This optimisation problem can be solved by introducing the Lagrange multiplier λ and by considering the cost function

$$\begin{aligned} J(\bar{\mathbf{W}}[k]) = \epsilon_y^2[k] + \lambda\left[\epsilon_v^2[k] - T\right] =& \mathcal{E}\Big\{\mathbf{x}^T[k]\big(\mathbf{I}_M - \bar{\mathbf{W}}^T[k]\big)^T\big(\mathbf{I}_M - \bar{\mathbf{W}}^T[k]\big)\mathbf{x}[k]\Big\} \\ &+\lambda\left[\mathcal{E}\Big\{\mathbf{v}^T[k]\bar{\mathbf{W}}[k]\bar{\mathbf{W}}^T[k]\mathbf{v}[k]\Big\} - T\right] . \end{aligned} \tag{3.31}$$

The solution of the minimisation problem (3.30) can be found by setting the derivative (see Appendix A.6)

$$\begin{aligned} \frac{\partial J(\bar{\mathbf{W}}[k])}{\partial \bar{\mathbf{W}}[k]} &= \mathcal{E}\Big\{2\mathbf{x}[k]\mathbf{x}^T[k]\bar{\mathbf{W}}[k] - 2\mathbf{x}[k]\mathbf{x}^T[k]\Big\} + \lambda\,\mathcal{E}\Big\{2\mathbf{v}[k]\mathbf{v}^T[k]\bar{\mathbf{W}}[k]\Big\} \\ &= 2\left(\bar{\mathbf{R}}_{xx}[k] + \lambda\bar{\mathbf{R}}_{vv}[k]\right)\bar{\mathbf{W}}[k] - 2\bar{\mathbf{R}}_{xx}[k] , \end{aligned} \tag{3.32}$$

to zero, yielding the solution

$$\begin{aligned} \bar{\mathbf{W}}[k] &= \left(\bar{\mathbf{R}}_{xx}[k] + \lambda\bar{\mathbf{R}}_{vv}[k]\right)^{-1}\bar{\mathbf{R}}_{xx}[k] & (3.33)\\ &= \left(\bar{\mathbf{R}}_{yy}[k] + (\lambda-1)\bar{\mathbf{R}}_{vv}[k]\right)^{-1}\left(\bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k]\right) . & (3.34) \end{aligned}$$

Using (3.15), the filter matrix $\bar{\mathbf{W}}[k]$ can also be written as

$$\boxed{\bar{\mathbf{W}}[k] = \bar{\mathbf{Q}}^{-T}[k]\ \mathrm{diag}\Big\{\frac{\bar{\sigma}_i^2[k] - \bar{\eta}_i^2[k]}{\bar{\sigma}_i^2[k] + (\lambda-1)\,\bar{\eta}_i^2[k]}\Big\}\ \bar{\mathbf{Q}}^T[k]} \tag{3.35}$$

This formula can be interpreted as attributing more or less importance to the noise using the factor λ. If $\lambda = 1$, then the MSE criterion is minimised and $\bar{\mathbf{W}}[k]$ is equal to (3.16). If $\lambda > 1$, then the noise level is assumed to be higher than the actual level, such that the residual noise level is reduced at the expense of increased signal distortion. On the contrary, taking $\lambda < 1$ assumes that the noise level is lower than the actual level, such that signal distortion is reduced at the expense of decreased noise reduction (in the extreme case, if $\lambda = 0$, then $\bar{\mathbf{W}}[k] = I_M$ and no filtering is performed). In the remainder of the thesis, we will generally assume MMSE estimation ($\lambda = 1$). For a specific speech-noise example, we have plotted the signal distortion energy $\epsilon_y^2[k]$ versus the residual noise energy $\epsilon_v^2[k]$ in Fig. 3.2, also indicating the MMSE solution. As can be seen from this figure, this function is monotonically decreasing and the factor λ trades off speech distortion and noise reduction. In Appendix C.1 it is shown that the maximum value for $\epsilon_v^2[k]$ is equal to $\mathrm{tr}\big\{\bar{\mathbf{R}}_{vv}[k]\big\}$, whereas the maximum value for $\epsilon_y^2[k]$ is equal to $\mathrm{tr}\big\{\bar{\mathbf{R}}_{xx}[k]\big\}$.

The Lagrange-multiplier λ can be related to the threshold T by satisfying the constraint $\epsilon_v^2[k] = T$, i.e.

$$T = \epsilon_v^2[k] = \mathcal{E}\Big\{\mathbf{v}^T[k]\bar{\mathbf{W}}[k]\bar{\mathbf{W}}^T[k]\mathbf{v}[k]\Big\} = \mathrm{tr}\Big\{\bar{\mathbf{W}}^T[k]\,\bar{\mathbf{R}}_{vv}[k]\,\bar{\mathbf{W}}[k]\Big\} \tag{3.36}$$

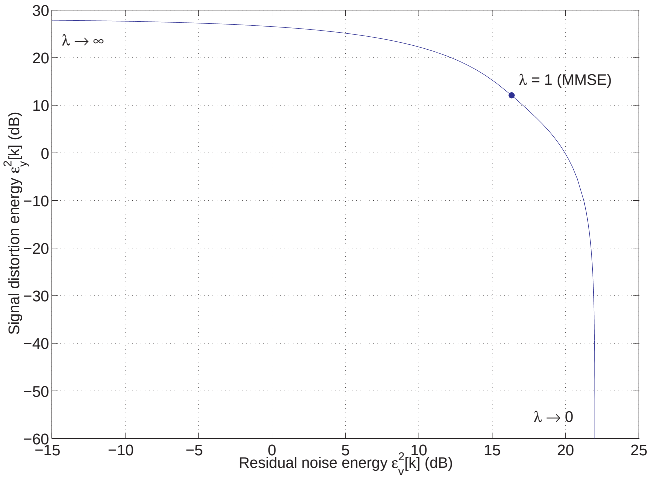


Figure 3.2: Signal distortion energy $\epsilon_y^2[k]$ versus residual noise energy $\epsilon_v^2[k]$

$$= \ \text{tr}\left\{\bar{\mathbf{Q}}[k] \ \text{diag}\left\{ \left(\frac{\bar{\sigma}_i^2[k]-\bar{\eta}_i^2[k]}{\bar{\sigma}_i^2[k]+(\lambda-1)\,\bar{\eta}_i^2[k]}\right)^2 \bar{\eta}_i^2[k]\right\} \bar{\mathbf{Q}}^T[k]\right\}, \qquad (3.37)$$

which is a non-linear relation, such that it is generally impossible to compute λ from T using a closed-form expression. Suffice it to say that similar expressions can be obtained when minimising the residual noise energy $\epsilon_v^2[k]$ while keeping the signal distortion energy $\epsilon_y^2[k]$ below some given threshold.

3.3 Practical computation using GSVD

In this section, we show that in practice the generalised singular value decomposition of a speech and a noise data matrix can be used for computing (an empirical estimate of) the optimal filter matrix. In Section 3.3.2 it is shown that different estimates are obtained for the speech components in the microphone signals, and a method is outlined for determining which estimate should be used. Section 3.3.3 discusses the batch and the recursive version of the GSVD-based optimal filtering technique and Section 3.3.4 gives a brief overview of other possible implementations.

3.3.1 Empirical estimates using data matrices

In practice, the matrix $\bar{\mathbf{Q}}[k]$ and the diagonal elements $\bar{\sigma}_i^2[k]$ and $\bar{\eta}_i^2[k]$ can be estimated by a generalised singular value decomposition (GSVD), cf. Appendix A.2, of a $P_k \times M$-dimensional speech data matrix $\mathbf{Y}[k]$, containing P speech data vectors, and a $Q_k \times M$-dimensional noise data matrix $\mathbf{V}[k]$, containing Q noise data vectors (with P and Q typically much larger than M)[3], i.e.

$$\mathbf{Y}[k] = \begin{bmatrix} \zeta[k-P_k+1]\,\mathbf{y}^T[k-P_k+1] \\ \vdots \\ \zeta[k-1]\,\mathbf{y}^T[k-1] \\ \zeta[k]\,\mathbf{y}^T[k] \end{bmatrix}, \tag{3.38}$$

$$\mathbf{V}[k] = \begin{bmatrix} (1-\zeta[k-Q_k+1])\,\mathbf{y}^T[k-Q_k+1] \\ \vdots \\ (1-\zeta[k-1])\,\mathbf{y}^T[k-1] \\ (1-\zeta[k])\,\mathbf{y}^T[k] \end{bmatrix} \tag{3.39}$$

$$= \begin{bmatrix} (1-\zeta[k-Q_k+1])\,\mathbf{v}^T[k-Q_k+1] \\ \vdots \\ (1-\zeta[k-1])\,\mathbf{v}^T[k-1] \\ (1-\zeta[k])\,\mathbf{v}^T[k] \end{bmatrix}, \tag{3.40}$$

where P_k and Q_k are chosen such that

$$\sum_{l=k-P_k+1}^{k} \zeta[l] = P \qquad \sum_{l=k-Q_k+1}^{k} (1-\zeta[l]) = Q\ . \tag{3.41}$$

Remember that $\zeta[k] = 1$ for speech-and-noise observations ($\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k]$), whereas $\zeta[k] = 0$ for noise-only observations ($\mathbf{y}[k] = \mathbf{v}[k]$). The correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ can now be approximated by the empirical correlation matrices

$$\mathbf{R}_{yy}[k] = \mathbf{Y}^T[k]\,\mathbf{Y}[k]/P, \quad \mathbf{R}_{vv}[k] = \mathbf{V}^T[k]\,\mathbf{V}[k]/Q\ , \tag{3.42}$$

which is an approximation because of the finite lengths P and Q.

The GSVD of the data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ is defined as

$$\left\{ \begin{array}{lcl} \mathbf{Y}[k] & = & \mathbf{U}_Y[k] \cdot \mathbf{\Sigma}_Y[k] \cdot \mathbf{Q}^T[k] \\ \mathbf{V}[k] & = & \mathbf{U}_V[k] \cdot \mathbf{\Sigma}_V[k] \cdot \mathbf{Q}^T[k]\ , \end{array} \right. \tag{3.43}$$

with $\mathbf{\Sigma}_Y[k] = \text{diag}\{\sigma_i[k]\}$, $\mathbf{\Sigma}_V[k] = \text{diag}\{\eta_i[k]\}$, $\mathbf{U}_Y[k]$ and $\mathbf{U}_V[k]$ orthogonal matrices, $\mathbf{Q}[k]$ an invertible but not necessarily orthogonal matrix containing the generalised singular vectors and $\sigma_i[k]/\eta_i[k]$ the generalised singular values.

[3] In the multi-microphone GSVD-based optimal filtering technique, we are considering larger data matrices than in the single-microphone case (cf. Section 2.3.2). Since typical values for P range from 4000 to 8000, i.e. longer than the average short-time stationarity of speech, and typical values for Q range from 20000 to 40000, the performance of the multi-microphone GSVD-based optimal filtering technique is largely dependent on the average, i.e. long-term, spectral and spatial characteristics of the speech and the noise sources (cf. Section 5.2.4). Hence, no short-time effects, such as residual musical noise, will occur in this multi-microphone technique.

The generalised singular vectors and singular values converge to the generalised eigenvectors and eigenvalues of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ when $P \to \infty$ and $Q \to \infty$.

Substituting these formulas into (3.14) gives an empirical estimate $\mathbf{W}_{WF}[k]$ for the optimal filter matrix $\bar{\mathbf{W}}_{WF}[k]$ at time k, i.e.

$$\boxed{\mathbf{W}_{WF}[k] = \mathbf{Q}^{-T}[k] \; \mathrm{diag}\Big\{1 - \frac{P}{Q}\frac{\eta_i^2[k]}{\sigma_i^2[k]}\Big\} \; \mathbf{Q}^T[k]} \tag{3.44}$$

showing that the optimal filter matrix estimate $\mathbf{W}_{WF}[k]$ can be written as a function of the generalised singular vectors and generalised singular values of the speech and the noise data matrices. Also the more general estimators discussed in Section 3.2.3 can be implemented using the GSVD of the speech and the noise data matrices.

In Section 3.2.1 it has been shown that theoretically the diagonal elements in (3.16) cannot become negative. However, since in practice the generalised singular values are estimated from the empirical correlation matrices, it may occur that some diagonal elements in (3.44) become negative. In [85] it has already been noted that negative values will always be obtained when an unbiased non-perfect estimator is used. Therefore these negative values, which are in fact zero estimates, will be set to zero.

3.3.2 Different estimates of speech components

Using the speech data matrix $\mathbf{Y}[k]$ and the optimal filter matrix $\mathbf{W}_{WF}[k]$, an estimate can be obtained for the $P_k \times M$ speech data matrix $\mathbf{X}[k]$, which is defined similarly as (3.38). Without any loss of generality, it is assumed in this section that all speech vectors in $\mathbf{Y}[k]$ are consecutive, i.e. $\zeta[l] = 1, \; l = k - P_k + 1 \ldots k$, implying that $P_k = P$. The estimated speech data matrix $\hat{\mathbf{X}}[k]$ can then be written as

$$\hat{\mathbf{X}}[k] = \begin{bmatrix} \hat{\mathbf{x}}_0^T[k-P+1] & \ldots & \hat{\mathbf{x}}_{N-1}^T[k-P+1] \\ \vdots & & \vdots \\ \hat{\mathbf{x}}_0^T[k-1] & \ldots & \hat{\mathbf{x}}_{N-1}^T[k-1] \\ \hat{\mathbf{x}}_0^T[k] & \ldots & \hat{\mathbf{x}}_{N-1}^T[k] \end{bmatrix} = \mathbf{Y}[k] \, \mathbf{W}_{WF}[k] \, . \tag{3.45}$$

Using a more explicit notation, we can rewrite the $P \times L$ sub-matrix $\hat{\mathbf{X}}_m[k]$ as

$$\hat{\mathbf{X}}_m[k] = \begin{bmatrix} \hat{\mathbf{x}}_m^T[k-P+1] \\ \vdots \\ \hat{\mathbf{x}}_m^T[k-1] \\ \hat{\mathbf{x}}_m^T[k] \end{bmatrix} \tag{3.46}$$

$$= \begin{bmatrix} \hat{x}^{k-L-P+2}_{m,k-P+1}[k-P+1] & \hat{x}^{k-L-P+2}_{m,k-P+1}[k-P] & \dots & \hat{x}^{k-L-P+2}_{m,k-P+1}[k-L-P+2] \\ \vdots & \vdots & & \vdots \\ \hat{x}^{k-L-1}_{m,k-2}[k-2] & \hat{x}^{k-L-1}_{m,k-2}[k-3] & \dots & \hat{x}^{k-L-1}_{m,k-2}[k-L-1] \\ \hat{x}^{k-L}_{m,k-1}[k-1] & \hat{x}^{k-L}_{m,k-1}[k-2] & \dots & \hat{x}^{k-L}_{m,k-1}[k-L] \\ \hat{x}^{k-L+1}_{m,k}[k] & \hat{x}^{k-L+1}_{m,k}[k-1] & \dots & \hat{x}^{k-L+1}_{m,k}[k-L+1] \end{bmatrix} ,$$

where $\hat{x}^{k-L+1}_{m,k}[k]$ is the estimate of the speech component $x_m[k]$ in the mth microphone signal at time k, obtained as a linear combination of the M noisy microphone samples $y_n[k-L+1]\dots y_n[k]$, $n = 0\dots N-1$. As can be seen from this matrix, several different estimates are available for the same speech sample, e.g. L different estimates are available for $x_m[k-L+1]$. If we subdivide the ith column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{WF}[k]$ into the L-dimensional filters $\mathbf{w}_{i,n}[k]$, $n = 0\dots N-1$, similarly as in (3.5), i.e.

$$\mathbf{w}_{WF,i}[k] = \begin{bmatrix} \mathbf{w}^T_{i,0} & \mathbf{w}^T_{i,1} & \dots & \mathbf{w}^T_{i,N-1} \end{bmatrix}^T , \tag{3.47}$$

then the L different estimates for e.g. $x_0[k-L+1]$ can be explicitly written as

$$\begin{bmatrix} \hat{x}^{k-L+1}_{0,k}[k-L+1] & \hat{x}^{k-L}_{0,k-1}[k-L+1] & \dots & \hat{x}^{k-2L+2}_{0,k-L+1}[k-L+1] \end{bmatrix}^T = \tag{3.48}$$

$$\underbrace{\left[\begin{array}{cccc|c|cccc} \boxed{\mathbf{w}^T_{L,0}} & 0 & \dots & 0 & \dots & \boxed{\mathbf{w}^T_{L,N-1}} & 0 & \dots & 0 \\ 0 & \boxed{\mathbf{w}^T_{L-1,0}} & \dots & 0 & \dots & 0 & \boxed{\mathbf{w}^T_{L-1,N-1}} & \dots & 0 \\ & & \ddots & & & & & \ddots & \\ 0 & 0 & \dots & \boxed{\mathbf{w}^T_{1,0}} & \dots & 0 & 0 & \dots & \boxed{\mathbf{w}^T_{1,N-1}} \end{array}\right]}_{\mathcal{W}^T_0[k]} \left[\begin{array}{c} y_0[k] \\ \vdots \\ y_0[k-2L+2] \\ \hline \vdots \\ \hline y_{N-1}[k] \\ \vdots \\ y_{N-1}[k-2L+2] \end{array}\right] .$$

with $\mathcal{W}_m[k]$ the filter matrix that is used for estimating speech components in the mth microphone signal. The question now arises which of the L available estimates in the mth microphone signal yields the lowest MSE. In addition, we have to decide from which of the N microphone signals we will use the speech estimates, leading to M possibilities. As already indicated in Section 3.2.1, the diagonal elements of the error covariance matrix $\mathbf{R}_{ee}[k]$ provide the answer. The ith diagonal element $\mathbf{R}^{ii}_{ee}[k]$ indicates how well the ith component of $\mathbf{x}[k]$ is estimated. The smallest element on the diagonal, say element i, therefore corresponds to the 'best' estimator, i.e. the column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{WF}[k]$. Using this filter, the enhanced speech signal can be computed as

$$\begin{bmatrix} \hat{x}_m[k-\Delta-P+1] \\ \vdots \\ \hat{x}_m[k-\Delta-1] \\ \hat{x}_m[k-\Delta] \end{bmatrix} = \mathbf{Y}[k]\,\mathbf{w}^i_{WF} , \tag{3.49}$$

with

$$m = \mathrm{div}(i-1, L), \quad \Delta = \mathrm{mod}(i-1, L) . \tag{3.50}$$

As discussed in Section 2.3.2, some single-microphone signal subspace-based techniques use an additional averaging step, thereby averaging out over all available speech estimates [43][49][121][138]. However it will be shown in Section 3.4.2 that this averaging step is unnecessary and even suboptimal. Other procedures [85][130], which are block-based, use an overlap-add procedure on the last row of $\hat{\mathbf{X}}_0[k]$, while the adaptive procedure in [220] only retains the first element of this row at each time step, thereby implicitly taking $i = 1$.

The optimal procedure for minimising the MSE thus consists in computing the error covariance matrix $\mathbf{R}_{ee}[k]$ at each time step and choosing the column corresponding to its smallest diagonal element. However this is a computationally very demanding procedure. Simulations have indicated that taking a fixed value $i = \frac{L}{2}$, i.e. using the optimal estimate of the delayed speech component in the first microphone signal $x_0[k - \frac{L}{2} + 1]$, does not significantly decrease the noise reduction performance and the speech intelligibility [56].

3.3.3 Batch and recursive algorithm

In the *batch version* of the algorithm, the speech and the noise data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ are constructed using all available speech and noise data vectors in the considered signal frame. The optimal filter matrix $\mathbf{W}_{WF}[k]$ (which is then actually independent of k) is computed using the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ in (3.44) and the enhanced signal is obtained by filtering the microphone signals with the filter $\mathbf{w}_{WF,i}[k]$. The batch version is not suitable for real-time implementation because of the large delay introduced by the frame-based processing.

In the *recursive version*, the speech and noise data matrices are updated for each time step k with the newly available speech or noise data vector (depending on the output of the VAD-algorithm). Depending on the specific implementation, a fixed length data window (with length P and Q for speech and noise respectively), or an exponential weighting window (with exponential weighting factors λ_y and λ_v, cf. Section 4.2.2) can be used. For each time k, the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ and the optimal filter matrix $\mathbf{W}_{WF}[k]$ are recomputed and the enhanced signal at time k is obtained by filtering the microphone signals with the filter $\mathbf{w}_{WF,i}[k]$. The recursive version introduces only a small processing delay equal to $\Delta = \frac{L}{2} - 1$ samples, and is able to track changing acoustic environments and signal statistics faster than the batch version. However, since at each time step the GSVD and the optimal filter need to be recalculated, the computational complexity is quite high. As will be shown in Chapter 4, this computational complexity can be drastically reduced by using recursive GSVD-updating algorithms. In Section 5.2.2 it will be shown using simulations that the batch and the recursive version of the GSVD-based optimal filtering technique nearly have the same performance.

3.3.4 Other implementations

Instead of using the discussed (fullband) GSVD-based implementation of the multi-channel Wiener expression (3.14), other implementations exist, which exhibit a lower computational complexity and/or a better performance. In [221][224] a (fast) QRD-based implementation is proposed, leading to a lower complexity scheme having nearly the same performance. However, in this QRD-based implementation it is not possible to incorporate the low-rank model of the speech signal. In [94] a stochastic gradient LMS-based implementation has been proposed, using circular data buffers and an instantaneous estimate for the gradient (3.8). The computational complexity of this implementation is very low, but the performance is also seriously degraded. In [242][240] subband implementations of the GSVD-based optimal filtering technique have been proposed, leading to lower complexity schemes with a better performance than the fullband implementation, since the MSE can then be optimised in each individual subband, which is perceptually more relevant.

3.4 Filter symmetry properties and averaging operation

In this section a number of symmetry properties are derived for the single- and multi-microphone optimal filter, which are valid for the white noise case as well as for the coloured noise case and for any function $f(\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k])$. Also the averaging operation of some single-microphone signal subspace-based algorithms is examined, leading to the conclusion that this averaging operation is unnecessary and often even suboptimal.

3.4.1 Single-microphone case

We refer to Appendices A.1 and A.5 for some definitions of structured matrices and properties of their eigenvectors. In the single-microphone case, the correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ are symmetric Toeplitz matrices. Hence, these matrices belong to the class of double symmetric matrices, which are symmetric with respect to both the main and the secondary diagonal and whose eigenvectors are either symmetric or skew-symmetric, cf. Theorem A.28.

Theorem 3.1 *If $\bar{\mathbf{W}}[k]$ is constructed using (3.28), then $\bar{\mathbf{W}}[k]$ satisfies*

$$\bar{\mathbf{W}}[k] = \mathbf{J}_L\, \bar{\mathbf{W}}[k]\, \mathbf{J}_L \qquad (\bar{\mathbf{W}}[k]^T = \mathbf{J}_L\, \bar{\mathbf{W}}[k]^T\, \mathbf{J}_L)\,, \tag{3.51}$$

with $\mathbf{J}_L$ the $L \times L$-dimensional reversal matrix, defined in (A.8). These properties hold in the white noise case as well as in the coloured noise case and for any function $f(\bar{\sigma}_i^2[k], \bar{\eta}_i^2[k])$.

Proof : Consider the joint diagonalisation of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ in (3.15). One can easily verify that

$$\bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{vv}[k] = \bar{\mathbf{Q}}^{-T}[k]\,\mathbf{\Lambda}_y^{-1}[k]\mathbf{\Lambda}_v[k]\,\bar{\mathbf{Q}}^T[k] \tag{3.52}$$

is the eigenvalue decomposition of $\bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{vv}[k]$. Because $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ are double-symmetric matrices, the following relations hold,

$$\mathbf{J}_L\,\bar{\mathbf{R}}_{yy}[k]\,\mathbf{J}_L = \bar{\mathbf{R}}_{yy}[k], \quad \mathbf{J}_L\,\bar{\mathbf{R}}_{vv}[k]\,\mathbf{J}_L = \bar{\mathbf{R}}_{vv}[k]\ , \tag{3.53}$$

such that also

$$\bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{vv}[k] = \mathbf{J}_L\,\mathbf{R}_{yy}^{-1}[k]\,\mathbf{R}_{vv}[k]\,\mathbf{J}_L\ . \tag{3.54}$$

Therefore the eigenvectors of $\bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{vv}[k]$, i.e. the columns of $\bar{\mathbf{Q}}^{-T}[k]$, satisfy the property (A.44),

$$\mathbf{J}_L\,\bar{\mathbf{Q}}^{-T}[k] = \bar{\mathbf{Q}}^{-T}[k]\,\mathrm{diag}\{\pm 1\}\ , \tag{3.55}$$

such that

$$\begin{aligned} \mathbf{J}_L\,\bar{\mathbf{W}}[k]\,\mathbf{J}_L &= \mathbf{J}_L\,\bar{\mathbf{Q}}^{-T}\,\mathrm{diag}\{f(\bar{\sigma}_i^2[k],\bar{\eta}_i^2[k])\}\,\bar{\mathbf{Q}}^T\,\mathbf{J}_L & (3.56)\\ &= \bar{\mathbf{Q}}^{-T}\,\mathrm{diag}\{f(\bar{\sigma}_i^2[k],\bar{\eta}_i^2[k])\}\,\bar{\mathbf{Q}}^T = \bar{\mathbf{W}}[k]\ . & (3.57) \end{aligned}$$

□

These symmetry properties imply that the ith row/column of $\bar{\mathbf{W}}[k]$ is equal to the $(L+1-i)$th row/column in reverse order. For L odd, the middle column in $\bar{\mathbf{W}}[k]$ is symmetric, and hence represents a *linear phase filter*. This linear phase property is an extension of the zero phase property that has already been attributed to SVD and rank truncation based estimators for the white noise case, if an additional averaging step is included [68] (cf. Section 3.4.2). The above linear phase property is however also valid for the coloured noise case as well as for a general function $f(\bar{\sigma}_i^2, \bar{\eta}_i^2)$.

3.4.2 Single-microphone averaging operation

As already indicated in Section 2.3.2, some single-microphone procedures [43] [49][121][138] use an additional averaging step for obtaining a final estimate from the different available estimates for $x_0[k-L+1]$. In the single-microphone case, (3.48) reduces to

$$\begin{bmatrix} \hat{x}_{0,k}^{k-L+1}[k-L+1] \\ \hat{x}_{0,k-1}^{k-L}[k-L+1] \\ \vdots \\ \hat{x}_{0,k-L+1}^{k-2L+2}[k-L+1] \end{bmatrix} = \underbrace{\begin{bmatrix} \boxed{\bar{\mathbf{w}}_{L,0}^T} & 0 & \dots & 0 \\ 0 & \boxed{\bar{\mathbf{w}}_{L-1,0}^T} & \dots & 0 \\ & & \ddots & \\ 0 \quad 0 & \dots & & \boxed{\bar{\mathbf{w}}_{1,0}^T} \end{bmatrix}}_{\bar{\mathcal{W}}_0^T[k]} \begin{bmatrix} y_0[k] \\ y_0[k-1] \\ \vdots \\ y_0[k-2L+2] \end{bmatrix} . \tag{3.58}$$

The filter $\bar{\mathbf{w}}_{1,0}$ is a completely causal filter, whereas the filter $\bar{\mathbf{w}}_{L,0}$ is a completely anti-causal filter and the other filters $\bar{\mathbf{w}}_{i,0}$, $i = 2 \ldots L-1$, consist of a combination of causal and anti-causal taps[4].

From $\bar{\mathbf{W}}[k]^T = \mathbf{J}_L \, \bar{\mathbf{W}}[k]^T \, \mathbf{J}_L$, with $\bar{\mathbf{W}}[k] = \begin{bmatrix} \bar{\mathbf{w}}_{1,0} & \ldots & \bar{\mathbf{w}}_{L-1,0} & \bar{\mathbf{w}}_{L,0} \end{bmatrix}$, it immediately follows that

$$\bar{\mathcal{W}}_0^T[k] = \mathbf{J}_L \, \bar{\mathcal{W}}_0^T[k] \, \mathbf{J}_{2L-1} \, . \tag{3.59}$$

The averaging operation can now be written as

$$\tilde{x}_{0,k}^{k-2L+2}[k-L+1] = \begin{bmatrix} \frac{1}{L} & \frac{1}{L} & \cdots & \frac{1}{L} \end{bmatrix} \begin{bmatrix} \hat{x}_{0,k}^{k-L+1}[k-L+1] \\ \hat{x}_{0,k-1}^{k-L}[k-L+1] \\ \vdots \\ \hat{x}_{0,k-L+1}^{k-2L+2}[k-L+1] \end{bmatrix} \tag{3.60}$$

$$= \underbrace{\begin{bmatrix} \frac{1}{L} & \frac{1}{L} & \cdots & \frac{1}{L} \end{bmatrix} \bar{\mathcal{W}}_0^T[k]}_{\tilde{\mathbf{w}}^T[k]} \begin{bmatrix} y_0[k] \\ y_0[k-1] \\ \vdots \\ y_0[k-2L+2] \end{bmatrix} , \tag{3.61}$$

with $\tilde{\mathbf{w}}[k]$ a $(2L-1)$-dimensional vector. The averaged value $\tilde{x}_{0,k}^{k-2L+2}[k-L+1]$ is estimated from $y_0[k-L+1]$ together with $L-1$ past samples and $L-1$ future samples. The filter $\tilde{\mathbf{w}}[k]$ is obtained by averaging over the available L-dimensional filters $\bar{\mathbf{w}}_{i,0}$, $i = 1 \ldots L$. From the symmetry property of $\bar{\mathcal{W}}_0[k]$, it is readily seen that $\tilde{\mathbf{w}}[k]$ represents a *zero phase filter.* The question now is whether $\tilde{\mathbf{w}}[k]$ has a better performance than the individual filters $\bar{\mathbf{w}}_{i,0}$ it is computed from. Specifically, $\tilde{\mathbf{w}}[k]$ should be compared with the symmetric middle row of $\bar{\mathbf{W}}[k]$ (if L is odd), which represents a linear phase filter that uses $\frac{L-1}{2}$ past samples and $\frac{L-1}{2}$ future samples.

First, it can be verified that $\tilde{\mathbf{w}}[k]$ is not the $(2L-1)$-dimensional optimal filter, i.e.

$$\tilde{x}_{0,k}^{k-2L+2}[k-L+1] \neq \hat{x}_{0,k}^{k-2L+2}[k-L+1] \, , \tag{3.62}$$

since $\tilde{x}_{0,k}^{k-2L+2}[k-L+1]$ is obtained by averaging out over a collection of L-dimensional optimal filters, whereas $\hat{x}_{0,k}^{k-2L+2}[k-L+1]$ is obtained by applying the optimal filter formulas to a $(2L-1)$-dimensional vector $\mathbf{y}_0[k]$. Secondly, simulations indicate that the obtained error variance for the $(2L-1)$-dimensional filter $\tilde{\mathbf{w}}[k]$ is consistently larger than the error variance for the 'best' L-dimensional filter $\bar{\mathbf{w}}_{i,0}$, obtained by considering the smallest diagonal element of the error covariance matrix $\bar{\mathbf{R}}_{ee}[k]$.

[4] Note that when the filter length $L \rightarrow \infty$ (and assuming certain stationarity conditions), the filters $\bar{\mathbf{w}}_{i,0}$, $i = 1 \ldots L$, are shifted versions of each other, implying that $\bar{\mathbf{W}}[k]$ is a Toeplitz matrix. Hence, since $\bar{\mathbf{W}}[k]$ is a Toeplitz matrix and $\bar{\mathbf{W}}[k] = \mathbf{J}_L \, \bar{\mathbf{W}}[k] \, \mathbf{J}_L$, according to (3.51), $\bar{\mathbf{W}}[k]$ is a symmetric Toeplitz matrix, However, in general this is not true for finite filter lengths.

Example 3.1 Consider the following simulation: the input signal $y_0[k]$ is constructed as the sum of two (stationary) unit-variance white noise signals $x_0[k]$ and $v_0[k]$,

$$y_0[k] = x_0[k] + \sigma_v v_0[k], \; k = 1 \dots P \, . \tag{3.63}$$

Both the optimal filter matrix $\mathbf{W}_{WF}[k]$, which consists of L-dimensional filters $\mathbf{w}_{WF,i}[k]$, $i = 1 \dots L$, and the $(2L-1)$-dimensional filter $\tilde{\mathbf{w}}[k]$ are computed from these signals. Also the enhanced signals $\hat{x}_{0,i}[k]$ and $\tilde{x}_0[k]$ are computed using the filters $\mathbf{w}_{WF,i}[k]$ and $\tilde{\mathbf{w}}[k]$. The error variances $\hat{\sigma}_i$, $i = 1 \dots L$, and $\tilde{\sigma}$ are defined as

$$\hat{\sigma}_i = \frac{1}{P} \sum_{k=1}^{P} \left(x_0[k] - \hat{x}_{0,i}[k] \right)^2 , \; i = 1 \dots L, \tag{3.64}$$

$$\tilde{\sigma} = \frac{1}{P} \sum_{k=1}^{P} \left(x_0[k] - \tilde{x}_0[k] \right)^2 \, . \tag{3.65}$$

For $L = 9$, $P = 10^5$ and $\sigma_v^2 = 2$, the error variances $\hat{\sigma}_i$, $i = 1 \dots L$, and $\tilde{\sigma}$ are compared in Fig. 3.3. As can be seen from this figure, the performance of the $(2L-1)$-dimensional filter $\tilde{\mathbf{w}}[k]$ is not always better than the individual L-dimensional filters $\mathbf{w}_{WF,i}[k]$ it is computed from. Moreover, there always seems to exist an L-dimensional filter $\mathbf{w}_{WF,i}[k]$ which leads to a lower error variance. △

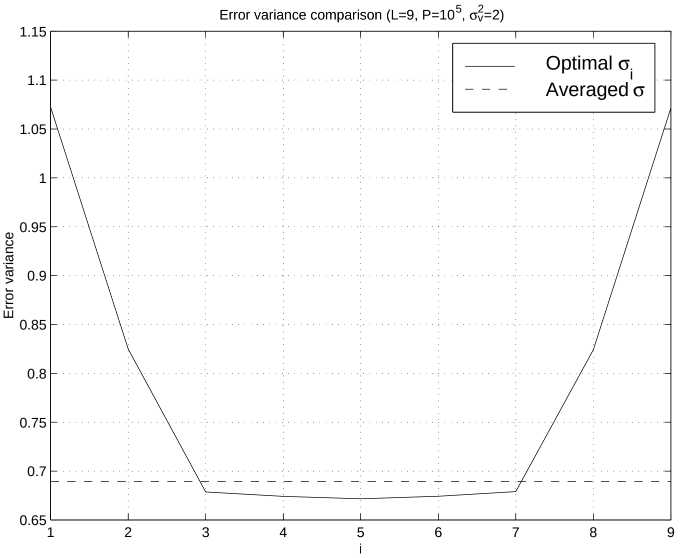


Figure 3.3: Error variance comparison between $(2L-1)$-dimensional filter $\tilde{\mathbf{w}}[k]$ and L-dimensional filters $\mathbf{w}_{WF,i}[k]$, $i = 1 \dots L$

Hence, averaging does not seem to be a well-founded operation, while on the other hand it increases computational complexity, since it requires $(2L-1)$-taps filtering instead of L-taps filtering. If minimal error variance is sought for, we suggest to use the L-dimensional filter corresponding to the smallest diagonal element in the error covariance matrix. However, as already indicated in Section 3.3, this is a computationally very demanding procedure, since for each time step k the error covariance matrix $\bar{\mathbf{R}}_{ee}[k]$ needs to be computed. Therefore, in practice we suggest to use the L-dimensional filter given by the middle column of $\bar{\mathbf{W}}_{WF}[k]$, which provides both low error variance (albeit generally not the lowest attainable error variance) and linear phase. It is unpredictable whether this filter or the averaged filter yields the lowest error variance.

3.4.3 Multi-microphone case

In the multi-microphone case, similar and additional symmetry properties can be derived, depending on the assumptions we make for the spatio-temporal correlation matrices $\bar{\mathbf{R}}_{xx}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$.

Without loss of generality we assume $N = 2$ in this section. However, the symmetry properties can be easily extended to the case of more than 2 microphones. We will subdivide the $2L \times 2L$ symmetric correlation matrices as

$$\bar{\mathbf{R}}_{xx}[k] = \begin{bmatrix} \bar{\mathbf{R}}_{xx}^{00}[k] & \bar{\mathbf{R}}_{xx}^{01}[k] \\ \bar{\mathbf{R}}_{xx}^{10}[k] & \bar{\mathbf{R}}_{xx}^{11}[k] \end{bmatrix}, \quad \bar{\mathbf{R}}_{vv}[k] = \begin{bmatrix} \bar{\mathbf{R}}_{vv}^{00}[k] & \bar{\mathbf{R}}_{vv}^{01}[k] \\ \bar{\mathbf{R}}_{vv}^{10}[k] & \bar{\mathbf{R}}_{vv}^{11}[k] \end{bmatrix}, \tag{3.66}$$

with $\bar{\mathbf{R}}_{xx}^{00}[k]$, $\bar{\mathbf{R}}_{xx}^{11}[k]$, $\bar{\mathbf{R}}_{vv}^{00}[k]$ and $\bar{\mathbf{R}}_{vv}^{11}[k]$ double-symmetric matrices, and

$$\bar{\mathbf{R}}_{xx}^{01}[k] = \left(\bar{\mathbf{R}}_{xx}^{10}[k]\right)^T, \quad \bar{\mathbf{R}}_{vv}^{01}[k] = \left(\bar{\mathbf{R}}_{vv}^{10}[k]\right)^T . \tag{3.67}$$

We also subdivide the filter matrix $\bar{\mathbf{W}}[k]$ as

$$\bar{\mathbf{W}}[k] = \begin{bmatrix} \bar{\mathbf{W}}^{00}[k] & \bar{\mathbf{W}}^{01}[k] \\ \bar{\mathbf{W}}^{10}[k] & \bar{\mathbf{W}}^{11}[k] \end{bmatrix} . \tag{3.68}$$

If we assume that the speech and the noise correlation matrices for both microphones are equal, i.e. $\bar{\mathbf{R}}_{xx}^{00}[k] = \bar{\mathbf{R}}_{xx}^{11}[k]$ and $\bar{\mathbf{R}}_{vv}^{00}[k] = \bar{\mathbf{R}}_{vv}^{11}[k]$, and that $\bar{\mathbf{R}}_{xx}^{01}[k]$ and $\bar{\mathbf{R}}_{vv}^{01}[k]$ are Toeplitz matrices, then

$$\mathbf{J}_{2L}\,\bar{\mathbf{R}}_{xx}[k]\,\mathbf{J}_{2L} = \bar{\mathbf{R}}_{xx}[k], \quad \mathbf{J}_{2L}\,\bar{\mathbf{R}}_{vv}[k]\,\mathbf{J}_{2L} = \bar{\mathbf{R}}_{vv}[k]\ , \tag{3.69}$$

such that the same symmetry properties as for the single-microphone case apply, cf. Theorem 3.1.

Moreover, if $\bar{\mathbf{R}}_{xx}^{01}[k]$ and $\bar{\mathbf{R}}_{vv}^{01}[k]$ are symmetric Toeplitz matrices, then also

$$\mathbf{S}_{2L}\,\bar{\mathbf{R}}_{xx}[k]\,\mathbf{S}_{2L} = \bar{\mathbf{R}}_{xx}[k], \quad \mathbf{S}_{2L}\,\bar{\mathbf{R}}_{vv}[k]\,\mathbf{S}_{2L} = \bar{\mathbf{R}}_{vv}[k]\ , \tag{3.70}$$

with $\mathbf{S}_{2L}$ the block-reversal matrix, defined in (A.13). Using Theorem A.35, it can be proved that the filter matrix $\bar{\mathbf{W}}[k]$, constructed using (3.28), satisfies the additional symmetry property

$$\bar{\mathbf{W}}[k] = \mathbf{S}_{2L}\,\bar{\mathbf{W}}[k]\,\mathbf{S}_{2L} \qquad (\bar{\mathbf{W}}^T[k] = \mathbf{S}_{2L}\,\bar{\mathbf{W}}^T[k]\,\mathbf{S}_{2L})\ , \tag{3.71}$$

such that

$$\mathbf{J}_L\,\bar{\mathbf{W}}^{00}[k]\,\mathbf{J}_L = \bar{\mathbf{W}}^{00}[k] = \bar{\mathbf{W}}^{11}[k]\ , \tag{3.72}$$

$$\mathbf{J}_L\,\bar{\mathbf{W}}^{01}[k]\,\mathbf{J}_L = \bar{\mathbf{W}}^{01}[k] = \bar{\mathbf{W}}^{10}[k]\ . \tag{3.73}$$

In this case, the middle columns (for L odd) of $\bar{\mathbf{W}}^{00}[k]$ and $\bar{\mathbf{W}}^{10}[k]$ again correspond to linear phase filters.

The same properties hold when the 2 noise components $v_0[k]$ and $v_1[k]$ are uncorrelated, since then $\bar{\mathbf{R}}_{vv}^{01}[k] = \bar{\mathbf{R}}_{vv}^{10}[k] = \mathbf{0}$. In the case of spatio-temporally white noise, the noise correlation matrix is equal to

$$\bar{\mathbf{R}}_{vv}[k] = \sigma_v^2 \begin{bmatrix} \mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_L \end{bmatrix}\ , \tag{3.74}$$

and the filter matrix $\bar{\mathbf{W}}[k]$ has the additional property of being symmetric, such that

$$\bar{\mathbf{W}}^{00}[k] = \left(\bar{\mathbf{W}}^{00}[k]\right)^T,\quad \bar{\mathbf{W}}^{01}[k] = \left(\bar{\mathbf{W}}^{01}[k]\right)^T\ . \tag{3.75}$$

3.5 Frequency-domain analysis

When analysing the multi-channel Wiener filter in the frequency-domain, it can indeed be shown that (under mild assumptions) this filter can be decomposed into a spectral and a spatial filtering term. We will discuss the power transfer functions for the speech and the noise components and we will simplify all expressions for a single speech source. We will also discuss the noise sensitivity of the GSC and the multi-channel Wiener filter and show under which conditions they are equal.

3.5.1 Multi-channel Wiener filter

In the frequency-domain analysis we assume that all signals are stationary and we consider the N-dimensional filter $\mathbf{W}(\omega)$,

$$\mathbf{W}(\omega) = \begin{bmatrix} W_0(\omega) & W_1(\omega) & \ldots & W_{N-1}(\omega) \end{bmatrix}^T\ , \tag{3.76}$$

which is equivalent to infinitely long filters $\mathbf{w}_n[k]$, $n = 0 \ldots N-1$, in the time-domain. Similarly as in the time-domain, the optimal frequency-domain filter $\mathbf{W}(\omega)$ minimises the MSE between the output signal $Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega)$

and the speech component $X_0(\omega)$ in the first microphone signal (or any other microphone signal)[5]. The frequency-domain cost function therefore is equal to

$$\begin{aligned} J_{MSE}(\mathbf{W}(\omega)) &= \mathcal{E}\{|X_0(\omega) - \mathbf{W}^H(\omega)\mathbf{Y}(\omega)|^2\} && (3.77)\\ &= \mathcal{E}\{|X_0(\omega)|^2\} + \mathcal{E}\{\mathbf{Y}^H(\omega)\mathbf{W}(\omega)\mathbf{W}^H(\omega)\mathbf{Y}(\omega)\} && (3.78)\\ &\quad -\mathcal{E}\{X_0(\omega)\mathbf{Y}^H(\omega)\mathbf{W}(\omega)\} - \mathcal{E}\{\mathbf{W}^H(\omega)\mathbf{Y}(\omega)X_0^*(\omega)\}\ , \end{aligned}$$

which is minimised by setting the derivative

$$\frac{\partial J_{MSE}(\mathbf{W}(\omega))}{\partial \mathbf{W}(\omega)} = -2\mathcal{E}\{\mathbf{Y}(\omega)X_0^*(\omega)\} + 2\mathcal{E}\{\mathbf{Y}(\omega)\mathbf{Y}^H(\omega)\}\,\mathbf{W}(\omega) \qquad (3.79)$$

to zero. Because the speech and the noise are assumed to be uncorrelated, i.e.

$$\mathcal{E}\{\mathbf{Y}(\omega)X_0^*(\omega)\} = \mathcal{E}\{\mathbf{X}(\omega)X_0^*(\omega)\}\ , \qquad (3.80)$$

the multi-channel Wiener filter in the frequency-domain is equal to

$$\boxed{\mathbf{W}_{WF}(\omega) = \bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1} \qquad (3.81)$$

with $\bar{\mathbf{R}}_{yy}(\omega)$ and $\bar{\mathbf{R}}_{xx}(\omega)$ similarly defined as in (2.36) and $\mathbf{e}_i$ an N-dimensional vector of which the ith element is equal to 1 and all other elements are equal to 0, i.e.

$$\mathbf{e}_i = \begin{bmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}^T\ . \qquad (3.82)$$

If we *assume that both the speech and the noise field are homogeneous*, i.e. that the PSD of the speech and the noise components $P_{x_n}(\omega) = P_x(\omega)$ and $P_{v_n}(\omega) = P_v(\omega)$, $n = 0 \ldots N-1$, then $P_{y_n}(\omega) = P_y(\omega) = P_x(\omega) + P_v(\omega)$, $n = 0 \ldots N-1$, such that $\bar{\mathbf{R}}_{yy}(\omega) = P_y(\omega)\mathbf{\Gamma}_y(\omega)$ and $\bar{\mathbf{R}}_{xx}(\omega) = P_x(\omega)\mathbf{\Gamma}_x(\omega)$, with the coherence matrices $\mathbf{\Gamma}_y(\omega)$ and $\mathbf{\Gamma}_x(\omega)$ similarly defined as in (2.37). The Wiener filter $\mathbf{W}_{WF}(\omega)$ can then be written as

$$\boxed{\mathbf{W}_{WF}(\omega) = \underbrace{\frac{P_x(\omega)}{P_x(\omega)+P_v(\omega)}}_{\text{spectral filtering}} \cdot \underbrace{\mathbf{\Gamma}_y^{-1}(\omega)\mathbf{\Gamma}_x(\omega)\,\mathbf{e}_1}_{\text{spatial filtering}}} \qquad (3.83)$$

As can be seen from this equation, the multi-channel Wiener filter consists of 2 terms [241][240] :

- a single-channel Wiener filter, depending on the spectral characteristics (PSD) of the speech and the noise sources;
- a spatial filtering operation, depending on the spatial characteristics (coherence) of the speech and the noise fields;

[5] Without loss of generality, we assume in the derivation of $\mathbf{W}(\omega)$ that the delay Δ is zero.

Hence, it will be observed in the simulations in Section 5.4.1 that the performance for a spectrally white noise source is better than the performance for a speech-like noise source at the same position (due to the additional spectral filtering) and that the performance for a single noise source is better than for three simultaneous noise sources at different positions (due to the spatial filtering term).

Let us reiterate that the GSC, on the contrary, only depends on the spatial characteristics of the noise field, cf. (2.153) – under the condition that no signal leakage into the noise references occurs.

3.5.2 Power Transfer Functions

Since $\bar{\mathbf{R}}_{yy}(\omega) = \bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}(\omega)$, we can write

$$\boldsymbol{\Gamma}_y(\omega) = \frac{P_x(\omega)}{P_y(\omega)} \left[\boldsymbol{\Gamma}_x(\omega) + \frac{P_v(\omega)}{P_x(\omega)}\boldsymbol{\Gamma}_v(\omega)\right] = \frac{\boldsymbol{\Gamma}_x(\omega) + \beta(\omega)\boldsymbol{\Gamma}_v(\omega)}{1+\beta(\omega)} , \tag{3.84}$$

with $\beta(\omega) = P_v(\omega)/P_x(\omega)$, such that (3.83) can be written as

$$\boxed{\mathbf{W}_{WF}(\omega) = \left[\boldsymbol{\Gamma}_x(\omega) + \beta(\omega)\boldsymbol{\Gamma}_v(\omega)\right]^{-1} \boldsymbol{\Gamma}_x(\omega)\,\mathbf{e}_1} \tag{3.85}$$

Using $Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega)$, the PSD $P_z(\omega)$ of the output signal is equal to

$$P_z(\omega) = \mathbf{W}^H(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{W}(\omega) = P_y(\omega)\mathbf{W}^H(\omega)\boldsymbol{\Gamma}_y(\omega)\mathbf{W}(\omega) , \tag{3.86}$$

such that the PTF of the speech component, cf. (2.39), is equal to

$$G_{x_0z_x}(\omega) = \frac{P_{z_x}(\omega)}{P_{x_0}(\omega)} = \mathbf{W}^H(\omega)\boldsymbol{\Gamma}_x(\omega)\mathbf{W}(\omega) , \tag{3.87}$$

and the PTF of the noise component, cf. (2.41), is equal to

$$G_{v_0z_v}(\omega) = \frac{P_{z_v}(\omega)}{P_{v_0}(\omega)} = \mathbf{W}^H(\omega)\boldsymbol{\Gamma}_v(\omega)\mathbf{W}(\omega) . \tag{3.88}$$

3.5.3 Single speech source

In the case of a single speech source $S(\omega)$ – the case which we will generally consider – the correlation matrix $\bar{\mathbf{R}}_{xx}(\omega)$ has rank 1, such that the coherence matrix $\boldsymbol{\Gamma}_x(\omega)$ can be written as

$$\boldsymbol{\Gamma}_x(\omega) = \mathbf{x}(\omega)\mathbf{x}^H(\omega) , \tag{3.89}$$

with $\mathbf{x}(\omega) = \sqrt{P_s(\omega)/P_x(\omega)}\,\mathbf{H}(\omega)$. Using the matrix inversion lemma (A.38), the matrix $\left[\boldsymbol{\Gamma}_x(\omega) + \beta(\omega)\boldsymbol{\Gamma}_v(\omega)\right]^{-1}$ can then be written as

$$\left[\boldsymbol{\Gamma}_x(\omega) + \beta(\omega)\boldsymbol{\Gamma}_v(\omega)\right]^{-1} = \frac{1}{\beta(\omega)} \left[\boldsymbol{\Gamma}_v^{-1}(\omega) - \frac{\boldsymbol{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)\mathbf{x}^H(\omega)\boldsymbol{\Gamma}_v^{-1}(\omega)}{\mathbf{x}^H(\omega)\boldsymbol{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega) + \beta(\omega)}\right] , \tag{3.90}$$

such that $\mathbf{W}_{WF}(\omega)$ in (3.85) can be written as

$$\begin{aligned}\mathbf{W}_{WF}(\omega) &= \frac{1}{\beta(\omega)}\left[\mathbf{\Gamma}_v^{-1}(\omega) - \frac{\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)}{\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)+\beta(\omega)}\right]\mathbf{x}(\omega)\mathbf{x}^H(\omega)\mathbf{e}_1 \\ &= \frac{\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)\mathbf{x}^H(\omega)\mathbf{e}_1}{\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)+\beta(\omega)}\ . \qquad (3.91)\end{aligned}$$

It will be proved in Section 3.5.4 that the noise sensitivity of the multi-channel Wiener filter for a single speech source is independent of the factor $\beta(\omega) = P_v(\omega)/P_x(\omega)$, and is equal to the noise sensitivity of the GSC (for a certain choice of the fixed beamformer and the blocking matrix in the GSC). Using the fact that $\mathbf{x}(\omega) = \sqrt{P_s(\omega)/P_x(\omega)}\,\mathbf{H}(\omega)$, the Wiener filter $\mathbf{W}_{WF}(\omega)$ can be written as

$$\boxed{\mathbf{W}_{WF}(\omega) = \frac{\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)H_0(\omega)}{\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)+P_v(\omega)/P_s(\omega)}} \qquad (3.92)$$

such that, using (2.39), the PTF for the speech component, i.e. speech distortion, is equal to

$$\begin{aligned}G_{x_0z_x}(\omega) &= \frac{\mathbf{W}^H(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega)\mathbf{W}(\omega)}{|H_0(\omega)|^2} \qquad (3.93)\\ &= \left[\frac{\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)}{\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)+P_v(\omega)/P_s(\omega)}\right]^2\ , \qquad (3.94)\end{aligned}$$

which depends on the spatial characteristics $\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)$ and the spectral characteristics $P_v(\omega)/P_s(\omega)$ of the speech and the noise sources. One can easily see that more *speech distortion* occurs at frequencies with a low SNR, i.e. high $P_v(\omega)/P_s(\omega)$, and when the spatial separation between the speech and the noise sources is poor, i.e. low $\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)$ [195]. Using (3.88), the PTF of the noise component is equal to

$$G_{v_0z_v}(\omega) = \mathbf{W}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{W}(\omega) = \frac{|H_0(\omega)|^2\,\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)}{\left[\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)+P_v(\omega)/P_s(\omega)\right]^2}\ , \qquad (3.95)$$

such that more *noise reduction* occurs at frequencies with a low SNR, i.e. high $P_v(\omega)/P_s(\omega)$, and when the speech and the noise sources are spatially well separated, i.e. high $\mathbf{H}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{H}(\omega)$.

3.5.4 Noise Sensitivity

In Section 5.4.3, the robustness of the multi-channel Wiener filter will be investigated for several types of deviations from the assumed signal model (microphone gain and position mismatch, look direction error). Since the multi-channel

Wiener filter does not rely on a-priori information about the signal model, it will be shown by simulations that it is more robust than the GSC.

However, robustness can also be analysed theoretically. In order to quantify the sensitivity to deviations in the assumed signal model, the *noise sensitivity* $\Phi(\omega)$ is often used [37][245], which is defined as the ratio of the white noise gain (cf. Section 2.5.1) to the power transfer function of the speech signal (cf. Section 2.2.4), i.e.

$$\Phi(\omega) = \frac{\mathbf{W}^H(\omega)\mathbf{W}(\omega)}{G_{x_0 z_x}(\omega)} = \frac{\mathbf{W}^H(\omega)\mathbf{W}(\omega)}{\mathbf{W}^H(\omega)\mathbf{\Gamma}_x(\omega)\mathbf{W}(\omega)} \ , \tag{3.96}$$

assuming a homogeneous speech sound field. The noise sensitivity for the GSC and the multi-channel Wiener filter have already been studied in [240], where it has been noted (but not proved) that the noise sensitivity of the GSC and the multi-channel Wiener filter are equal under certain situations. In this section, we will prove under which conditions these noise sensitivities are equal.

Using (2.155) and (2.156), the noise sensitivity of the GSC is equal to

$$\Phi_{GSC}(\omega) = \frac{\mathbf{W}_q^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)}{\mathbf{W}_q^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{\Gamma}_x(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{W}_q(\omega)} \tag{3.97}$$

with $\mathbf{W}_q(\omega)$ the fixed beamformer (assumed to be orthogonal to the blocking matrix). Using (3.85) and (3.87), the noise sensitivity of the multi-channel Wiener filter $\Phi_{WF}(\omega)$ is equal to

$$\frac{\mathbf{e}_1^H\mathbf{\Gamma}_x(\omega)\left[\mathbf{\Gamma}_x(\omega)+\beta(\omega)\mathbf{\Gamma}_v(\omega)\right]^{-1}\left[\mathbf{\Gamma}_x(\omega)+\beta(\omega)\mathbf{\Gamma}_v(\omega)\right]^{-1}\mathbf{\Gamma}_x(\omega)\mathbf{e}_1}{\mathbf{e}_1^H\mathbf{\Gamma}_x(\omega)\left[\mathbf{\Gamma}_x(\omega)+\beta(\omega)\mathbf{\Gamma}_v(\omega)\right]^{-1}\mathbf{\Gamma}_x(\omega)\left[\mathbf{\Gamma}_x(\omega)+\beta(\omega)\mathbf{\Gamma}_v(\omega)\right]^{-1}\mathbf{\Gamma}_x(\omega)\mathbf{e}_1} \ . \tag{3.98}$$

At first sight, these noise sensitivities do not share many similarities. However, first assume that *a single speech source* is present, i.e. $\mathbf{\Gamma}_x(\omega) = \mathbf{x}(\omega)\mathbf{x}^H(\omega)$ (cf. Section 3.5.3). Using (3.91), the noise sensitivity $\Phi_{WF}(\omega)$ can be written as

$$\Phi_{WF}(\omega) = \frac{\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)}{\left[\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)\right]^2} \ , \tag{3.99}$$

which is independent of the spectral factor $\beta(\omega) = P_v(\omega)/P_x(\omega)$. Secondly, assume that the fixed beamformer $\mathbf{W}_q(\omega)$ of the GSC is *a matched filter*, i.e. $\mathbf{W}_q(\omega) = \alpha\mathbf{H}(\omega)$, with $\mathbf{H}(\omega)$ the N-dimensional vector of acoustical transfer functions. Since $\mathbf{H}(\omega) = \sqrt{P_x(\omega)/P_s(\omega)}\,\mathbf{x}(\omega)$ (cf. Section 3.5.3), it can be shown that the noise sensitivity $\Phi_{GSC}(\omega)$ is also equal to

$$\Phi_{GSC}(\omega) = \frac{\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)}{\left[\mathbf{x}^H(\omega)\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{x}(\omega)\right]^2} \ . \tag{3.100}$$

Hence, *the noise sensitivity of the GSC and the multi-channel Wiener filter are equal in the case of a single speech source and if the fixed beamformer of the GSC is a matched filter.*

In [240] it has been shown that the noise sensitivity $\Phi(\omega)$ is large for low frequencies ω, small microphone distances d, and in complicated noise scenarios (e.g. multiple noise sources, noise source close to the speech source, diffuse noise). A high noise sensitivity implies an increased sensitivity of the signal enhancement algorithm to deviations from the assumed (nominal) signal model.

3.6 Combined noise and echo reduction

3.6.1 Introduction

In some speech communication applications, such as teleconferencing and hands-free mobile telephony in car environments, a specific type of noise is present, namely *far-end echo signals* coming from the remote site (see Fig. 1.1). Since the far-end signals emitted by the loudspeakers are readily available, these signals can be used as a reference for the noise sources (unlike the noise sources which we have considered until now, for which no reference is available). In this section we will assume that a single far-end echo source is present, although the presented results can be extended to the case of multiple far-end echo sources. The nth microphone signal $y_n[k]$ can then be written as (cf. Section 2.2)

$$y_n[k] = h_n[k] \otimes s[k] + v_n^u[k] + h_{n,0}^f[k] \otimes f_0[k] = x_n[k] + \underbrace{v_n^u[k] + v_n^f[k]}_{v_n[k]} , \quad (3.101)$$

with $v_n^u[k]$ the noise component from the unknown noise sources in the nth microphone signal, $h_{n,0}^f[k]$ the acoustic impulse response between the far-end loudspeaker and the nth microphone, and $f_0[k]$ the far-end echo signal emitted by the loudspeaker. Since the signal $f_0[k]$ is assumed to be known, it can be used in the signal enhancement algorithms, such that we can design not only the filters $w_n[k]$, $n = 0 \dots N-1$ (on the microphone signals), but also an additional filter $w^f[k]$ (on the far-end echo signal), cf. (2.24),

$$z[k] = \sum_{n=0}^{N-1} w_n[k] \otimes y_n[k] + w^f[k] \otimes f_0[k] \; . \quad (3.102)$$

The goal of combined noise and echo reduction [152][158][174] is to cancel the far-end echo components using the filter $w^f[k]$ and to reduce the noise components (and possibly also far-end echo components to some extent) using the filters $w_n[k]$. Several solutions have been proposed, which can be classified into three kinds of algorithms: *multi-channel echo cancellation followed by noise reduction* (see Fig. 3.4), *noise reduction followed by single-channel echo cancellation* (see Fig. 3.5) and *integrated noise and echo reduction* (see Fig. 3.6).

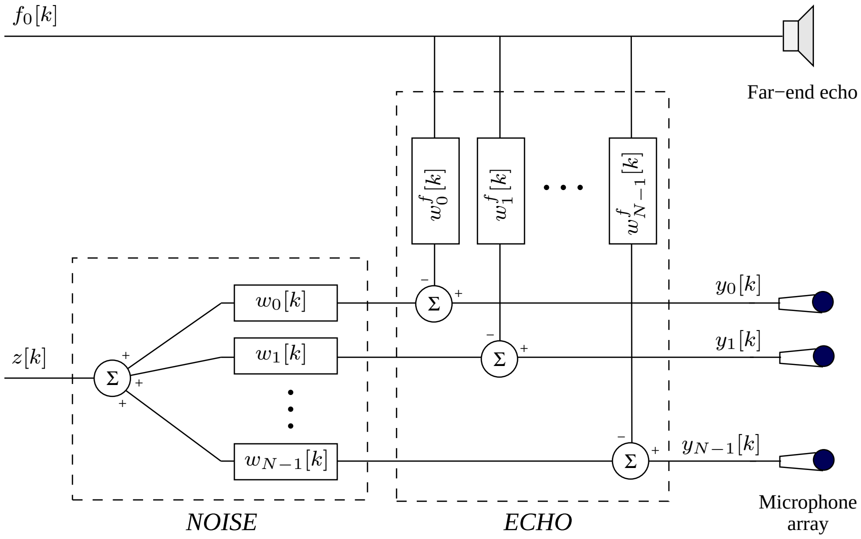


Figure 3.4: Combined noise and echo reduction scheme using multi-channel echo cancellation

In the scheme presented in Fig. 3.4, first the echo components $v_n^f[k]$ in all microphone signals are cancelled using N filters $w_n^f[k]$, $n = 0 \ldots N-1$. The goal of each filter $w_n^f[k]$ is to model the acoustic impulse response $h_n^f[k]$, which can be achieved by using an adaptive filtering algorithm (cf. Section 1.4.2). Next, a multi-microphone noise reduction technique (e.g. GSC, multi-channel Wiener filter) is applied to the echo-free microphone signals, yielding the same noise reduction performance as if no echo source were present. However, the computational complexity of this noise and echo reduction scheme is quite high, since N adaptive filters (typically with large filter lengths) are required.

In the scheme presented in Fig. 3.5, a single-channel echo canceller $w^f[k]$ is used after a multi-microphone noise reduction technique, thereby reducing the computational complexity compared to the scheme in Fig. 3.4. A possible implementation has been discussed in [125][126][151], where echo cancellation is combined with the GSC. Since the multi-microphone noise reduction technique just considers the far-end echo source as an additional (unknown) noise source, it now has to reduce both the noise and the far-end echo components. Therefore, it is obvious that the noise reduction performance for the unknown noise sources is worse than if no echo source were present. Of course, the noise reduction technique now also reduces the far-end echo components to some extent. The adaptive filter $w^f[k]$ has to model $\sum_{n=0}^{N-1} w_n[k] \otimes h_n^f[k]$, with $w_n[k]$, $n = 0 \ldots N-1$, generally also adaptive filters. This cascade of 2 adaptive filtering algorithms may give rise to problems when the adaptive filter $w^f[k]$ can not track the changes of the adaptive filters $w_n[k]$. Hence, an integrated

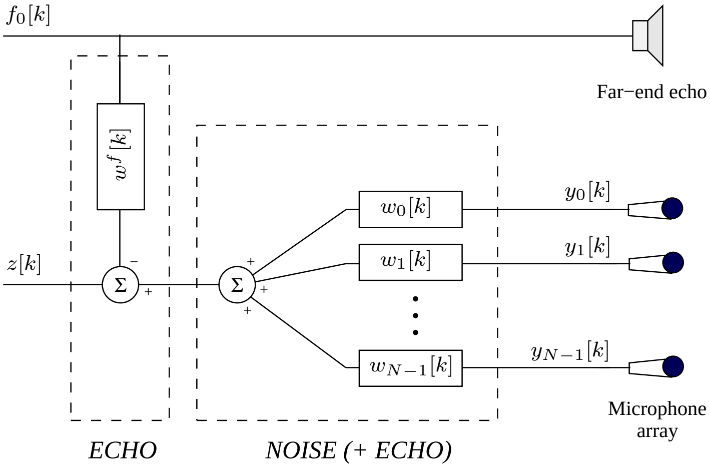


Figure 3.5: Combined noise and echo reduction scheme using single-channel echo cancellation

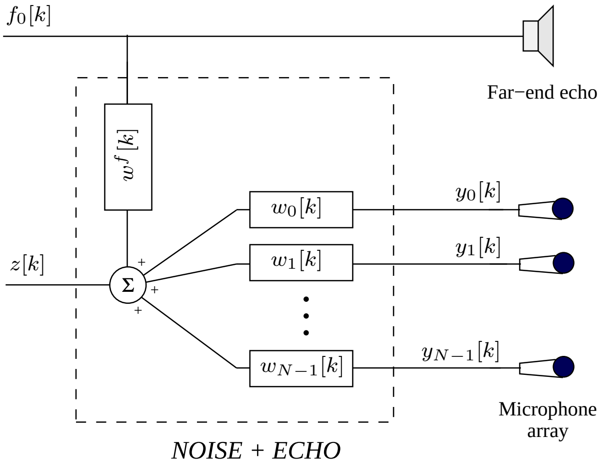


Figure 3.6: Combined noise and echo reduction scheme using an integrated approach

noise and echo reduction approach is called for, which is presented in Fig. 3.6. It will be shown that by using a multi-channel Wiener filter, theoretically *this scheme can obtain the same noise reduction performance for the unknown noise sources as if no echo source were present*, and the far-end echo components can be completely cancelled.

3.6.2 Integrated multi-channel Wiener filter approach

When *considering the far-end echo signal $f_0[k]$ in Fig. 3.6 merely as an additional input signal*, we can again use the multi-channel Wiener filter discussed in Sections 3.2 and 3.3 for computing an optimal estimate of the speech components $x_n[k]$, by just redefining the filter vector in (3.5) as

$$\mathbf{w}_t[k] = \begin{bmatrix} \mathbf{w}[k] \\ \mathbf{w}^f[k] \end{bmatrix} , \tag{3.103}$$

and by redefining the data vector in (3.6) as

$$\mathbf{y}_t[k] = \begin{bmatrix} \mathbf{y}[k] \\ \mathbf{f}[k] \end{bmatrix} . \tag{3.104}$$

The L_f-dimensional vectors $\mathbf{w}^f[k]$ and $\mathbf{f}[k]$ are equal to

$$\begin{aligned} \mathbf{w}^f[k] &= \begin{bmatrix} w_0^f[k] & w_1^f[k] & \ldots & w_{L_f-1}^f[k] \end{bmatrix}^T , & (3.105) \\ \mathbf{f}[k] &= \begin{bmatrix} f_0[k] & f_0[k-1] & \ldots & f_0[k-L_f+1] \end{bmatrix}^T , & (3.106) \end{aligned}$$

with the filter length L_f typically larger than L, since the filter $\mathbf{w}^f[k]$ needs to model a long acoustic impulse response. The input vector $\mathbf{y}_t[k]$ consists of a speech and a noise component, i.e.

$$\mathbf{y}_t[k] = \mathbf{x}_t[k] + \mathbf{v}_t[k] = \begin{bmatrix} \mathbf{x}[k] \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{v}[k] \\ \mathbf{f}[k] \end{bmatrix} , \tag{3.107}$$

where we obviously can assume that no speech components are present in the far-end echo signal $f_0[k]$.

In the frequency-domain, the nth microphone signal $Y_n(\omega)$ can be written as

$$Y_n(\omega) = X_n(\omega) + V_n(\omega) = X_n(\omega) + V_n^u(\omega) + V_n^f(\omega) , \tag{3.108}$$

with $V_n^u(\omega)$ the noise component from the unknown noise sources and $V_n^f(\omega)$ the noise component from the far-end echo source. The noise component $V_n^f(\omega)$ is equal to $H_{n,0}^f(\omega)F_0(\omega)$, with $H_{n,0}^f(\omega)$ the acoustic transfer function between the far-end loudspeaker and the nth microphone and $F_0(\omega)$ the far-end echo signal emitted by the loudspeaker. The stacked vector of microphone signals can then be written as

$$\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega) = \mathbf{X}(\omega) + \mathbf{V}_u(\omega) + \mathbf{V}_f(\omega) , \tag{3.109}$$

with $\mathbf{V}_f(\omega) = \mathbf{H}_f(\omega)F_0(\omega)$ and

$$\mathbf{H}_f(\omega) = \begin{bmatrix} H_{0,0}^f(\omega) & H_{1,0}^f(\omega) & \ldots & H_{N-1,0}^f(\omega) \end{bmatrix}^T . \tag{3.110}$$

Since the speech, noise and far-end echo components are assumed to be mutually uncorrelated, the correlation matrix $\bar{\mathbf{R}}_{yy}(\omega)$ can be written as

$$\bar{\mathbf{R}}_{yy}(\omega) = \bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega) + \bar{\mathbf{R}}_{vv}^f(\omega) = \bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega) + P_f(\omega)\mathbf{H}_f(\omega)\mathbf{H}_f^H(\omega) , \tag{3.111}$$

with $P_f(\omega) = \mathcal{E}\{|F_0(\omega)|^2\}$. Similarly as in (3.103) and (3.104), the $(N+1)$-dimensional filter vector in (3.76) can be redefined as

$$\mathbf{W}_t(\omega) = \begin{bmatrix} \mathbf{W}(\omega) \\ W^f(\omega) \end{bmatrix} , \tag{3.112}$$

with $W^f(\omega)$ the filter applied to the far-end echo signal $F_0(\omega)$, whereas the $(N+1)$-dimensional data vector can be redefined as

$$\mathbf{Y}_t(\omega) = \begin{bmatrix} \mathbf{Y}(\omega) \\ F_0(\omega) \end{bmatrix} = \begin{bmatrix} \mathbf{X}(\omega) \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{V}(\omega) \\ F_0(\omega) \end{bmatrix} = \mathbf{X}_t(\omega) + \mathbf{V}_t(\omega) \; . \tag{3.113}$$

Using (3.81), the Wiener filter for combined noise and echo reduction now is equal to

$$\mathbf{W}_{WF}^t(\omega) = \begin{bmatrix} \mathbf{W}_{WF}(\omega) \\ W_{WF}^f(\omega) \end{bmatrix} = \bar{\mathbf{R}}_{y_t y_t}^{-1}(\omega)\bar{\mathbf{R}}_{x_t x_t}(\omega)\,\mathbf{e}_1 \; , \tag{3.114}$$

with the matrices $\bar{\mathbf{R}}_{y_t y_t}(\omega)$ and $\bar{\mathbf{R}}_{x_t x_t}(\omega)$ defined as

$$\bar{\mathbf{R}}_{y_t y_t}(\omega) \;=\; \mathcal{E}\{\mathbf{Y}_t(\omega)\mathbf{Y}_t^H(\omega)\} \tag{3.115}$$
$$\bar{\mathbf{R}}_{x_t x_t}(\omega) \;=\; \mathcal{E}\{\mathbf{X}_t(\omega)\mathbf{X}_t^H(\omega)\} \; . \tag{3.116}$$

In Appendix C.2 it is shown that $\mathbf{W}_{WF}(\omega)$ is equal to

$$\boxed{\mathbf{W}_{WF}(\omega) = \left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1} \tag{3.117}$$

which is the same formula for the multi-channel Wiener filter as if no echo source were present, cf. (3.81), implying that the echo source has no influence on $\mathbf{W}_{WF}(\omega)$[6]. In Appendix C.2 it is also shown that $W_{WF}^f(\omega)$ is equal to

$$\boxed{W_{WF}^f(\omega) = -\mathbf{H}_f^H(\omega)\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1 = -\mathbf{H}_f^H(\omega)\mathbf{W}_{WF}(\omega)} \tag{3.118}$$

such that the total filter operation on the far-end echo signal $F(\omega)$ is equal to

$$\mathbf{W}_{WF}^{t,H}(\omega)\begin{bmatrix} \mathbf{H}_f(\omega) \\ 1 \end{bmatrix} = \mathbf{W}_{WF}^H(\omega)\mathbf{H}_f(\omega) + W_{WF}^{f,*}(\omega) = 0 \; , \tag{3.119}$$

implying that the far-end echo components in the microphone signals are completely cancelled by the multi-channel Wiener filter $\mathbf{W}_{WF}^t(\omega)$. This is not surprising, since infinitely long filters are assumed in this theoretical analysis,

[6] Note that the multi-channel Wiener filter for the scheme in Fig. 3.5 is equal to $\mathbf{W}_{WF}(\omega) = [\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega) + \bar{\mathbf{R}}_{vv}^f(\omega)]^{-1}\,\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1$, giving rise to a different noise reduction for the unknown noise sources and a different speech distortion than if no echo source were present.

and the filter $W^f_{WF}(\omega)$ therefore can always model $-\mathbf{H}^H_f(\omega)\mathbf{W}_{WF}(\omega)$, whatever the filter matrix $\mathbf{W}_{WF}(\omega)$ is. However, in practice the filters have a finite length (and the signals are not perfectly uncorrelated), such that the far-end echo source will have an influence on $\mathbf{W}_{WF}$ and the far-end echo components (generally) will not be completely cancelled.

The described combined noise and echo reduction technique has been realised using a GSVD-based implementation [47], and using a QRD-based implementation [223]. In [47] quite short filter lengths have been used (also for $w^f[k]$) due to the large computational complexity of the GSVD-based implementation. Because the QRD-based implementation has a smaller computational complexity, longer filter lengths have been used in [223], giving rise to an improved noise and echo reduction performance. However, since echo cancellation is not the main topic of this thesis, we will not further consider far-end echo sources.

3.7 Conclusion

In this chapter we have discussed a class of unconstrained optimal filtering techniques for multi-microphone speech enhancement, which combine the spatio-temporal information of the speech and the noise sources. The optimal filter in the MSE sense is the multi-channel Wiener filter, which produces an MMSE estimate of the speech components in the microphone signals. We have shown that the multi-channel Wiener filter belongs to a more general class of estimators where it is possible to trade off speech distortion and noise reduction. Although different possibilities exist for implementing the multi-channel Wiener filter, we have considered a GEVD-based implementation, enabling to easily incorporate the low-rank model of the speech signal . We have shown that the described class of optimal filtering techniques hence can be considered a multi-microphone extension of the single-microphone subspace-based techniques.

In Section 3.3 it has been shown that an empirical estimate of the optimal filter matrix can be computed using the GSVD of a speech and a noise data matrix. These data matrices are constructed based on the output of a VAD algorithm, which is the only a-priori information the GSVD-based optimal filtering technique relies on. We have shown that different estimates are obtained for the same speech component, and a procedure has been given for determining which estimate should be used (in practice typically the delayed speech component $x_0[k - \frac{L}{2} + 1]$ is selected). Both for the batch and for the recursive version of the GSVD-based optimal filtering technique, the computational complexity is quite high. Therefore, in Chapter 4 several techniques will be discussed for reducing this complexity.

In Section 3.4 we have derived a number of symmetry properties for the optimal filter, which are valid for the white noise case as well as for the coloured noise case and for any weighting function. In addition, we have examined the averaging operation, leading to the conclusion that this averaging operation is unnecessary and may even be suboptimal.

The unconstrained optimal filtering technique has been analysed in the time-domain (Section 3.2) and in the frequency-domain (Section 3.5). We have shown that the multi-channel Wiener filter in the frequency-domain can indeed be decomposed into a spectral and a spatial filtering term (under mild assumptions). We have simplified the expressions for the case of a single speech source and we have derived conditions under which the noise sensitivity of the GSC and the multi-channel Wiener filter are equal. We have shown that more speech distortion occurs at frequencies with a low SNR and when the spatial separation between the speech and the noise sources is poor and that more noise reduction occurs at frequencies with a low SNR and when the speech and the noise sources are spatially well separated.

In Section 3.6 we have shown that the unconstrained optimal filtering technique can also be used for combined noise and echo reduction. When assuming infinite-length filters, we have proved that the far-end echo source has no influence on the filters applied to the microphone signals and the far-end echo components in the microphone signals can be completely cancelled.

Chapter 4

Complexity reduction using recursive GSVD and ANC postprocessing stage

In this chapter, several techniques are discussed for reducing the computational complexity of the GSVD-based optimal filtering technique described in the previous chapter. First, several techniques are discussed for efficiently calculating the GSVD of the speech and the noise data matrix, making the GSVD-based optimal filtering technique amenable to real-time implementation. Secondly, it is shown that the GSVD-based optimal filtering technique can be incorporated in a GSC-type structure by adding an adaptive noise cancellation (ANC) post-processing stage. It will be shown by simulations in Chapter 5 that the same noise reduction performance can then be achieved with shorter filter lengths for the optimal filter, resulting in a lower overall complexity.

4.1 Introduction

In Chapter 3 a class of unconstrained optimal filtering techniques for multi-microphone speech enhancement has been discussed, which can be realised in practice using a GSVD-based implementation (cf. Section 3.3). In the recursive version of this GSVD-based optimal filtering technique (cf. Section 3.3.3), for each time step the speech and the noise data matrices are updated with the newly available speech or noise data vector, depending on the output of the VAD-algorithm. Since at each time step the GSVD and the optimal filter need to be recalculated, the computational complexity is quite high.

Section 4.2 describes several techniques for reducing the complexity by using recursive and square root-free Jacobi-type GSVD-updating algorithms, and by using sub-sampling. Instead of recomputing the GSVD from scratch for each time step, recursive algorithms compute the GSVD at time k using the decomposition at time $k-1$. Sub-sampling in this context means that the GSVD and the optimal filter are not updated for every sample. The computational complexity is summarised for realistic parameter values, showing that the complexity can be significantly reduced such that the recursive GSVD-based optimal filtering technique indeed becomes suitable for real-time implementation.

Section 4.3 describes how the GSVD-based optimal filtering technique can be incorporated in a GSC-type structure by creating speech and noise reference signals and by using these signals in an ANC postprocessing stage. The output of the GSVD-based optimal filtering technique is used as speech reference signal, whereas different possibilities exist for creating a noise reference. It will be shown by simulations in Chapter 5 that the same noise reduction performance can then be achieved with shorter filter lengths for the optimal filter, resulting in a lower overall complexity.

4.2 Recursive GSVD and sub-sampling

As already stated, the recursive version of the GSVD-based optimal filtering technique needs to recompute the GSVD of the speech and the noise data matrices for each time step, leading to a high computational complexity, even when using short filter lengths L. This section describes several techniques for drastically reducing the computational complexity by using recursive GSVD-updating algorithms and sub-sampling. Section 4.2.1 describes a Jacobi-type algorithm for computing the GSVD of two matrices using Givens rotations. This Jacobi-type algorithm lends itself well to a recursive GSVD-updating algorithm with a significantly lower computational complexity, described in Section 4.2.2. Section 4.2.3 discusses the square-root free implementation of this recursive GSVD-updating algorithm. For stationary acoustic environments the complexity can be further reduced by using sub-sampling, which is described in Section 4.2.4. A summary of the overall computational complexity of the considered algorithms for realistic parameter values is given in Section 4.2.5.

4.2.1 Jacobi-type algorithm for computing the GSVD

For conciseness the time index k will be omitted in this section. The GSVD of the $P \times M$-dimensional matrix $\mathbf{Y}$ and the $Q \times M$-dimensional matrix $\mathbf{V}$, cf. (3.43), can be computed as follows (for details we refer to [110][168][205][206] [262][263]). First, the matrices $\mathbf{Y}$ and $\mathbf{V}$ are reduced to upper-triangular form by a QR-decomposition,

$$\mathbf{Y} = \mathbf{Q}_Y \cdot \mathbf{R}_Y, \quad \mathbf{V} = \mathbf{Q}_V \cdot \mathbf{R}_V \;, \tag{4.1}$$

where $\mathbf{R}_Y$ and $\mathbf{R}_V$ are $M \times M$-dimensional upper-triangular matrices, and $\mathbf{Q}_Y$ and $\mathbf{Q}_V$ have orthonormal columns, i.e. $\mathbf{Q}_Y^T \cdot \mathbf{Q}_Y = \mathbf{Q}_V^T \cdot \mathbf{Q}_V = \mathbf{I}_M$. The GSVD of $\mathbf{Y}$ and $\mathbf{V}$ readily follows from the GSVD of $\mathbf{R}_Y$ and $\mathbf{R}_V$. The GSVD of the square matrices $\mathbf{R}_Y$ and $\mathbf{R}_V$ is computed by carrying out an iterative procedure, where a series of orthogonal Givens transformations are applied to $\mathbf{R}_Y$ and $\mathbf{R}_V$ in order to yield $M \times M$-dimensional upper-triangular factors $\mathbf{S}_Y$ and $\mathbf{S}_V$ with parallel rows, i.e.

$$\begin{cases} \mathbf{U}_{R_Y}^T \cdot \mathbf{R}_Y \cdot \mathbf{Q}_R = \mathbf{S}_Y = \mathbf{\Sigma}_Y \cdot \mathbf{R} \\ \mathbf{U}_{R_V}^T \cdot \mathbf{R}_V \cdot \mathbf{Q}_R = \mathbf{S}_V = \mathbf{\Sigma}_V \cdot \mathbf{R} \;, \end{cases} \tag{4.2}$$

with $\mathbf{U}_{R_Y}$, $\mathbf{U}_{R_V}$ and $\mathbf{Q}_R$ $M \times M$-dimensional orthogonal matrices, $\mathbf{\Sigma}_Y$ and $\mathbf{\Sigma}_V$ $M \times M$-dimensional diagonal matrices and $\mathbf{R}$ a $M \times M$-dimensional upper-triangular matrix. Combining (4.1) and (4.2), the GSVD of $\mathbf{Y}$ and $\mathbf{V}$ can be written as

$$\begin{cases} \mathbf{Y} = \mathbf{Q}_Y \cdot \mathbf{R}_Y = \mathbf{U}_Y \cdot \mathbf{S}_Y \cdot \mathbf{Q}_R^T \triangleq \mathbf{U}_Y \cdot \mathbf{\Sigma}_Y \cdot \mathbf{Q}^T \\ \mathbf{V} = \mathbf{Q}_V \cdot \mathbf{R}_V = \mathbf{U}_V \cdot \mathbf{S}_V \cdot \mathbf{Q}_R^T \triangleq \mathbf{U}_V \cdot \mathbf{\Sigma}_V \cdot \mathbf{Q}^T \;, \end{cases} \tag{4.3}$$

with $\mathbf{U}_Y = \mathbf{Q}_Y \mathbf{U}_{R_Y}$, $\mathbf{U}_V = \mathbf{Q}_V \mathbf{U}_{R_V}$ and $\mathbf{Q}^T = \mathbf{R}\,\mathbf{Q}_R^T$. The algorithm for computing the matrices $\mathbf{U}_{R_Y}$, $\mathbf{U}_{R_V}$, $\mathbf{S}_Y$, $\mathbf{S}_V$ and $\mathbf{Q}_R$ is presented in Table 4.1 (typically only $\mathbf{S}_Y$, $\mathbf{S}_V$ and $\mathbf{Q}_R$ are stored).

```
1. Initialisation:   $\mathbf{S}_Y \Leftarrow \mathbf{R}_Y$     $\mathbf{U}_{R_Y} \Leftarrow \mathbf{I}_M$     $\mathbf{Q}_R \Leftarrow \mathbf{I}_M$
                     $\mathbf{S}_V \Leftarrow \mathbf{R}_V$     $\mathbf{U}_{R_V} \Leftarrow \mathbf{I}_M$

2. Iterative GSVD-procedure:
     for $j = 1 \ \ldots \ \alpha M$ (sweeps)
       for $i = 1 \ \ldots \ M-1$ (GSVD-steps)
         $\mathbf{S}_Y \Leftarrow \mathbf{\Theta}_{i,j}^T \cdot \mathbf{S}_Y \cdot \mathbf{Q}_{i,j}$     $\mathbf{U}_{R_Y} \Leftarrow \mathbf{U}_{R_Y} \cdot \mathbf{\Theta}_{i,j}$     (4.4)
         $\mathbf{S}_V \Leftarrow \mathbf{\Phi}_{i,j}^T \cdot \mathbf{S}_V \cdot \mathbf{Q}_{i,j}$     $\mathbf{U}_{R_V} \Leftarrow \mathbf{U}_{R_V} \cdot \mathbf{\Phi}_{i,j}$     (4.5)
         $\mathbf{Q}_R \Leftarrow \mathbf{Q}_R \cdot \mathbf{Q}_{i,j}$                                        (4.6)
       end
     end
```

Table 4.1: Algorithm for computing the GSVD of $\mathbf{R}_Y$ and $\mathbf{R}_V$

The orthogonal matrices $\mathbf{\Theta}_{i,j}$ and $\mathbf{\Phi}_{i,j}$ in (4.4) and (4.5) represent plane Givens rotations with rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ in the $(i, i+1)$-plane, i.e.

$$\mathbf{\Theta}_{i,j} = \begin{bmatrix} \mathbf{I}_{i-1} & & & \\ & -\sin\theta_{i,j} & \cos\theta_{i,j} & \\ & \cos\theta_{i,j} & \sin\theta_{i,j} & \\ & & & \mathbf{I}_{M-i-1} \end{bmatrix}, \tag{4.7}$$

$$\mathbf{\Phi}_{i,j} = \begin{bmatrix} \mathbf{I}_{i-1} & & & \\ & -\sin\phi_{i,j} & \cos\phi_{i,j} & \\ & \cos\phi_{i,j} & \sin\phi_{i,j} & \\ & & & \mathbf{I}_{M-i-1} \end{bmatrix}. \tag{4.8}$$

In each iteration, the computation of the rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ and $\mathbf{Q}_{i,j}$, essentially reduces to the GSVD of the elementary 2×2-dimensional blocks $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$ on the main diagonal, where $\{\mathbf{A}\}_{i,i+1}$ denotes the 2×2-dimensional matrix on the intersection of the rows $\{i, i+1\}$ and the columns $\{i, i+1\}$ of the matrix $\mathbf{A}$. The pivot index i repeatedly takes up all possible values $i = 1 \ldots M-1$ on the main diagonal. Here, one such sequence is referred to as a sweep ($= M-1$ GSVD-steps).

Since the GSVD of the upper-triangular matrices $\mathbf{S}_Y$ and $\mathbf{S}_V$ corresponds to the SVD of the upper-triangular matrix $\mathbf{S}_C = \mathbf{S}_Y\mathbf{S}_V^{-1}$, it is possible to implicitly apply a Jacobi-type SVD-algorithm to $\mathbf{S}_C$ without explicitly having to compute $\mathbf{S}_V^{-1}$ and $\mathbf{S}_C$ [168]. The GSVD of the 2×2-dimensional blocks $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$ corresponds to the SVD of the 2×2-dimensional block $\{\mathbf{S}_C\}_{i,i+1}$, since it can be easily proved that for the upper-triangular matrices $\mathbf{S}_C$, $\mathbf{S}_Y$ and $\mathbf{S}_V$, the following relation holds,

$$\{\mathbf{S}_C\}_{i,i+1} = \{\mathbf{S}_Y\}_{i,i+1} \cdot \{\mathbf{S}_V^{-1}\}_{i,i+1} = \{\mathbf{S}_Y\}_{i,i+1} \cdot \{\mathbf{S}_V\}_{i,i+1}^{-1}, \tag{4.9}$$

such that

$$\{\mathbf{S}_C\}_{i,i+1} = \begin{bmatrix} s_C^{i,i} & s_C^{i,i+1} \\ 0 & s_C^{i+1,i+1} \end{bmatrix} = \begin{bmatrix} \frac{s_Y^{i,i}}{s_V^{i,i}} & \frac{s_Y^{i,i+1} s_V^{i,i} - s_Y^{i,i} s_V^{i,i+1}}{s_V^{i,i} s_V^{i+1,i+1}} \\ 0 & \frac{s_Y^{i+1,i+1}}{s_V^{i+1,i+1}} \end{bmatrix}. \tag{4.10}$$

Calculating the SVD of $\{\mathbf{S}_C\}_{i,i+1}$ comes down to calculating the Givens rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ in (4.4) and (4.5) such that

$$\underbrace{\begin{bmatrix} \tilde{s}_C^{i,i} & 0 \\ 0 & \tilde{s}_C^{i+1,i+1} \end{bmatrix}}_{\{\mathbf{\Sigma}\}_{i,i+1}} = \tag{4.11}$$

$$\underbrace{\begin{bmatrix} -\sin\theta_{i,j} & \cos\theta_{i,j} \\ \cos\theta_{i,j} & \sin\theta_{i,j} \end{bmatrix}}_{\{\mathbf{\Theta}_{i,j}^T\}_{i,i+1}} \cdot \underbrace{\begin{bmatrix} s_C^{i,i} & s_C^{i,i+1} \\ 0 & s_C^{i+1,i+1} \end{bmatrix}}_{\{\mathbf{S}_C\}_{i,i+1}} \cdot \underbrace{\begin{bmatrix} -\sin\phi_{i,j} & \cos\phi_{i,j} \\ \cos\phi_{i,j} & \sin\phi_{i,j} \end{bmatrix}}_{\{\mathbf{\Phi}_{i,j}\}_{i,i+1}}.$$

Different possibilities for calculating these rotation angles have been discussed in [31][168], and generally consist of a symmetrisation step and a diagonalisation step. The orthogonal transformations $\{\mathbf{\Theta}_{i,j}\}_{i,i+1}$ and $\{\mathbf{\Phi}_{i,j}\}_{i,i+1}$ are seen to parallelise the rows of $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$, i.e.

$$\{\mathbf{\Theta}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_Y\}_{i,i+1} = \{\mathbf{\Sigma}\}_{i,i+1} \cdot \{\mathbf{\Phi}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_V\}_{i,i+1} \ , \tag{4.12}$$

which then allows for a joint upper-triangularising orthogonal transformation $\{\mathbf{Q}_{i,j}\}_{i,i+1}$ in order to obtain the GSVD of $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$,

$$\underbrace{\{\mathbf{\Theta}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_Y\}_{i,i+1} \cdot \{\mathbf{Q}_{i,j}\}_{i,i+1}}_{\text{upper-triangular}} = \{\mathbf{\Sigma}\}_{i,i+1} \cdot \underbrace{\{\mathbf{\Phi}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_V\}_{i,i+1} \cdot \{\mathbf{Q}_{i,j}\}_{i,i+1}}_{\text{upper-triangular}} . \tag{4.13}$$

In each iteration, the sum-of-squares of the off-diagonal elements of $\mathbf{S}_C$ is reduced by $(s_C^{i,i+1})^2$, and it has been shown that after αM sweeps (with α typically $3 \ldots 5$) the algorithm converges, i.e. the sum-of-squares of the off-diagonal elements of $\mathbf{S}_C$ is a very small number, such that $\mathbf{S}_C$ can be considered to be a diagonal matrix.

Table 4.2 summarises the *computational complexity*, defined as the number of additions and multiplications, of the GSVD-calculation in iteration step i and for one sweep, i.e. $i = 1 \ldots M-1$. The number of square-root operations is also stated in Table 4.2. The computational complexity for one sweep is (approximately) equal to $18M^2$. Since computing a full GSVD first requires a QR-decomposition of the $P \times M$-dimensional matrix $\mathbf{Y}$ and the $Q \times M$-dimensional matrix $\mathbf{V}$ and requires αM sweeps in the iterative GSVD-procedure, the total computational complexity is equal to

$$\underbrace{3M^2(P+Q-2M/3)}_{\text{QR-decomposition}} + \underbrace{18\alpha M^3}_{\text{GSVD-procedure}} = (18\alpha - 2)M^3 + 3M^2(P+Q) \ . \tag{4.14}$$

For typical values of P, Q and M, the complexity of this algorithm is too high to be suitable for real-time implementation (cf. Table 4.4).

	Step i	*One sweep* $(i = 1 \ldots M-1)$
Calculation $\theta_{i,j}$ and $\phi_{i,j}$	$43 + 3\sqrt{\cdot}$	$43(M-1) + 3(M-1)\sqrt{\cdot}$
Calculation $\mathbf{Q}_{i,j}$	$5 + 1\sqrt{\cdot}$	$5(M-1) + (M-1)\sqrt{\cdot}$
Multiplication $\mathbf{\Theta}_{i,j}^T \cdot \mathbf{S}_Y$	$6(M-i)+2$	$3M^2 - M - 2$
Multiplication $\mathbf{\Phi}_{i,j}^T \cdot \mathbf{S}_V$	$6(M-i)+2$	$3M^2 - M - 2$
Multiplication $\tilde{\mathbf{S}}_Y \cdot \mathbf{Q}_{i,j}$	$6i$	$3M^2 - 3M$
Multiplication $\tilde{\mathbf{S}}_V \cdot \mathbf{Q}_{i,j}$	$6i$	$3M^2 - 3M$
Multiplication $\mathbf{Q}_R \cdot \mathbf{Q}_{i,j}$	$6M$	$6M^2 - 6M$
Total		$18M^2 + 34M - 52 + 4(M-1)\sqrt{\cdot}$

Table 4.2: Computational complexity for one sweep in GSVD-calculation

4.2.2 Recursive GSVD-updating algorithm

Instead of recomputing the GSVD from scratch, recursive GSVD-updating algorithms compute the GSVD at time k using the decomposition at time $k-1$. In [181][183][184] a Jacobi-type (G)SVD-updating algorithm has been described. Suppose that at time $k-1$ the upper-triangular factors are reduced to $\mathbf{S}_Y[k-1]$ and $\mathbf{S}_V[k-1]$ having approximately parallel rows, cf. (4.2),

$$\left\{\begin{array}{rcl} \mathbf{Y}[k-1] & = & \mathbf{U}_Y[k-1]\cdot\mathbf{S}_Y[k-1]\cdot\mathbf{Q}_R^T[k-1] \\ & \triangleq & \mathbf{U}_Y[k-1]\cdot\mathbf{\Sigma}_Y[k-1]\cdot\mathbf{Q}^T[k-1] \\ \mathbf{V}[k-1] & = & \mathbf{U}_V[k-1]\cdot\mathbf{S}_V[k-1]\cdot\mathbf{Q}_R^T[k-1] \\ & \triangleq & \mathbf{U}_V[k-1]\cdot\mathbf{\Sigma}_V[k-1]\cdot\mathbf{Q}^T[k-1]\ , \end{array}\right. \tag{4.15}$$

of which only the upper-triangular matrices $\mathbf{S}_Y[k-1]$, $\mathbf{S}_V[k-1]$ and the orthogonal matrix $\mathbf{Q}_R[k-1]$ are stored and updated.

At time k, a new data vector $\mathbf{y}[k]$ is present, such that we need to recompute the GSVD of the updated data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$, which are constructed by exponentially weighting $\mathbf{Y}[k-1]$ and $\mathbf{V}[k-1]$ (when using fixed length data windows, also a down-date has to be performed, which is not numerically stable). If $\mathbf{y}[k]$ is classified by the VAD-algorithm as a speech-and-noise vector ($\zeta[k]=1$), only the speech data matrix $\mathbf{Y}[k]$ is updated, i.e.

$$\mathbf{Y}[k]=\left[\begin{array}{c}\lambda_y\cdot\mathbf{Y}[k-1]\\ \mathbf{y}^T[k]\end{array}\right],\quad \mathbf{V}[k]=\mathbf{V}[k-1]\ , \tag{4.16}$$

whereas if $\mathbf{y}[k]$ is classified as a noise-only vector ($\zeta[k]=0$), only the noise data matrix $\mathbf{V}[k]$ is updated, i.e.

$$\mathbf{Y}[k]=\mathbf{Y}[k-1],\quad \mathbf{V}[k]=\left[\begin{array}{c}\lambda_v\cdot\mathbf{V}[k-1]\\ \mathbf{y}^T[k]\end{array}\right]\ , \tag{4.17}$$

with λ_y an exponential weighting factor for speech and λ_v an exponential weighting factor for noise (if $\lambda=1$, no weighting is performed). Assuming that $\mathbf{y}[k]$ is classified as a speech-and-noise vector, the speech data matrix $\mathbf{Y}[k]$ can be rewritten as

$$\mathbf{Y}[k]=\left[\begin{array}{cc}\boxed{\mathbf{U}_Y[k-1]} & \boxed{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\\ \boxed{0\ \ \ldots\ \ 0} & \boxed{1}\end{array}\right]\cdot\left[\begin{array}{c}\lambda_y\cdot\mathbf{S}_Y[k-1]\\ \mathbf{y}^T[k]\cdot\mathbf{Q}_R[k-1]\end{array}\right]\cdot\mathbf{Q}_R^T[k-1]\ . \tag{4.18}$$

First, the upper-triangular factor is restored by performing a QR-update with the transformed input vector $\tilde{\mathbf{y}}^T[k]=\mathbf{y}^T[k]\cdot\mathbf{Q}_R[k-1]$. QR-updating can be performed by using orthogonal Givens rotations, zeroing the elements on the bottom row, yielding the upper-triangular matrix $\tilde{\mathbf{S}}_Y[k]$,

$$\mathbf{Y}[k] = \underbrace{\left[\begin{array}{c|c} \boxed{\mathbf{U}_Y[k-1]} & \boxed{\begin{array}{c}0\\ \vdots\\ 0\end{array}} \\ \boxed{0 \; \ldots \; 0} & \boxed{1} \end{array}\right] \cdot \tilde{\mathbf{Q}}_Y[k]}_{\tilde{\mathbf{U}}_Y[k]} \cdot \tilde{\mathbf{S}}_Y[k] \cdot \mathbf{Q}_R^T[k-1] \,. \tag{4.19}$$

In this equation, $\tilde{\mathbf{Q}}_Y[k]$ is an $(M+1) \times M$-dimensional matrix with orthogonal columns, which does not need to be computed explicitly. Note that the matrix $\mathbf{Q}_R[k-1]$ is not altered by the QR-update. If $\mathbf{y}[k]$ is classified as a noise-only vector, a similar procedure needs to be performed for $\mathbf{V}[k]$ instead of for $\mathbf{Y}[k]$, i.e.

$$\mathbf{V}[k] = \underbrace{\left[\begin{array}{c|c} \boxed{\mathbf{U}_V[k-1]} & \boxed{\begin{array}{c}0\\ \vdots\\ 0\end{array}} \\ \boxed{0 \; \ldots \; 0} & \boxed{1} \end{array}\right] \cdot \tilde{\mathbf{Q}}_V[k]}_{\tilde{\mathbf{U}}_V[k]} \cdot \tilde{\mathbf{S}}_V[k] \cdot \mathbf{Q}_R^T[k-1] \,. \tag{4.20}$$

Secondly, the iterative GSVD-procedure is resumed in order to further parallelise the rows of the square upper-triangular matrices $\tilde{\mathbf{S}}_Y[k]$ and $\tilde{\mathbf{S}}_V[k]$. A fixed number of sweeps (s) is performed, where the pivot index i takes up r consecutive values. Typically one sweep is performed ($s = 1$), where the pivot index takes up all possible values along the main diagonal ($r = M - 1$).

The complete procedure at time k, where only the square $M \times M$-dimensional upper-triangular matrices $\mathbf{S}_Y[k]$ and $\mathbf{S}_V[k]$ and the $M \times M$-dimensional orthogonal matrix $\mathbf{Q}_R[k]$ are stored and updated, is summarised in Table 4.3. The computational complexity of one GSVD-update then is equal to

$$\underbrace{0.5M^2}_{\text{weighting}} + \underbrace{2M^2}_{\text{input vector}} + \underbrace{3M^2}_{\text{QR-update}} + \underbrace{s \cdot r/(M-1) \cdot 18M^2}_{\text{GSVD-procedure}} \,, \tag{4.21}$$

such that for $s = 1$ and $r = M - 1$ the complexity amounts to $23.5M^2$.

The optimal filter matrix $\mathbf{W}_{WF}[k]$ in (3.44) can now be computed as

$$\mathbf{W}_{WF}[k] = \mathbf{Q}^{-T}[k] \,\mathrm{diag}\left\{1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\eta_i^2[k]}{\sigma_i^2[k]}\right\} \mathbf{Q}^T[k] \,, \tag{4.29}$$

where the factor P/Q has been replaced by $(1-\lambda_v^2)/(1-\lambda_y^2)$, because exponential weighting factors λ_y and λ_v are used. Upon convergence of the recursive GSVD-updating algorithm, it follows from (4.15) that

$$\mathbf{Q}^T[k] = \mathbf{\Sigma}_Y^{-1}[k] \cdot \mathbf{S}_Y[k] \cdot \mathbf{Q}_R^T[k] \,, \tag{4.30}$$

```
1. matrix-vector multiplication and QR-update
      if $\zeta[k] = 1$ (speech-and-noise)
            $\mathbf{S}_Y[k] \Leftarrow \tilde{\mathbf{Q}}_Y^T[k] \cdot \begin{bmatrix} \lambda_y \cdot \mathbf{S}_Y[k-1] \\ \mathbf{y}^T[k] \cdot \mathbf{Q}_R[k-1] \end{bmatrix}$      (4.22)
            $\mathbf{S}_V[k] \Leftarrow \mathbf{S}_V[k-1]$      (4.23)
      else if $\zeta[k] = 0$ (noise-only)
            $\mathbf{S}_Y[k] \Leftarrow \mathbf{S}_Y[k-1]$      (4.24)
            $\mathbf{S}_V[k] \Leftarrow \tilde{\mathbf{Q}}_V^T[k] \cdot \begin{bmatrix} \lambda_v \cdot \mathbf{S}_V[k-1] \\ \mathbf{y}^T[k] \cdot \mathbf{Q}_R[k-1] \end{bmatrix}$      (4.25)
      end
      $\mathbf{Q}_R[k] \Leftarrow \mathbf{Q}_R[k-1]$
2. GSVD-update procedure
      $r_{k+1} = \mathrm{mod}(r_k + r - 1, M - 1) + 1$
      for $j = 1 \ \ldots \ s$ (sweeps)
        for $i = r_k \ \ldots \ r_{k+1} - 1$ (GSVD-steps)
              $\mathbf{S}_Y[k] \Leftarrow \mathbf{\Theta}_{i,j}^T[k] \cdot \mathbf{S}_Y[k] \cdot \mathbf{Q}_{i,j}[k]$      (4.26)
              $\mathbf{S}_V[k] \Leftarrow \mathbf{\Phi}_{i,j}^T[k] \cdot \mathbf{S}_V[k] \cdot \mathbf{Q}_{i,j}[k]$      (4.27)
              $\mathbf{Q}_R[k] \Leftarrow \mathbf{Q}_R[k] \cdot \mathbf{Q}_{i,j}[k]$      (4.28)
        end
      end
```

Table 4.3: Algorithm for recursive GSVD-updating using Jacobi-rotations

and $s_Y^{i,i}[k]/s_V^{i,i}[k] = \sigma_i[k]/\eta_i[k]$, since $\mathbf{S}_Y[k]$ and $\mathbf{S}_V[k]$ have parallel rows[1]. Hence, $\mathbf{W}_{WF}[k]$ can be computed as

$$\begin{aligned}\mathbf{W}_{WF}[k] &= \mathbf{Q}_R[k] \cdot \mathbf{S}_Y^{-1}[k] \cdot \mathbf{\Sigma}_Y[k] \, \mathrm{diag}\Big\{1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\eta_i^2[k]}{\sigma_i^2[k]}\Big\} \, \mathbf{\Sigma}_Y^{-1}[k] \cdot \\ &\qquad \mathbf{S}_Y[k] \cdot \mathbf{Q}_R^T[k] \qquad (4.31) \\ &= \mathbf{Q}_R[k] \cdot \mathbf{S}_Y^{-1}[k] \, \mathrm{diag}\Big\{1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\big(s_V^{i,i}[k]\big)^2}{\big(s_Y^{i,i}[k]\big)^2}\Big\} \, \mathbf{S}_Y[k] \cdot \mathbf{Q}_R^T[k] \, . \qquad (4.32)\end{aligned}$$

[1] When the recursive GSVD-updating algorithm has not reached convergence, $\mathbf{S}_Y[k]$ and $\mathbf{S}_V[k]$ only have approximately parallel rows, such that (4.32) is an approximation of the filter matrix $\mathbf{W}_{WF}[k]$.

Since only the ith column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{WF}[k]$ needs to be computed (cf. Section 3.3.2), this column can be computed as the solution of the linear set of equations

$$\boxed{\mathbf{S}_Y[k]\cdot\underbrace{\mathbf{Q}_R^T[k]\cdot\mathbf{w}_{WF,i}[k]}_{\tilde{\mathbf{w}}[k]}=\underbrace{\text{diag}\Big\{1-\frac{(1-\lambda_v^2)}{(1-\lambda_y^2)}\frac{\left(s_V^{i,i}[k]\right)^2}{\left(s_Y^{i,i}[k]\right)^2}\Big\}\cdot\mathbf{S}_Y[k]\cdot\mathbf{q}_{R,i}[k]}_{\tilde{\mathbf{q}}[k]}} \tag{4.33}$$

with $\mathbf{q}_{R,i}[k]$ the ith column of $\mathbf{Q}_R^T[k]$. The calculation of $\mathbf{w}_{WF,i}[k]$ consists of computing $\tilde{\mathbf{q}}[k]$, requiring M^2 operations (multiplication of triangular matrix with vector), solving the equation $\mathbf{S}_Y[k]\cdot\tilde{\mathbf{w}}[k]=\tilde{\mathbf{q}}[k]$ by back-substitution, requiring M^2 operations, and computing $\mathbf{w}_{WF,i}[k]$ as

$$\mathbf{w}_{WF,i}[k]=\mathbf{Q}_R[k]\cdot\tilde{\mathbf{w}}[k]\ , \tag{4.34}$$

requiring $2M^2$ operations. Hence, the total computational complexity for computing $\mathbf{w}_{WF,i}[k]$ from the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ amounts to $4M^2$.

4.2.3 Square root-free implementation

The computational complexity can be further reduced by using a square root-free implementation for the QR-updates and for the calculation of the elementary 2×2-dimensional GSVDs. The calculation of the rotation angles for a QR-update and for an elementary 2×2 -dimensional GSVD respectively requires one and three square roots [31][168] (cf. Table 4.2). Gentleman has developed a square root-free procedure for QR-updating where a one-sided factorisation of the upper-triangular $\mathbf{R}$-matrix is used [105]. However, since the above GSVD-schemes as such do not lend themselves to square root-free implementation, alternative schemes based on approximate formulas for the calculation of the rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ have to be considered [31]. When combined with a generalised Gentleman procedure with a two-sided factorisation of the upper-triangular $\mathbf{S}$-factor, these schemes eventually yield square root-free SVD-updating algorithms [185], which can easily be extended to square root-free GSVD-updating algorithms [183].

For a square root-free QR-decomposition and QR-update of e.g. the speech data matrix $\mathbf{Y}$, the matrices $\mathbf{Q}_Y$ and $\mathbf{R}_Y$ are factorised as

$$\mathbf{R}_Y=\mathbf{D}^{\frac{1}{2}}\cdot\bar{\mathbf{R}}_Y,\quad \mathbf{Q}_Y=\bar{\mathbf{Q}}_Y\cdot\mathbf{D}^{\frac{1}{2}}\ , \tag{4.35}$$

with $\mathbf{D}$ a diagonal matrix, performing a row scaling for the upper-triangular factor $\mathbf{R}_Y$ and a column scaling for the orthogonal matrix $\mathbf{Q}_Y$. In a square root-free QR-update only the diagonal matrix $\mathbf{D}$ and the upper-triangular factor $\bar{\mathbf{R}}_Y$ are stored and updated, without calculating square roots, i.e. $\mathbf{D}^{\frac{1}{2}}$ is never computed explicitly.

For a square root-free 2×2-dimensional GSVD with approximate formulas, the relevant transformation formula (4.11) becomes

$$\begin{bmatrix} \tilde{s}_C^{i,i} & \tilde{s}_C^{i,i+1} \\ 0 & \tilde{s}_C^{i+1,i+1} \end{bmatrix} = \tag{4.36}$$

$$\begin{bmatrix} -\sin\theta_{i,j} & \cos\theta_{i,j} \\ \cos\theta_{i,j} & \sin\theta_{i,j} \end{bmatrix} \cdot \begin{bmatrix} s_C^{i,i} & s_C^{i,i+1} \\ 0 & s_C^{i+1,i+1} \end{bmatrix} \cdot \begin{bmatrix} -\sin\phi_{i,j} & \cos\phi_{i,j} \\ \cos\phi_{i,j} & \sin\phi_{i,j} \end{bmatrix},$$

where $s_C^{i,i+1}$ is now only approximately annihilated, i.e. $|\tilde{s}_C^{i,i+1}| \leq |s_C^{i,i+1}|$. Several approximate formulas exist for calculating $\tan\theta_{i,j}$ and $\tan\phi_{i,j}$ without square roots. For details we refer to [31][185].

For the square root-free GSVD-update procedure, the matrices $\mathbf{U}_Y$, $\mathbf{U}_V$, $\mathbf{S}_Y$, $\mathbf{S}_V$ and $\mathbf{Q}_R$ are factorised as

$$\begin{aligned} \mathbf{S}_Y &= (\mathbf{D}_{row}^Y)^{\frac{1}{2}} \cdot \bar{\mathbf{S}}_Y \cdot (\mathbf{D}_{col})^{\frac{1}{2}},\ \mathbf{S}_V = (\mathbf{D}_{row}^V)^{\frac{1}{2}} \cdot \bar{\mathbf{S}}_V \cdot (\mathbf{D}_{col})^{\frac{1}{2}} && (4.37) \\ \mathbf{U}_Y &= \bar{\mathbf{U}}_Y \cdot (\mathbf{D}_{row}^Y)^{\frac{1}{2}},\ \mathbf{U}_V = \bar{\mathbf{U}}_V \cdot (\mathbf{D}_{row}^V)^{\frac{1}{2}},\ \mathbf{Q}_R = \bar{\mathbf{Q}}_R \cdot (\mathbf{D}_{col})^{\frac{1}{2}}\ , && (4.38) \end{aligned}$$

with $\mathbf{D}_{row}^Y$, $\mathbf{D}_{row}^V$ and $\mathbf{D}_{col}$ diagonal (row and column scaling) matrices. In a GSVD-update only the diagonal matrices $\mathbf{D}_{row}^Y$, $\mathbf{D}_{row}^V$ and $\mathbf{D}_{col}$, the upper-triangular matrices $\bar{\mathbf{S}}_Y$ and $\bar{\mathbf{S}}_V$, and the matrix $\bar{\mathbf{Q}}_R$ are stored and updated, without calculating any square roots [183][185]. The actual complexity reduction results from the fact that the 2×2-dimensional row and column transformation matrices in the update formulas contain ones along the diagonal (or anti-diagonal), hereby halving the number of multiplications required (for details we refer to [185]). If we substitute these factorisations into (4.15), the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ can be written as

$$\begin{cases} \mathbf{Y}[k] &= \bar{\mathbf{U}}_Y[k] \cdot \mathbf{D}_{row}^Y[k] \cdot \bar{\mathbf{S}}_Y[k] \cdot \mathbf{D}_{col}[k] \cdot \bar{\mathbf{Q}}_R^T[k] \\ \mathbf{V}[k] &= \bar{\mathbf{U}}_V[k] \cdot \mathbf{D}_{row}^V[k] \cdot \bar{\mathbf{S}}_V[k] \cdot \mathbf{D}_{col}[k] \cdot \bar{\mathbf{Q}}_R^T[k]\ , \end{cases} \tag{4.39}$$

such that the optimal filter $\mathbf{W}_{WF}[k]$ in (4.32) can be computed as

$$\begin{aligned} \mathbf{W}_{WF}[k] &= \bar{\mathbf{Q}}_R[k] \cdot \bar{\mathbf{S}}_Y^{-1}[k] \cdot \mathrm{diag}\Big\{1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{d_{row}^{V,i}[k]}{d_{row}^{Y,i}[k]} \frac{(\bar{s}_V^{i,i}[k])^2}{(\bar{s}_Y^{i,i}[k])^2}\Big\} \cdot \\ & \quad \bar{\mathbf{S}}_Y[k] \cdot \mathbf{D}_{col}[k] \cdot \bar{\mathbf{Q}}_R^T[k]\ . \end{aligned} \tag{4.40}$$

Similarly to (4.21), the computational complexity of one square root-free GSVD-update is equal to

$$\underbrace{0.5M^2}_{\text{weighting}} + \underbrace{2M^2}_{\text{input vector}} + \underbrace{2M^2}_{\text{sqrt-free QR-update}} + \underbrace{s \cdot r/(M-1) \cdot 12M^2}_{\text{sqrt-free GSVD-procedure}}\ , \tag{4.41}$$

such that for $s = 1$ and $r = M - 1$ the complexity amounts to $16.5M^2$, which is less expensive than the 'conventional' non square root-free GSVD-updating procedure ($23.5M^2$). The computational complexity for computing the column $\mathbf{w}_{WF,i}[k]$ from the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ is again equal to $4M^2$.

	Non-recursive/Batch	Recursive	Square root-free
	$\frac{16M^3+3M^2(P+Q)}{s_g}$	$\frac{23.5M^2}{s_g}+\frac{4M^2}{s_f}$	$\frac{16.5M^2}{s_g}+\frac{4M^2}{s_f}$
$s_f = s_g = 1$	7504 Gflops	2.8 Gflops	2.1 Gflops
$s_f = s_g = 20$	375 Gflops	141 Mflops	105 Mflops

Table 4.4: Summary of overall complexity of GSVD-based optimal filtering technique for batch and recursive versions using realistic parameter values

4.2.4 Sub-sampling techniques

For stationary acoustic environments, the computational complexity can be reduced without any loss in performance by using sub-sampling techniques. In this context sub-sampling means that the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ and the optimal filter $\mathbf{w}_{WF,i}[k]$ are not updated for every sample, but that the GSVD is updated every s_g samples and that the optimal filter is updated every s_f samples. In Section 5.2.3 it will be shown that the convergence speed towards the converged optimal filter (for stationary environments) is slower if higher sub-sampling factors are used, implying that the amount of sub-sampling needs to be limited in non-stationary acoustic environments.

4.2.5 Overall computational complexity

Table 4.4 summarises the total computational complexity in floating point operations per second (flops) for the batch and the recursive version of the GSVD-based optimal filtering technique, assuming that $s = 1$, $r = M - 1$ and $\alpha = 1$. The numerical results are obtained for $N = 4$ microphones, filter length $L = 20$ ($M = 80$), sampling frequency $f_s = 16\,\text{kHz}$, data window lengths $P = 4000$ and $Q = 20000$ (for the batch version) and are shown both in case of no sub-sampling and in case of sub-sampling with $s_g = s_f = 20$. By using the recursive version of the GSVD-based optimal filtering technique, the computational complexity can be significantly reduced such that the algorithm becomes suitable for real-time implementation.

4.3 ANC postprocessing stage

As already discussed in Section 2.5.3, the GSC-structure depicted in Fig. 2.10 is a widely used structure in adaptive beamforming, where speech and noise reference signals are created and then used in an adaptive noise cancellation (ANC) stage [17][34][37][100][113][116][123][128][191][194][254][264]. The objective is to create a speech reference signal having a higher SNR than the original microphone signals and to create one or more noise reference signals containing as little speech energy as possible. A multi-channel adaptive filter

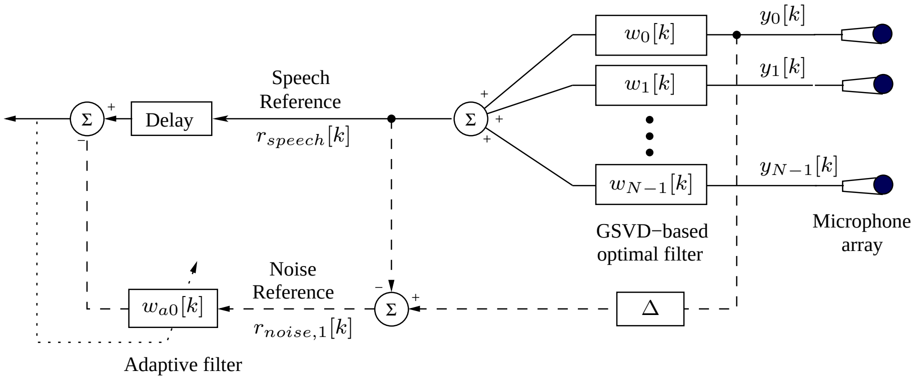


Figure 4.1: GSVD-based optimal filtering technique incorporated in GSC-type structure with ANC postprocessing stage

then removes the remaining correlation between the (residual) noise component in the speech reference signal and the noise reference signals. In order to avoid signal cancellation and distortion, signal leakage into the noise reference (e.g. caused by reverberation, microphone mismatch, look direction error and spatially distributed sources) needs to be minimised and the effect of the signal leakage on the ANC adaptive filters needs to be limited (cf. Section 2.5.3). For adaptive beamformers, signal leakage can be reduced by e.g. using a spatial filter designed blocking matrix [194][191], whereas the effect of the signal leakage on the ANC adaptive filters can be limited by e.g. using a speech-controlled adaptation algorithm [254][113][128][194].

However, instead of using a fixed beamformer to create the speech reference signal, it is also possible to use the GSVD-based optimal filtering technique. The complete noise reduction scheme, incorporating the GSVD-based optimal filtering technique in a GSC-type structure with an ANC postprocessing stage, is depicted in Fig. 4.1. The output signal of the GSVD-based optimal filter is used as the *speech reference* signal, cf. (3.19),

$$r_{speech}[k] = \hat{x}_m[k-\Delta] = \mathbf{y}^T[k]\,\mathbf{w}_{WF,i}[k]\ , \tag{4.42}$$

which is the optimal estimate for the speech component in the mth microphone signal (with delay Δ), obtained by filtering the microphone signals with $\mathbf{w}_{WF,i}[k]$, with $i = mL + \Delta + 1$. The residual noise level in the speech reference signal depends on the filter length L used for the GSVD-based optimal filter. For the creation of a *noise reference* different possibilities exist. An obvious choice consists in simply subtracting the speech reference signal from the delayed mth microphone signal, i.e.

$$r_{noise,1}[k] = y_m[k-\Delta] - r_{speech}[k] = y_m[k-\Delta] - \hat{x}_m[k-\Delta]\ . \tag{4.43}$$

Indeed, if $\mathbf{W}_{WF}[k]$ is the optimal filter matrix for estimating the speech components in the microphone signals, i.e.

$$\hat{\mathbf{x}}[k] = \mathbf{W}_{WF}^T[k]\,\mathbf{y}[k]\ , \tag{4.44}$$

then it is easily shown that $(\mathbf{I}_M - \mathbf{W}_{WF}[k])$ is the optimal filter matrix for estimating the noise components in the microphone signals, i.e.

$$\hat{\mathbf{v}}[k] = (\mathbf{I}_M - \mathbf{W}_{WF}^T[k])\,\mathbf{y}[k]\ . \tag{4.45}$$

The ith element of $\hat{\mathbf{v}}[k]$ is equal to the optimal estimate of the noise component in the mth microphone signal (with delay Δ),

$$\hat{v}_m[k-\Delta] = \mathbf{y}^T[k]\,(\mathbf{e}_i - \mathbf{w}_{WF,i}[k]) = y_m[k-\Delta] - \hat{x}_m[k-\Delta]\ , \tag{4.46}$$

with $\mathbf{e}_i$ defined in Section 3.5.1. The creation of this noise reference signal for the microphone signal $y_0[k]$ is depicted in Fig. 4.1. The adaptive filter $w_{a0}[k]$ with filter length L_{ANC} is used for reducing the remaining correlation between the residual noise component in the speech reference signal $r_{speech}[k]$, which is typically delayed with $\frac{L_{ANC}}{2}$ samples, and the noise reference signal $r_{noise,1}[k]$. Instead of only calculating a noise reference for one microphone signal, it is also possible to calculate noise references for all microphone signals, i.e.

$$\mathbf{r}_{noise,2}[k] = \begin{bmatrix} \hat{v}_0[k-\Delta] \\ \hat{v}_1[k-\Delta] \\ \vdots \\ \hat{v}_{N-1}[k-\Delta] \end{bmatrix} = \begin{bmatrix} y_0[k-\Delta] - \hat{x}_0[k-\Delta] \\ y_1[k-\Delta] - \hat{x}_1[k-\Delta] \\ \vdots \\ y_{N-1}[k-\Delta] - \hat{x}_{N-1}[k-\Delta] \end{bmatrix}\ . \tag{4.47}$$

In order to construct $\mathbf{r}_{noise,2}[k]$, optimal estimates for the speech components in all microphone signals $\hat{x}_m[k-\Delta]$, $m = 0 \ldots N-1$, need to be computed, leading to an increased computational complexity.

Also for the ANC postprocessing stage of the GSVD-based optimal filtering technique, signal leakage into the noise references will occur, since the estimate of the speech component $\hat{x}_m[k-\Delta]$ is generally not exactly equal to $x_m[k-\Delta]$. Signal leakage can be reduced by using longer filter lengths L for the GSVD-based optimal filter and the effect of the signal leakage on the ANC adaptive filters can be limited by using a speech-controlled (VAD) adaptation algorithm, where the ANC adaptive filters are only allowed to adapt during noise-only periods [46][113][128][194] [254].

In Section 5.2.5 the noise reduction improvement and additional speech distortion of the ANC postprocessing stage will be investigated experimentally for different filter lengths of the GSVD-based optimal filter and the ANC adaptive filter and for the two different noise references $r_{noise,1}[k]$ and $\mathbf{r}_{noise,2}[k]$. It will be shown that the SNR of the enhanced signal improves with increasing filter lengths and increasing number of noise reference signals. It will also be shown that the decrease in noise reduction performance due to short filter lengths L can be fully compensated by adding the ANC postprocessing stage, at a lower overall computational complexity. The ANC postprocessing stage can therefore be used either for increasing the noise reduction performance or for computational complexity reduction without decreasing the performance. The ANC postprocessing stage will however give rise to a slight increase in speech distortion, which can be limited by using longer filter lengths for the GSVD-based optimal filter and for the ANC adaptive filter.

4.4 Conclusion

In this chapter we have discussed several techniques for reducing the overall computational complexity of the GSVD-based optimal filtering technique.

In Section 4.2 several techniques have been discussed for efficiently calculating the GSVD of the speech and the noise data matrix. Instead of recomputing the GSVD from scratch at each time step, the GSVD can be updated using a recursive Jacobi-type updating algorithm. The computational complexity can be further reduced by using a square root-free implementation and by using sub-sampling, where in this context sub-sampling means that the GSVD and the optimal filter are not updated for every sample. The computational complexity has been summarised for realistic parameter values, showing that the complexity can be significantly reduced such that the recursive GSVD-based optimal filtering technique indeed becomes suitable for real-time implementation.

In Section 4.3 it has been shown how to incorporate the GSVD-based optimal filtering technique into a GSC-type structure with an ANC postprocessing stage. The output of the GSVD-based optimal filtering technique is used as a speech reference signal, whereas different possibilities exist for creating a noise reference. In Chapter 5 it will be shown that the ANC postprocessing stage can either be used for increasing the noise reduction performance or for reducing the computational complexity without decreasing the performance. In order to limit the effect of signal leakage in the noise reference on the ANC adaptive filters, these adaptive filters are only adapted during noise-only periods.

Chapter 5

Simulation results and control algorithm

For several simulated acoustic environments and for a real-life recording this chapter discusses the performance of the GSVD-based implementation of the multi-channel optimal filtering technique, in which the low-rank model of the speech signal is implicitly incorporated. The performance of the GSVD-based optimal filtering technique is compared with standard fixed and adaptive beamforming techniques and in addition, robustness issues such as the effect of speech detection errors and deviations from the assumed signal model are analysed.

In Section 5.2 the performance, i.e. unbiased SNR improvement and speech distortion, of the GSVD-based optimal filtering technique with and without ANC postprocessing stage is analysed for several algorithmic parameters (batch vs. recursive version, filter lengths, sub-sampling) and for different reverberation times. In addition, the spatial directivity pattern for simple acoustic scenarios is discussed and it is shown that the GSVD-based optimal filtering technique can also be used for suppressing a spectrally non-stationary noise source.

Section 5.3 analyses the effect of speech detection errors on the performance of the GSVD-based optimal filtering technique. It is shown that speech detection errors mainly have an influence on the speech distortion, and not on the SNR improvement (unless the ANC postprocessing stage is added). This section also evaluates the performance of the GSVD-based optimal filtering technique, combined with several VAD algorithms, for different noise types, showing that the log-energy and the log-likelihood VAD yield the best performance.

In Section 5.4 the performance of the GSVD-based optimal filtering technique is compared with standard fixed and adaptive beamforming techniques for simulated acoustic scenarios and for a real-life recording. The SNR improvement of the GSVD-based optimal filtering technique with ANC postprocessing stage clearly outperforms the SNR improvement of the GSC. This section also compares the robustness for several deviations from the assumed signal model (microphone gain and position, look direction error), showing that the GSVD-based optimal filtering technique is more robust than the GSC.

5.1 Implementation issues

This section first describes the used simulation environment and discusses some implementation issues for the GSVD-based optimal filtering technique and for the fixed and adaptive beamforming techniques.

5.1.1 Simulation environment

The simulation room is depicted in Fig. 5.1 and has dimensions 6 m × 3 m × 2.5 m. It consists of a microphone array, a speech source and 3 noise sources. Unless otherwise indicated, we will only use the noise source at position 1 (only in Section 5.4.1, three simultaneous noise sources at different positions will be used). In our simulations we have used a linear equi-spaced microphone array with $N = 4$ microphones and the distance d between two adjacent microphones is 5 cm. The speech source is located at 1.3 m from the centre of the microphone array at an angle of 56°.

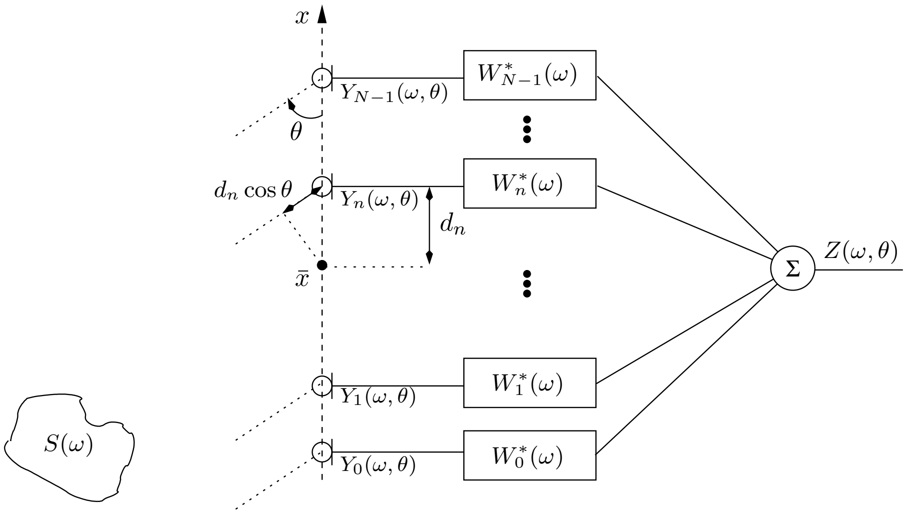


Figure 5.1: Simulation environment

The used signals are a 16 kHz clean speech signal, consisting of English sentences from the 'Hearing in Noise Test' [190], and 3 different noise signals: stationary white noise, stationary speech noise from the NOISEX-92 database [267], having the same long-term spectrum as speech, and a non-stationary music signal. The speech and the noise components received at the nth microphone are filtered versions of the clean speech and noise signals with simulated acoustic impulse responses, constructed using the image method (cf. Section 1.3.3) for different reverberation times T_{60}. We will use (1.2) for computing the reverberation time, where it is assumed that the absorption coefficients α_i are equal for each room surface S_i. By using simulated acoustic impulse responses, we can easily compare the performance for different reverberation conditions.

Since all described algorithms (GSVD-based optimal filter, ANC postprocessing stage, fixed and adaptive beamforming) amount to linear filtering operations, the speech and the noise components of the output signal and all intermediate signals can be easily obtained by applying the computed filters to the speech and the noise components of the microphone signals. The performance of the GSVD-based optimal filtering technique will be described by the unbiased SNR improvement and by the average speech distortion, defined in (2.32) and (2.40).

In our simulations we have constructed the noisy microphone signals such that the unbiased SNR of the first microphone signal $y_0[k]$ equals 0 dB. Figures 5.2a and 5.2b depict the speech component $x_0[k]$ and the noisy microphone signal $y_0[k]$ for reverberation time $T_{60} = 300$ ms when using speech noise. Figure 5.2c shows the enhanced signal $z[k]$ after the recursive GSVD-based optimal filtering technique with ANC postprocessing stage using all noise reference signals[1].

5.1.2 GSVD-based optimal filtering technique

Both for the batch and the recursive implementation of the GSVD-based optimal filtering technique, a VAD algorithm determines when speech is present. Figure 5.2a shows the output of a perfect VAD algorithm on the speech component of the first microphone signal. In practice, the VAD should be tuned such that especially the speech-and-noise periods are correctly classified, since the effect of adding speech vectors to the noise data matrix is more harmful than adding noise vectors to the speech data matrix, cf. Section 5.3.2. In the simulations we will generally assume that a perfect VAD is available. However, in Section 5.3.2 the effect of speech detection errors on the performance of the GSVD-based optimal filtering technique is investigated, and in Section 5.4.2 simulations have been performed for a real-life recording using a (non-perfect) energy-based VAD.

In the *batch GSVD-based optimal filtering technique*, the speech and the noise data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ are constructed from the noisy microphone signals

[1] For this specific simulation, sound files, spectrograms and power transfer functions are available at `http://www.esat.kuleuven.ac.be/~doclo/SA00061/audio.html`

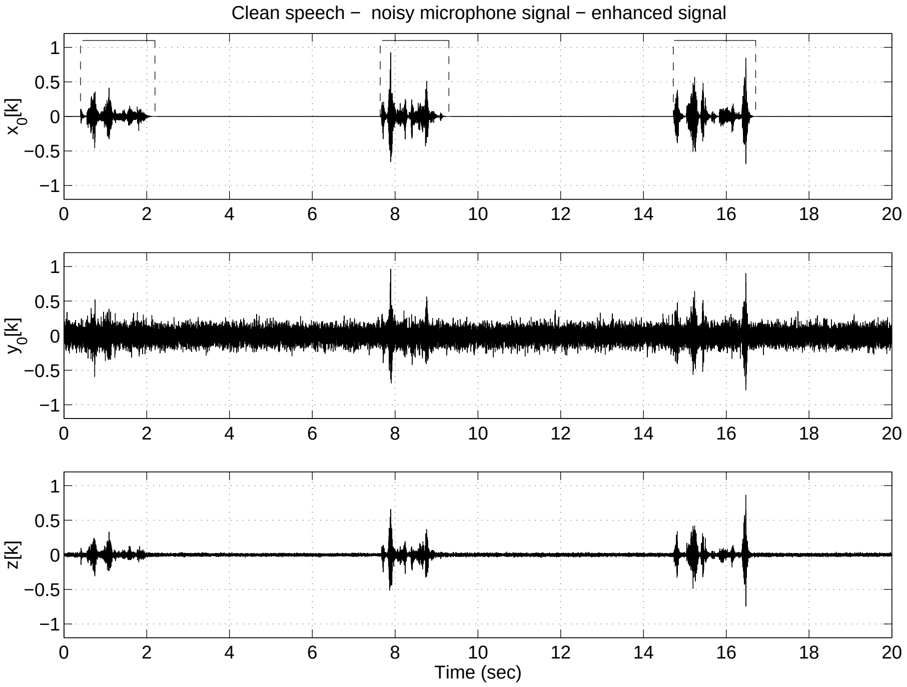


Figure 5.2: (**a**) Speech component $x_0[k]$ and VAD, (**b**) Noisy microphone signal $y_0[k]$ (speech noise, SNR=0 dB, $T_{60} = 300\,\text{ms}$), (**c**) Enhanced signal $z[k]$ using recursive GSVD-based optimal filtering technique with ANC postprocessing stage ($L = 20$, $L_{ANC} = 400$, no sub-sampling, all noise references)

$y_n[k]$, $n = 0 \ldots N-1$, using all available speech and noise samples. The filter length of the optimal filter is denoted by L. The optimal filter matrix $\mathbf{W}_{WF}[k]$ is computed using (3.44), where all negative diagonal elements are put to zero. The stacked filter $\mathbf{w}[k]$ is determined as the ith column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{WF}[k]$, with the fixed value $i = \frac{L}{2}$ (cf. Section 3.3.2). The enhanced signal $z[k]$ is obtained by filtering the microphone signals with $\mathbf{w}_n[k], n = 0 \ldots N-1$.

In the *recursive GSVD-based optimal filtering technique*, the data matrices are updated according to (4.16) or (4.17), with $\lambda_y = 0.99999$ and $\lambda_v = 0.999995$. Using the recursive techniques of Section 4.2, the GSVD and the optimal filter are updated for every sample. The ith column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{WF}[k]$, with $i = \frac{L}{2}$, is computed using (4.33) or (4.40) and the enhanced signal at time k is computed by filtering the microphone signals with $\mathbf{w}_n[k], n = 0 \ldots N-1$. When using sub-sampling, the GSVD and the optimal filter are updated respectively for every s_g and s_f samples. In order to avoid initial effects (initially no knowledge about either the speech nor the noise data matrix is available), signal segments twice as long as for the batch version are processed and only the last half is used for computing the performance measures.

For the *ANC postprocessing stage*, two possible noise references will be investigated: $r_{noise,1}[k]$ in (4.43) with 1 noise reference signal, and $\mathbf{r}_{noise,2}[k]$ in (4.47) with $N = 4$ noise reference signals. The adaptive filter used in the ANC postprocessing stage is a time-domain NLMS algorithm (cf. Section 2.5.3). The filter length of the adaptive filter is denoted by L_{ANC} and the step size is $\mu = 0.05$. The desired signal of the adaptive filter is delayed by $\frac{L_{ANC}}{2}$ samples in order for the adaptive filter to model some acausal taps. As already mentioned in Section 4.3, in order to limit signal cancellation and distortion, a speech-controlled adaptation algorithm will be used, where the ANC adaptive filters are only allowed to adapt during noise-only periods.

5.1.3 Fixed and adaptive beamforming techniques

In Section 5.4, the performance of the GSVD-based optimal filtering technique is compared with fixed and adaptive beamforming techniques. The fixed *delay-and-sum (DS) beamformer*, discussed in Section 2.5.2, spatially aligns the microphone signals to the direction of the speech source. The delays δ_n in (2.117) are computed as $\delta_n = -\frac{nd\cos\theta_x}{c} f_s$, with $\theta_x = 56°$. The standard *Generalised Sidelobe Canceller (GSC)*, discussed in Section 2.5.3 and depicted in Fig. 2.10, uses the output signal of a DS beamformer as speech reference signal, and creates the noise reference $\mathbf{r}^{GSC}_{noise}[k]$ in (2.149) using a Griffiths-Jim blocking matrix. When using 1 noise reference signal, only the first element of $\mathbf{r}^{GSC}_{noise}[k]$ is considered. The used adaptive filter is a time-domain NLMS algorithm, with filter length denoted by L_{ANC} and step size $\mu = 0.1$. The speech reference signal is delayed by $\frac{L_{ANC}}{2}$ samples in order for the adaptive filter to model some acausal taps. In order to limit the effect of the signal leakage on the adaptive filters, a speech-controlled adaptation algorithm will be used, where the ANC adaptive filters are only allowed to adapt during noise-only periods.

As has already been mentioned in Section 2.5.3, it is possible to reduce the amount of signal leakage in the noise reference by using a spatial filter designed blocking matrix [191][194] instead of the Griffiths-Jim blocking matrix. We have designed a spatial filter for the blocking matrix using the non-linear design criterion for far-field broadband beamformers, discussed in Section 8.3.4, with stopband specifications $(\Omega_s, \Theta_s) = \big(0{-}7500\,\text{Hz}, (\theta_x{-}20°){-}(\theta_x{+}20°)\big)$, and passband specifications $(\Omega_p, \Theta_p) = (0{-}7500\,\text{Hz}, 0°{-}(\theta_x{-}20°)$ and $(\theta_x{+}20°){-}180°)$. We have designed this spatial filter with $L = 20$ taps per microphone and using $N{-}1$ microphones, such that we are able to create 2 independent noise reference signals [191]. The fixed beamformer for creating the speech reference signal has inverse stopband and passband specifications and is designed to be orthogonal to the blocking matrix, which can be achieved by imposing linear constraints in the design procedure [191]. The spatial directivity pattern of the fixed beamformer and the spatial filter designed blocking matrix are depicted in Fig. 5.3.

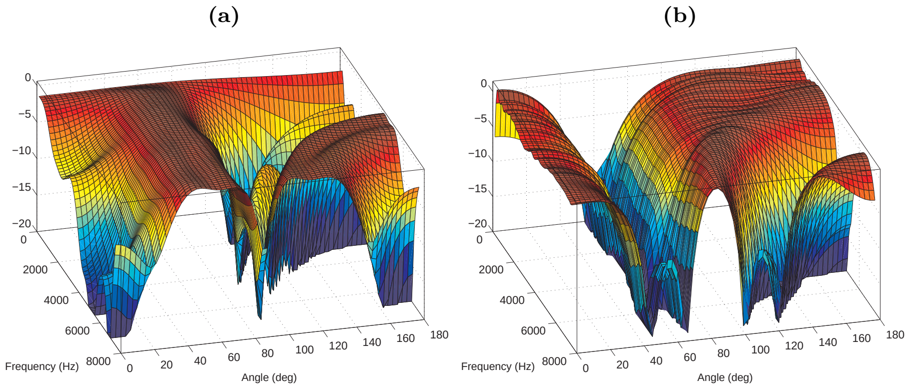


Figure 5.3: Spatial directivity pattern of (**a**) fixed beamformer (speech reference) and (**b**) blocking matrix (noise reference)

Although the amount of signal leakage into the noise reference will be reduced, it can never be completely avoided (certainly not in highly reverberant acoustic environments). Therefore, we will still use a speech-controlled adaptation algorithm, switching off the adaptation during speech-and-noise periods.

The speech distortion measure, defined in (2.40), is not really useful for fixed and adaptive beamformers, since this speech distortion measure considers the PTF of the speech component in the first microphone signal $x_0[k]$, whereas the DS beamformer and the GSC try to recover the speech signal $s[k]$.

5.2 Performance of optimal filtering technique

This section discusses the performance (SNR improvement and speech distortion) of the GSVD-based optimal filtering technique with and without ANC postprocessing stage. In Section 5.2.1 the spatial directivity pattern for simple acoustic scenarios is discussed, showing that the GSVD-based optimal filtering technique then exhibits the desired beamforming behaviour. In Section 5.2.2 the performance of the batch and the recursive version is compared for different filter lengths L and reverberation times T_{60}, showing that these two versions nearly have the same performance. In Section 5.2.3 the effect of several parameters in the recursive GSVD-updating algorithms is analysed, showing that a smaller number of GSVD-steps (or sweeps) and sub-sampling can be used for stationary acoustic environments. In Section 5.2.4 it is shown that the GSVD-based optimal filtering technique can also be used for suppressing a spectrally non-stationary noise source. In Section 5.2.5 the effect of the ANC postprocessing stage is analysed, showing that the ANC postprocessing stage gives rise to an SNR improvement and a small increase in speech distortion.

5.2.1 Spatial directivity pattern

When considering no multi-path propagation ($T_{60} = 0$), it can be shown that the GSVD-based optimal filtering technique exhibits the desired beamforming behaviour for simple acoustic scenarios.

First, consider spatio-temporally white noise (e.g. sensor noise), i.e. the noise component $v_n[k]$ in the nth microphone signal is temporally white and is uncorrelated with the noise components in all other microphone signals. The speech source impinges on the microphone array at an angle $\theta_x = 45°$. Figure 5.4a shows the amplitude $|H(\omega, \theta)|$ of the spatial directivity pattern, defined in (2.108), for the frequencies $\omega_i = 2\pi \cdot 200\, i,\ i = 1 \ldots 40$. For most frequencies, $|H(\omega, \theta)|$ attains its maximum for $\theta_x = 45°$, implying that the GSVD-based optimal filtering technique automatically finds the direction of the speech source. However for low frequencies the spatial selectivity is rather poor.

Secondly, consider two localised white noise sources that impinge on the microphone array at angles $\theta_{v1} = 60°$ and $\theta_{v2} = 150°$. The speech source is located at broadside, i.e. $\theta_x = 90°$. Fig. 5.4b shows the spatial directivity pattern $|H(\omega, \theta)|$ for the frequencies $\omega_i = 2\pi \cdot 200\, i,\ i = 1 \ldots 40$. As can be seen from this figure, $|H(\omega, \theta)|$ is approximately equal to zero for $\theta = 60°$ and $\theta = 150°$, i.e. the directions of the two noise sources. Although difficult to see on this figure, $|H(\omega, \theta)|$ in the direction of the speech source ($\theta = 90°$) is not equal to 1, as is the case for the GSC, but depends on the spectral content of the speech and the noise components, in accordance with (3.83).

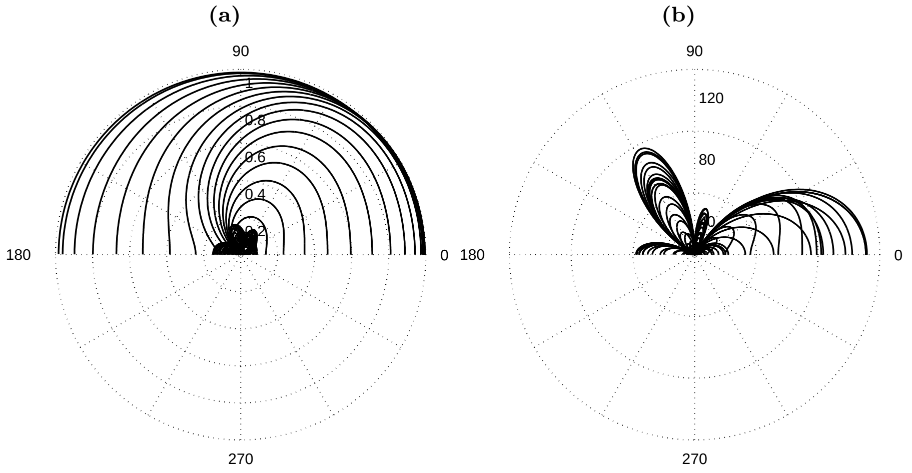


Figure 5.4: Spatial directivity pattern $|H(\omega, \theta)|$ for (**a**) spatio-temporally white noise and speech source at $\theta_x = 45°$ ($L = 10$) and (**b**) localised white noise sources at $\theta_{v1} = 60°$ and $\theta_{v2} = 150°$ and speech source at $\theta_x = 90°$ ($L = 20$)

We can conclude that the GSVD-based optimal filtering technique has the desired beamforming behaviour for both simple scenarios. For more realistic reverberant environments, it is rather difficult to interpret these spatial directivity patterns, since the GSVD-based optimal filtering technique computes an estimate for the speech component in one microphone signal, thereby reducing the additive noise but not the reverberation of the speech signal.

5.2.2 Batch vs. recursive processing

Figure 5.5 compares the unbiased SNR and the speech distortion (SD) of the enhanced signal $z[k]$ for the batch and for the recursive version of the GSVD-based optimal filtering technique (without ANC postprocessing stage). The noisy microphone signals have been constructed using a speech noise source at position 1, and the simulations have been performed for different filter lengths L and for different reverberation times T_{60}. Low reverberation corresponds to highly correlated signals, whereas high reverberation corresponds to highly uncorrelated (diffuse) signals. No sub-sampling has been used in the recursive version of the GSVD-based optimal filtering technique.

As can be seen from Fig. 5.5a and Fig. 5.5b, the unbiased SNR increases and the speech distortion decreases for higher filter lengths L and for lower reverberation times T_{60}. This can be explained from the fact that in highly reverberant acoustic environments the GSVD-based optimal filtering technique will trade off noise reduction and cancellation of the reverberant part of the speech signal, in order to make an optimal estimate of the speech component $x_0[k]$. As can also be seen from these figures, the unbiased SNR and the speech distortion of the batch and the recursive version are practically equal for all reverberation times and filter lengths.

5.2.3 Recursive GSVD-updating algorithms

As has been discussed in Sections 4.2.2 and 4.2.3, different implementations of the recursive GSVD-updating algorithm exist: a 'conventional' implementation and an approximate square root-free implementation, both with the possibility to perform s sweeps and r GSVD-steps. Figure 5.6 shows the unbiased SNR of the enhanced signal $z[k]$ for different implementations of the recursive GSVD-updating algorithms and for a different number of sweeps and GSVD-steps. The noisy microphone signals have been constructed using a speech noise source at position 1 and for reverberation time $T_{60} = 300\,\text{ms}$. The simulations have been performed with $L = 20$, without sub-sampling and without ANC postprocessing stage. Figure 5.6 shows that there is practically no difference in unbiased SNR between the 'conventional' and the square root-free implementation. When performing more than one sweep, the SNR only marginally improves. When performing less than $M - 1$ GSVD-steps, the SNR gradually decreases. However, when using a small number of GSVD-steps (or sweeps), the optimal filter will adapt slower in non-stationary acoustic environments.

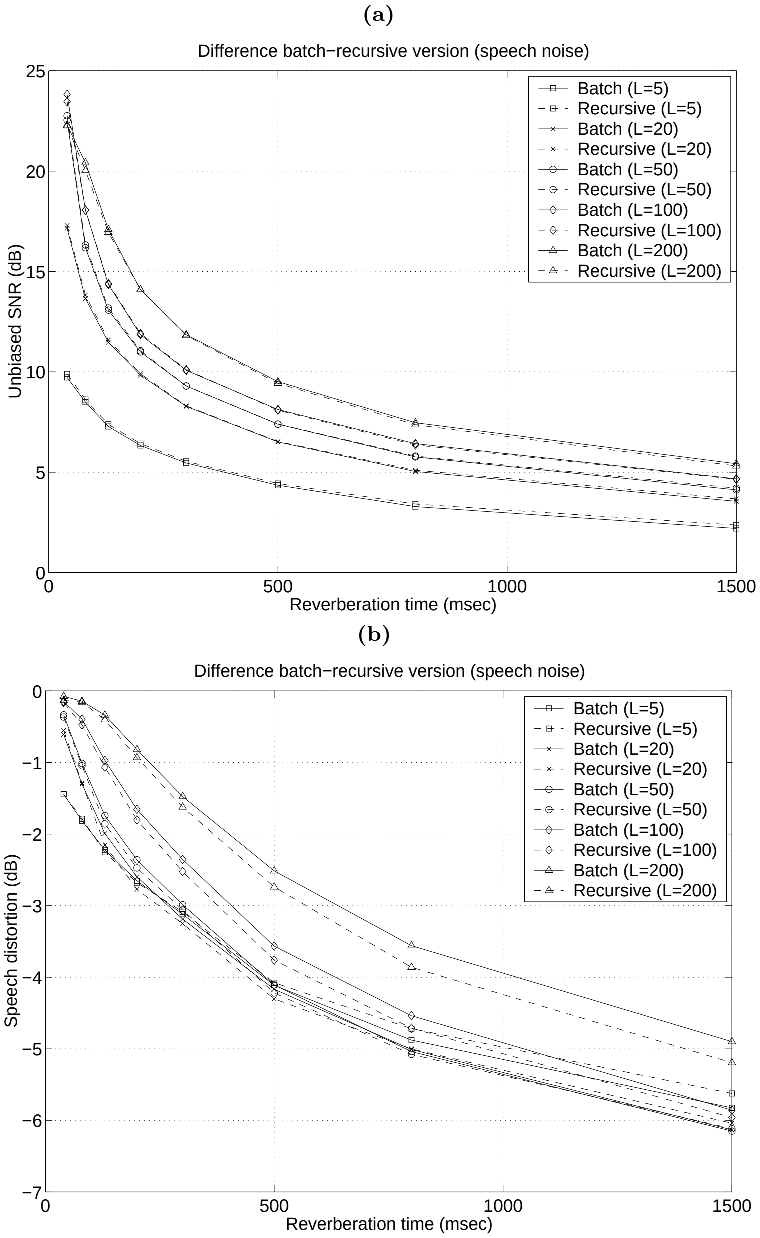


Figure 5.5: Comparison of (**a**) unbiased SNR and (**b**) speech distortion (SD) for batch and recursive version of GSVD-based filtering technique for different filter lengths L and reverberation times T_{60} (speech noise, no sub-sampling)

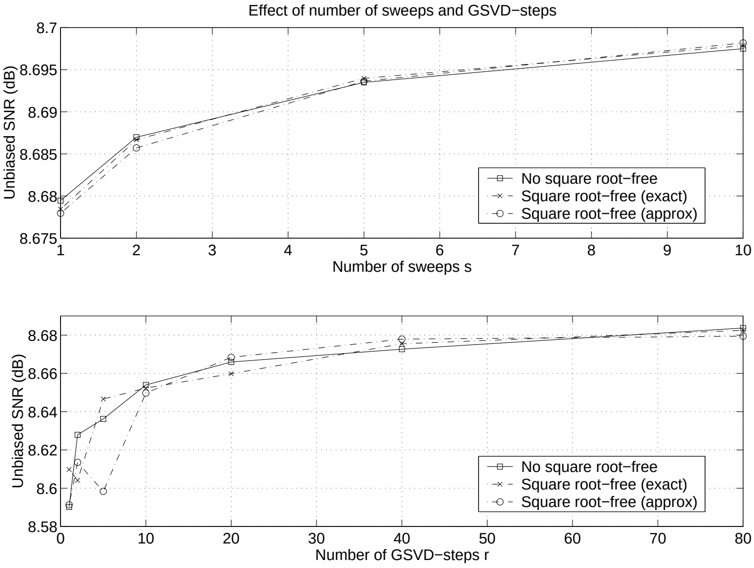


Figure 5.6: Effect of number of sweeps, GSVD-steps and square root-free implementation on unbiased SNR for recursive GSVD-based optimal filtering technique (speech noise, $T_{60} = 300\,\text{ms}$, $L = 20$, no sub-sampling)

In Section 4.2.4 the use of sub-sampling has been discussed for reducing the computational complexity. Figure 5.7 shows the energy of the residual noise $z_v^2[k]$ in the enhanced signal $z[k]$ for different values of the sub-sampling factors $s_g = s_f$. The noisy microphone signals have been constructed using a stationary speech noise source. Figure 5.7 shows that a higher sub-sampling factor results in a slower convergence towards the converged optimal filter (corresponding to highest unbiased SNR), implying that the amount of sub-sampling has to be limited in non-stationary acoustic environments.

5.2.4 Spectrally non-stationary noise source

In this section simulations are performed using a spectrally non-stationary noise source, i.e. a noise source at a fixed position but with a changing spectrum. Since we are considering quite long data blocks in the GSVD-based optimal filtering technique, i.e. λ_v close to 1 or large Q, the noise reduction performance is mainly dependent on the average, i.e. long-term, spectral (and spatial) characteristics of the noise source, cf. (3.83). This implies that the GSVD-based optimal filtering technique can also be used for suppressing non-stationary noise sources.

In the simulations, the used non-stationary noise source has been created by filtering a white noise source with a time-varying FIR-filter, represented by

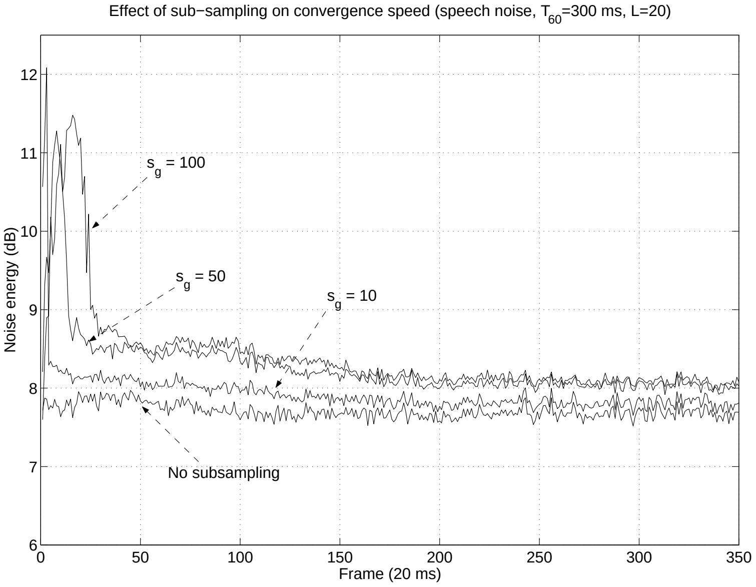


Figure 5.7: Effect of sub-sampling factors on convergence speed for recursive GSVD-based optimal filtering technique (speech noise, $T_{60} = 300\,\text{ms}$, $L = 20$)

the 10-dimensional vector $\mathbf{f}[k]$. The filter $\mathbf{f}[k]$ varies between a low-pass filter $\mathbf{f}_L$ (with cut-off frequency 2400 Hz) and a high-pass filter $\mathbf{f}_H$ (with cut-off frequency 1600 Hz) at different rates, i.e.

$$\mathbf{f}[k] = \nu[k]\,\mathbf{f}_H + (1 - \nu[k])\,\mathbf{f}_L\ , \tag{5.1}$$

with $0 \leq \nu[k] \leq 1$ a time-varying parameter, determining how fast the filter $\mathbf{f}[k]$ varies in time. The frequency response of $\mathbf{f}_L$, $\mathbf{f}_H$ and of a number of intermediate filters $\mathbf{f}[k]$ is plotted in Fig. 5.8a. The non-stationary noise source is filtered with the simulated acoustic impulse responses between the position of the noise source and the microphone array. The reverberation time is $T_{60} = 300\,\text{ms}$. A non-stationarity factor indicates how many times the filter $\mathbf{f}[k]$ varies between the low-pass and the high-pass filter (and back) over the total signal. Figure 5.8b compares the unbiased SNR of the enhanced signal $z[k]$ of the batch GSVD-based optimal filtering technique for different filter lengths L at different levels of non-stationarity. As can be seen from this figure, the unbiased SNR is practically independent of the non-stationarity factor. Therefore we can conclude that the GSVD-based optimal filtering technique is only dependent on the average spectral (and spatial) characteristics of the noise source.

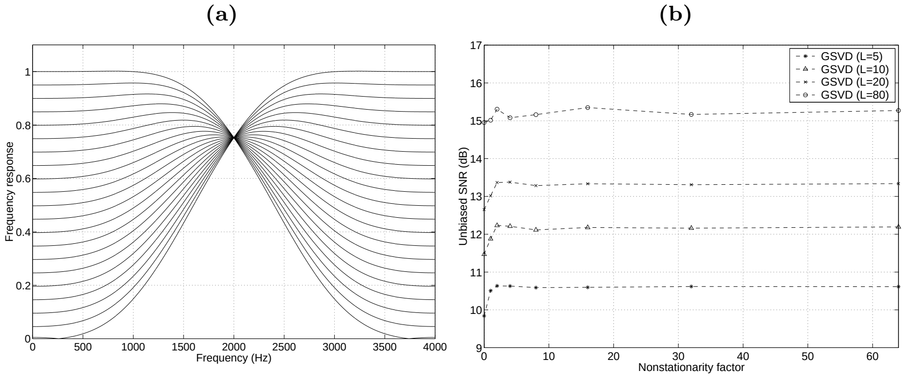


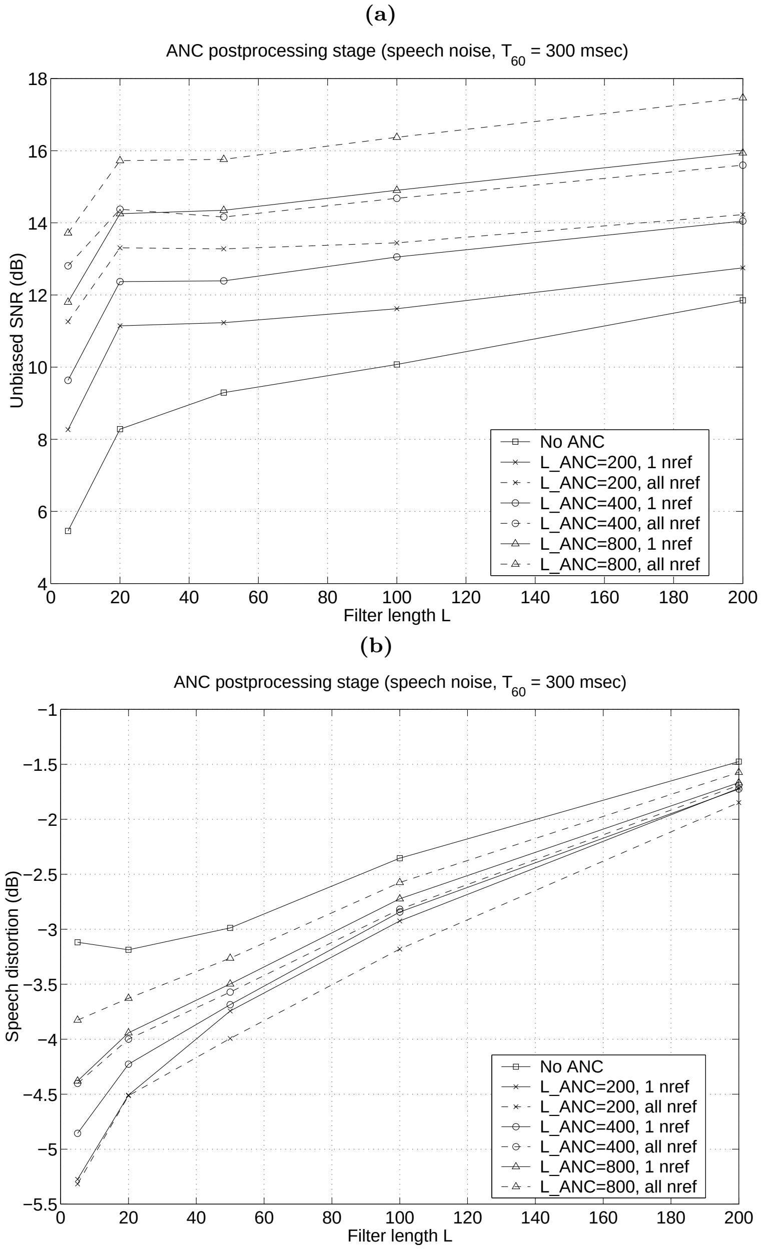
Figure 5.8: (**a**) Frequency response of time-varying filter $\mathbf{f}[k]$, (**b**) Comparison of unbiased SNR for non-stationary noise source ($T_{60} = 300\,\text{ms}$, batch version)

5.2.5 Effect of ANC postprocessing stage

In this section, the effect of the ANC postprocessing stage, discussed in Section 4.3, is analysed. It will be shown that the ANC postprocessing stage can either be used for increasing the noise reduction performance or for computational complexity reduction without decreasing the performance. The ANC postprocessing stage however also gives rise to a slight increase in speech distortion, which can be limited by using longer filter lengths.

Figure 5.9 investigates the effect of the ANC postprocessing stage on the noise reduction performance and on the speech distortion for different filter lengths L and L_{ANC} and for a different number of noise reference signals. The noisy microphone signals have been constructed using a speech noise source at position 1 and for reverberation time $T_{60} = 300\,\text{ms}$, and simulations have been performed with the batch version of the GSVD-based optimal filtering technique.

Figure 5.9a shows that the unbiased SNR of the enhanced signal improves with increasing filter lengths L and L_{ANC} and with increasing number of noise reference signals. In addition, this figure shows that the same noise reduction performance can be obtained either with large filter lengths L without ANC postprocessing stage or with short filter lengths L and using the ANC postprocessing stage. Since the total computational complexity is $\mathcal{O}(L^2) + \mathcal{O}(L_{ANC})$, using short filter lengths L with ANC postprocessing stage gives rise to a lower computational complexity. The ANC postprocessing stage can therefore be used either for increasing the noise reduction performance or for computational complexity reduction without decreasing the performance. Figure 5.9b shows that the ANC postprocessing stage however also gives rise to a small increase in speech distortion. However, speech distortion can be limited by using longer filter lengths L (since signal leakage into the noise reference is then reduced) and longer filter lengths L_{ANC}.


(a)

(b)

Figure 5.9: Effect of ANC postprocessing stage on unbiased SNR and speech distortion for different filter lengths and for different number of noise references (speech noise, $T_{60} = 300\,\mathrm{ms}$, batch version)

5.3 Control algorithm: VAD

Many single- and multi-microphone speech enhancement techniques require a voice activity detection (VAD) algorithm to classify the incoming samples into speech-and-noise samples and noise-only samples. The main reason is that these techniques require an estimate of the spectral and/or spatial characteristics of the noise components, which can be estimated during speech inactivity and which are assumed to remain valid during subsequent speech-and-noise periods. E.g. the GSVD-based optimal filtering technique requires a VAD in order to estimate the spatio-temporal noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$, i.e. the VAD algorithm determines whether a data vector belongs to the speech or the noise data matrix. Other techniques, such as the GSC and the ANC postprocessing stage, typically use a speech-controlled adaptation algorithm, where the ANC adaptive filters are only allowed to adapt during noise-only periods. The performance of most speech enhancement algorithms is strongly influenced by the correct classification between speech-and-noise and noise-only periods, hence the need for a robust VAD algorithm. Since in fact the GSVD-based optimal filtering technique uses no other a priori information than the output of a VAD algorithm, it is expected to be quite sensitive to speech detection errors (cf. Section 5.4.3). The design of a robust VAD is particularly difficult when the SNR is low ($< 0\,\text{dB}$) and when the background noise is highly non-stationary (e.g. music, cocktail party). However, these are the normal operational conditions for hands-free mobile telephony and voice-controlled speech applications.

In Section 5.3.1 an overview is given of several (single-microphone) VAD algorithms, whose performance has been tested for different noise types and signal-to-noise ratios in [50]. In this thesis we will not consider multi-microphone VAD algorithms, although it is expected that the performance and the robustness of the VAD algorithms will increase when using more than 1 microphone signal. In Section 5.3.2 the average effect of (manually introduced) VAD-errors on the performance of the GSVD-based optimal filtering technique is analysed, both theoretically and experimentally. It will be shown that speech detection errors mainly have an influence on the speech distortion, and not on the SNR improvement (unless the ANC postprocessing stage is added) and that speech detection errors in the beginning of a speech segment have a larger effect than at the end of a speech segment. In Section 5.3.3 we evaluate the performance of the GSVD-based optimal filtering technique, combined with the VAD algorithms from Section 5.3.1, for different noise types, showing that the log-energy and the log-likelihood VAD algorithms yield the best performance.

5.3.1 VAD algorithms

The following (existing) single-microphone VAD algorithms have been analysed in combination with the GSVD-based optimal filtering technique :

1. *Log-likelihood based method* [72][235][236] : this VAD algorithm is based on the log-likelihood ratio, which is defined as

$$\log \Lambda = \frac{1}{L} \sum_{l=0}^{L-1} \log \Lambda_l \, , \quad \Lambda_l = \frac{1}{1+\xi_l} \exp \left\{ \frac{\gamma_l \xi_l}{1+\xi_l} \right\} \, , \tag{5.2}$$

   with Λ_l the likelihood ratio and ξ_l and γ_l the a priori and the a posteriori SNR in frequency band l ($l = 0 \ldots L-1$). In order to estimate these statistics, a noise spectrum adaptation algorithm has been developed.

2. *Log-energy based methods* [46][142][255][256][260] : Energy-based methods assume that the short-time energy of a speech-and-noise segment is higher than the short-time energy of a noise-only segment. By continuously monitoring the signal energy on a frame-by-frame basis, the start and the endpoint of speech can be found when the short-time energy is higher than a certain (fixed or adaptive) threshold value. For this technique to work in highly non-stationary noise, e.g. other speech signals, the energy of the desired speaker must however be sufficiently larger than the energy of the undesired speakers.

3. *Zero Crossing Rate-based methods* [42][143][226] : these methods are based on the computation of the zero-crossing rate, i.e. the average number of sign changes in the noisy speech signal $y_0[k]$, i.e.

$$\zeta_c[k] = \sum_{l=0}^{L-1} |\text{sign}(\text{sign}(y_0[k-l]) - \text{sign}(y_0[k-l-1]))| \, .$$

   The zero crossing rate for noise is assumed to be considerably higher than for speech. This assumption is however only accurate at high SNR. At low SNR problems occur especially in the presence of periodic background noise and speech with a high zero crossing rate (e.g. voiced speech).

4. *Short-time amplitude based method* : in [124] a double-talk detection algorithm has been presented that computes short-time and long-time estimates of the background noise and the far-end signal. This algorithm can also be modified into a VAD algorithm [50].

5. *Sample-based VAD method* [226] : this rule-based algorithm continuously generates speech and noise metrics from the noisy input signal, based on statistical assumptions about the characteristics of the speech and the noise signal. The algorithm is sample-based, hence the decision delay is small. However, as a consequence the computational complexity is higher than for the other VAD algorithms, which are typically frame-based.

6. *Spectral entropy based method* [131][229] : Spectral entropy-based methods first estimate the probability density function (pdf) of the spectrum for each frame of the signal. Using this pdf the spectral entropy measure

can be computed [229]. The (weighted) spectral entropy is assumed to be higher for speech than for background noise (except if the background noise is another speech-like signal). In [131] the spectral entropy and the log-energy features have been combined to form a new feature that possesses the advantages of both VAD algorithms.

7. *Geometric VAD algorithm* [204][250] : this method builds on the differences between the probability distribution properties of the amplitudes of the speech and the noise components. A so-called modified amplitude pdf (MAPD) is defined and it is observed that the speech and the noise samples can be partially separated on this MAPD plot.

In [50] the performance of all presented VAD algorithms has been analysed for different noise types (stationary white noise and speech noise; non-stationary car noise, babble noise and traffic noise) at different signal-to-noise ratios (from $-5\,\text{dB}$ to $15\,\text{dB}$). Different performance measures have been used: onset error, offset error, global error, noise detected as speech, speech detected as noise. In general, the log-likelihood and the log-energy VAD algorithms provide the best performance, whereas the performance of the zero crossing rate, short-time amplitude and sample-based VAD methods simply is unsatisfactory.

5.3.2 Effect of VAD-errors on performance

Before analysing the performance of the GSVD-based optimal filtering technique in combination with the VAD algorithms from the previous section, we will first theoretically analyse its performance when manually introducing a certain amount of speech detection errors.

Theoretical analysis

In [239] the average effect of speech detection errors on the performance of the GSC and the multi-channel Wiener filter has been theoretically analysed. This can be done by modifying the coherence matrices $\mathbf{\Gamma}_x(\omega)$ and $\mathbf{\Gamma}_v(\omega)$ in the formulas for the GSC and the multi-channel Wiener filter, cf. (2.155) and (3.85).

If the blocking matrix $\mathbf{C}_a$ of the GSC would be perfect such that the noise references do not contain any speech components, then the performance of the GSC would be independent of the VAD. However, in practice signal leakage into the noise references occurs. Since the ANC adaptive filters are only adapted during noise-only periods, the performance of the GSC – apart from a slower convergence – will not degrade when noise is wrongly detected as speech. On the contrary, when speech is wrongly detected as noise, the ANC may additionally cancel the speech signal. The impact of speech detection errors on the performance of the GSC therefore depends on the quality of the noise references: the larger the ratio of speech leakage to noise in the noise references, the larger the impact of speech detection errors. Simulations show that the performance of the GSC is strongly affected by speech detection errors.

Since the multi-channel Wiener filter does not require any other a priori information, its reliance on the VAD algorithm would be expected to be crucial. However, it can be shown that the *unbiased SNR improvement of the multichannel Wiener filter is not degraded* by speech detection errors, neither when speech is wrongly detected as noise nor when noise is wrongly detected as speech. When speech is wrongly detected as noise, *the speech distortion* however *increases* with $20\log_{10}(1-\delta)\,\text{dB}$, with δ the percentage of the 'noise-only' samples that contain speech components. This additional speech distortion is independent of the noise scenario and the input SNR. For error rates $\delta < 0.2$, speech distortion however remains limited. When noise is wrongly detected as speech, the speech distortion only slightly increases with increasing error rate δ. The multi-channel Wiener filter with ANC postprocessing stage is more strongly affected by speech detection errors, especially when $\delta > 0.2$. The SNR improvement then decreases and the speech distortion increases with increasing error rate δ and increasing input SNR. However, simulations have shown that the multi-channel Wiener filter with or without ANC postprocessing stage preserves its potential benefit over the GSC for a reasonable speech detection error rate of 0.2 or less, even when the GSC is supplied with a noise sensitivity constraint [239].

Experimental validation

In [50] the effect of manually introducing speech detection errors on the performance of the GSVD-based optimal filtering technique has been experimentally verified. The noisy microphone signals have been constructed using a white noise source at position 1 and for reverberation time $T_{60} = 200\,\text{ms}$. We have investigated the performance of the batch and the recursive version of the GSVD-based optimal filtering technique, both with and without ANC postprocessing stage. The simulations have been performed with $L = 20$, and for the recursive version the subsampling factors are $s_f = s_g = 5$ and different weighting factors $\{\lambda_x, \lambda_v\} = \big[\{1, 1\}, \{0.99999, 0.999995\}, \{0.9999, 0.99995\}, \{0.999, 0.9995\}\big]$ have been used. We will only consider speech detection errors where speech is wrongly classified as noise (since the effect of these errors is the largest). Speech detection error rates ranging from $\delta = 0.1$ to $\delta = 0.5$ have been applied at the beginning and at the end of the speech segments. For these experiments, speech distortion has been measured by the distance measure D,

$$D = 10\log_{10}\frac{\sum x_0^2[k]}{\sum(z_x[k] - x_0[k])^2}\ , \tag{5.3}$$

which is computed during speech-and-noise periods. A large distance measure D corresponds to little speech distortion.

Figure 5.10 shows the unbiased SNR improvement and the speech distance measure for all considered error rates δ and for all algorithms. From this figure, it can indeed be seen that for the batch GSVD-based optimal filtering technique (without ANC postprocessing stage), the SNR improvement is practically unaffected by the speech detection errors and the speech distortion only substantially increases when $\delta > 0.2$. For the recursive version, essentially the same conclusions hold, but speech detection errors will have a larger effect for smaller weighting factors, since the effective speech and noise data windows in that case are smaller. This is especially noticeable for $\{\lambda_x, \lambda_v\} = \{0.999, 0.9995\}$.

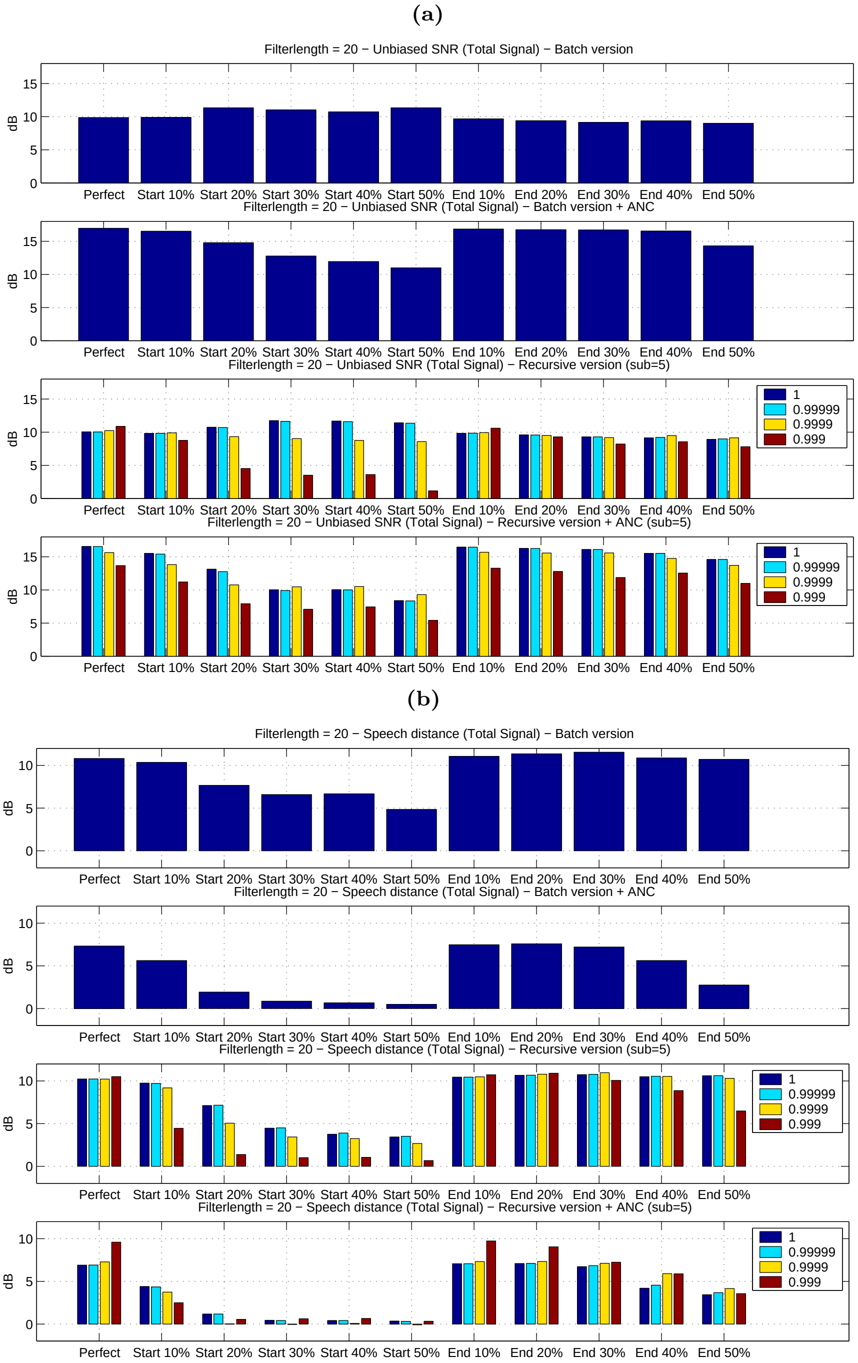
From this figure it can also be seen that the GSVD-based optimal filtering technique (both the batch and the recursive version) with ANC postprocessing stage is more strongly affected by speech detection errors, especially when $\delta > 0.2$. The SNR improvement decreases and the speech distortion increases with increasing error rate δ. For all considered algorithms, speech detection errors at the beginning of a speech segment seem to have a larger effect on the SNR improvement and the speech distortion than speech detection errors at the end of a speech segment. This is probably due to the higher energy portions at the beginning of the speech segments. However, when only considering the misclassified parts of the speech segments, also speech detection errors at the end of a speech segment clearly have a (negative) influence on the SNR improvement and the speech distortion in these misclassified parts.

5.3.3 Combination of GSVD-based optimal filtering technique and VAD algorithms

In this section, we evaluate the performance (unbiased SNR improvement and speech distance measure) of the GSVD-based optimal filtering technique in combination with the different VAD algorithms discussed in Section 5.3.1. The output of the VAD on the noisy first microphone signal $y_0[k]$ is used. The performance is evaluated for several noise types (white noise, speech noise, car noise, babble noise and traffic noise), such that a mean performance can be calculated. The simulations have been performed with $L = 20$, and for the recursive version the subsampling factors are $s_f = s_g = 10$ and the weighting factors are $\{\lambda_x, \lambda_v\} = \{0.9999, 0.99995\}$.

In general, the best SNR improvement and speech distortion is obtained for the stationary white noise and for the (low-frequency) car noise, whereas the (stationary) speech noise proves to be more difficult, since it has the same long-term spectrum as the desired speech signal, and the non-stationary babble noise and traffic noise are the most difficult noise types.

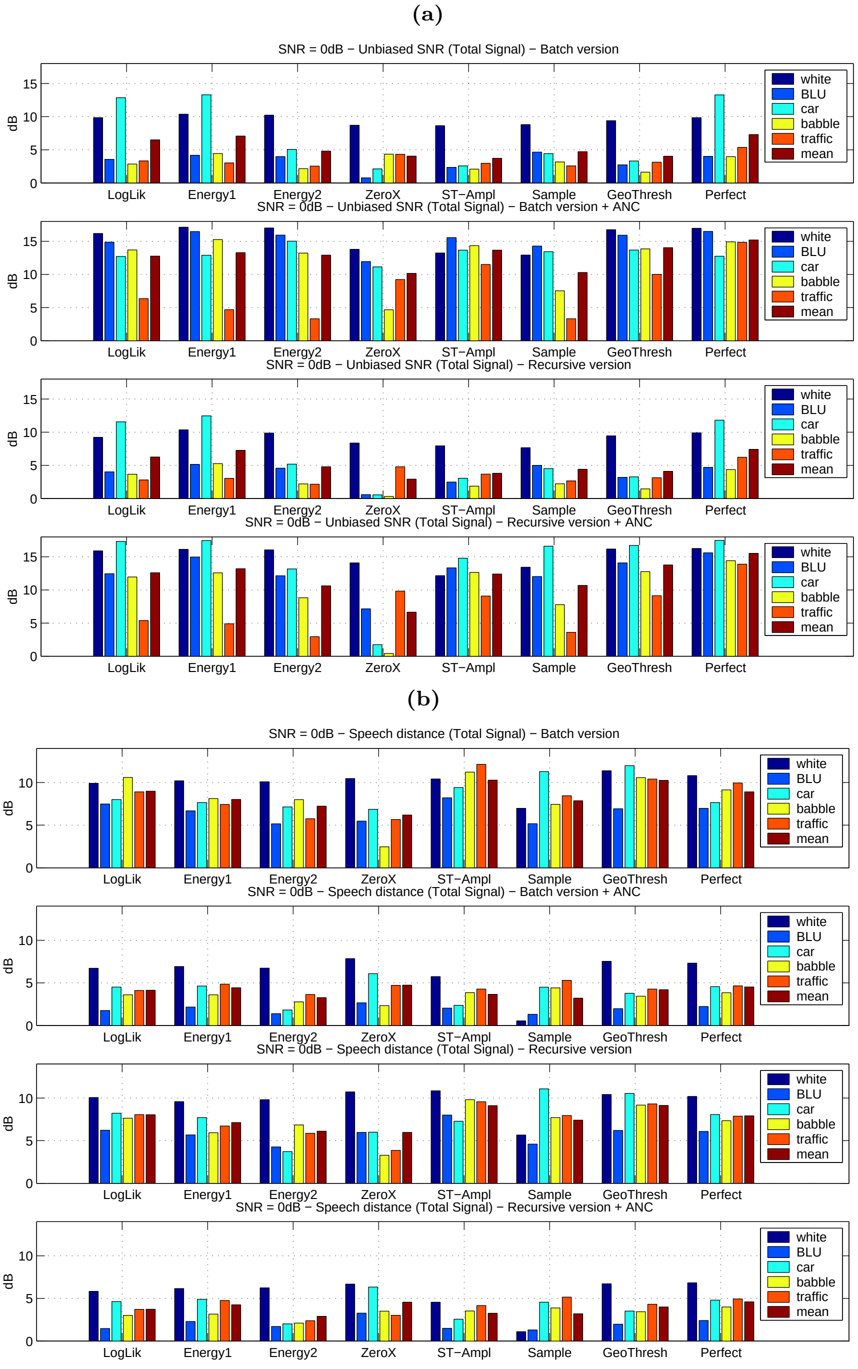
Figure 5.11 shows the unbiased SNR improvement and the speech distance measure for all considered noise types and for all algorithms. If we analyse the unbiased SNR improvement for the recursive version, then the log-energy, the log-likelihood and the geometric VAD provide the best performance, whereas the zero-crossing VAD has the worst performance. If we analyse the speech distortion for the recursive version, then the zero-crossing and the geometric VAD provide the least speech distortion, whereas the log-likelihood and the log-energy VAD perform quite reasonably (in comparison with the perfect VAD).



(a)

(b)

Figure 5.10: Effect of speech detection errors on (**a**) unbiased SNR improvement and (**b**) speech distance measure D



(a)

SNR = 0dB − Unbiased SNR (Total Signal) − Batch version

SNR = 0dB − Unbiased SNR (Total Signal) − Batch version + ANC

SNR = 0dB − Unbiased SNR (Total Signal) − Recursive version

SNR = 0dB − Unbiased SNR (Total Signal) − Recursive version + ANC

(b)

SNR = 0dB − Speech distance (Total Signal) − Batch version

SNR = 0dB − Speech distance (Total Signal) − Batch version + ANC

SNR = 0dB − Speech distance (Total Signal) − Recursive version

SNR = 0dB − Speech distance (Total Signal) − Recursive version + ANC

Figure 5.11: (**a**) Unbiased SNR improvement and (**b**) speech distance measure D for different VAD algorithms and different noise types

VAD-algorithm	*Unbiased SNR*	*Speech distortion*
Log-likelihood	Good	Good
Log-energy (1/2)	Good/Moderate	Good/Moderate
Zero-Crossing	Bad	Good
Short-Time Amplitude	Moderate	Good
Sample-based	Moderate	Moderate
Geometric	Good	Good

Table 5.1: Classification of VAD algorithms based on unbiased SNR improvement and speech distortion (different noise types)

VAD-algorithm	*Unbiased SNR*	*Speech distortion*
Log-likelihood	Good	Good
Log-energy (1/2)	Good/Moderate	Good/Moderate
Zero-Crossing	Bad	Bad
Short-Time Amplitude	Good	Good
Sample-based	Bad	Moderate
Geometric	Bad	Good

Table 5.2: Classification of VAD algorithms based on unbiased SNR improvement and speech distortion during misclassified part of speech segments

These results are summarised in Table 5.1. When analysing the performance only in the misclassified parts of the speech segments, the best unbiased SNR performance is produced by the log-likelihood, the log-energy and the short-time amplitude VAD, whereas the worst performance is produced by the zero-crossing, the sample-based and the geometric VAD. However, the geometric VAD produces not much speech distortion, as do the log-likelihood, the log-energy and the short-time amplitude VAD. These results are summarised in Table 5.2. In conclusion, the best trade-off between unbiased SNR improvement and speech distortion for the different noise types is provided by the log-likelihood and the log-energy VAD algorithms.

5.4 Performance comparison with beamforming techniques

In this section the performance of the GSVD-based optimal filtering technique is compared with standard fixed and adaptive beamforming techniques. Section 5.4.1 compares the performance for simulated acoustic scenarios (assuming a perfect VAD), whereas Section 5.4.2 compares the performance for a real-life recording (using a non-perfect VAD). The SNR improvement of the GSVD-based optimal filtering technique with ANC postprocessing stage clearly outperforms the SNR improvement of the GSC for all reverberation times and all considered acoustic scenarios. Section 5.4.3 compares the robustness for several deviations from the assumed signal model (microphone gain and position, look direction error), showing that the GSVD-based optimal filtering technique is more robust than the GSC.

5.4.1 Simulated acoustic scenarios

In this section, the noise reduction performance and the speech distortion of the GSVD-based optimal filtering technique with and without ANC postprocessing stage is compared with standard beamforming techniques for three simulated acoustic scenarios: white noise source at position 1 (Fig. 5.12), speech noise source at position 1 (Fig. 5.13) and 3 simultaneous noise sources (white+speech+music) at the 3 noise positions (Fig. 5.14). In all scenarios the noisy microphone signals are constructed such that the unbiased SNR of $y_0[k]$ is 0 dB. The following multi-microphone signal enhancement techniques are compared: DS-beamformer, GSC ($L_{ANC} = 400$, 1 and all noise reference signals), GSC with spatial filter designed blocking matrix ($L_{ANC} = 400$), recursive GSVD-based optimal filtering technique ($L = 20$, no sub-sampling) with and without ANC postprocessing stage ($L_{ANC} = 400$, 1 and all noise reference signals). This comparison is performed for different reverberation conditions. The situation of the 3 simultaneous noise sources in a highly reverberant environment can actually be considered quite a good approximation of a diffuse noise field.

Figures 5.12a, 5.13a and 5.14a show that for low T_{60} the SNR improvement of the GSC-based techniques is better than the SNR improvement of the GSVD-based optimal filtering technique without ANC postprocessing stage. When adding the ANC postprocessing stage using all noise reference signals, the SNR improvement of the GSVD-based optimal filtering technique clearly outperforms the SNR improvement of the GSC (both using Griffiths-Jim and spatial filter designed blocking matrix) for all reverberation times and all considered acoustic scenarios. In addition, the performance for the white noise source is better than for the speech noise source and the performance for a single noise source is better than for 3 simultaneous noise sources at different positions. This can be explained by the fact that the GSVD-based optimal filter can be decomposed as the combination of a spatial filtering operation, depending on the spatial characteristics (coherence) of the speech and the noise field, and a single-channel Wiener filter, depending on the spectral characteristics of the speech and the noise sources, cf. (3.83) in Section 3.5.1.

Figures 5.12b, 5.13b and 5.14b show the speech distortion introduced by the recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage for different reverberation times. More speech distortion occurs for higher reverberation times and when using more noise reference signals. This can also be seen from Fig. 5.15, where the PSD and the PTF of the speech and the noise components, defined in (2.39) and (2.41), have been plotted for 3 different reverberation times. For reverberation time $T_{60} = 300$ ms, Fig. 5.15a shows the PSD of the speech and the noise components of the first microphone

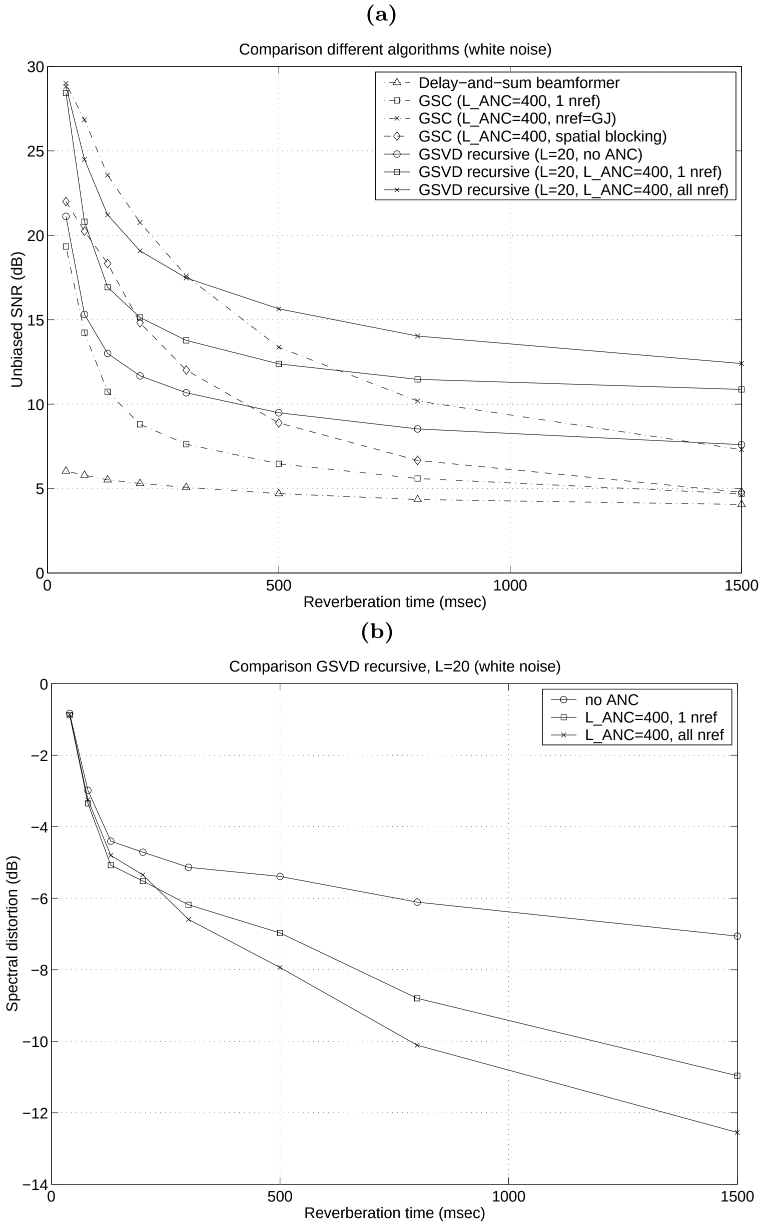


Figure 5.12: Comparison of (**a**) unbiased SNR and (**b**) speech distortion for DS, GSC and recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage (white noise, $L = 20$, $L_{ANC} = 400$, no sub-sampling)

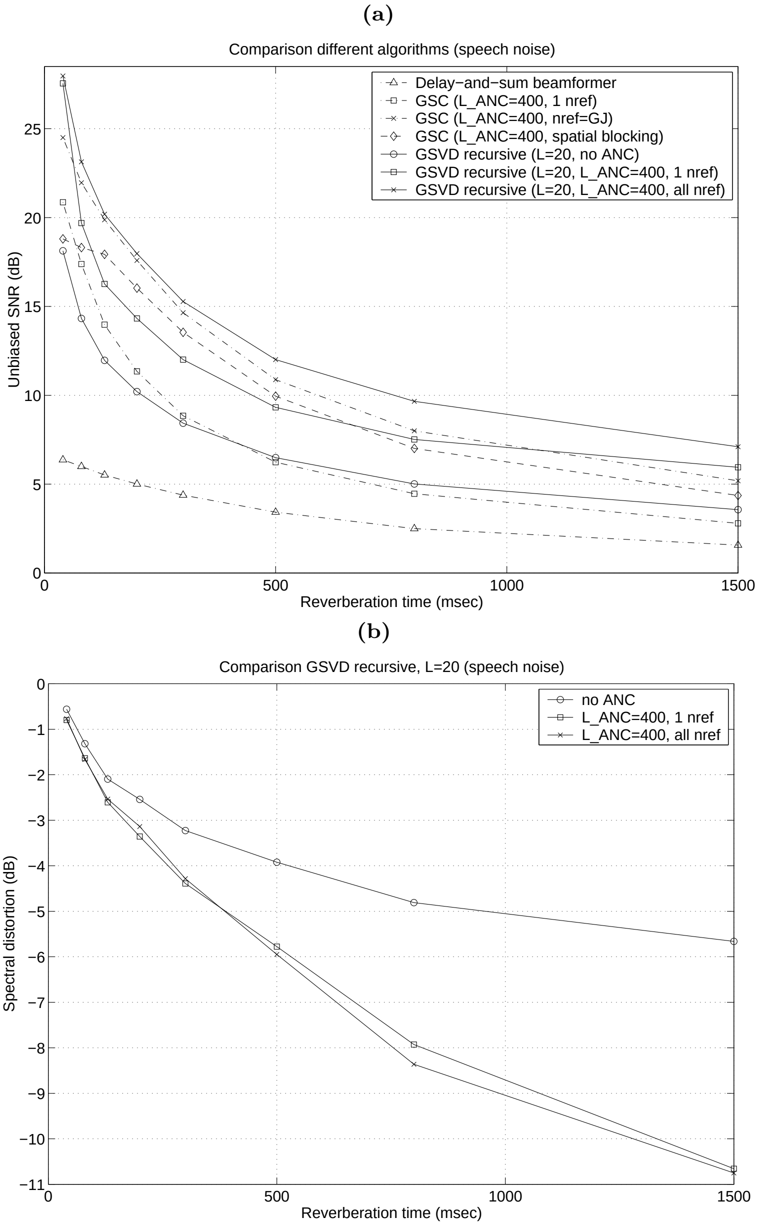


Figure 5.13: Comparison of (**a**) unbiased SNR and (**b**) speech distortion for DS, GSC and recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage (speech noise, $L = 20$, $L_{ANC} = 400$, no subsampling)

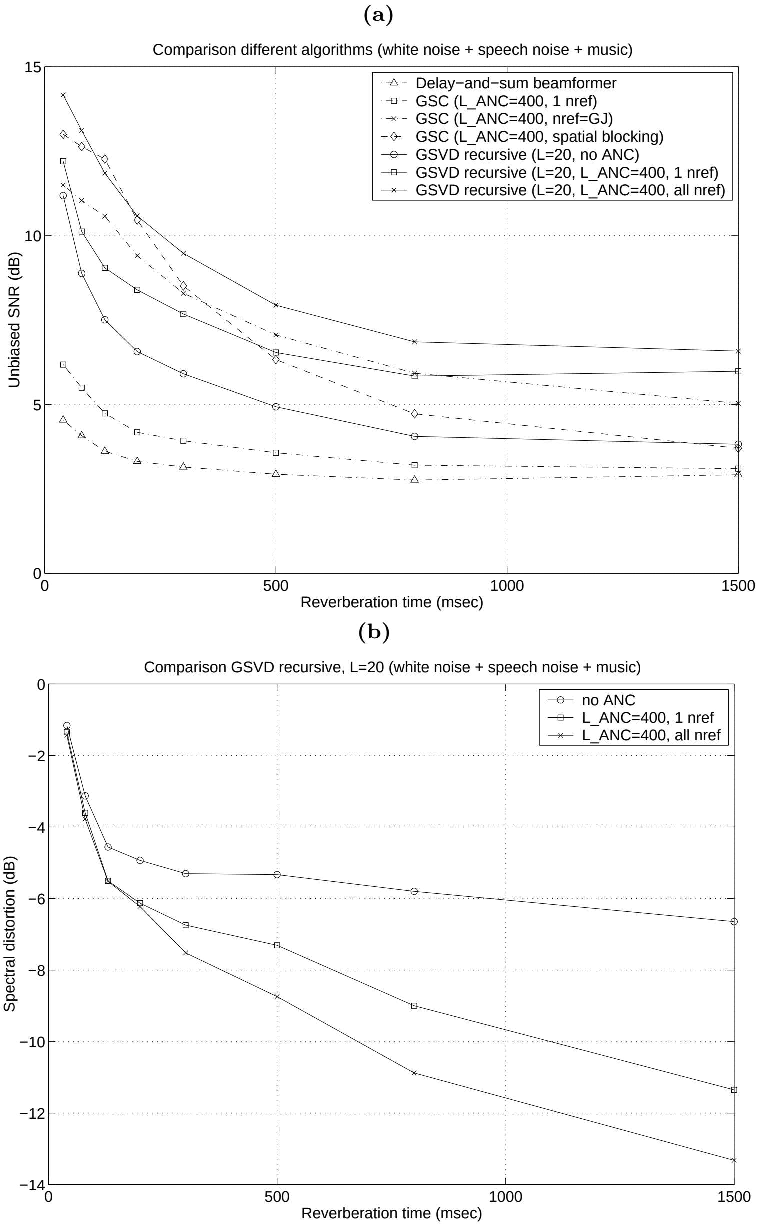


Figure 5.14: Comparison of (**a**) unbiased SNR and (**b**) speech distortion for DS, GSC and recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage (3 noise sources, $L = 20$, $L_{ANC} = 400$, no subsampling)

signal and Figure 5.15c shows the PTF for the speech and the noise components of the output signal of the recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage. As can be seen from Fig. 5.15c, speech distortion is limited, mainly occurs in frequency regions having a low input SNR and is slightly higher when using an ANC postprocessing stage (which however also reduces a large amount of noise). Figures 5.15b and 5.15d show the PTFs for reverberation times $T_{60} = 130\,\text{ms}$ and $T_{60} = 800\,\text{ms}$. By comparing these figures, one can see that more speech distortion and less noise reduction occurs for higher reverberation times (both in the GSVD-based optimal filter and in the ANC postprocessing stage).

5.4.2 Real-life recording and energy-based VAD

We have also compared the performance of the different multi-microphone speech enhancement algorithms for a real-life recording, performed in the Speech Lab at our department[2]. The reverberation time of the used room is approximately 500 ms. We have used a linear equi-spaced microphone array with $N = 3$ omni-directional microphones (Sennheiser ME-102) and inter-microphone distance $d = 5\,\text{cm}$. The speech source is located at approximately 1 m from the centre of the microphone array at an angle of 110°, whereas 3 noise sources are located at different positions in the room. We have used the same speech signal as for the simulated acoustic environments and for all noise sources we have used speech noise from the NOISEX-92 database.

The parameters for the speech enhancement algorithms are the same as for the simulated acoustic environments (cf. Sections 5.1.2 and 5.1.3). We have used a non-perfect energy-based VAD, cf. Section 5.3.1, on the noisy microphone signal $y_0[k]$. For the design of the spatial filter designed blocking matrix (and the associated fixed beamformer), we have again considered the non-linear criterion, but we have now used a robust design procedure, taking into account some errors in the microphone characteristics (cf. Chapter 10). We have used a uniform gain and phase probability density function (pdf), assuming a maximum gain and phase deviation of $\pm 2\,\text{dB}$ and $\pm 10°$.

The unbiased SNR of the first microphone signal is 0 dB, and the unbiased SNRs of the output signal for the DS beamformer, the GSC with Griffiths-Jim blocking matrix and the spatial filter designed blocking matrix respectively are 0.46 dB, 7.43 dB and 6.67 dB. For the recursive GSVD-based optimal filtering technique, the unbiased SNR is 6.25 dB, and when adding the ANC postprocessing stage using all noise reference signals, the unbiased SNR is 9.02 dB. The GSVD-based optimal filtering technique also introduces some amount of speech distortion, which is higher when adding the ANC postprocessing stage.

5.4.3 Robustness issues

[2] For this recording, sound files, spectrograms and power transfer functions are available at `http://www.esat.kuleuven.ac.be/~doclo/SA00061/audio.html`

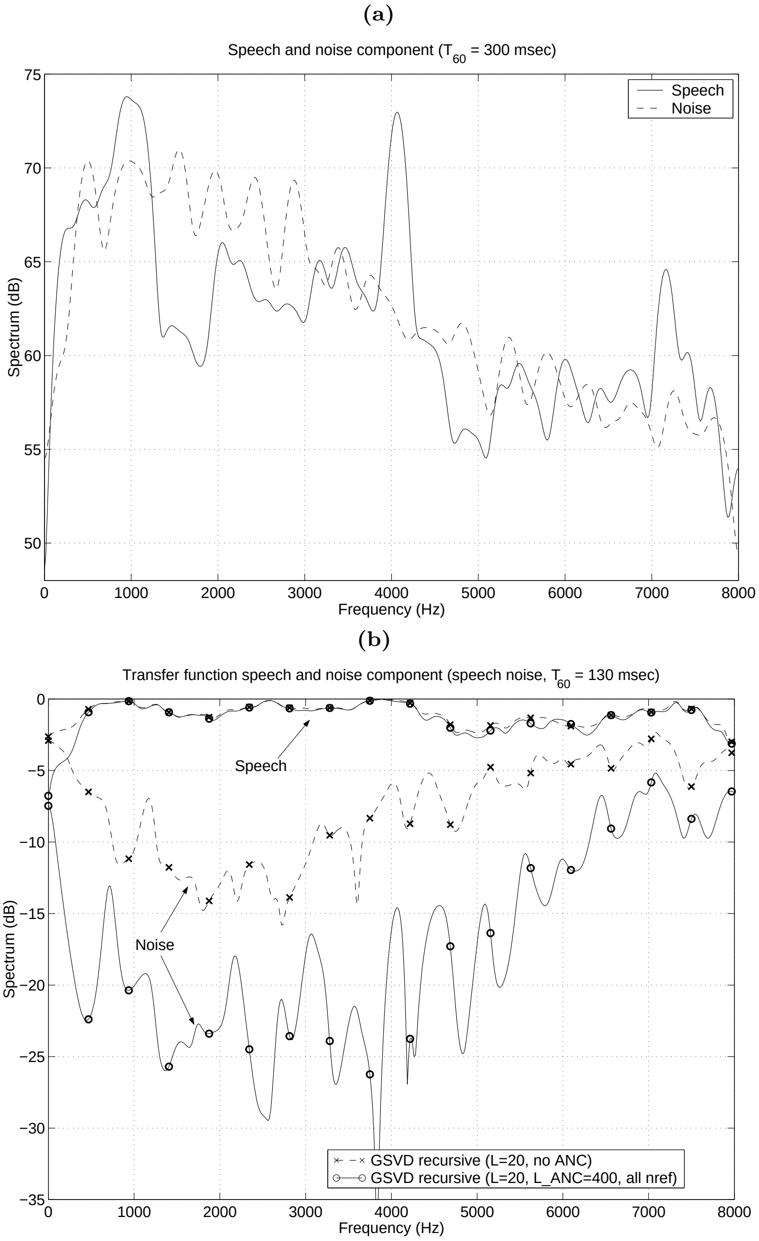


Figure 5.15: (**a**) PSD of speech and noise components of first microphone signal ($T_{60} = 300\,\text{ms}$), (**b**) PTF of speech and noise components for recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage for $T_{60} = 130\,\text{ms}$ (speech noise, $L = 20$, no sub-sampling, $L_{ANC} = 400$, all noise references)

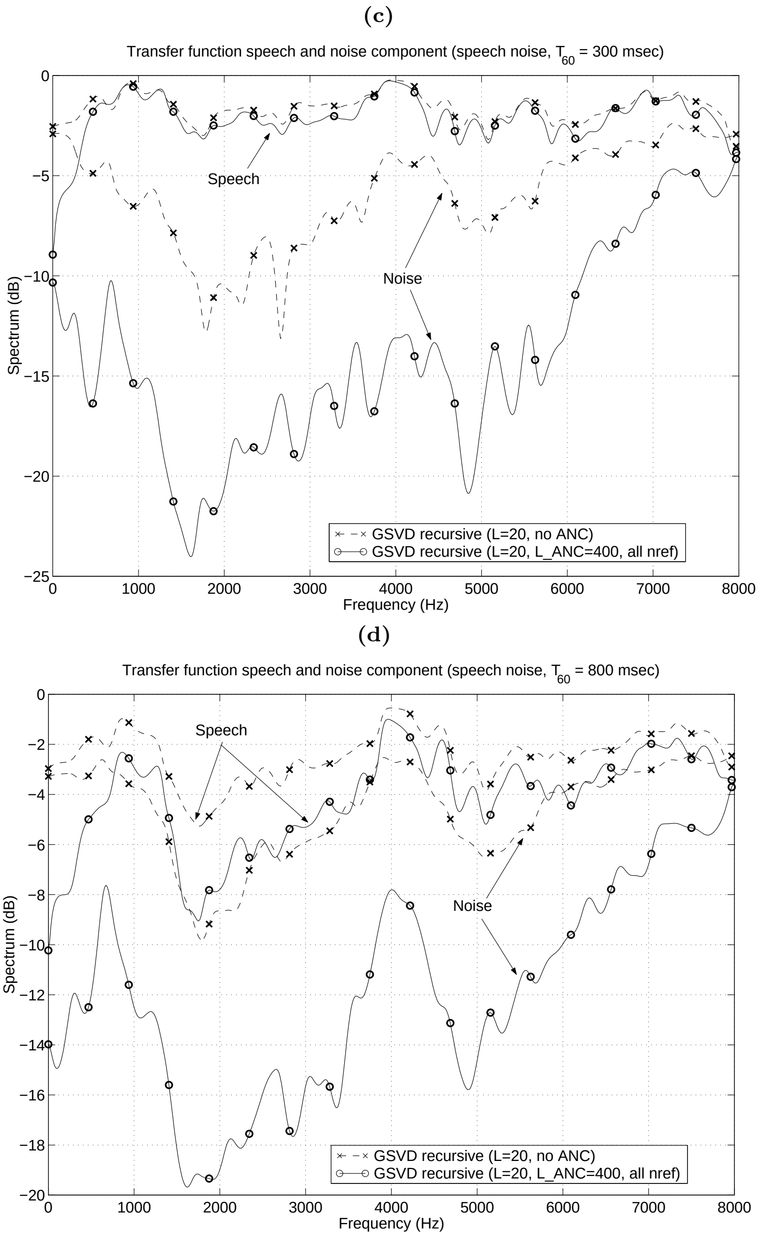


Figure 5.15: PTF of speech and noise components for recursive GSVD-based optimal filtering technique with and without ANC postprocessing stage for (**c**) $T_{60} = 300\,\text{ms}$, (**d**) $T_{60} = 800\,\text{ms}$ (speech noise, $L = 20$, no sub-sampling, $L_{ANC} = 400$, all noise references)

As mentioned in Section 1.3.4, in a real microphone array setup different kinds of imperfections occur (e.g. microphone gain and phase mismatch, deviations from the assumed array geometry). Multi-microphone signal enhancement techniques should be robust against (small) deviations from the assumed signal model. In this section, we investigate the robustness of the GSVD-based optimal filtering technique and the GSC for several types of deviations.

Many multi-microphone noise reduction techniques, e.g. GSC, rely on a priori assumptions about the position of the speech source and about the microphone array configuration. These techniques therefore tend to be rather sensitive to deviations from the assumed signal model, as e.g. encountered when incorrectly estimating the direction of the speech source or when using uncalibrated microphone arrays. Since the GSVD-based optimal filtering technique does not make any a priori assumptions about the location of the speaker and the microphone characteristics, it is expected to be more robust to deviations.

In [240] robustness has been analysed theoretically, i.e. using infinitely long filters (cf. Section 3.5) and not taking into account reverberation, for 3 types of model errors: (a) microphone gain and phase mismatch, (b) microphone displacement, and (c) incorrect estimation of the direction of the speech source. These types of deviations can be analysed by modifying the coherence matrices $\mathbf{\Gamma}_x(\omega)$ and/or $\mathbf{\Gamma}_v(\omega)$, cf. [240]. It has been shown that the GSC is extremely sensitive to microphone gain and phase mismatch (and to a smaller extent to the other two types of deviations) when the noise sensitivity $\Phi(\omega)$, cf. Section 3.5.4, is high. It has also been shown that the multi-channel Wiener filter is more robust than the GSC (even when adding a noise sensitivity constraint to the GSC [37][127][128][134]) for all 3 types of deviations. It can e.g. be proved that the performance of the multi-channel Wiener filter is independent of a deviation in microphone gain and phase. Suppose that the microphone characteristics of the nth microphone are independent of the angle of incidence θ and can be described by the function

$$A_n(\omega) = a_n(\omega)e^{-j\psi_n(\omega)} \ , \tag{5.4}$$

with $a_n(\omega)$ and $\psi_n(\omega)$ the frequency-dependent gain and phase. The nth microphone signal $\tilde{Y}_n(\omega)$ can then be written as $\tilde{Y}_n(\omega) = A_n(\omega)\, Y_n(\omega)$, with $Y_n(\omega)$ the microphone signal assuming a flat frequency response equal to 1. The vector of microphone signals $\tilde{\mathbf{Y}}(\omega)$ can now be written as $\tilde{\mathbf{Y}}(\omega) = \mathbf{A}(\omega)\,\mathbf{Y}(\omega)$, with $\mathbf{A}(\omega) = \mathrm{diag}\{A_n(\omega)\}$, such that the correlation matrix $\tilde{\mathbf{R}}_{yy}(\omega)$ is equal to

$$\tilde{\mathbf{R}}_{yy}(\omega) = \mathcal{E}\{\tilde{\mathbf{Y}}(\omega)\tilde{\mathbf{Y}}^H(\omega)\} = \mathbf{A}(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{A}^H(\omega) \ . \tag{5.5}$$

Using (5.5) and a similar definition for $\tilde{\mathbf{R}}_{xx}(\omega)$, the multi-channel Wiener filter in (3.81) can then be written as

$$\begin{aligned} \tilde{\mathbf{W}}_{WF}(\omega) &= \tilde{\mathbf{R}}_{yy}^{-1}(\omega)\tilde{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1 = \mathbf{A}^{-H}(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{R}}_{xx}(\omega)\mathbf{A}^H(\omega)\,\mathbf{e}_1 && (5.6) \\ &= A_0^*(\omega)\mathbf{A}^{-H}(\omega)\mathbf{W}_{WF}(\omega) \ , && (5.7) \end{aligned}$$

such that the output signal $\tilde{Z}(\omega)$ is equal to

$$\tilde{Z}(\omega) = \tilde{\mathbf{W}}_{WF}^{H}(\omega)\tilde{\mathbf{Y}}(\omega) = A_0(\omega)\mathbf{W}_{WF}^{H}(\omega)\mathbf{Y}(\omega) = A_0(\omega)Z(\omega) \;, \tag{5.8}$$

which is a scaled version of $Z(\omega)$. Therefore the power transfer functions for the speech and the noise components and hence also the performance measures (unbiased SNR, speech distortion) of the multi-channel Wiener filter are independent of the microphone gain and phase. For the GSC, the filter in (2.155) now is equal to (assuming it is only calculated during noise-only periods)

$$\tilde{\mathbf{W}}(\omega) = \frac{\mathbf{W}_q^H(\omega)\mathbf{W}_q(\omega)}{\mathbf{W}_q^H(\omega)\mathbf{A}^{-H}(\omega)\boldsymbol{\Gamma}_v^{-1}(\omega)\mathbf{A}^{-1}(\omega)\mathbf{W}_q(\omega)}\mathbf{A}^{-H}(\omega)\boldsymbol{\Gamma}_v^{-1}(\omega)\mathbf{A}^{-1}(\omega)\mathbf{W}_q(\omega). \tag{5.9}$$

However, since $\mathbf{C}_a(\omega)\tilde{\mathbf{X}}(\omega) \neq \mathbf{0}$ and $\mathbf{C}_a(\omega)\tilde{\mathbf{R}}_{xx}(\omega) \neq \mathbf{0}$, signal leakage and hence signal distortion occurs.

For simulated acoustic environments (with reverberation), we have compared the robustness of the GSVD-based optimal filtering technique with the GSC in [51]. It has been shown that for all three considered deviations the GSVD-based optimal filtering technique is more robust than the GSC. In this section we present results for microphone gain mismatch and for microphone displacement. Fig. 5.16a shows the difference in unbiased SNR of the output signal between the GSVD-based optimal filtering technique and the GSC for different gains a_1 of the second microphone (this gain is assumed to be frequency-independent). For most reverberation times – especially higher reverberation times – the larger the deviation of the gain a_1 from the assumed nominal gain $a_1^{nom} = 1$, the larger the difference in performance between the two techniques. Hence, the GSVD-based optimal filtering technique is more robust than the GSC for microphone gain mismatch. Fig. 5.16b shows the difference in unbiased SNR of the output signal for different positions $\mathbf{p}_1$ of the second microphone. The larger the deviation of the microphone position $\mathbf{p}_1$ from the assumed nominal position $\mathbf{p}_1^{nom} = 3.05\,\text{m}$, the larger the difference in performance. Hence, the GSVD-based optimal filtering technique is also more robust than the GSC for microphone displacement.

5.5 Conclusion

In this chapter the performance (unbiased SNR improvement, speech distortion and robustness) of the GSVD-based implementation of the multi-channel optimal filtering technique has been analysed for several acoustic environments and has been compared with standard fixed and adaptive beamforming techniques.

In Section 5.2 the performance of the GSVD-based optimal filtering technique has been analysed for several algorithmic parameters. For higher filter lengths L and for lower reverberation times T_{60}, the unbiased SNR increases and the

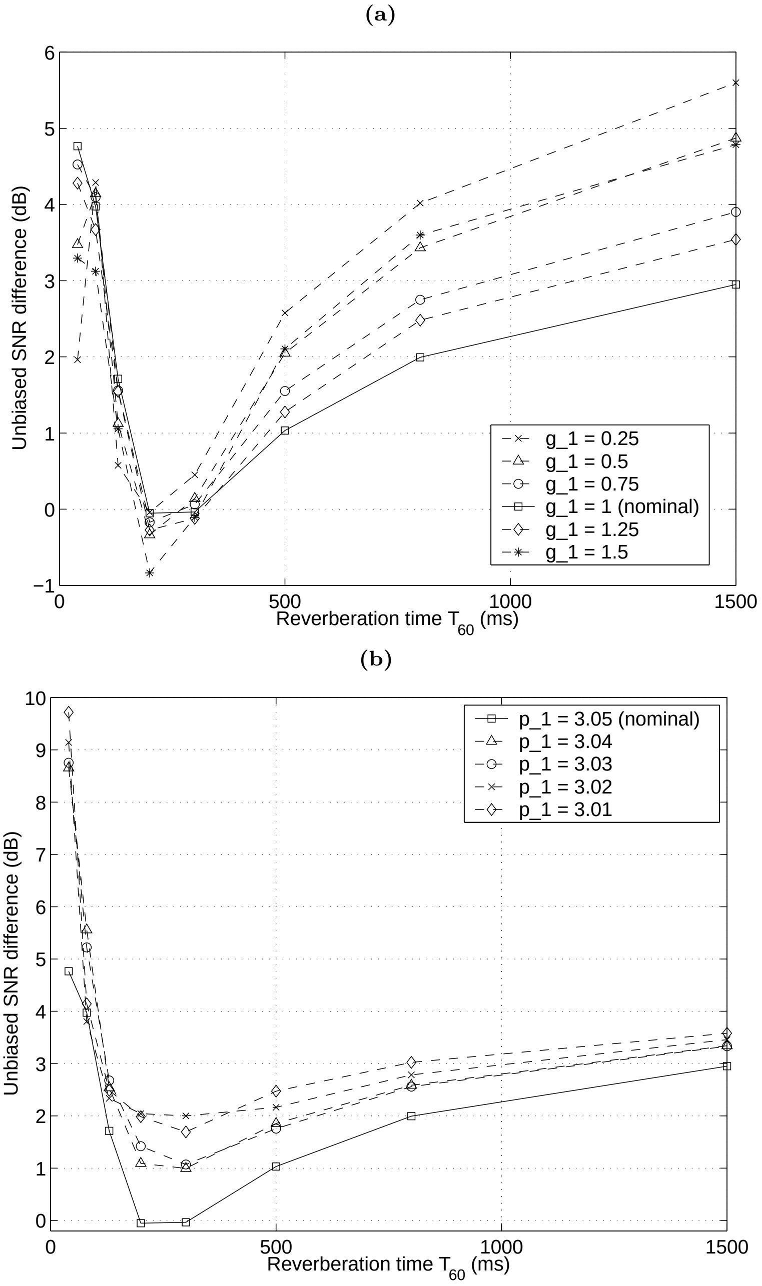


Figure 5.16: Unbiased SNR difference between GSVD-based optimal filtering ($L = 80$) and GSC for (**a**) different microphone gain a_1, (**b**) different microphone position $\mathbf{p}_1$

speech distortion decreases. It has been shown that the batch and the recursive version (both 'conventional' and square root-free implementation) nearly have the same performance. For stationary acoustic environments, a small number of GSVD-steps (and sweeps) and a higher sub-sampling factor can be used without decreasing the performance. The ANC postprocessing stage can either be used for increasing the noise reduction performance or for computational complexity reduction without decreasing the performance. The ANC postprocessing stage however also gives rise to a slight increase in speech distortion, which can be limited by using longer filter lengths. It has also been shown that the GSVD-based optimal filtering technique exhibits the desired beamforming behaviour for simple acoustic scenarios and that this technique can be used for suppressing a spectrally non-stationary noise source.

Since the GSVD-based optimal filtering technique uses no other a priori information than the output of a VAD algorithm, it is expected to be quite sensitive to speech detection errors. However, in Section 5.3 it has been shown (both theoretically and experimentally) that the unbiased SNR improvement is not degraded by speech detection errors, but that the speech distortion increases with increasing error rate δ. For error rates $\delta < 0.2$, speech distortion remains limited (also when adding the ANC postprocessing stage). When evaluating the performance of the GSVD-based optimal filtering technique in combination with different VAD algorithms, it has been shown that the best performance for different noise types is achieved using the log-likelihood and the log-energy VAD algorithms.

In Section 5.4 the performance of the GSVD-based optimal filtering technique has been compared with standard beamforming techniques for various acoustic scenarios (single and multiple noise sources, real-life recording). The SNR improvement of the GSVD-based optimal filtering technique with ANC postprocessing stage outperforms the SNR improvement of the GSC for all reverberation times and for all considered acoustic scenarios. More speech distortion occurs for higher reverberation times and when adding the ANC postprocessing stage using multiple noise reference signals. In addition, the robustness of the GSC and the GSVD-based optimal filter has been analysed for several deviations from the assumed signal model. It has been shown that the performance of the GSVD-based optimal filter is independent of a deviation in the microphone gain and phase and that the GSVD-based optimal filter is more robust than the GSC for microphone mismatch, microphone displacement and look direction error.

Part II

Multi-Microphone Dereverberation and Source Localisation

Chapter 6

Robust Time-Delay Estimation for Acoustic Source Localisation

In this chapter two adaptive algorithms are presented for robust time-delay estimation (TDE) in acoustic environments where a large amount of additive background noise and reverberation is present.

Section 6.1 gives an introduction to the acoustic source localisation problem and gives a brief overview of existing methods for acoustic source localisation and time-delay estimation. Generally we will consider time-delay estimation between two microphone signals in this chapter.

Section 6.2 discusses the batch, i.e. non-adaptive, estimation of the complete acoustic impulse responses from the microphone signals. It is shown that if the length of the acoustic impulse responses is either known or can be overestimated, the complete acoustic impulse responses can be identified from the eigenvalue decomposition (EVD) of the speech correlation matrix (in the noiseless case and in the spatio-temporally white noise case) or from the generalised eigenvalue decomposition (GEVD) of the speech and the noise correlation matrices (in the coloured noise case).

These batch impulse response estimation procedures form the basis for deriving stochastic gradient algorithms which iteratively estimate the (generalised) eigenvector corresponding to the smallest (generalised) eigenvalue. These adaptive EVD and GEVD algorithms are discussed in Section 6.3. In [9] it has been shown that the adaptive EVD algorithm can be used for TDE, remarkably even when underestimating the length of the acoustic impulse responses. We will show that this result can be extended to the spatio-temporally coloured noise case, by using an adaptive GEVD algorithm (and an adaptive prewhitening algorithm) for TDE. In Section 6.4 it is shown that these adaptive TDE algorithms can be straightforwardly extended to the case of more than two microphones.

Section 6.5 describes the simulation results, using different reverberation conditions (ideal and realistic), different SNRs and different microphone configurations. For all conditions it is shown that the time-delays can be estimated more robustly using the adaptive GEVD algorithm than using the adaptive EVD algorithm and the adaptive prewhitening algorithm.

6.1 Introduction

In many applications, such as teleconferencing, hands-free voice-controlled systems and hearing aids, it is desirable to localise the dominant speaker. By using a microphone array, it is possible to determine the *position* of this speaker, such that the microphone array then can be electronically steered using fixed (and adaptive) beamforming techniques, cf. Section 2.5 and Part III. In multimedia teleconferencing systems, the position of the dominant speaker can be used not only for microphone array beamforming, but also for automatic video camera steering [132][271] and for determining binaural cues for stereo imaging.

It has been shown that it is possible to calculate the position of a speaker, e.g. using maximum likelihood or least-squares methods, when the *time-delays* between the different microphone signals are known [23][30][133][276]. However, accurate estimation of the time-delays between the different microphone signals is not an easy task because of the room reverberation, the acoustic background noise and the non-stationary character and the low-rank model of the speech signal. Generally, room reverberation is considered to be the main problem for TDE [29], but acoustic background noise can also considerably decrease the performance of time-delay estimators. Whereas highly noisy situations are not very common in teleconferencing applications, they frequently occur in e.g. hearing aid applications.

Most TDE methods are based on the generalised cross-correlation (GCC) or the cross-power spectrum phase (CSP) between the microphone signals [106][154] [200][217]. However, since most of these methods assume an ideal room model without reverberation, i.e. only a direct path between the speech source and the microphone array, they can not handle reverberation well. In order to make TDE more robust to reverberation, a cepstral pre-filtering technique has been proposed in [246] and techniques have been developed that use a more realistic room model incorporating reverberation [9][33]. In [9] an adaptive EVD algorithm has been developed for *(partial) estimation of two acoustic impul-*

se responses, using a stochastic gradient algorithm that iteratively estimates the eigenvector corresponding to the smallest eigenvalue. From the estimated acoustic impulse responses, the time-delay can be calculated as the time difference between the first peak (direct path) of the two impulse responses or as the peak of the correlation function between the two impulse responses. Since only the time difference between the first peak (direct path) of the acoustic impulse responses is required, it is therefore not necessary for TDE to estimate the complete acoustic impulse responses.

The adaptive EVD algorithm for TDE performs much better in highly reverberant environments than the GCC-based methods. However, the adaptive EVD algorithm is - strictly speaking - only valid if either no noise or if spatio-temporally white noise is present. In this chapter we extend the adaptive EVD algorithm to the spatio-temporally coloured noise case, by deriving an (adaptive) stochastic gradient algorithm for the GEVD or by prewhitening the noisy microphone signals. In addition, we extend all adaptive TDE algorithms to the case of more than two microphones.

6.2 Batch estimation of two impulse responses

This section discusses the batch, i.e. non-adaptive, estimation of the complete acoustic impulse responses from the recorded microphone signals. The techniques discussed in this section are based on the subspace method, e.g. used in [1][7][101][117][187][252][265] for different applications. We will briefly review these well-known techniques, because they form the basis for deriving the stochastic gradient algorithms which iteratively estimate the (generalised) eigenvector corresponding to the smallest (generalised) eigenvalue. These adaptive techniques will be used for TDE in practice (cf. Section 6.3).

Section 6.2.1 discusses the estimation of the acoustic impulse responses using correlation matrices for the noiseless case, whereas the spatio-temporally white noise and coloured noise case are discussed in Sections 6.2.2 and 6.2.3. Section 6.2.4 discusses the practical computation using data matrices and Section 6.2.5 gives some simulation results.

Recall from the recording model of Section 2.2 that each microphone signal $y_n[k]$ consists of a filtered version of the speech signal $s[k]$ and additive noise, i.e.

$$y_n[k] = h_n[k] \otimes s[k] + v_n[k] = x_n[k] + v_n[k] \ . \tag{6.1}$$

The goal is to estimate the acoustic impulse responses $h_n[k]$ from the microphone signals $y_n[k]$ without any a priori knowledge about the speech signal $s[k]$. After estimating the complete impulse responses, it is then trivial to compute the time-delays between the direct paths. The acoustic impulse response $h_n[k]$

can generally be modelled using an FIR-filter $\mathbf{h}_n$ of length K, cf. (2.13),

$$\mathbf{h}_n = \begin{bmatrix} h_{n,0} & h_{n,1} & \ldots & h_{n,K-1} \end{bmatrix}^T . \tag{6.2}$$

Since $h_m[k] \otimes x_n[k] = h_m[k] \otimes h_n[k] \otimes s[k] = h_n[k] \otimes x_m[k]$, the relation

$$\mathbf{x}_{n,K}^T[k]\, \mathbf{h}_m = \mathbf{x}_{m,K}^T[k]\, \mathbf{h}_n\ , \quad m, n = 0 \ldots N-1\ , \tag{6.3}$$

holds [9], with the K-dimensional data vector $\mathbf{x}_{n,K}[k]$ defined as

$$\mathbf{x}_{n,K}[k] = \begin{bmatrix} x_n[k] & x_n[k-1] & \ldots & x_n[k-K+1] \end{bmatrix}^T . \tag{6.4}$$

Although we do not explicitly attribute a time index k to the impulse responses, this does not imply that these are time-invariant. In addition, in the remainder of this section we will assume $N = 2$, although it is quite easy to extend all algorithms to the case of more than two microphones (cf. Section 6.4).

6.2.1 Noiseless case

Similarly as in Section 3.4.3, the $2L \times 2L$-dimensional correlation matrix $\bar{\mathbf{R}}_{xx}[k]$ is defined as

$$\bar{\mathbf{R}}_{xx}[k] = \mathcal{E}\{\mathbf{x}[k]\, \mathbf{x}^T[k]\} = \begin{bmatrix} \bar{\mathbf{R}}_{xx}^{00}[k] & \bar{\mathbf{R}}_{xx}^{01}[k] \\ \bar{\mathbf{R}}_{xx}^{10}[k] & \bar{\mathbf{R}}_{xx}^{11}[k] \end{bmatrix} , \tag{6.5}$$

with the $2L$-dimensional stacked data vector $\mathbf{x}[k]$ equal to

$$\mathbf{x}[k] = \begin{bmatrix} \mathbf{x}_{0,L}[k] \\ \mathbf{x}_{1,L}[k] \end{bmatrix} , \tag{6.6}$$

and the $L \times L$-dimensional sub-matrix $\bar{\mathbf{R}}_{xx}^{mn}[k]$ equal to

$$\bar{\mathbf{R}}_{xx}^{mn}[k] = \mathcal{E}\{\mathbf{x}_{m,L}[k]\, \mathbf{x}_{n,L}^T[k]\} . \tag{6.7}$$

Using (3.24), i.e. $\bar{\mathbf{R}}_{xx}[k] = \mathcal{H}[k]\, \bar{\mathbf{R}}_{ss}[k]\, \mathcal{H}^T[k]$, with $\mathcal{H}[k]$ a $2L \times (K+L-1)$-dimensional matrix, one can see that when the true acoustic impulse response length K is overestimated, i.e. $L \geq K$, the correlation matrix $\bar{\mathbf{R}}_{xx}[k]$ has rank $K+L-1$ and its null-space has dimension $L-K+1$, provided that [187]

1. the acoustic impulse responses $\mathbf{h}_0$ and $\mathbf{h}_1$ do not have common zeros;
2. the $(K+L-1) \times (K+L-1)$-dimensional correlation matrix $\bar{\mathbf{R}}_{ss}[k]$ of the clean speech signal $s[k]$ has full rank. Although it has been assumed in Section 3.2.2 that this matrix has rank R, with $R \leq K+L-1$, the low-rank model of the speech signal is only approximately valid, i.e. we can assume that $\bar{\mathbf{R}}_{ss}[k]$ has $K+L-1-R$ eigenvalues which are very small, but which are not exactly equal to zero.

If $L = K$, the null-space of $\bar{\mathbf{R}}_{xx}[k]$ has dimension 1, and the $2K$-dimensional vector

$$\bar{\mathbf{v}} = \begin{bmatrix} -\mathbf{h}_1 \\ \mathbf{h}_0 \end{bmatrix} \tag{6.8}$$

belongs to this null-space, since, using (6.3), $\bar{\mathbf{R}}_{xx}[k]\bar{\mathbf{v}} = \mathbf{0}$. Consider the EVD of $\bar{\mathbf{R}}_{xx}[k]$, cf. Appendix A.2,

$$\bar{\mathbf{R}}_{xx}[k] = \bar{\mathbf{V}}_x \bar{\mathbf{\Delta}}_x \bar{\mathbf{V}}_x^T \ , \tag{6.9}$$

with $\bar{\mathbf{V}}_x$ an orthogonal matrix, containing the eigenvectors, and $\bar{\mathbf{\Delta}}_x$ a diagonal matrix, containing the eigenvalues. Hence, the unit-norm eigenvector, corresponding to the only zero eigenvalue of $\bar{\mathbf{R}}_{xx}[k]$, contains a scaled version of the two acoustic impulse responses $\mathbf{h}_0$ and $\mathbf{h}_1$.

If $L > K$, the null-space of $\bar{\mathbf{R}}_{xx}[k]$ is spanned by $L - K + 1$ eigenvectors, corresponding to the $L-K+1$ zero eigenvalues. All these eigenvectors contain a different filtered version of the acoustic impulse responses. By extracting the common part of the eigenvectors, which can e.g. be done by performing a QR-decomposition of the full null-space or by using a least-squares approach [101][117], the correct acoustic impulse responses of length K can be identified.

If $L < K$, the rank of $\bar{\mathbf{R}}_{xx}[k]$ is equal to $2L$ and the null-space of $\bar{\mathbf{R}}_{xx}[k]$ is empty, such that generally the acoustic impulse responses can not be correctly identified.

6.2.2 Spatio-temporally white noise

When additive noise is present, consider the $2L \times 2L$-dimensional speech and noise correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$, defined as in (6.5), i.e.

$$\bar{\mathbf{R}}_{yy}[k] = \begin{bmatrix} \bar{\mathbf{R}}_{yy}^{00}[k] & \bar{\mathbf{R}}_{yy}^{01}[k] \\ \bar{\mathbf{R}}_{yy}^{10}[k] & \bar{\mathbf{R}}_{yy}^{11}[k] \end{bmatrix}, \quad \bar{\mathbf{R}}_{vv}[k] = \begin{bmatrix} \bar{\mathbf{R}}_{vv}^{00}[k] & \bar{\mathbf{R}}_{vv}^{01}[k] \\ \bar{\mathbf{R}}_{vv}^{10}[k] & \bar{\mathbf{R}}_{vv}^{11}[k] \end{bmatrix}, \tag{6.10}$$

with the $L \times L$-dimensional sub-matrices $\bar{\mathbf{R}}_{yy}^{mn}[k]$ and $\bar{\mathbf{R}}_{vv}^{mn}[k]$ equal to

$$\bar{\mathbf{R}}_{yy}^{mn}[k] = \mathcal{E}\{\mathbf{y}_{m,L}[k]\,\mathbf{y}_{n,L}^T[k]\}, \quad \bar{\mathbf{R}}_{vv}^{mn}[k] = \mathcal{E}\{\mathbf{v}_{m,L}[k]\,\mathbf{v}_{n,L}^T[k]\} \ , \tag{6.11}$$

and the L-dimensional vectors $\mathbf{y}_{n,L}[k]$ and $\mathbf{v}_{n,L}[k]$ defined as in (6.4). Since the speech and the noise components are assumed to be uncorrelated, we can write

$$\bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{R}}_{xx}[k] + \bar{\mathbf{R}}_{vv}[k] \ . \tag{6.12}$$

If the noise is spatio-temporally white, i.e. $\bar{\mathbf{R}}_{vv}[k] = \bar{\sigma}_v^2\,\mathbf{I}_{2L}$, with $\bar{\sigma}_v^2$ the noise power, the acoustic impulse responses can still be identified from the EVD of the speech correlation matrix, i.e.

$$\bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{V}}_y \bar{\mathbf{\Delta}}_y \bar{\mathbf{V}}_y^T \ . \tag{6.13}$$

Using (6.9) and (6.12), we can write $\bar{\mathbf{R}}_{yy}[k]$ in the spatio-temporally white noise case as

$$\bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{V}}_x(\bar{\mathbf{\Delta}}_x + \bar{\sigma}_v^2\,\mathbf{I}_{2L})\bar{\mathbf{V}}_x^T\ , \tag{6.14}$$

such that $\bar{\mathbf{V}}_y = \bar{\mathbf{V}}_x$ and $\bar{\mathbf{\Delta}}_y = \bar{\mathbf{\Delta}}_x + \bar{\sigma}_v^2\,\mathbf{I}_{2L}$. If $L = K$, only one of the diagonal elements of $\bar{\mathbf{\Delta}}_y$ is equal to $\bar{\sigma}_v^2$ (smallest eigenvalue), and the eigenvector in $\bar{\mathbf{V}}_y$, corresponding to this eigenvalue, again contains a scaled version of the acoustic impulse responses. If $L > K$, the procedure for estimating the impulse responses of length K is similar to the procedure in the noiseless case, and is now based on the $L-K+1$ eigenvectors in $\bar{\mathbf{V}}_y$ corresponding to the eigenvalues which are equal to $\bar{\sigma}_v^2$.

6.2.3 Spatio-temporally coloured noise

If spatio-temporally coloured noise is present, the acoustic impulse responses can not be identified from the EVD of $\bar{\mathbf{R}}_{yy}[k]$, but they can still be identified from the GEVD of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ or from the EVD of the pre-whitened speech correlation matrix $\tilde{\mathbf{R}}_{yy}[k]$. In both cases, the noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$ needs to be known in advance or we have to be able to estimate $\bar{\mathbf{R}}_{vv}[k]$ from noise-only periods, requiring a VAD-algorithm (cf. Section 5.3).

1. *GEVD-procedure*: The GEVD of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ is equal to, cf. (3.15),

$$\left\{\begin{array}{lcl} \bar{\mathbf{R}}_{yy}[k] & = & \bar{\mathbf{Q}}\,\bar{\mathbf{\Lambda}}_y\,\bar{\mathbf{Q}}^T \\ \bar{\mathbf{R}}_{vv}[k] & = & \bar{\mathbf{Q}}\,\bar{\mathbf{\Lambda}}_v\,\bar{\mathbf{Q}}^T\ , \end{array}\right. \tag{6.15}$$

   with $\bar{\mathbf{Q}}$ a $2L \times 2L$-dimensional invertible, but not necessarily orthogonal, matrix and $\bar{\mathbf{\Lambda}}_y$ and $\bar{\mathbf{\Lambda}}_v$ diagonal matrices. From (6.12) and (6.15), it follows that

$$\bar{\mathbf{R}}_{vv}^{-1}[k]\,\bar{\mathbf{R}}_{xx}[k] = \bar{\mathbf{R}}_{vv}^{-1}[k]\,(\bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k]) = \bar{\mathbf{Q}}^{-T}(\bar{\mathbf{\Lambda}}_v^{-1}\bar{\mathbf{\Lambda}}_y - \mathbf{I}_{2L})\,\bar{\mathbf{Q}}^T\ .$$

   Since $\bar{\mathbf{R}}_{vv}^{-1}[k]\,\bar{\mathbf{R}}_{xx}[k]$ has rank $K+L-1$ ($\bar{\mathbf{R}}_{vv}[k]$ is assumed to be of full rank), $L-K+1$ diagonal elements of the diagonal matrix $\bar{\mathbf{\Lambda}}_v^{-1}\bar{\mathbf{\Lambda}}_y$ are equal to 1. Hence, $L-K+1$ columns $\bar{\mathbf{q}}$ of $\bar{\mathbf{Q}}^{-T}$ exist for which

$$\bar{\mathbf{R}}_{vv}^{-1}[k]\,\bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{q}} = \mathbf{0}\ , \tag{6.16}$$

   such that $\bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{q}} = \mathbf{0}$. If $L = K$, the null-space of $\bar{\mathbf{R}}_{xx}[k]$ has dimension 1, and the $2K$-dimensional vector $\bar{\mathbf{q}}$ contains a scaled version of the acoustic impulse responses. If $L > K$, the $L-K+1$ vectors $\bar{\mathbf{q}}$ again contain different filtered versions of the acoustic impulse responses, and the procedure for estimating the correct acoustic impulse responses of length K is similar to the procedure in the noiseless case.

2. *Prewhitening procedure*: The $2L \times 2L$-dimensional pre-whitened speech correlation matrix $\tilde{\mathbf{R}}_{yy}[k]$ is defined as

$$\tilde{\mathbf{R}}_{yy}[k] \triangleq \bar{\mathbf{R}}_{vv}^{-T/2}[k]\,\bar{\mathbf{R}}_{yy}[k]\,\bar{\mathbf{R}}_{vv}^{-1/2}[k]\ , \tag{6.17}$$

with $\bar{\mathbf{R}}_{vv}^{1/2}[k]$ the $2L \times 2L$-dimensional upper-triangular Cholesky-factor [110] of the noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$, i.e. $\bar{\mathbf{R}}_{vv}[k] = \bar{\mathbf{R}}_{vv}^{T/2}[k]\,\bar{\mathbf{R}}_{vv}^{1/2}[k]$. From the EVD of $\tilde{\mathbf{R}}_{yy}[k]$, i.e.

$$\tilde{\mathbf{R}}_{yy}[k] = \tilde{\mathbf{V}}_y \tilde{\mathbf{\Lambda}}_y \tilde{\mathbf{V}}_y^T\ , \tag{6.18}$$

it follows, using (6.12), that the pre-whitened matrix $\tilde{\mathbf{R}}_{xx}[k]$ can be written as

$$\tilde{\mathbf{R}}_{xx}[k] \triangleq \bar{\mathbf{R}}_{vv}^{-T/2}[k]\,\bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{R}}_{vv}^{-1/2}[k] = \tilde{\mathbf{V}}_y(\tilde{\mathbf{\Lambda}}_y - \mathbf{I}_{2L})\tilde{\mathbf{V}}_y^T\ . \tag{6.19}$$

Since $\tilde{\mathbf{R}}_{xx}[k]$ has rank $K+L-1$, this implies that $L-K+1$ diagonal elements of the diagonal matrix $\tilde{\mathbf{\Lambda}}_y$ are equal to 1 and $L-K+1$ columns $\tilde{\mathbf{v}}$ of $\tilde{\mathbf{V}}_y$ exist for which

$$\tilde{\mathbf{R}}_{xx}[k]\,\tilde{\mathbf{v}} = \bar{\mathbf{R}}_{vv}^{-T/2}[k]\,\bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{R}}_{vv}^{-1/2}[k]\,\tilde{\mathbf{v}} = \mathbf{0}\ , \tag{6.20}$$

such that $\bar{\mathbf{R}}_{xx}[k]\,\bar{\mathbf{R}}_{vv}^{-1/2}[k]\,\tilde{\mathbf{v}} = \mathbf{0}$. If $L = K$, the null-space of $\bar{\mathbf{R}}_{xx}[k]$ has dimension 1, and the vector $\bar{\mathbf{R}}_{vv}^{-1/2}[k]\,\tilde{\mathbf{v}}$ contains a scaled version of the acoustic impulse responses. If $L > K$, the $L-K+1$ vectors $\bar{\mathbf{R}}_{vv}^{-1/2}[k]\,\tilde{\mathbf{v}}$ again contain different filtered versions of the acoustic impulse responses, and the procedure for estimating the correct acoustic impulse responses of length K is similar to the procedure in the noiseless case.

It is readily verified that the batch GEVD-procedure and the batch prewhitening procedure are in fact equivalent, since

$$\tilde{\mathbf{\Lambda}}_y = \bar{\mathbf{\Lambda}}_v^{-1}\bar{\mathbf{\Lambda}}_y, \quad \bar{\mathbf{Q}}^{-T} = \bar{\mathbf{R}}_{vv}^{-1/2}[k]\tilde{\mathbf{V}}_y\ . \tag{6.21}$$

However, the adaptive versions of both algorithms, which will be used in practice for TDE and which are discussed in Section 6.3, can produce different results.

6.2.4 Practical computation

As for the GSVD-based optimal filtering technique discussed in Chapter 3, in practice we do not work with correlation matrices but with data matrices. The $P \times 2L$-dimensional speech data matrix $\mathbf{Y}[k]$ is defined as, cf. (3.38),

$$\mathbf{Y}[k] = \begin{bmatrix} \mathbf{y}^T[k-P+1] \\ \vdots \\ \mathbf{y}^T[k-1] \\ \mathbf{y}^T[k] \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{0,L}^T[k-P+1] & \mathbf{y}_{1,L}^T[k-P+1] \\ \vdots & \\ \mathbf{y}_{0,L}^T[k-1] & \mathbf{y}_{1,L}^T[k-1] \\ \mathbf{y}_{0,L}^T[k] & \mathbf{y}_{1,L}^T[k] \end{bmatrix}\ , \tag{6.22}$$

assuming without loss of generality that all speech vectors are consecutive. The number of speech vectors P is typically much larger than L, such that the empirical speech correlation matrix can be computed as $\mathbf{R}_{yy}[k] = \mathbf{Y}^T[k]\mathbf{Y}[k]/P$. The $Q \times 2L$-dimensional noise data matrix $\mathbf{V}[k]$ is defined similarly as in (6.22).

1. *GSVD-procedure*: Instead of computing the GEVD of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$, we compute the GSVD of the data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$, cf. (3.43),

$$\begin{cases} \mathbf{Y}[k] &= \mathbf{U}_Y\, \mathbf{\Sigma}_Y\, \mathbf{Q}^T \\ \mathbf{V}[k] &= \mathbf{U}_V\, \mathbf{\Sigma}_V\, \mathbf{Q}^T \ , \end{cases} \tag{6.23}$$

   with $\mathbf{U}_Y$ and $\mathbf{U}_V$ orthogonal matrices, $\mathbf{\Sigma}_Y$ and $\mathbf{\Sigma}_V$ diagonal matrices and $\mathbf{Q}$ an invertible, but not necessarily orthogonal, matrix. The acoustic impulse responses are estimated from the columns of $\mathbf{Q}^{-T}$.

2. *Prewhitening procedure*: The pre-whitened speech data matrix $\tilde{\mathbf{Y}}[k]$ is defined as

$$\tilde{\mathbf{Y}}[k] = \mathbf{Y}[k]\, \mathbf{R}_{vv}^{-1/2}[k] \ , \tag{6.24}$$

   where the Cholesky-factor $\mathbf{R}_{vv}^{1/2}[k]$ can be computed using the QR-decomposition of the noise data matrix, i.e. $\mathbf{V}[k] = \mathbf{Q}_V[k]\, \mathbf{R}_{vv}^{1/2}[k]$. The SVD of $\tilde{\mathbf{Y}}[k]$ is defined as

$$\tilde{\mathbf{Y}}[k] = \tilde{\mathbf{U}}_Y\, \tilde{\mathbf{\Sigma}}_Y\, \tilde{\mathbf{V}}_Y^T \ , \tag{6.25}$$

   with $\tilde{\mathbf{U}}_Y$ and $\tilde{\mathbf{V}}_Y$ orthogonal matrices and $\tilde{\mathbf{\Sigma}}_Y$ a diagonal matrix. The acoustic impulse responses are estimated from the columns of $\mathbf{R}_{vv}^{-1/2}[k]\, \tilde{\mathbf{V}}_Y$.

6.2.5 Simulation results

We have filtered a 16 kHz speech segment of 160000 samples (10 sec) with 2 artificially generated impulse responses ($K = 20$), which are depicted in Fig. 6.1a. Stationary speech noise, having the same long-term spectrum as speech, has been added and the SNR of $y_0[k]$ is 10 dB.

Figures 6.1b and 6.1c show the estimated impulse responses for the SVD-procedure and for the GSVD-procedure, using all microphone samples and $L = K$ (for the GSVD-procedure, the noise components $v_n[k]$ are assumed to be available in order to compute $\mathbf{R}_{vv}^{1/2}[k]$). As can be clearly seen, the impulse responses are almost correctly estimated using the GSVD-procedure, but not using the SVD-procedure, because this procedure assumes that spatio-temporally white noise is present. However, since the assumption of uncorrelated speech and noise segments is not perfectly satisfied in practice due to the data-based estimation, i.e. $\mathbf{X}^T[k]\mathbf{V}[k] \neq \mathbf{0}$, small estimation errors occur in the GSVD-procedure and it appears that the estimation procedure is quite sensitive to this assumption. During the simulations we have noticed that the better this independence assumption is satisfied, i.e. the higher the SNR and the longer the speech and the noise segments, the smaller the estimation error becomes. This has also been observed in [101].

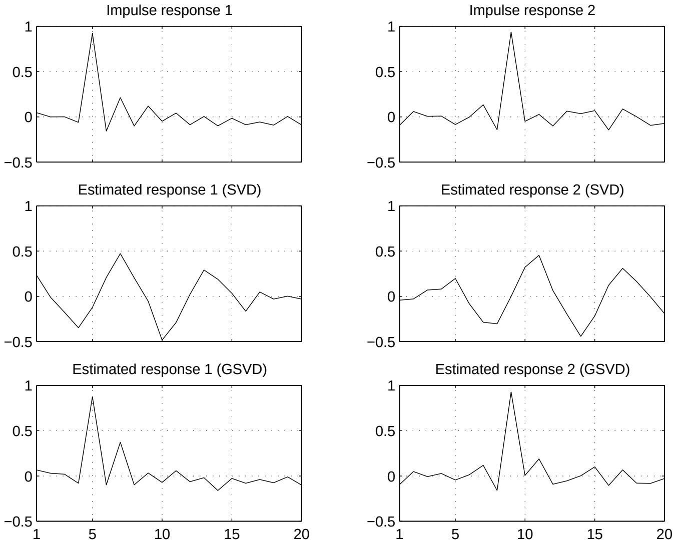


Figure 6.1: (**a**) Impulse responses $\mathbf{h}_0$ and $\mathbf{h}_1$, (**b**) Estimated impulse responses for SVD-procedure and (**c**) GSVD-procedure

6.3 Adaptive procedure for TDE

In practice, acoustic impulse responses may have thousands of taps, depending on the amount of room reverberation. Because of the (approximate) low-rank model of the speech signal, correspondingly large correlation matrices $\bar{\mathbf{R}}_{ss}[k]$ of the clean speech signal $s[k]$ will be rank-deficient or at least ill-conditioned [92][178]. Therefore it is quite difficult in practice to identify the complete acoustic impulse responses, especially when a large amount of background noise is present [101]. However, if we *underestimate* the length of the impulse responses ($L < K$), *the impulse responses estimated with the batch procedures are biased* and do not necessarily exhibit any resemblance to the actual impulse responses. This makes it difficult (and impossible in practice) to calculate the correct time-delays from these estimated impulse responses.

In [9] an adaptive EVD algorithm has been presented, which iteratively estimates the eigenvector corresponding to the smallest eigenvalue. *Remarkably*, even when underestimating the length of the acoustic impulse responses, simulations show that *this adaptive EVD algorithm is still able to identify the main peak of the impulse responses*, where it is assumed that this main peak corresponds to the first peak (direct path) of the acoustic impulse response. Obviously, for TDE only the time difference between the first peak of the acoustic impulse responses is required.

Strictly speaking, the adaptive EVD algorithm is only valid when no noise or when spatio-temporally white noise is present. In this section, we will therefore extend the adaptive EVD algorithm to the coloured noise case, by deriving a stochastic gradient algorithm for the procedures presented in Section 6.2.3, i.e. an algorithm which iteratively estimates the generalised eigenvector corresponding to the smallest generalised eigenvalue. Using simulations with spatio-temporally coloured noise, it will be shown that - as for the adaptive EVD algorithm - it is possible to correctly estimate the time-delays with the adaptive GEVD algorithm, even when underestimating the length of the acoustic impulse responses (cf. Section 6.5).

In the remainder of this chapter we assume that the length of the acoustic impulse responses is underestimated ($L < K$), and hence derive algorithms which estimate the one-dimensional subspace corresponding to the smallest (generalised) eigenvalue[1].

6.3.1 Adaptive EVD algorithm [9]

The eigenvector corresponding to the smallest eigenvalue of the empirical correlation matrix $\mathbf{R}_{yy}[k]$ can be iteratively estimated by minimising the cost function $\mathbf{v}^T\mathbf{R}_{yy}[k]\mathbf{v}$, subject to the constraint $\mathbf{v}^T\mathbf{v} = 1$. A cheap procedure consists in minimising the mean square value of the error signal $e[k]$, defined as

$$e[k] = \frac{\mathbf{v}^T[k]\,\mathbf{y}[k]}{\|\mathbf{v}[k]\|}\ , \tag{6.26}$$

with $\mathbf{y}[k] = \left[\begin{array}{cc} \mathbf{y}_{0,L}^T[k] & \mathbf{y}_{1,L}^T[k] \end{array}\right]^T$. This can e.g. be done using a gradient-descent LMS-procedure, where normalisation is included in each iteration step in order to avoid roundoff error propagation [218], i.e.

$$\mathbf{v}[k+1] = \frac{\mathbf{v}[k] - \mu e[k]\frac{\partial e[k]}{\partial \mathbf{v}[k]}}{\|\mathbf{v}[k] - \mu e[k]\frac{\partial e[k]}{\partial \mathbf{v}[k]}\|}\ , \tag{6.27}$$

with μ the step size of the adaptive algorithm. The gradient of $e[k]$ is equal to

$$\frac{\partial e[k]}{\partial \mathbf{v}[k]} = \frac{1}{\|\mathbf{v}[k]\|}\left(\mathbf{y}[k] - e[k]\frac{\mathbf{v}[k]}{\|\mathbf{v}[k]\|}\right)\ . \tag{6.28}$$

In [9] it has been assumed that the smallest eigenvalue of $\mathbf{R}_{yy}[k]$ is very small (in the noiseless case), such that the gradient eventually reduces to $\frac{\partial e[k]}{\partial \mathbf{v}[k]} \approx \mathbf{y}[k]$,

[1] It would also be possible to use the recursive updating procedures presented in Section 4.2. However, since we only need to estimate/update a one-dimensional subspace, namely the (generalised) eigenvector corresponding to the smallest (generalised) eigenvalue, stochastic gradient algorithms constitute a far less computationally complex alternative to updating the full (generalised) eigenvalue decomposition.

and the update formulas become

$$\boxed{\begin{array}{lcl} e[k] & = & \mathbf{v}^T[k]\,\mathbf{y}[k] \\ \mathbf{v}[k+1] & = & \dfrac{\mathbf{v}[k]-\mu e[k]\mathbf{y}[k]}{\|\mathbf{v}[k]-\mu e[k]\mathbf{y}[k]\|} \end{array}} \tag{6.29}$$

In [9] it has been indicated that a good initialisation of $\mathbf{v}$ and a proper choice of the parameters L and μ are essential for a good convergence behaviour. It has also been shown by simulations that the adaptive EVD algorithm performs more robustly in highly reverberant environments than GCC-based methods.

6.3.2 Adaptive GEVD and prewhitening algorithm

For the batch GEVD and prewhitening procedures, described in Section 6.2.3, it is also possible to derive stochastic gradient algorithms, which iteratively estimate the generalised eigenvector corresponding to the smallest generalised eigenvalue. It will be assumed that the empirical noise correlation matrix $\mathbf{R}_{vv}[k]$ (or its Cholesky-factor) is either known or is updated during noise-only periods. Since the noise correlation matrix can not be updated during speech-and-noise periods, we assume that the noise is stationary enough, such that the noise correlation matrix computed during noise-only periods can be used in the update formulas during subsequent speech-and-noise periods.

1. *Adaptive GEVD algorithm*: The generalised eigenvector corresponding to the smallest generalised eigenvalue of the empirical correlation matrices $\mathbf{R}_{yy}[k]$ and $\mathbf{R}_{vv}[k]$ can be iteratively estimated by minimising the cost function $\mathbf{q}^T\mathbf{R}_{yy}[k]\mathbf{q}$, subject to the constraint $\mathbf{q}^T\mathbf{R}_{vv}[k]\mathbf{q}=1$. A cheap procedure consists in minimising the mean square value of the error signal $e[k]$, defined as

$$e[k]=\frac{\mathbf{q}^T[k]\mathbf{y}[k]}{\sqrt{\mathbf{q}^T[k]\mathbf{R}_{vv}[k]\mathbf{q}[k]}}=\frac{\mathbf{q}^T[k]\mathbf{y}[k]}{\|\mathbf{R}_{vv}^{1/2}[k]\,\mathbf{q}[k]\|}\;, \tag{6.30}$$

which can e.g. be done using a gradient-descent LMS-procedure, i.e.

$$\mathbf{q}[k+1]=\mathbf{q}[k]-\mu e[k]\frac{\partial e[k]}{\partial \mathbf{q}[k]}\;, \tag{6.31}$$

with μ the step size of the adaptive algorithm. The gradient of $e[k]$ is equal to

$$\frac{\partial e[k]}{\partial \mathbf{q}[k]}=\frac{1}{\sqrt{\mathbf{q}^T[k]\mathbf{R}_{vv}[k]\mathbf{q}[k]}}\left(\mathbf{y}[k]-e[k]\frac{\mathbf{R}_{vv}[k]\mathbf{q}[k]}{\sqrt{\mathbf{q}^T[k]\mathbf{R}_{vv}[k]\mathbf{q}[k]}}\right)\;. \tag{6.32}$$

Substituting (6.30) and (6.32) into (6.31) gives

$$\mathbf{q}[k+1]=\mathbf{q}[k]-\frac{\mu}{\mathbf{q}^T[k]\mathbf{R}_{vv}[k]\mathbf{q}[k]}\Big(\mathbf{y}[k]\mathbf{y}^T[k]\mathbf{q}[k]-e^2[k]\mathbf{R}_{vv}[k]\mathbf{q}[k]\Big)\;, \tag{6.33}$$

such that, when taking expectation after convergence, we obtain

$$\bar{\mathbf{R}}_{yy}[k]\mathbf{q}[\infty] = \mathcal{E}\{e^2[k]\}\bar{\mathbf{R}}_{vv}[k]\mathbf{q}[\infty] \ . \tag{6.34}$$

This is exactly what is desired, i.e. $\mathbf{q}[\infty]$ is the generalised eigenvector corresponding to the smallest generalised eigenvalue of the correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$.

Since the smallest generalised eigenvalue is equal to 1 (cf. Section 6.2.3), we can not further simplify the expression in (6.33). In order to avoid roundoff error propagation, we include an additional normalisation in each iteration step, such that the update formulas can be written as

$$\boxed{\begin{array}{lcl} e[k] & = & \mathbf{q}^T[k]\mathbf{y}[k] \\ \mathbf{q}'[k+1] & = & \mathbf{q}[k] - \mu e[k]\big\{\mathbf{y}[k] - e[k]\mathbf{R}_{vv}[k]\mathbf{q}[k]\big\} \\ \mathbf{q}[k+1] & = & \dfrac{\mathbf{q}'[k+1]}{\sqrt{\mathbf{q}'^T[k+1]\mathbf{R}_{vv}[k]\mathbf{q}'[k+1]}} \end{array}} \tag{6.35}$$

2. *Adaptive prewhitening algorithm*: The batch prewhitening procedure can be made adaptive by using pre-whitened data vectors $\tilde{\mathbf{y}}[k] = \mathbf{R}_{vv}^{-T/2}[k]\,\mathbf{y}[k]$ in the adaptive EVD-procedure. The update formulas then become

$$\boxed{\begin{array}{lcl} e[k] & = & \tilde{\mathbf{v}}^T[k]\tilde{\mathbf{y}}[k] \\ \tilde{\mathbf{v}}[k+1] & = & \dfrac{\tilde{\mathbf{v}}[k] - \mu e[k]\big(\tilde{\mathbf{y}}[k] - e[k]\tilde{\mathbf{v}}[k]\big)}{\|\tilde{\mathbf{v}}[k] - \mu e[k]\big(\tilde{\mathbf{y}}[k] - e[k]\tilde{\mathbf{v}}[k]\big)\|} \end{array}} \tag{6.36}$$

Note that the gradient $\frac{\partial e[k]}{\partial \tilde{\mathbf{v}}[k]}$ can not be approximated by $\tilde{\mathbf{y}}[k]$ (as in the adaptive EVD algorithm), since the smallest eigenvalue of $\tilde{\mathbf{R}}_{yy}[k]$ is not equal to zero. The impulse response at time k is estimated as $\mathbf{R}_{vv}^{-1/2}[k]\,\tilde{\mathbf{v}}[k]$. If the empirical noise correlation matrix $\mathbf{R}_{vv}[k]$ is not known in advance, the Cholesky-factor $\mathbf{R}_{vv}^{-1/2}[k]$ can be updated by inverse QR-updating of the noise data matrix during noise-only periods.

The computational complexity of the adaptive GEVD and the adaptive prewhitening algorithm is higher than the complexity of the adaptive EVD algorithm, since in each iteration step two additional matrix-vector multiplications (with the noise correlation matrix or with the inverse Cholesky factor) have to be performed. Reducing the computational complexity of these algorithms is a topic of further research. One could e.g. replace the empirical noise correlation matrix in the adaptive GEVD algorithm by an instantaneous estimate $\mathbf{v}[k']\mathbf{v}^T[k']$, where $\mathbf{v}[k']$ is a noise data vector which is stored in a buffer during noise-only periods and which is used in the update equations during subsequent speech-and-noise periods. In addition, the computational complexity of all presented adaptive TDE algorithms can be reduced by using sub-sampling (cf. Section 4.2.4), i.e. the estimated impulse response vectors are not updated for every time step, at the expense of slower convergence and tracking.

6.4 Extension to more than two microphones

All previously presented (batch and adaptive) algorithms can be easily extended to the case of more than two microphones, either by constructing $P(N-1)\times NL$-dimensional data matrices, considering the time-delays between every microphone and the first microphone, or by constructing $P\,C_N^2 \times NL$-dimensional data matrices (with C_N^2 all possible combinations of 2 out of N, i.e. $C_N^2 = N(N-1)/2$), considering the time-delays between every combination of 2 microphones. E.g. if $N=3$, the speech data matrix $\mathbf{Y}[k]$ in (6.22) can be redefined by replacing each vector $\mathbf{y}^T[k]$ by the matrix

$$\begin{bmatrix} \mathbf{0} & \mathbf{y}_{0,L}^T[k] & \mathbf{y}_{1,L}^T[k] \\ \mathbf{y}_{0,L}^T[k] & \mathbf{0} & \mathbf{y}_{2,L}^T[k] \end{bmatrix}, \tag{6.37}$$

considering time-delays between every microphone and the first microphone, or by the matrix

$$\begin{bmatrix} \mathbf{0} & \mathbf{y}_{0,L}^T[k] & \mathbf{y}_{1,L}^T[k] \\ \mathbf{y}_{0,L}^T[k] & \mathbf{0} & \mathbf{y}_{2,L}^T[k] \\ \mathbf{y}_{1,L}^T[k] & -\mathbf{y}_{2,L}^T[k] & \mathbf{0} \end{bmatrix}, \tag{6.38}$$

considering time-delays between every combination of 2 microphones. The noise data matrix $\mathbf{V}[k]$ is constructed similarly. It can be easily verified that, if $L=K$ and for the noiseless case, the NK-dimensional vector consisting of the impulse responses

$$\mathbf{v} = \begin{bmatrix} -\mathbf{h}_{N-1} \\ \vdots \\ -\mathbf{h}_1 \\ \mathbf{h}_0 \end{bmatrix} \tag{6.39}$$

belongs to the null-space of the speech data matrix. Therefore all previously presented (batch and adaptive) algorithms can be used with the redefined data matrices and data vectors. For the adaptive algorithms, several updates now need to be performed in each iteration step, either with $N-1$ or with C_N^2 data vectors. However, the computational complexity can be reduced by only performing an update with one data vector in each iteration step, i.e. by using consecutive rows of the matrices (6.37) or (6.38) in each iteration step.

6.5 Simulations

We have performed several simulations, analysing the performance of the different adaptive TDE algorithms (EVD, GEVD, prewhitening) for different reverberation conditions (ideal and realistic), different SNRs and different microphone configurations. In all simulations the sampling frequency $f_s = 16\,\text{kHz}$ and

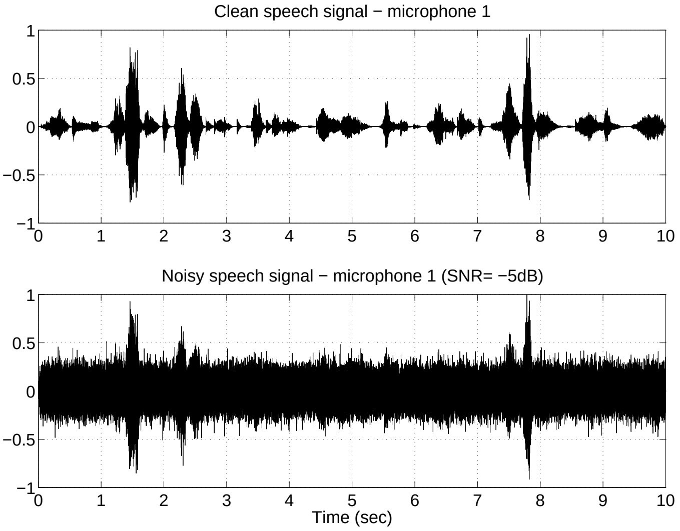


Figure 6.2: (**a**) Speech component $x_0[k]$, (**b**) Noisy microphone signal $y_0[k]$ (SNR=-5 dB)

the length of the used signals is 160000 samples (10 sec). We have used a continuous clean speech signal (see Fig. 6.2a), such that no VAD is required and we can continuously estimate the time-delays. We have calculated the noise correlation matrix $\mathbf{R}_{vv}[k]$ in advance from the noise components $v_n[k]$ of the microphone signals, which are assumed to be known. The time-delay between the different microphone signals is computed using the peak of the correlation function between the different estimated acoustic impulse responses.

6.5.1 No reverberation, 2-microphone case

In a first simulation, we have assumed no reverberation and $N = 2$ microphones. We have used a coloured noise signal, constructed by filtering white noise with the five-tap FIR filter $\begin{bmatrix} 1 & -4 & 6 & 4 & 0.5 \end{bmatrix}$. The microphone signals are constructed such that the time-delay between the speech components is -8 samples, whereas the time-delay between the noise components is 5 samples. We have performed simulations using the adaptive EVD, prewhitening and GEVD algorithms for different SNRs (-5 dB, 0 dB, 5 dB). The used filter length $L = 40$, the sub-sampling factor for the update formulas is 10 and the step size μ of the adaptive algorithms is chosen such that the optimal performance is obtained, i.e. such that most of the estimated time-delays are close to the correct time-delay (in this case $\mu = 1e - 7$ for all algorithms).

Figure 6.3 shows the TDE convergence plots for the different adaptive algorithms for different SNRs. The correct time-delay is indicated by the dashed

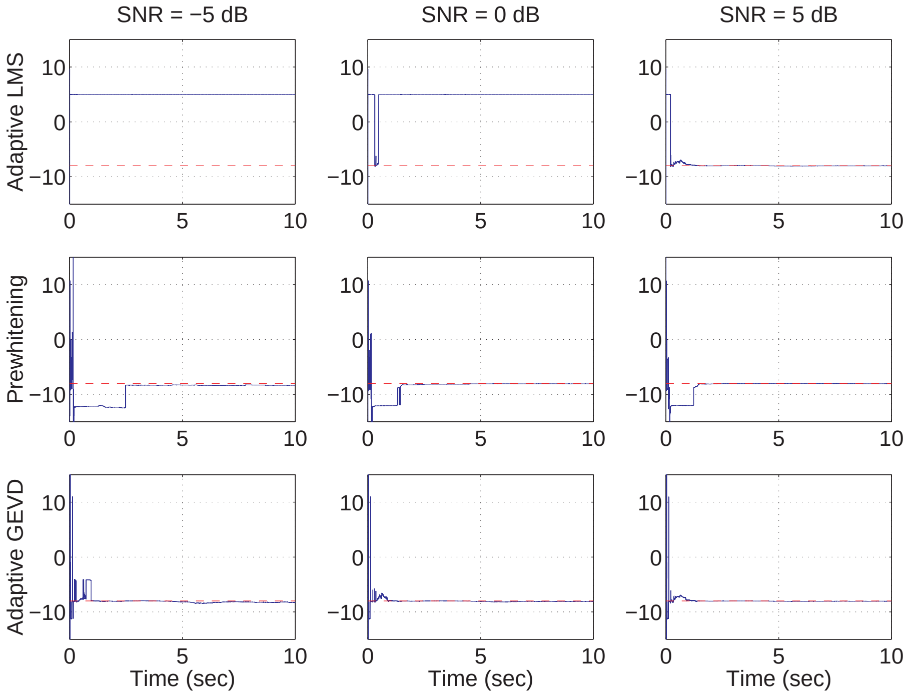


Figure 6.3: TDE convergence plots for adaptive EVD, prewhitening and GEVD algorithms for different SNRs without reverberation ($N = 2$, $L = 40$, subsampling $= 10$, $\mu = 1e - 7$)

line. As can be seen, the adaptive EVD algorithm converges to the correct time-delay for SNR $= 5$ dB, but converges to the (wrong) time-delay of the noise source for lower SNRs. Both the adaptive prewhitening and the adaptive GEVD algorithm converge to the correct time-delay for all SNRs. The adaptive GEVD algorithm converges faster than the adaptive prewhitening algorithm.

6.5.2 Realistic conditions, 2-microphone case

In order to simulate realistic reverberation conditions, we have simulated a room with dimensions 5 m × 4 m × 2 m, having a reverberation time $T_{60} = 250$ msec. The room consists of a microphone array with $N = 2$ omni-directional microphones at positions $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1.5 & 1 & 1 \end{bmatrix}$, a speech source at position $\begin{bmatrix} 2 & 2 & 1.7 \end{bmatrix}$ and a noise source at position $\begin{bmatrix} 4 & 1.5 & 1 \end{bmatrix}$. The used noise signal is stationary speech noise. The speech and the noise components of the nth microphone signal are filtered versions of the clean speech and noise signals with simulated acoustic impulse responses, constructed using the image method (cf. Section 1.3.3) with $K = 1000$. The two acoustic impulse responses

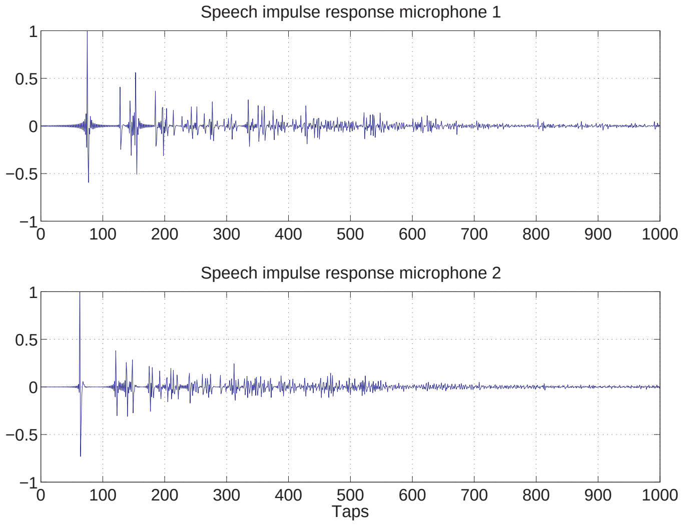


Figure 6.4: Acoustic impulse responses for the speech source

for the speech source are plotted in Fig. 6.4. The exact time-delay between the speech components of the microphone signals is −12.18 samples, which has been obtained by a simple geometrical calculation.

We have performed simulations using the adaptive EVD, prewhitening and GEVD algorithms for different SNRs (−5 dB, 0 dB) and different sub-sampling factors (1, 10). The noisy microphone signal $y_0[k]$ with SNR = −5 dB is plotted in Fig. 6.2b. The used filter length $L = 40$ and for each algorithm we have chosen the step size μ which gives rise to the best result.

Figure 6.5 shows the TDE convergence plots for SNR = −5 dB and sub-sampling factor 1, i.e. no sub-sampling. The correct time-delay is indicated by the dashed line. As can be seen, the adaptive EVD algorithm does not converge to the correct time-delay (except for the signal segment between 1.5 and 3 sec, where the segmental SNR is quite high, see Fig. 6.2), whereas both the adaptive prewhitening and the adaptive GEVD algorithm converge to the correct time-delay.

Figure 6.6 shows the TDE convergence plots for SNR = −5 dB and sub-sampling factor 10 (i.e. the estimated impulse responses are updated every 10 samples). Again, the adaptive EVD algorithm does not converge to the correct time-delay, whereas both the adaptive prewhitening and the adaptive GEVD algorithm converge to the correct time-delay. By comparing Fig. 6.5 and 6.6, it can be observed that the adaptive prewhitening and the adaptive GEVD algorithms exhibit a slower convergence for sub-sampling factor 10 than for sub-sampling factor 1.

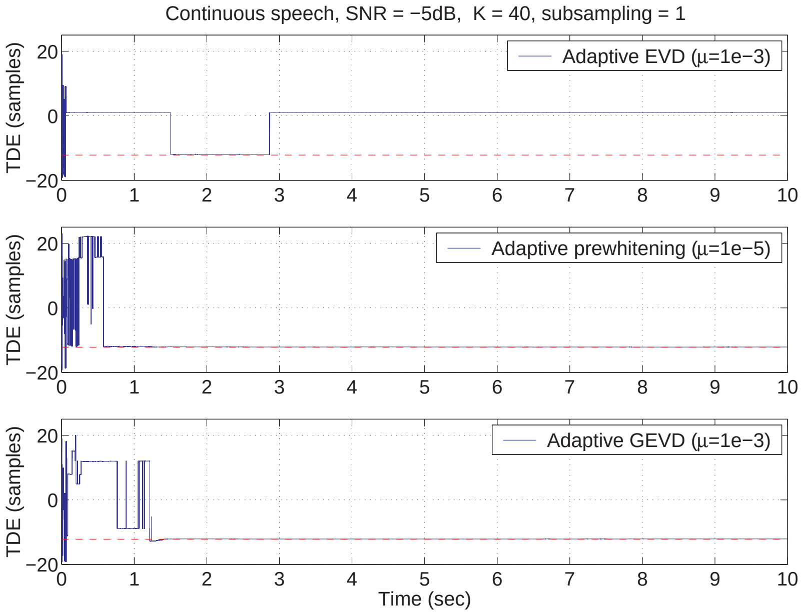


Figure 6.5: TDE convergence plots of adaptive EVD, prewhitening and GEVD algorithms ($N = 2$, $L = 40$, SNR $= -5$ dB, $T_{60} = 250$ msec, sub-sampling $= 1$)

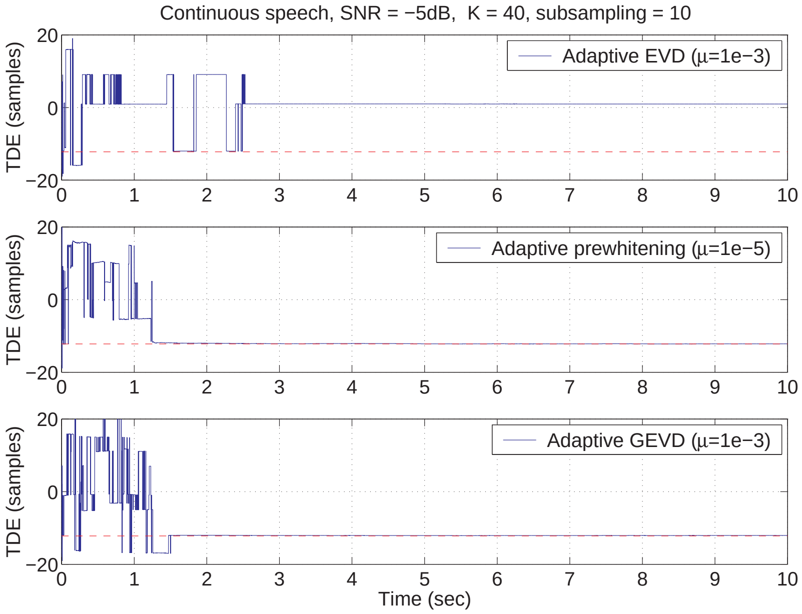


Figure 6.6: TDE convergence plots of adaptive EVD, prewhitening and GEVD algorithms ($N = 2$, $L = 40$, SNR $= -5$ dB, $T_{60} = 250$ msec, sub-sampling $= 10$)

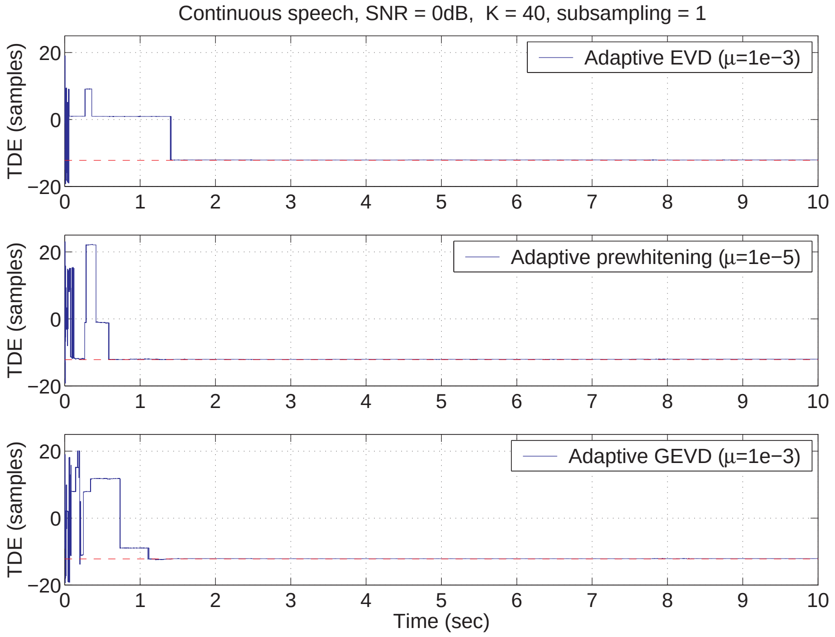


Figure 6.7: TDE convergence plots of adaptive EVD, prewhitening and GEVD algorithms ($N = 2$, $L = 40$, SNR $= 0\,$dB, $T_{60} = 250\,$msec, sub-sampling $= 1$)

Figure 6.7 shows the TDE convergence plots for SNR $= 0\,$dB and sub-sampling factor 1. In this case all algorithms converge to the correct time-delay, but both the adaptive prewhitening and the adaptive GEVD algorithm converge faster than the adaptive EVD algorithm. Note that it is still quite remarkable that the adaptive EVD algorithm converges to the correct time-delay for SNR $=$ $0\,$dB, without any knowledge of the noise characteristics. This can be partly explained by the room reverberation, which approximately turns the noise field into a diffuse sound field.

6.5.3 Realistic conditions, 3-microphone case

For the same acoustical conditions as in Section 6.5.2, we have performed simulations using $N = 3$ microphones, where the position of the third microphone is $\left[\begin{array}{ccc} 1 & 1 & 1.5 \end{array}\right]$. We consider the time-delays between every combination of 2 microphones and in each iteration step we have performed updates using all three data vectors from (6.38). The exact time-delay between the speech components of the first and the second microphone signal is -12.18 samples, between the first and the third microphone signal -7.04 samples, and between the second and the third microphone signal 5.14 samples. We have performed simulations for different SNRs ($-5\,$dB, $0\,$dB), the used filter length $L = 40$, the

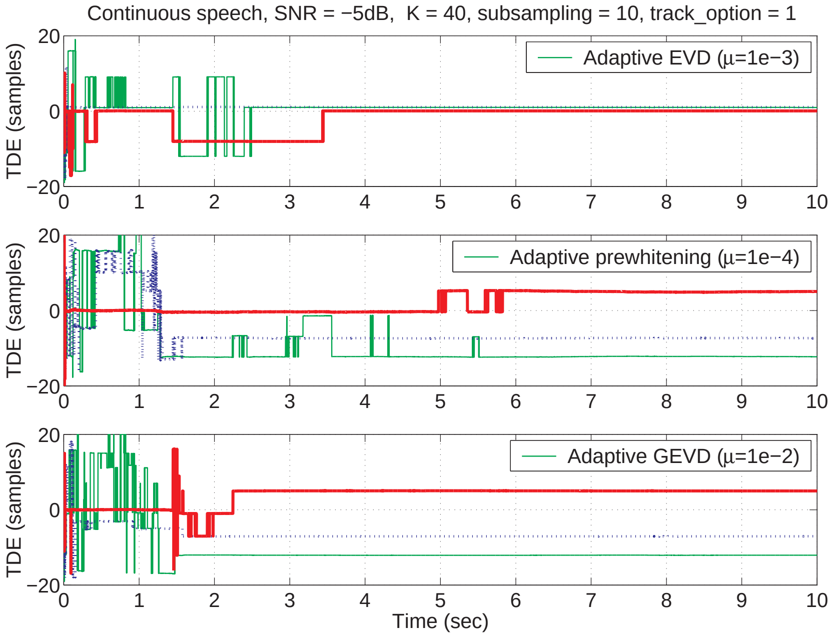


Figure 6.8: TDE convergence plots of adaptive EVD, prewhitening and GEVD algorithms ($N = 3$, $L = 40$, SNR $= -5$ dB, $T_{60} = 250$ msec, sub-sampling $=$ 10). TDE mic1-mic2 solid line, TDE mic1-mic3 dotted line, TDE mic2-mic3 thick solid line.

sub-sampling factor is 1 and for each algorithm we have chosen the step size μ which gives rise to the best result.

Figure 6.8 shows the TDE convergence plots for SNR $= -5$ dB. As can be seen, the adaptive EVD algorithm does not converge to the correct time-delays, whereas both the adaptive prewhitening and the adaptive GEVD algorithm converge to the correct time-delays. The adaptive GEVD algorithm exhibits a better and a faster convergence than the adaptive prewhitening algorithm.

Figure 6.9 shows the TDE convergence plots for SNR $=$ 0 dB. In this case all algorithms converge to the correct time-delays, although the time-delay between the second and the third microphone signal is only correctly estimated by the adaptive EVD algorithm in signal segments with a high segmental SNR.

From these simulations, we can conclude that for all SNRs, sub-sampling factors and microphone configurations, the adaptive prewhitening and the adaptive GEVD algorithms converge more robustly to the correct time-delays than the adaptive EVD algorithm, certainly in low SNR conditions. In addition, the adaptive GEVD algorithm exhibits a slightly better and faster convergence than the adaptive prewhitening algorithm.

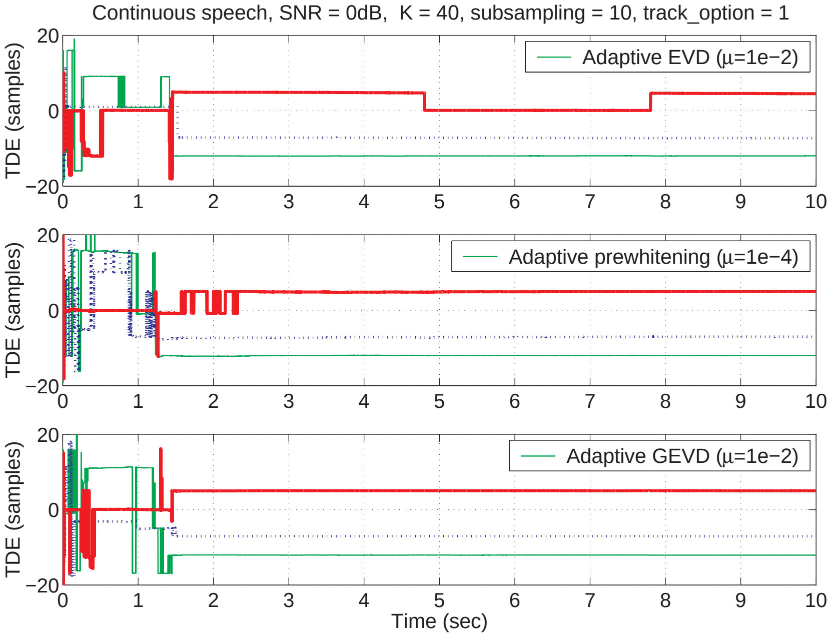


Figure 6.9: TDE convergence plots of adaptive EVD, prewhitening and GEVD algorithms ($N = 3$, $L = 40$, SNR = 0 dB, $T_{60} = 250$ msec, sub-sampling = 10). TDE mic1-mic2 solid line, TDE mic1-mic3 dotted line, TDE mic2-mic3 thick solid line.

6.6 Conclusion

In this chapter we have presented two adaptive algorithms for robust TDE in adverse acoustic environments, where a large amount of reverberation and additive noise is present. We have extended a recently developed adaptive EVD algorithm for TDE to noisy environments, by using an adaptive GEVD or by prewhitening the microphone signals. For the adaptive GEVD, we have derived a stochastic gradient algorithm which iteratively estimates the generalised eigenvector corresponding to the smallest generalised eigenvalue. In addition, we have extended all TDE algorithms to the case of more than two microphones. It has been shown by simulations that for all conditions the time-delays can be estimated more robustly using the adaptive GEVD algorithm than using the adaptive EVD algorithm and the adaptive prewhitening algorithm.

Chapter 7

Combined noise reduction and dereverberation

In this chapter a combined frequency-domain noise reduction and dereverberation technique is discussed which produces an MMSE estimate of the clean dereverberated speech signal. It is shown that this combined technique provides a trade-off between the noise reduction and the dereverberation objectives.

Section 7.1 gives a brief introduction of the dereverberation and the combined noise reduction and dereverberation problem.

Section 7.2 describes a frequency-domain technique for estimating the acoustic transfer functions from the microphone signals which are corrupted by spatially coloured noise. This technique is an extension of the frequency-domain technique presented in [2], which is only optimal in the case of spatially white noise. However, unlike the time-domain techniques presented in the previous chapter, these frequency-domain techniques require some prior knowledge about the acoustic transfer functions.

In Section 7.3 it is shown that using the estimated acoustic transfer functions, dereverberation can be performed with a normalised matched filtering approach. It is also shown that the MMSE estimate of the clean dereverberated speech signal can be obtained by matched filtering of the MMSE estimates of the speech components in the microphone signals. Hence, by integrating the normalised matched filter with the multi-channel Wiener filter, discussed in Chapter 3, we obtain a combined noise reduction and dereverberation technique. Since both algorithms essentially require the same decomposition, i.e. a GSVD of a speech and a noise data matrix, they can easily be combined.

Section 7.4 discusses some practical implementation issues. Since essentially a convolution in the frequency-domain is performed, the corresponding time-domain filters need to be constrained in order to avoid circular convolutions.

Section 7.5 describes the simulation results, showing that the GSVD-based noise reduction technique yields the best SNR and the GSVD-based dereverberation technique has the best dereverberation performance, while the combined noise reduction and dereverberation technique provides a trade-off between both objectives.

7.1 Introduction

As already indicated in Section 2.2.2, the objective of multi-microphone signal enhancement can be either noise reduction (not caring about residual reverberation), dereverberation (not caring about residual noise), or combined noise reduction and dereverberation. For *dereverberation* the total speech transfer function should be equal to 1 (or more realistically a delay), while for *combined noise reduction and dereverberation* the total speech transfer function should approximate a delay and the energy of the residual noise should be minimised at the same time.

Many multi-microphone dereverberation algorithms require an estimate of the acoustic impulse responses, either in the time-domain or in the frequency-domain [2][91][101][117][180]. By using the batch or the adaptive estimation techniques discussed in Chapter 6, a time-domain estimate of the acoustic impulse responses can be obtained. This estimate can then be used in an inverse filtering or a matched filtering algorithm for multi-microphone dereverberation, cf. Section 2.6. However, as already indicated in Section 6.3, because of the length of the acoustic impulse responses and the low-rank model of the speech signal, it is quite difficult in practice to identify the complete acoustic impulse responses, especially when a large amount of background noise is present. Moreover, time-domain subspace techniques appear to be quite sensitive to underestimation of the length of the acoustic impulse responses, such that the length of the acoustic impulse responses needs to be known in advance, which is often not possible in practice.

Hence, frequency-domain techniques have been proposed for estimating the acoustic transfer functions. In [2] a procedure has been proposed for identifying and tracking the acoustic transfer functions in the frequency-domain. Although frequency-domain estimation techniques have the advantage to be less sensitive to order estimation errors, an (unknown) scaling ambiguity arises in each frequency bin. In order to eliminate this scaling ambiguity, prior knowledge about the acoustic transfer functions is required, which clearly is a disadvantage and which limits the practical use of these frequency-domain estimation techniques.

Strictly speaking, the procedure in [2] is only valid when spatially white noise is present. In Section 7.2 we extend this procedure to the spatially coloured noise case. In Section 7.3 it is shown that after estimating the acoustic transfer functions, dereverberation can be performed with a normalised matched filtering approach. It is also shown that combined noise reduction and dereverberation can be performed by integrating this matched filtering approach for dereverberation with the multi-channel Wiener filter for noise reduction. Note that both for dereverberation and for combined noise reduction and dereverberation again some prior knowledge about the acoustic transfer functions is required. Eliminating this need for prior knowledge is a topic of further research.

7.2 Estimation of acoustic transfer functions

In this section a frequency-domain technique is presented for estimating the acoustic transfer functions when spatially coloured noise is present. Moreover, in the spatially white noise case a computationally efficient subspace tracking algorithm can be used for estimating and tracking the acoustic transfer functions [2]. It is however not trivial to extend this subspace tracking algorithm to the coloured noise case.

7.2.1 Frequency-domain signal model

Consider again Fig. 2.1, depicting a microphone array which records a speech source and background noise. In the frequency-domain, the stacked vector of microphone signals $\mathbf{Y}(\omega)$ can be written as (2.27), i.e.

$$\begin{aligned} \mathbf{Y}(\omega) &= \mathbf{H}(\omega)S(\omega) + \mathbf{V}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega) && (7.1) \\ &= \begin{bmatrix} H_0(\omega) \\ H_1(\omega) \\ \vdots \\ H_{N-1}(\omega) \end{bmatrix} S(\omega) + \begin{bmatrix} V_0(\omega) \\ V_1(\omega) \\ \vdots \\ V_{N-1}(\omega) \end{bmatrix}, && (7.2) \end{aligned}$$

with $H_n(\omega)$ the acoustic transfer function between the speech source and the nth microphone. Although we assume here that the acoustic transfer functions $H_n(\omega)$ are time-invariant, in Section 7.2.3 a subspace tracking algorithm is discussed which is able to track time-variations of the acoustic transfer functions.

Using (2.31), the output signal $Z(\omega)$ of a multi-microphone signal enhancement algorithm can be written as

$$Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega) = \underbrace{\mathbf{W}^H(\omega)\mathbf{H}(\omega)}_{F(\omega)} S(\omega) + \mathbf{W}^H(\omega)\mathbf{V}(\omega)\ , \qquad (7.3)$$

with $F(\omega)$ the total speech transfer function and

$$\mathbf{W}(\omega) = \left[\begin{array}{cccc} W_0(\omega) & W_1(\omega) & \dots & W_{N-1}(\omega) \end{array} \right]^T . \tag{7.4}$$

The filter $\mathbf{W}(\omega)$ can be designed with different objectives in mind:

- The objective of *dereverberation* is to compute the filter $\mathbf{W}_d(\omega)$ such that the total speech transfer function $F(\omega) = \mathbf{W}_d^H(\omega)\mathbf{H}(\omega) = 1$ (or a delay). Clearly, the normalised matched filter $\mathbf{W}_d(\omega) = \mathbf{H}(\omega)/\|\mathbf{H}(\omega)\|^2$ is a possible solution (cf. Section 7.3.1).
- The multi-channel Wiener filter $\mathbf{W}_{WF}(\omega)$ discussed in Section 3.5 produces an MMSE estimate of the speech component $X_n(\omega)$ in one (or all) of the microphone signals and can therefore be used for *noise reduction*, but not for dereverberation (cf. Section 7.3.2).
- The goal of *combined noise reduction and dereverberation* is to compute the filter $\mathbf{W}_c(\omega)$ such that the output signal $Z(\omega)$ is the MMSE estimate of the clean speech signal $S(\omega)$, thereby both reducing reverberation and background noise, but also introducing some speech distortion (cf. Section 7.3.3).

Both for dereverberation and for combined noise reduction and dereverberation, an estimate of the acoustic transfer function vector $\mathbf{H}(\omega)$ is required, cf. Section 7.3. This section discusses a frequency-domain technique for estimating $\mathbf{H}(\omega)$ without any knowledge of the speech signal $S(\omega)$. This frequency-domain estimation technique is quite similar to the batch time-domain estimation technique discussed in Section 6.2.3, now using the generalised eigenvector corresponding to the *largest* generalised eigenvalue of the frequency-domain speech and noise correlation matrices.

Using (7.1) and assuming that the speech and the noise components are uncorrelated, the $N \times N$-dimensional frequency-domain speech correlation matrix $\bar{\mathbf{R}}_{yy}(\omega) = \mathcal{E}\{\mathbf{Y}(\omega)\mathbf{Y}^H(\omega)\}$ is equal to

$$\bar{\mathbf{R}}_{yy}(\omega) = \bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}(\omega) = P_s(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega) + \bar{\mathbf{R}}_{vv}(\omega) , \tag{7.5}$$

with $P_s(\omega) = \mathcal{E}\{|S(\omega)|^2\}$. In case of a single speech source, the correlation matrix $\bar{\mathbf{R}}_{xx}(\omega)$ has rank 1. The noise correlation matrix $\bar{\mathbf{R}}_{vv}(\omega)$ can be estimated during noise-only periods and reduces to $\bar{\sigma}_v^2(\omega)\,\mathbf{I}_N$ for spatially white noise.

The transfer function vector $\mathbf{H}(\omega)$ can be computed using the GEVD of the speech and the noise correlation matrices $\bar{\mathbf{R}}_{yy}(\omega)$ and $\bar{\mathbf{R}}_{vv}(\omega)$, cf. (6.15),

$$\left\{ \begin{array}{lcl} \bar{\mathbf{R}}_{yy}(\omega) & = & \bar{\mathbf{Q}}(\omega)\,\bar{\mathbf{\Lambda}}_y(\omega)\,\bar{\mathbf{Q}}^H(\omega) \\ \bar{\mathbf{R}}_{vv}(\omega) & = & \bar{\mathbf{Q}}(\omega)\,\bar{\mathbf{\Lambda}}_v(\omega)\,\bar{\mathbf{Q}}^H(\omega) , \end{array} \right. \tag{7.6}$$

with $\bar{\mathbf{Q}}(\omega)$ an $N \times N$-dimensional invertible, but not necessarily orthogonal matrix and $\bar{\mathbf{\Lambda}}_y(\omega)$ and $\bar{\mathbf{\Lambda}}_v(\omega)$ diagonal matrices. Since the correlation matrix

$$\bar{\mathbf{R}}_{xx}(\omega) = \bar{\mathbf{R}}_{yy}(\omega) - \bar{\mathbf{R}}_{vv}(\omega) = \bar{\mathbf{Q}}(\omega)\left[\bar{\mathbf{\Lambda}}_y(\omega) - \bar{\mathbf{\Lambda}}_v(\omega)\right]\bar{\mathbf{Q}}^H(\omega) \tag{7.7}$$

has rank 1, it is equal to $\bar{\mathbf{R}}_{xx}(\omega) = \bar{\sigma}_x^2(\omega)\bar{\mathbf{q}}(\omega)\bar{\mathbf{q}}^H(\omega)$, with $\bar{\mathbf{q}}(\omega)$ the N-dimensional principal generalised eigenvector, corresponding to the largest generalised eigenvalue. Using (7.5), $\bar{\mathbf{R}}_{xx}(\omega)$ can be written as

$$\bar{\mathbf{R}}_{xx}(\omega) = P_s(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega) = \bar{\sigma}_x^2(\omega)\bar{\mathbf{q}}(\omega)\bar{\mathbf{q}}^H(\omega)\ , \tag{7.8}$$

such that the vector $\mathbf{H}(\omega)$ can be estimated up to a phase shift $e^{j\phi(\omega)}$ as

$$\boxed{\mathbf{H}(\omega) = \frac{\|\mathbf{H}(\omega)\|}{\|\bar{\mathbf{q}}(\omega)\|}\bar{\mathbf{q}}(\omega)e^{j\phi(\omega)}} \tag{7.9}$$

We will assume that the human auditory system is not very sensitive to this phase shift. As can be seen from (7.9), the vector $\mathbf{H}(\omega)$ can only be estimated up to a frequency-dependent scaling factor, resulting in an ambiguity which can only be resolved if the norm $\|\mathbf{H}(\omega)\|$ is known. Hence, unlike the time-domain estimation techniques presented in Section 6.2, where the (unknown) scaling factor is frequency-independent and hence irrelevant, frequency-domain subspace-based estimation techniques require some prior knowledge about the acoustic transfer functions.

7.2.2 Practical computation

In practice, the continuous speech and noise spectra $Y_n(\omega)$ and $V_n(\omega)$ are approximated by their DFT-components, which can be efficiently computed using an FFT algorithm (cf. Section 2.1). The lth (frequency-)component of the DFT of the mth frame of $y_n[k]$ is equal to

$$Y_n(l,m) = \sum_{k=0}^{L-1} y_n[mL+k]\, e^{-j2\pi kl/L}\ , \quad l = 0 \ldots L-1\ , \tag{7.10}$$

with L the size of the DFT and for the time being considering no overlap between the frames. The stacked vector of microphone signals $\mathbf{Y}(l,m)$ is equal to

$$\mathbf{Y}(l,m) = \begin{bmatrix} Y_0(l,m) & Y_1(l,m) & \ldots & Y_{N-1}(l,m) \end{bmatrix}^T\ . \tag{7.11}$$

If we assume that $X_n(l,m) = H_n(l)S(l,m)$, which is actually only true when $L \to \infty$, then the vector $\mathbf{Y}(l,m)$ can be written as

$$\mathbf{Y}(l,m) = \mathbf{X}(l,m)+\mathbf{V}(l,m) = \mathbf{H}(l)S(l,m)+\mathbf{V}(l,m)\,, \quad l = 0 \ldots L-1\ . \tag{7.12}$$

The acoustic transfer function vector $\mathbf{H}(l)$ for the lth frequency-component can now be estimated (up to a phase shift) as the generalised singular vector

$\mathbf{q}(l,m)$, corresponding to the largest generalised singular vector in the GSVD of the frequency-domain speech and the noise data matrices $\mathcal{Y}(l,m)$ and $\mathcal{V}(l,m)$, which are defined as

$$\mathcal{Y}(l,m) = \begin{bmatrix} \mathbf{Y}^H(l,m-P+1) \\ \vdots \\ \mathbf{Y}^H(l,m-1) \\ \mathbf{Y}^H(l,m) \end{bmatrix} \quad \mathcal{V}(l,m) = \begin{bmatrix} \mathbf{V}^H(l,m-Q+1) \\ \vdots \\ \mathbf{V}^H(l,m-1) \\ \mathbf{V}^H(l,m) \end{bmatrix} .$$

7.2.3 White noise case: subspace tracking algorithm

In the spatially white noise case, the matrix $\bar{\mathbf{Q}}(\omega)$ is orthogonal, and the acoustic transfer function vector $\mathbf{H}(\omega)$ can be estimated from the principal eigenvector $\bar{\mathbf{q}}(\omega)$ of $\bar{\mathbf{R}}_{yy}(\omega)$, corresponding to its largest eigenvalue. Since $\|\bar{\mathbf{q}}(\omega)\| = 1$, the expression in (7.9) reduces to

$$\mathbf{H}(\omega) = \|\mathbf{H}(\omega)\|\, \bar{\mathbf{q}}(\omega) e^{j\phi(\omega)} . \tag{7.13}$$

For the practical computation of $\mathbf{q}(l,m)$, a subspace tracking procedure can be used which adaptively estimates and tracks the principal singular vector of $\mathcal{Y}(l,m)$, corresponding to its largest singular value[1]. In the literature different subspace tracking procedures have been proposed [41][199][281]. In [281], where the PAST (Projection Approximation Subspace Tracking) algorithm is derived, it has been shown that the vector $\mathbf{q}(l)$, minimising the cost function

$$J\big(\mathbf{q}(l)\big) = \mathcal{E}\{\|\mathbf{Y}(l) - \mathbf{q}(l)\mathbf{q}^H(l)\mathbf{Y}(l)\|_F^2\} \tag{7.14}$$

is equal to the principal eigenvector of $\bar{\mathbf{R}}_{yy}(l) = \mathcal{E}\{\mathbf{Y}(l)\mathbf{Y}^H(l)\}$. By using a gradient-descent procedure and approximating the gradient by its instantaneous value, the following adaptive subspace tracking algorithm is obtained

$$\begin{aligned} z(l,m) &= \mathbf{q}^H(l,m)\mathbf{Y}(l,m) && (7.15) \\ \mathbf{q}(l,m+1) &= \mathbf{q}(l,m) + \mu\left[2\mathbf{Y}(l,m) - \mathbf{Y}(l,m)\mathbf{q}^H(l,m)\mathbf{q}(l,m) - \right. \\ &\qquad \left. \mathbf{q}(l,m)z(l,m)\right] z^*(l,m) , && (7.16) \end{aligned}$$

with μ the step size of the adaptive algorithm. By additionally assuming that $\mathbf{q}^H(l,m)\mathbf{q}(l,m) = 1$, which is true upon convergence, the expression in (7.16) reduces to

$$\boxed{\mathbf{q}(l,m+1) = \mathbf{q}(l,m) + \mu\left[\mathbf{Y}(l,m) - \mathbf{q}(l,m)z(l,m)\right] z^*(l,m)} \tag{7.17}$$

[1] Note that in this case we can not use the adaptive (time-domain) algorithms presented in Section 6.3, since these algorithms estimate and track the subspace corresponding to the *smallest* (generalised) singular value. For estimating the acoustic transfer function vector in the frequency-domain, we require an adaptive algorithm which tracks the subspace corresponding to the *largest* (generalised) singular value.

which is equal to Oja's learning rule [199] and which strongly resembles the LMS-algorithm, cf. (2.160). The computational complexity of this adaptive subspace tracking algorithm is only $\mathcal{O}(N)$.

Strictly speaking, the described subspace tracking procedure is only valid for spatially white noise. For the spatially coloured noise case, the full GSVD of $\mathcal{Y}(l,m)$ and $\mathcal{V}(l,m)$ needs to be updated, which can e.g. be done using the recursive GSVD-updating algorithms discussed in Section 4.2.2. However, the computational complexity of these recursive (full) GSVD-updating algorithms is much larger than the complexity of the subspace tracking procedure in (7.15) and (7.17). Extending this adaptive subspace tracking procedure to the coloured noise case remains a topic of further research.

7.3 Noise reduction and dereverberation

Using the estimated acoustic transfer function vector $\mathbf{H}(\omega)$, it is shown in Section 7.3.1 that dereverberation can be performed with a normalised matched filtering approach. In Section 7.3.2 the multi-channel Wiener filter for noise reduction is briefly reviewed. In Section 7.3.3 it is shown that combined noise reduction and dereverberation can be performed by integrating the normalised matched filter for dereverberation with the multi-channel Wiener filter.

7.3.1 Speech dereverberation

As has been shown in Section 7.2.1, a possible dereverberation filter $\mathbf{W}_d(\omega)$ is the *normalised matched filter* $\mathbf{W}_d(\omega) = \mathbf{H}(\omega)/\|\mathbf{H}(\omega)\|^2$. Using (7.9) and assuming $\phi(\omega) = 0$, this filter can be computed using the principal generalised eigenvector $\bar{\mathbf{q}}(\omega)$ as

$$\boxed{\mathbf{W}_d(\omega) = \frac{\bar{\mathbf{q}}(\omega)}{\|\bar{\mathbf{q}}(\omega)\|\|\mathbf{H}(\omega)\|}} \tag{7.18}$$

However, as can be seen from this expression, prior knowledge about the acoustic transfer functions, i.e. the norm $\|\mathbf{H}(\omega)\|$, is required for computing $\mathbf{W}_d(\omega)$. Although it has been indicated in [2] that $\|\mathbf{H}(\omega)\|$ is less affected by small speaker movements than the individual transfer functions $H_n(\omega)$, this norm will nevertheless drastically change when the speaker moves around in the room. Hence, the practical use of this dereverberation algorithm is limited to e.g. desktop or car applications, where the speaker position is roughly fixed and $\|\mathbf{H}(\omega)\|$ can be measured beforehand. Using this normalised matched filtering approach, the output signal $Z_d(\omega)$ is equal to

$$Z_d(\omega) = \mathbf{W}_d^H(\omega)\mathbf{Y}(\omega) = S(\omega) + \frac{\bar{\mathbf{q}}^H(\omega)}{\|\bar{\mathbf{q}}(\omega)\|\|\mathbf{H}(\omega)\|}\mathbf{V}(\omega)\,. \tag{7.19}$$

As can be seen, the speech component of the output signal $Z_d(\omega)$ is equal to the clean dereverberated signal $S(\omega)$. However, since no attention has been given to the residual noise component, it is possible that the noise components of the microphone signals are even amplified by the dereverberation filter $\mathbf{W}_d(\omega)$. In the case of spatially white noise, the matrix $\bar{\mathbf{Q}}(\omega)$ is orthogonal, such that $\|\bar{\mathbf{q}}(\omega)\| = 1$ and the filter $\mathbf{W}_d(\omega)$ reduces to

$$\mathbf{W}_d^w(\omega) = \frac{\bar{\mathbf{q}}(\omega)}{\|\mathbf{H}(\omega)\|} \ . \tag{7.20}$$

Among all filters $\mathbf{W}_d(\omega)$ which perform perfect dereverberation, i.e. $F(\omega) = 1$, the filter which leads to the smallest residual noise energy is given by

$$\min_{\mathbf{W}_d(\omega)} \mathbf{W}_d^H(\omega)\bar{\mathbf{R}}_{vv}(\omega)\mathbf{W}_d(\omega), \quad \text{subject to } F(\omega) = \mathbf{W}_d^H(\omega)\mathbf{H}(\omega) = 1 \ . \tag{7.21}$$

This problem formulation is very similar to the superdirective or the MVDR beamformer formulation (cf. Section 2.5.3), now using the actual acoustic transfer function vector $\mathbf{H}(\omega)$ instead of the steering vector $\mathbf{d}(\omega, \theta_x)$ for free-field conditions. Similarly as the derivation in Appendix B.3, it can be shown that the solution of this optimisation problem is given by

$$\tilde{\mathbf{W}}_d(\omega) = \frac{\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)}{\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)} \ . \tag{7.22}$$

Using (7.9), this filter can be computed as

$$\boxed{\tilde{\mathbf{W}}_d(\omega) = \frac{\|\bar{\mathbf{q}}(\omega)\|}{\|\mathbf{H}(\omega)\|} \frac{\bar{\mathbf{R}}_{vv}^{-1}(\omega)\bar{\mathbf{q}}(\omega)}{\bar{\mathbf{q}}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\bar{\mathbf{q}}(\omega)}} \tag{7.23}$$

again requiring prior knowledge of the norm $\|\mathbf{H}(\omega)\|$.

7.3.2 Noise reduction

In Section 3.5 the multi-channel Wiener filter $\mathbf{W}_{WF}(\omega)$ has been discussed. This filter produces an MMSE estimate of the speech component in one of the microphone signals and can therefore be used for noise reduction, but not for dereverberation. Using (3.81), it can be seen that the $N \times N$-dimensional multi-channel Wiener filter matrix $\mathcal{W}_{WF}(\omega)$, which makes an MMSE estimate of the speech components $\mathbf{X}(\omega)$ in *all* microphone signals, is given by

$$\mathcal{W}_{WF}(\omega) = \bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{R}}_{xx}(\omega) \ . \tag{7.24}$$

Using the rank-1 definition for $\bar{\mathbf{R}}_{xx}(\omega)$ in (7.5), this filter can be written as

$$\mathcal{W}_{WF}(\omega) = P_s(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega) \ , \tag{7.25}$$

such that the MMSE estimate of the (reverberated) speech components $\hat{\mathbf{X}}(\omega)$ in all microphone signals is equal to

$$\hat{\mathbf{X}}(\omega) = \mathcal{W}_{WF}^H(\omega)\mathbf{Y}(\omega) = P_s(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\mathbf{Y}(\omega)\ . \tag{7.26}$$

Using the GEVD of $\bar{\mathbf{R}}_{yy}(\omega)$ and $\bar{\mathbf{R}}_{vv}(\omega)$ in (7.6), the filter matrix $\mathcal{W}_{WF}(\omega)$ in (7.24) can be computed as

$$\boxed{\begin{aligned} \mathcal{W}_{WF}(\omega) &= \bar{\mathbf{Q}}^{-H}(\omega)\bar{\mathbf{\Lambda}}_y^{-1}(\omega)\big(\bar{\mathbf{\Lambda}}_y(\omega) - \bar{\mathbf{\Lambda}}_v(\omega)\big)\bar{\mathbf{Q}}^H(\omega) \\ &= \frac{\bar{\sigma}_x^2(\omega)}{\bar{\sigma}_{y1}^2(\omega)}\tilde{\mathbf{q}}(\omega)\bar{\mathbf{q}}^H(\omega) \end{aligned}} \tag{7.27}$$

with $\bar{\sigma}_{y1}^2(\omega)$ the principal generalised eigenvalue of $\bar{\mathbf{R}}_{yy}(\omega)$ and $\tilde{\mathbf{q}}(\omega)$ the corresponding column of $\bar{\mathbf{Q}}^{-H}(\omega)$ [2].

Using the matrix inversion lemma (A.38), the matrix $\bar{\mathbf{R}}_{yy}^{-1}(\omega)$ in (7.5) can be written as

$$\bar{\mathbf{R}}_{yy}^{-1}(\omega) = \bar{\mathbf{R}}_{vv}^{-1}(\omega) - \frac{P_s(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)}{1 + P_s(\omega)\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)}\ , \tag{7.28}$$

such that, similarly to (3.92), the filter matrix $\mathcal{W}_{WF}(\omega)$ in (7.25) is equal to

$$\mathcal{W}_{WF}(\omega) = \frac{P_s(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)\mathbf{H}^H(\omega)}{1 + P_s(\omega)\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)} \tag{7.29}$$

7.3.3 Combined noise reduction and dereverberation

The objective of combined noise reduction and dereverberation is to compute the filter $\mathbf{W}_c(\omega)$ such that the output signal

$$Z_c(\omega) = \mathbf{W}_c^H(\omega)\mathbf{Y}(\omega) \tag{7.30}$$

is the MMSE estimate of the clean dereverberated speech signal $S(\omega)$, thereby taking into account both noise reduction and dereverberation. The optimal filter $\mathbf{W}_c(\omega)$ is equal to

$$\mathbf{W}_c(\omega) = \bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{r}}_{ys}(\omega)\ , \tag{7.31}$$

with the N-dimensional vector $\bar{\mathbf{r}}_{ys}(\omega) = \mathcal{E}\{\mathbf{Y}(\omega)S(\omega)\} = P_s(\omega)\mathbf{H}(\omega)$, such that the filter $\mathbf{W}_c(\omega)$ can be written as

$$\mathbf{W}_c(\omega) = P_s(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\mathbf{H}(\omega)\ . \tag{7.32}$$

The MMSE estimate of $S(\omega)$ now is equal to

$$\hat{S}(\omega) = \mathbf{W}_c^H(\omega)\mathbf{Y}(\omega) = P_s(\omega)\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\mathbf{Y}(\omega)\ . \tag{7.33}$$

[2] Recall that $\bar{\mathbf{q}}(\omega)$ is a column of the matrix $\bar{\mathbf{Q}}(\omega)$.

When comparing (7.26) and (7.33), we notice that

$$\boxed{\hat{\mathbf{X}}(\omega) = \mathbf{H}(\omega)\hat{S}(\omega)} \tag{7.34}$$

which implies that the MMSE estimate $\hat{S}(\omega)$ of the clean speech signal can be obtained by applying the normalised matched filter $\mathbf{W}_d(\omega)$ for dereverberation (cf. Section 7.3.1) to the MMSE estimate $\hat{\mathbf{X}}(\omega)$ of the speech components. Since essentially the same decomposition is used both for dereverberation and for noise reduction, these two procedures can be easily combined. Using (7.18) and (7.27), the combined filter $\mathbf{W}_c(\omega)$ can be computed as

$$\boxed{\begin{aligned} \mathbf{W}_c(\omega) &= \mathcal{W}_{WF}(\omega)\mathbf{W}_d(\omega) = \frac{\bar{\sigma}_x^2(\omega)}{\bar{\sigma}_{y1}^2(\omega)}\tilde{\mathbf{q}}(\omega)\bar{\mathbf{q}}^H(\omega)\frac{\bar{\mathbf{q}}(\omega)}{\|\bar{\mathbf{q}}(\omega)\|\|\mathbf{H}(\omega)\|} \\ &= \frac{\|\bar{\mathbf{q}}(\omega)\|}{\|\mathbf{H}(\omega)\|}\frac{\bar{\sigma}_x^2(\omega)}{\bar{\sigma}_{y1}^2(\omega)}\tilde{\mathbf{q}}^H(\omega) \end{aligned}} \tag{7.35}$$

In the case of spatially white noise, the matrix $\bar{\mathbf{Q}}(\omega)$ is orthogonal, such that $\tilde{\mathbf{q}}(\omega) = \bar{\mathbf{q}}(\omega)$ and $\|\bar{\mathbf{q}}(\omega)\| = 1$, such that the filter $\mathbf{W}_c(\omega)$ reduces to

$$\mathbf{W}_c^w(\omega) = \frac{\bar{\sigma}_x^2(\omega)}{\bar{\sigma}_{y1}^2(\omega)}\frac{\bar{\mathbf{q}}^H(\omega)}{\|\mathbf{H}(\omega)\|}\,, \tag{7.36}$$

which is equal to the normalised matched filter $\mathbf{W}_d^w(\omega)$ for spatially white noise, up to the spectral weighting term $\bar{\sigma}_x^2(\omega)/\bar{\sigma}_{y1}^2(\omega)$.

Using (7.28) in (7.32), the combined filter $\mathbf{W}_c(\omega)$ can also be written as

$$\begin{aligned} \mathbf{W}_c(\omega) &= \frac{P_s(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)}{1 + P_s(\omega)\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)} && (7.37) \\ &= \frac{P_s(\omega)}{P_s(\omega) + \left[\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)\right]^{-1}} \underbrace{\frac{\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)}{\mathbf{H}^H(\omega)\bar{\mathbf{R}}_{vv}^{-1}(\omega)\mathbf{H}(\omega)}}_{\tilde{\mathbf{W}}_d(\omega)}\,, && (7.38) \end{aligned}$$

such that the combined filter $\mathbf{W}_c(\omega)$ can be decomposed as the product of the dereverberation filter $\tilde{\mathbf{W}}_d(\omega)$ and a scalar factor. This scalar factor can be interpreted as a postfilter for the dereverberation filter [232], giving rise to an improved noise reduction performance at the expense of speech distortion.

7.4 Practical implementation issues

In (7.10) we have assumed non-overlapping frames. However, in practice we will use frames of length L with an overlap of $L-R$ samples for computing the filters

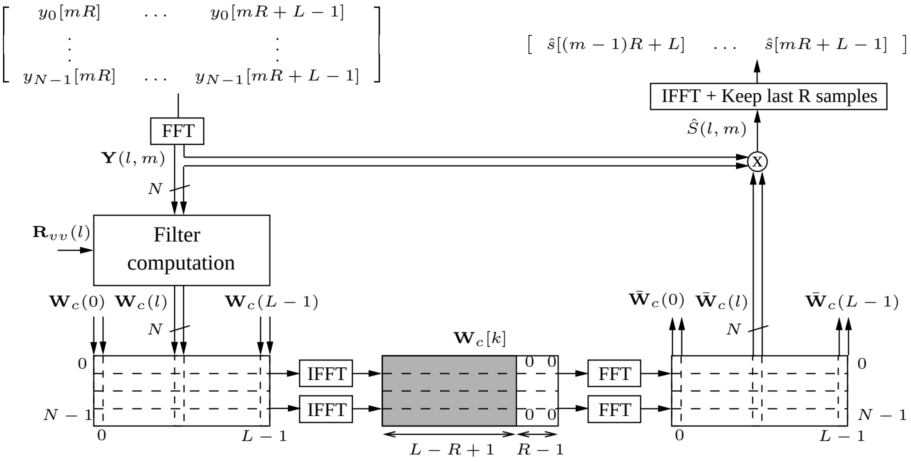

$y_0[mR]$ ... $y_0[mR+L-1]$
$y_{N-1}[mR]$ ... $y_{N-1}[mR+L-1]$
FFT
$\mathbf{Y}(l,m)$
N
$\mathbf{R}_{vv}(l)$
Filter computation
$\mathbf{W}_c(0)$ $\mathbf{W}_c(l)$ $\mathbf{W}_c(L-1)$
0
$N-1$
0
$L-1$
IFFT
IFFT
$\mathbf{w}_c[k]$
0 0
0 0
$L-R+1$
$R-1$
FFT
FFT
$\bar{\mathbf{W}}_c(0)$ $\bar{\mathbf{W}}_c(l)$ $\bar{\mathbf{W}}_c(L-1)$
X
$\hat{S}(l,m)$
IFFT + Keep last R samples
$[\ \hat{s}[(m-1)R+L]\ \ ...\ \ \hat{s}[mR+L-1]\ ]$

Figure 7.1: Practical block-processing implementation for combined noise reduction and dereverberation

and for filtering the microphone signals. For such a block-processing scheme it is well known that the underlying fast convolutions in the frequency-domain should be constrained to be linear [231]. Hence, in order to avoid circular convolutions, we put the last $R-1$ taps of the corresponding time-domain filters to zero and only keep the last R samples of the filtered microphone signals in an overlap-save procedure. This procedure is depicted in Fig. 7.1 for the combined noise reduction and dereverberation algorithm.

The N-dimensional stacked microphone signals $\mathbf{Y}(l,m)$, $l=0\ldots L{-}1$, are computed as the FFT of the frames $\left[\, y_n[mR]\ \ldots\ y_n[mR+L-1]\,\right]$, $n=0\ldots N{-}1$. The N-dimensional frequency-domain filters $\mathbf{W}_c(l)$, $l=0\ldots L-1$, are computed using (7.35). The corresponding L-dimensional time-domain filters are obtained as the IFFT of $\mathbf{W}_c(l)$, $l=0\ldots L-1$. These time-domain filters are constrained by putting the last $R-1$ taps to zero and are transformed back to the constrained frequency-domain filters $\bar{\mathbf{W}}_c(l)$, $l=0\ldots L-1$. The enhanced speech signal is computed as $\hat{S}(l,m)=\bar{\mathbf{W}}_c^H(l)\mathbf{Y}(l,m)$. From the IFFT of $\hat{S}(l,m)$, $l=0\ldots L{-}1$, the last R samples $\left[\,\hat{s}[(m-1)R+L]\ \ldots\ \hat{s}[mR+L-1]\,\right]$ are kept in an overlap-save procedure.

7.5 Simulations

In our simulations we have filtered a 16kHz continuous speech signal and a white noise source with acoustic impulse responses that are constructed using the image method ($K=1000$). The room dimensions are 3 m × 3 m × 4 m, the position of the speech source is $\left[\,1\ 2\ 2\,\right]$ and the position of the noise source

	SNR (dB)	$\mathrm{SNR}_{\mathrm{w}}$ (dB)	DI (dB)
Noisy microphone signal $y_0[k]$	0	2.88	4.74
GSVD-based noise reduction $\hat{x}_0[k]$	17.81	**16.82**	4.73
SVD-based dereverberation $\hat{s}_d^w[k]$	11.99	-0.30	1.86
GSVD-based dereverberation $\hat{s}_d[k]$	15.10	2.30	**0.86**
Noise and dereverberation $\hat{s}[k]$	**20.15**	10.12	1.35

Table 7.1: Dereverberation and noise reduction performance measures for the different algorithms ($L = 1024, R = 16$)

is $\begin{bmatrix} 0.5 & 1 & 1 \end{bmatrix}$. We have used an array of $N = 4$ omni-directional microphones and the distance between adjacent microphones is $d = 2\,\mathrm{cm}$. The positions of the microphones are $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$, $\begin{bmatrix} 1.02 & 1 & 1 \end{bmatrix}$, $\begin{bmatrix} 1.04 & 1 & 1 \end{bmatrix}$, and $\begin{bmatrix} 1.06 & 1 & 1 \end{bmatrix}$. The reverberation time of the room is $T_{60} = 400\,\mathrm{msec}$. We have used these acoustical conditions, since the smaller the microphone distance and the reverberation time, the more the frequency-domain signals are spatially correlated. The unbiased SNR of the first microphone signal $y_0[k]$ is $0\,\mathrm{dB}$. In all algorithms we have used a frame length (FFT-size) $L = 1024$ and overlap $R = 16$.

As objective measures for the noise reduction performance we use the unbiased SNR, defined in (2.32), and the frequency-weighted signal-to-noise ratio SNR_w, defined in [112], which is a weighted subband SNR. As an objective measure for dereverberation we use the dereverberation index (DI), defined in (2.43).

Table 7.1 gives an overview of the objective noise reduction and dereverberation performance measures for the different algorithms. Figure 7.2 plots the noisy microphone signal $y_0[k]$ and the enhanced microphone signal $\hat{x}_0[k]$ using the GSVD-based noise reduction technique. As can be seen in Table 7.1, this technique produces the highest SNR_w, but does not achieve any dereverberation, since the DI of $\hat{x}_0[k]$ and $y_0[k]$ are almost equal (DI $\approx 4.7\,\mathrm{dB}$).

Figures 7.3a and 7.3b show the amplitude responses of the total speech transfer function $F(\omega)$ between $S(\omega)$ and the speech component in the output signal for the SVD-based and the GSVD-based dereverberation algorithms. Figures 7.4a and 7.4b depict the time-domain output signals $\hat{s}_d^w[k]$ and $\hat{s}_d[k]$ for these algorithms. As can be seen, the filter $\mathbf{W}_d(\omega)$ computed using the GEVD of $\bar{\mathbf{R}}_{yy}(\omega)$ and $\bar{\mathbf{R}}_{vv}(\omega)$ produces the flattest amplitude response (DI = 0.86), and has a better dereverberation performance than the filter $\mathbf{W}_d^w(\omega)$ computed using the EVD of $\bar{\mathbf{R}}_{yy}(\omega)$ (DI = 1.86). However, as can be seen in Table 7.1, the noise reduction performance of both algorithms is quite poor, since SNR_w is smaller than for the noisy microphone signal.

Figures 7.3c and 7.4c show the amplitude response of the total speech transfer function $F(\omega)$ and the time-domain output signal $\hat{s}[k]$ for the combined noise reduction and dereverberation technique. From Table 7.1 it can be seen that

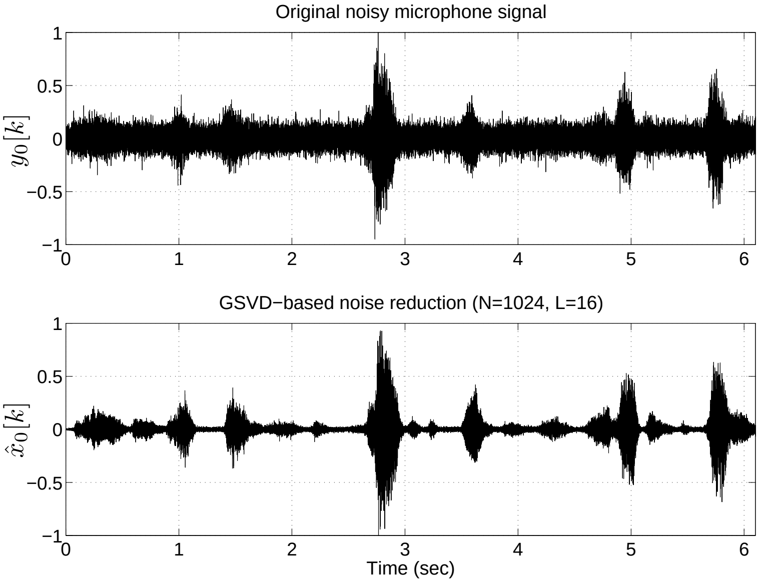


Figure 7.2: (**a**) Noisy microphone signal $y_0[k]$, (**b**) Enhanced microphone signal $\hat{x}_0[k]$ with GSVD-based noise reduction technique ($L = 1024, R = 16$)

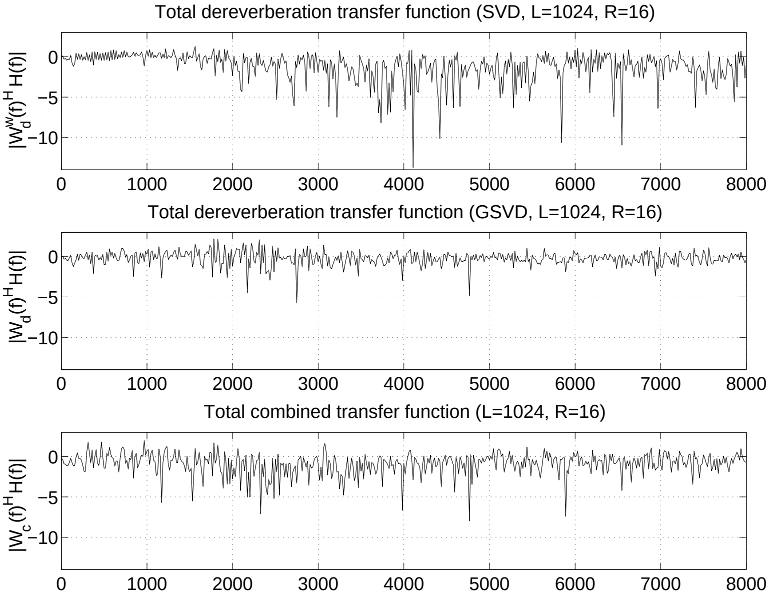


Figure 7.3: Total speech transfer function $F(\omega)$ computed with (**a**) SVD-based dereverberation technique, (**b**) GSVD-based dereverberation technique, (**c**) combined noise reduction and dereverberation technique ($L = 1024, R = 16$)

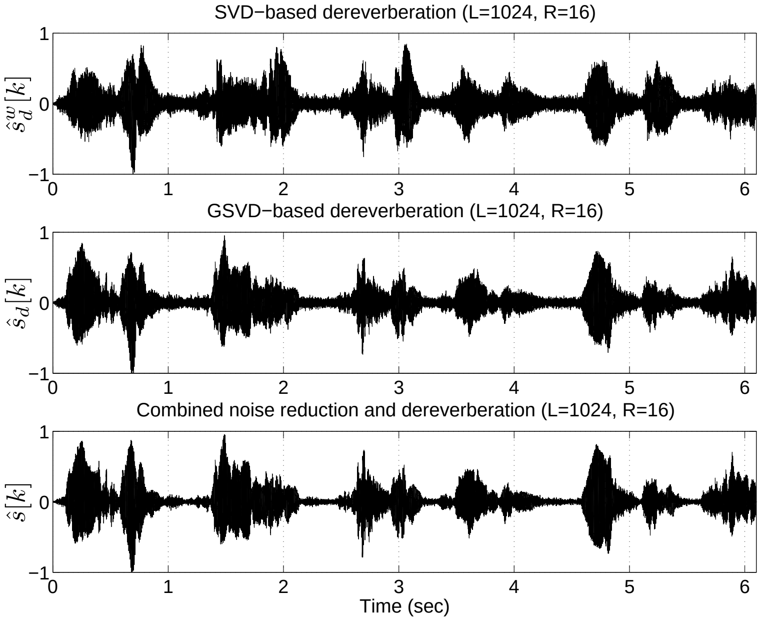


Figure 7.4: (**a**) SVD-based dereverberated signal $\hat{s}_d^w[k]$, (**b**) GSVD-based dereverberated signal $\hat{s}_d[k]$, (**b**) Enhanced signal $\hat{s}[k]$ with combined dereverberation and noise reduction technique ($L = 1024, R = 16$)

the dereverberation performance is not as good as for the GSVD-based dereverberation technique (but it produces a better SNR), while its noise reduction performance is not as good as for the GSVD-based noise reduction technique (but is has a better dereverberation index DI). It is therefore clear that the combined noise reduction and dereverberation technique makes a trade-off between the noise reduction and the dereverberation objectives.

7.6 Conclusion

In this chapter we have presented frequency-domain GSVD-based signal enhancement techniques for noise reduction and for dereverberation. It has been shown that the optimal MMSE estimate of the clean speech signal can be obtained by matched filtering of the MMSE estimate of the speech components in the microphone signals. By simulations it has been shown that the combined noise reduction and dereverberation algorithm makes a trade-off between the dereverberation and noise reduction objectives.

Part III

Broadband Beamformer Design

Chapter 8

Far-Field Broadband Beamforming

This chapter discusses several design procedures for designing *broadband beamformers* with an *arbitrary desired spatial directivity pattern* for a given *arbitrary microphone array configuration*, using an *FIR filter-and-sum structure*. In this chapter, we assume that the speech source is in the *far-field* of the microphone array and that the microphones are (perfect) omni-directional microphones with a flat frequency response equal to 1. In Chapter 9 we will discuss near-field and mixed near-field far-field broadband beamformers and in Chapter 10 we will discuss robust broadband beamformers, taking into account the (frequency- and angle-dependent) microphone characteristics.

Section 8.1 gives a brief overview of fixed beamforming for speech applications. In Section 8.2 the far-field broadband beamforming problem is introduced and some definitions and notational conventions are given.

Section 8.3 discusses several cost functions that can be used for designing far-field broadband beamformers. In general we would like to use the non-linear cost function that minimises the error between the amplitudes of the actual and the desired spatial directivity pattern. However, for this cost function no closed-form solution is available and an iterative non-linear optimisation procedure is required, giving rise to a high computational complexity. Hence, we will also consider other cost functions with a lower computational complexity that can be solved using non-iterative optimisation techniques, such as the weighted least-squares (LS) and the maximum energy array cost function. For all considered cost functions, we first discuss the general design procedure for an arbitrary spatial directivity pattern and we then focus on the specific design case of a beamformer having a passband and a stopband region. For all cost functions, it will also be shown how linear constraints can be imposed on the filter coefficients.

In section 8.4 two novel non-iterative cost functions are discussed, which are both based on eigenfilters. In the conventional eigenfilter technique a reference point is required, whereas in the eigenfilter technique based on a TLS (Total Least Squares) error criterion, this reference point is not required.

In Section 8.5 different linear constraints are considered which can be imposed on the filter coefficients. Point constraints, line constraints and derivative constraints will be discussed.

Section 8.6 gives simulation results for the different cost functions and design cases. It is shown that among the considered non-iterative design procedures the TLS eigenfilter technique has the best performance, i.e. best resembling the performance of the non-linear design procedure but having a significantly lower computational complexity.

8.1 Introduction

Well-known multi-microphone signal enhancement techniques are fixed and adaptive beamforming (cf. Sections 2.5.2 and 2.5.3). Adaptive beamformers, generally have a better noise reduction performance than fixed beamformers and are able to adapt to changing acoustic environments. However, adaptive beamformers are quite sensitive to modelling errors [37][240], cf. Section 5.4.3, resulting in speech distortion and cancellation if no countermeasures are taken [100][128][191][194][254]. Therefore, fixed beamforming techniques (with a fixed spatial directivity pattern) are sometimes preferred because they do not require a control algorithm and because of their easy implementation and low computational complexity. Fixed beamformers are frequently used in highly reverberant environments, in applications where the position of the speech source is assumed to be known (e.g. hearing aids [146][234][245]), for creating multiple beams [150][259] and for creating the speech reference signal in a GSC.

In general, fixed beamforming techniques try to obtain spatial focusing on the speech source, thereby reducing reverberation and suppressing background noise not coming from the same direction as the speech source. In order to obtain some robustness against estimation errors for the direction of the speech source (cf. Chapter 6) and – small – speaker movements, a region of angles around the direction of the speech source should be passed without distortion. It should even be possible to design a broadband beamformer with an arbitrary spatial directivity pattern. However, using most fixed beamformers discussed in Section 2.5.2, such as DS beamforming, differential microphones [75], superdirective microphone arrays [16][36][146] and frequency-invariant beamforming [274][275], it is not possible to design arbitrary spatial directivity patterns for an arbitrary microphone array configuration. Differential microphones e.g. require a small-size microphone array, superdirective microphone arrays are designed using an assumption about the noise field, and for frequency-invariant beamformers the desired spatial directivity pattern is equal for all frequencies.

Using the most general beamformer structure, i.e. the FIR filter-and-sum structure (cf. Section 2.5.2), it is possible to design a fixed broadband beamformer whose spatial directivity pattern optimally fits a (predefined) desired spatial directivity pattern, by minimising some specific cost function. Several design procedures exist, which are e.g. based on weighted least-squares (LS) filter design [167], a maximum energy array [155] or non-linear optimisation techniques [144][157][159][192]. Although in general we would like to use the non-linear design procedure, this procedure gives rise to a high computational complexity, since it requires an iterative optimisation technique. In this chapter two novel non-iterative design procedures are presented, which are based on eigenfilters. In the conventional eigenfilter technique, a reference point is required, whereas in the eigenfilter technique based on a TLS error criterion, this reference point is not required. Eigenfilters have already been used for designing 1-D linear-phase FIR filters [209][253] and for designing 2-D and spatial filters [32][208][209]. In this chapter we extend their usage to the design of far-field broadband beamformers. It will be shown by simulations that the TLS eigenfilter technique has a better performance than the weighted LS, the maximum energy array and the conventional eigenfilter technique.

Many broadband beamformer design procedures either perform the design individually for separate frequencies or approximate the double integrals that arise in the design by a finite sum over a grid of frequencies and angles. However, in this thesis we will always calculate such integrals exactly over the frequency-angle plane and hence perform a true broadband design. Note that in typical speech communication applications, broadband design implies a design over several octaves (e.g. $300-3500\,\text{Hz}$ with sampling frequency $f_s = 8\,\text{kHz}$).

8.2 Far-field beamforming: configuration

In this chapter, we assume that the sources are in the far-field of the microphone array, such that planar wave propagation and equal signal attenuation for all microphones can be assumed. For the near-field case, we refer to Chapter 9.

Consider the linear microphone array depicted in Fig. 8.1, with N microphones and d_n the distance between the nth microphone and the centre of the microphone array. The *spatial directivity pattern* $H(\omega,\theta)$ for a source $S(\omega)$ at an angle θ from the microphone array is defined as, cf. Section 2.5.1,

$$H(\omega,\theta) = \frac{Z(\omega,\theta)}{\bar{Y}(\omega,\theta)} = \frac{\sum_{n=0}^{N-1} W_n(\omega) Y_n(\omega,\theta)}{\bar{Y}(\omega,\theta)} \,, \tag{8.1}$$

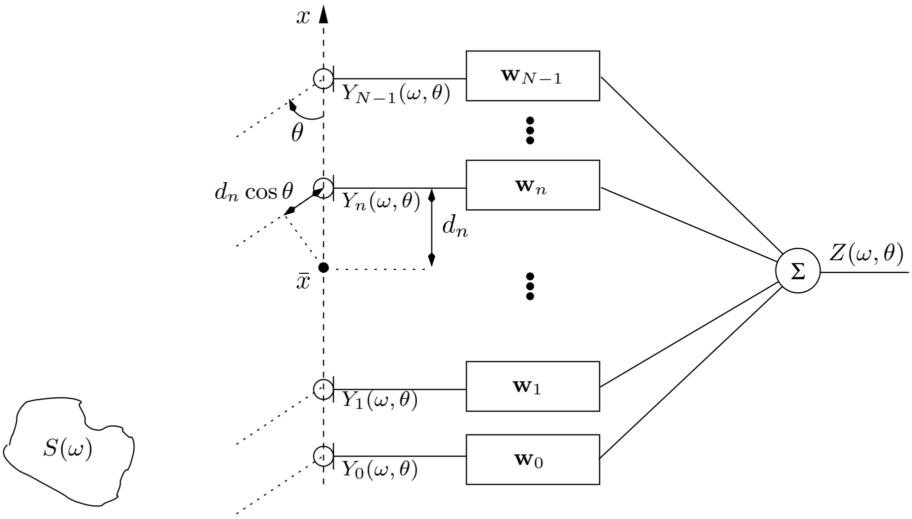


Figure 8.1: Linear microphone array configuration

with $W_n(\omega)$ the frequency response of the real-valued L-dimensional FIR filter $\mathbf{w}_n$,

$$W_n(\omega) = \sum_{l=0}^{L-1} w_{n,l}\, e^{-jl\omega} = \mathbf{w}_n^T \mathbf{e}(\omega) \ , \tag{8.2}$$

with

$$\mathbf{w}_n = \begin{bmatrix} w_{n,0} \\ w_{n,1} \\ \vdots \\ w_{n,L-1} \end{bmatrix} \quad \mathbf{e}(\omega) = \begin{bmatrix} 1 \\ e^{-j\omega} \\ \vdots \\ e^{-j(L-1)\omega} \end{bmatrix} . \tag{8.3}$$

Under far-field conditions, the microphone signals $Y_n(\omega,\theta)$, $n = 0 \ldots N-1$, are delayed versions of the signal $\bar{Y}(\omega,\theta)$ received at the centre of the microphone array, i.e.

$$Y_n(\omega,\theta) = \bar{Y}(\omega,\theta) e^{-j\omega\tau_n(\theta)}, \ -\pi \leq \omega \leq \pi, \ -\pi \leq \theta \leq \pi \ , \tag{8.4}$$

with the delay $\tau_n(\theta)$ in number of samples equal to

$$\tau_n(\theta) = \frac{d_n \cos\theta}{c} f_s \ , \tag{8.5}$$

with c the speed of sound ($c = 340\frac{m}{s}$) and f_s the sampling frequency.

In fact, for a random two-dimensional (planar) microphone configuration (see

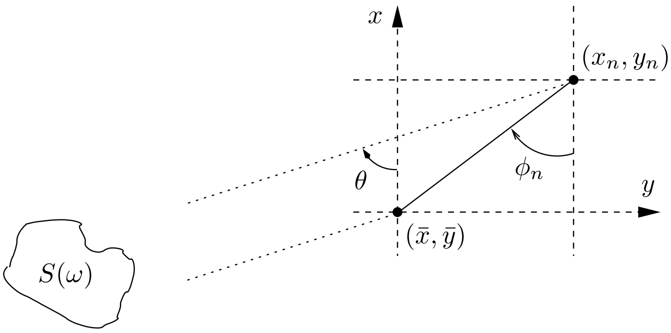


Figure 8.2: Two-dimensional microphone array configuration

Fig. 8.2), the microphone signals $Y_n(\omega,\theta)$ can also be written as (8.4), with

$$\tau_n(\theta) = \frac{d_n \cos(\theta - \phi_n)}{c} f_s, \; d_n = \sqrt{(x_n - \bar{x})^2 + (y_n - \bar{y})^2}, \; \tan\phi_n = \frac{y_n - \bar{y}}{x_n - \bar{x}} \; ,$$

with $(\bar{x}, \bar{y})$ the centre of the planar array. Without loss of generality we will assume a linear array ($\phi_n = 0$ or $\phi_n = \pi$) in the remainder of the text.

Combining (8.1) and (8.4), the spatial directivity pattern $H(\omega,\theta)$ can be written as

$$\boxed{H(\omega,\theta) = \sum_{n=0}^{N-1} W_n(\omega) e^{-j\omega\tau_n(\theta)} = \sum_{n=0}^{N-1} \mathbf{w}_n^T \mathbf{e}(\omega) e^{-j\omega\tau_n(\theta)} = \mathbf{w}^T \mathbf{g}(\omega,\theta)} \tag{8.6}$$

with the M-dimensional filter vector $\mathbf{w}$ and steering vector $\mathbf{g}(\omega,\theta)$ equal to

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_{N-1} \end{bmatrix} \quad \mathbf{g}(\omega,\theta) = \begin{bmatrix} \mathbf{e}(\omega) e^{-j\omega\tau_0(\theta)} \\ \mathbf{e}(\omega) e^{-j\omega\tau_1(\theta)} \\ \vdots \\ \mathbf{e}(\omega) e^{-j\omega\tau_{N-1}(\theta)} \end{bmatrix} . \tag{8.7}$$

The steering vector $\mathbf{g}(\omega,\theta)$ can be decomposed into a real and an imaginary part, $\mathbf{g}(\omega,\theta) = \mathbf{g}_R(\omega,\theta) + j\mathbf{g}_I(\omega,\theta)$. Using (8.6), the spatial directivity spectrum $|H(\omega,\theta)|^2$ can be written as

$$|H(\omega,\theta)|^2 = H(\omega,\theta)H^*(\omega,\theta) = \mathbf{w}^T \mathbf{G}(\omega,\theta)\mathbf{w} \; , \tag{8.8}$$

with

$$\mathbf{G}(\omega,\theta) = \mathbf{g}(\omega,\theta)\mathbf{g}^H(\omega,\theta) \; . \tag{8.9}$$

The matrix $\mathbf{G}(\omega,\theta)$ can be decomposed into a real and an imaginary part, $\mathbf{G}(\omega,\theta) = \mathbf{G}_R(\omega,\theta) + j\mathbf{G}_I(\omega,\theta)$. Since $\mathbf{G}_I(\omega,\theta)$ is anti-symmetric (cf. Appendix E.2), the spatial directivity spectrum $|H(\omega,\theta)|^2$ is equal to (E.15),

$$\boxed{|H(\omega,\theta)|^2 = \mathbf{w}^T \mathbf{G}_R(\omega,\theta)\mathbf{w}} \tag{8.10}$$

8.3 Broadband beamforming procedures

8.3.1 Overview

The design of a broadband beamformer consists of the calculation of the filter $\mathbf{w}$, such that $H(\omega, \theta)$ optimally fits a desired spatial directivity pattern $D(\omega, \theta)$, where $D(\omega, \theta)$ is an arbitrary two-dimensional function in ω and θ. Several design procedures exist, depending on the specific cost function which is optimised. In this section three different cost functions will be considered:

- a weighted least-squares (LS) cost function J_{LS}, minimising the weighted least-squares error between the actual and the desired spatial directivity pattern, which can be written as a quadratic function (cf. Section 8.3.2);
- a maximum energy array cost function J_{ME}, maximising the energy ratio between the passband and the stopband region. Maximising this cost function leads to a generalised eigenvalue problem (cf. Section 8.3.3);
- a non-linear cost function J_{NL}, minimising the error between the amplitudes of the actual and the desired spatial directivity pattern, not taking into account the phase of the spatial directivity patterns. Minimising this cost function leads to a non-linear optimisation problem, which can be solved using iterative optimisation techniques (cf. Section 8.3.4).

In general we would like to use the non-linear cost function J_{NL}. However, since optimising this cost function requires an iterative non-linear optimisation technique (cf. Section 8.3.4), giving rise to a large computational complexity, we will also consider non-iterative design procedures with a lower computational complexity. In Section 8.4 two non-iterative eigenfilter-based cost functions will be defined and in Section 8.6 the performance of all considered non-iterative design procedures will be compared with the non-linear design procedure.

We will consider the design of broadband beamformers over the total frequency-angle plane of interest, i.e. we will not split up the fullband problem into separate smallband problems for different frequencies. Moreover, we will not approximate the double integrals over the frequency-angle plane by a finite Riemann-sum over a grid of frequencies and angles, as e.g has been done in [144] for the non-linear cost function. For all cost functions, we will first discuss the general design procedure for an arbitrary function $D(\omega, \theta)$, and we will then focus on the specific design case of a broadband beamformer having a desired response $D(\omega, \theta) = 0$ in the stopband region (Ω_s, Θ_s) and $D(\omega, \theta) = 1$ in the passband region (Ω_p, Θ_p). For the specific design case, the weighting function is $F(\omega, \theta) = 1$ in the passband and $F(\omega, \theta) = \alpha$ in the stopband. We will also discuss how linear constraints of the form $\mathbf{Cw} = \mathbf{b}$ (cf. Section 8.5) can be imposed on the filter $\mathbf{w}$.

8.3.2 Weighted least-squares

The weighted least-squares (LS) cost function is a well-known cost function from literature, which can e.g. be used for designing FIR filters [167], 2D-filters [208] and broadband beamformers.

General design

Considering the LS error $|H(\omega,\theta)-D(\omega,\theta)|^2$, the weighted LS cost function is defined as

$$J_{LS}(\mathbf{w}) = \int_\Theta\int_\Omega F(\omega,\theta)|H(\omega,\theta)-D(\omega,\theta)|^2 d\omega d\theta\ , \tag{8.11}$$

where both the phase and the amplitude of $H(\omega,\theta)$ are taken into account. $F(\omega,\theta)$ is a positive real weighting function, assigning more or less importance to certain frequencies or angles. Using $F(\omega,\theta)$ it is e.g. possible to use a speech-intelligibility motivated frequency weighting [198]. The weighted LS cost function can be written as

$$\begin{aligned} J_{LS}(\mathbf{w}) &= \int_\Theta\int_\Omega F(\omega,\theta)|H(\omega,\theta)|^2 d\omega d\theta + \int_\Theta\int_\Omega F(\omega,\theta)|D(\omega,\theta)|^2 d\omega d\theta \\ &\quad -2\int_\Theta\int_\Omega F(\omega,\theta)\,\mathrm{Re}\{D(\omega,\theta)H^*(\omega,\theta)\}\ . \end{aligned} \tag{8.12}$$

Using (8.10) and the fact that

$$\mathrm{Re}\{D(\omega,\theta)H^*(\omega,\theta)\} = \mathbf{w}^T\left[D_R(\omega,\theta)\mathbf{g}_R(\omega,\theta)+D_I(\omega,\theta)\mathbf{g}_I(\omega,\theta)\right]\ , \tag{8.13}$$

this cost function can be rewritten as the quadratic function

$$\boxed{J_{LS}(\mathbf{w}) = \mathbf{w}^T\mathbf{Q}_{LS}\mathbf{w} - 2\mathbf{w}^T\mathbf{a} + d_{LS}} \tag{8.14}$$

with

$$\begin{aligned} \mathbf{Q}_{LS} &= \int_\Theta\int_\Omega F(\omega,\theta)\mathbf{G}_R(\omega,\theta)d\omega d\theta && (8.15)\\ \mathbf{a} &= \int_\Theta\int_\Omega F(\omega,\theta)\left[D_R(\omega,\theta)\mathbf{g}_R(\omega,\theta)+D_I(\omega,\theta)\mathbf{g}_I(\omega,\theta)\right]d\omega d\theta && (8.16)\\ d_{LS} &= \int_\Theta\int_\Omega F(\omega,\theta)|D(\omega,\theta)|^2 d\omega d\theta\ . && (8.17) \end{aligned}$$

The weighted LS cost function $J_{LS}(\mathbf{w})$ is minimised by setting the derivative $\frac{\partial J_{LS}(\mathbf{w})}{\partial \mathbf{w}}$ equal to 0, such that the solution $\mathbf{w}_{LS}$ is given by

$$\boxed{\mathbf{w}_{LS} = \mathbf{Q}_{LS}^{-1}\,\mathbf{a}} \tag{8.18}$$

Specific design case

For the specific design case where $D(\omega, \theta) = 1$ and $F(\omega, \theta) = 1$ in the passband and $D(\omega, \theta) = 0$ and $F(\omega, \theta) = \alpha$ in the stopband, equations (8.15), (8.16) and (8.17) can be written as

$$\mathbf{Q}_{LS} = \underbrace{\int_{\Theta_p}\int_{\Omega_p} \mathbf{G}_R(\omega,\theta) d\omega d\theta}_{\mathbf{Q}_e^p} + \alpha \underbrace{\int_{\Theta_s}\int_{\Omega_s} \mathbf{G}_R(\omega,\theta) d\omega d\theta}_{\mathbf{Q}_e^s} \tag{8.19}$$

$$\mathbf{a} = \int_{\Theta_p}\int_{\Omega_p} \mathbf{g}_R(\omega,\theta) d\omega d\theta \tag{8.20}$$

$$d_{LS} = \int_{\Theta_p}\int_{\Omega_p} 1 \, d\omega d\theta \; . \tag{8.21}$$

The quantity $\mathbf{w}^T\mathbf{Q}_e^p\mathbf{w}$ is equal to the energy in the passband, whereas $\mathbf{w}^T\mathbf{Q}_e^s\mathbf{w}$ is equal to the energy in the stopband. The calculation of the integrals in (8.20) and (8.19) is discussed in Appendix E.1 and E.2.

Linear constraints

Different linear constraints of the form $\mathbf{Cw} = \mathbf{b}$, with $\mathbf{C}$ a $J \times M$-dimensional matrix and $\mathbf{b}$ a J-dimensional vector, will be discussed in Section 8.5. When imposing linear constraints on the weighted LS criterion, the constrained optimisation problem has the form

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^T\mathbf{Q}_{LS}\mathbf{w} - 2\mathbf{w}^T\mathbf{a} + d_{LS}, \quad \text{subject to } \mathbf{Cw} = \mathbf{b}} \tag{8.22}$$

This constrained minimisation problem can be transformed into an unconstrained minimisation problem, cf. Appendix D.1 (similar to the derivation of the Generalised Sidelobe Canceller, cf. Section 2.5.3). The solution $\mathbf{w}_{LS}^c$ of the constrained minimisation problem is equal to (D.10),

$$\boxed{\mathbf{w}_{LS}^c = \mathbf{Q}_{LS}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{Q}_{LS}^{-1}\mathbf{C}^T)^{-1}(\mathbf{b} - \mathbf{C}\mathbf{Q}_{LS}^{-1}\mathbf{a}) + \mathbf{Q}_{LS}^{-1}\mathbf{a}} \tag{8.23}$$

8.3.3 Maximum energy array

In [155] a so-called maximum energy cost function has been defined. Since in the design of a maximum energy array broadband beamformer it is assumed that a passband region and a stopband region are present, we can only consider the specific design case for this design procedure.

Specific design case

The maximum energy cost function $J_{ME}(\mathbf{w})$ is defined as the ratio of the energy in one frequency-angle region (passband) and the energy in another frequency-angle region (stopband), i.e.

$$J_{ME}(\mathbf{w}) = \frac{\int_{\Theta_p}\int_{\Omega_p} |H(\omega,\theta)|^2 d\omega d\theta}{\int_{\Theta_s}\int_{\Omega_s} |H(\omega,\theta)|^2 d\omega d\theta} \,. \tag{8.24}$$

Maximising this ratio can actually be considered as a broadband generalisation of the (smallband) superdirective beamformer formulation [16]. Using (8.19), this cost function can be written as

$$\boxed{J_{ME}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{Q}_e^p\mathbf{w}}{\mathbf{w}^T\mathbf{Q}_e^s\mathbf{w}}} \tag{8.25}$$

with $\mathbf{Q}_e^p$ and $\mathbf{Q}_e^s$ defined in (8.19). The filter $\mathbf{w}_{ME}$ which maximises $J_{ME}(\mathbf{w})$ is equal to the generalised eigenvector corresponding to the maximum generalised eigenvalue in the generalised eigenvalue decomposition (GEVD) of $\mathbf{Q}_e^p$ and $\mathbf{Q}_e^s$. However, as will be shown in the simulations, the spatial directivity pattern corresponding to this filter mainly amplifies the high frequencies, since it is easier to obtain a large directivity for high frequencies than for low frequencies (cf. delay-and-sum beamformer). Hence, a frequency-dependent angle integration interval has to be used with a larger integration interval at low frequencies [155], or alternatively, linear constraints have to be imposed (cf. Section 8.6).

Linear constraints

When imposing linear constraints of the form $\mathbf{Cw} = \mathbf{b}$, the constrained optimisation problem can be written as

$$\boxed{\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{Q}_e^p\mathbf{w}}{\mathbf{w}^T\mathbf{Q}_e^s\mathbf{w}}, \quad \text{subject to } \; \mathbf{Cw} = \mathbf{b}} \tag{8.26}$$

with $\mathbf{b}$ generally not equal to $\mathbf{0}$. This constrained ratio maximisation problem can be rewritten as the extended constrained ratio maximisation problem

$$\max_{\hat{\mathbf{w}}} \frac{\hat{\mathbf{w}}^T\hat{\mathbf{Q}}_e^p\hat{\mathbf{w}}}{\hat{\mathbf{w}}^T\hat{\mathbf{Q}}_e^s\hat{\mathbf{w}}}, \quad \text{subject to } \; \hat{\mathbf{C}}\hat{\mathbf{w}} = \mathbf{0} \,, \tag{8.27}$$

with the extended vector $\hat{\mathbf{w}}$ and matrices $\hat{\mathbf{C}}$, $\hat{\mathbf{Q}}_e^p$ and $\hat{\mathbf{Q}}_e^s$ defined as

$$\hat{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix}, \quad \hat{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & \mathbf{b} \end{bmatrix}, \quad \hat{\mathbf{Q}}_e^p = \begin{bmatrix} \mathbf{Q}_e^p & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix}, \quad \hat{\mathbf{Q}}_e^s = \begin{bmatrix} \mathbf{Q}_e^s & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix},$$

The constrained optimisation problem (8.27) can be transformed into the unconstrained optimisation problem

$$\max_{\tilde{\mathbf{w}}} \frac{\tilde{\mathbf{w}}^T\mathbf{B}\hat{\mathbf{Q}}_e^p\mathbf{B}^T\tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^T\mathbf{B}\hat{\mathbf{Q}}_e^s\mathbf{B}^T\tilde{\mathbf{w}}} \,, \tag{8.28}$$

with $\hat{\mathbf{w}} = \mathbf{B}^T\tilde{\mathbf{w}}$ and $\mathbf{B}$ the $(M+1-J) \times (M+1)$-dimensional null space of $\hat{\mathbf{C}}$ and $\tilde{\mathbf{w}}$ an $(M+1-K)$-dimensional vector. The solution $\tilde{\mathbf{w}}_{ME}$ of the unconstrained optimisation problem (8.28) is the generalised eigenvector of $\mathbf{B}\hat{\mathbf{Q}}_e^p\mathbf{B}^T$ and $\mathbf{B}\hat{\mathbf{Q}}_e^s\mathbf{B}^T$, corresponding to the maximum generalised eigenvalue, such that the solution $\hat{\mathbf{w}}_{ME}^c$ of the constrained optimisation problem (8.27) is equal to

$$\hat{\mathbf{w}}_{ME}^c = \mathbf{B}^T\tilde{\mathbf{w}}_{ME} \ . \tag{8.29}$$

After scaling the last element of $\hat{\mathbf{w}}_{ME}^c$ to -1, the actual solution $\mathbf{w}_{ME}^c$ of (8.26) is obtained as the first M elements of $\hat{\mathbf{w}}_{ME}^c$. The fact that the matrices $\hat{\mathbf{Q}}_e^p$ and $\hat{\mathbf{Q}}_e^s$ are singular does not give rise to problems, since the matrices $\mathbf{B}\hat{\mathbf{Q}}_e^p\mathbf{B}^T$ and $\mathbf{B}\hat{\mathbf{Q}}_e^s\mathbf{B}^T$ are in general not singular.

Single linear constraint

The constrained ratio maximisation problem (8.26) is however simplified when a single linear constraint is present, i.e. $J = 1$. In this case the solution is given by the scaled generalised eigenvector $\upsilon\mathbf{u}_{max}$, corresponding to the maximum generalised eigenvalue of $\mathbf{Q}_e^p$ and $\mathbf{Q}_e^s$, where the scaling factor υ is determined such that the linear constraint equation $\mathbf{C}\mathbf{w} = b$ (with $\mathbf{C}$ a row vector and b a scalar) is satisfied, i.e.

$$\upsilon = \frac{b}{\mathbf{C}\mathbf{u}_{max}} \ . \tag{8.30}$$

8.3.4 Non-linear criterion

Different non-linear cost functions for broadband beamforming have been proposed in literature, leading to a minimax problem [157][192] or requiring iterative optimisation techniques [144][159]. In this section we will slightly modify the non-linear cost function presented in [144], such that the double integrals arising in the optimisation problem only need to be computed once.

General design

Instead of minimising the LS error $|H(\omega,\theta) - D(\omega,\theta)|^2$, it is also possible to minimise the error between the amplitudes $|H(\omega,\theta)| - |D(\omega,\theta)|$, because in general the phase of the spatial directivity pattern is of no relevance. This problem formulation leads to the cost function [144]

$$\bar{J}_{NL}(\mathbf{w}) = \int_\Theta \int_\Omega F(\omega,\theta)\left[|H(\omega,\theta)| - |D(\omega,\theta)|\right]^2 d\omega d\theta \ , \tag{8.31}$$

which can be rewritten as

$$\begin{aligned}\bar{J}_{NL}(\mathbf{w}) \ = \ & \int_\Theta \int_\Omega F(\omega,\theta)|H(\omega,\theta)|^2 d\omega d\theta + \int_\Theta \int_\Omega F(\omega,\theta)|D(\omega,\theta)|^2 d\omega d\theta - \\ & 2\int_\Theta \int_\Omega F(\omega,\theta)|D(\omega,\theta)||H(\omega,\theta)| d\omega d\theta \end{aligned} \tag{8.32}$$

$$= \mathbf{w}^T\mathbf{Q}_{LS}\mathbf{w} + d_{LS} - \underbrace{2\int_\Theta\int_\Omega F(\omega,\theta)|D(\omega,\theta)||H(\omega,\theta)|d\omega d\theta}_{J_{abs}(\mathbf{w})}\ , \tag{8.33}$$

with $\mathbf{Q}_{LS}$ and d_{LS} defined in (8.15) and (8.17). Minimising $\bar{J}_{NL}(\mathbf{w})$ leads to a non-linear optimisation problem, which can be solved using iterative optimisation techniques. These optimisation techniques generally involve several evaluations of $\bar{J}_{NL}(\mathbf{w})$ in each iteration step. A complexity problem now arises in the computation of $J_{abs}(\mathbf{w})$. Without loss of generality, assume that $F(\omega,\theta) = 1$ and $|D(\omega,\theta)| = 1$ over some frequency-angle region and that $D(\omega,\theta) = 0$ elsewhere. $J_{abs}(\mathbf{w})$ can then be written using (8.10) as

$$J_{abs}(\mathbf{w}) = 2\int_{\Theta_p}\int_{\Omega_p}|H(\omega,\theta)|d\omega d\theta = 2\int_{\Theta_p}\int_{\Omega_p}\sqrt{\mathbf{w}^T\mathbf{G}_R(\omega,\theta)\mathbf{w}}\,d\omega d\theta\ . \tag{8.34}$$

Because of the square root, the filter coefficients can not be extracted from the double integral (cf. Appendix E.4), and *for every* $\mathbf{w}$ *the double integrals need to be recomputed numerically*, which is a computationally very demanding procedure. However, by slightly modifying the non-linear cost function, it is possible to overcome this computational problem.

Instead of minimising the error between the amplitudes $|H(\omega,\theta)|$ and $|D(\omega,\theta)|$, we propose a *novel non-linear criterion* which minimises the error between the square of the amplitudes $|H(\omega,\theta)|^2$ and $|D(\omega,\theta)|^2$, i.e.

$$\boxed{J_{NL}(\mathbf{w}) = \int_\Theta\int_\Omega F(\omega,\theta)\big[|H(\omega,\theta)|^2 - |D(\omega,\theta)|^2\big]^2 d\omega d\theta} \tag{8.35}$$

which is also independent of the phase of the spatial directivity patterns. The cost function $J_{NL}(\mathbf{w})$ can be written (without square-roots) as

$$\begin{aligned} J_{NL}(\mathbf{w}) &= \int_\Theta\int_\Omega F(\omega,\theta)\left(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\right)^2 d\omega d\theta + \int_\Theta\int_\Omega F(\omega,\theta)\,|D(\omega,\theta)|^4\,d\omega d\theta \\ &\quad -2\int_\Theta\int_\Omega F(\omega,\theta)\,|D(\omega,\theta)|^2\left(\mathbf{w}^T\mathbf{G}_R(\omega,\theta)\mathbf{w}\right)d\omega d\theta \qquad (8.36)\\ &= J_{sum}(\mathbf{w}) + d_{NL} - 2\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w}\ , \qquad (8.37)\end{aligned}$$

with

$$J_{sum}(\mathbf{w}) = \int_\Theta\int_\Omega F(\omega,\theta)\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta \tag{8.38}$$

$$d_{NL} = \int_\Theta\int_\Omega F(\omega,\theta)|D(\omega,\theta)|^4 d\omega d\theta \tag{8.39}$$

$$\mathbf{Q}_{NL} = \int_\Theta\int_\Omega F(\omega,\theta)|D(\omega,\theta)|^2\mathbf{G}_R(\omega,\theta)d\omega d\theta\ . \tag{8.40}$$

In Appendix E.4 it is shown that the function $J_{sum}(\mathbf{w})$ – and therefore the total cost function $J_{NL}(\mathbf{w})$ – can be evaluated without having to calculate double integrals for every $\mathbf{w}$, since $\mathbf{w}$ can be extracted from the double integrals.

Specific design case

For the specific design case where $F(\omega,\theta)=1$ and $D(\omega,\theta)=1$ in the passband and $D(\omega,\theta)=0$ and $F(\omega,\theta)=\alpha$ in the stopband, equations (8.38), (8.39) and (8.40) can be written as

$$\begin{aligned} J_{sum}(\mathbf{w}) &= \underbrace{\int_{\Theta_p}\int_{\Omega_p}\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta}_{J^p_{sum}(\mathbf{w})} + \alpha\underbrace{\int_{\Theta_s}\int_{\Omega_s}\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta}_{J^s_{sum}(\mathbf{w})} \\ d_{NL} &= \int_{\Theta_p}\int_{\Omega_p} 1\ d\omega d\theta = d_{LS} \qquad (8.41) \\ \mathbf{Q}_{NL} &= \int_{\Theta_p}\int_{\Omega_p}\mathbf{G}_R(\omega,\theta)d\omega d\theta = \mathbf{Q}^p_e\ . \qquad (8.42) \end{aligned}$$

We refer to appendix E.4 for the calculation of $J_{sum}(\mathbf{w})$.

Non-linear optimisation technique

Minimising $J_{NL}(\mathbf{w})$ requires an iterative non-linear optimisation technique. We have used the MATLAB-function `fminunc` [35], which finds the minimum of an unconstrained multi-variable function (both a medium-scale quasi-Newton method with cubic polynomial line search and a large-scale subspace trust region method have been used [93]). In order to improve numerical robustness, the gradient

$$\frac{\partial J_{NL}(\mathbf{w})}{\partial\mathbf{w}} = \frac{\partial J_{sum}(\mathbf{w})}{\partial\mathbf{w}} - 4\mathbf{Q}_{NL}\mathbf{w} \qquad (8.43)$$

can be supplied analytically. In Appendix E.4 the calculation of the gradient $\frac{\partial J_{sum}(\mathbf{w})}{\partial\mathbf{w}}$ is discussed and it is shown that, cf. (E.70) and (E.71),

$$\frac{\partial J_{sum}(\mathbf{w})}{\partial\mathbf{w}} = 4\mathbf{Q}_{sum}(\mathbf{w})\cdot\mathbf{w} \qquad (8.44)$$

with

$$\mathbf{Q}_{sum}(\mathbf{w}) = \mathrm{Re}\left\{\int_{\Theta}\int_{\Omega}\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}(\omega,\theta)d\omega d\theta\right\}. \qquad (8.45)$$

Hence, the total gradient $\frac{\partial J_{NL}(\mathbf{w})}{\partial\mathbf{w}}$ can be written as

$$\frac{\partial J_{NL}(\mathbf{w})}{\partial\mathbf{w}} = 4\big(\mathbf{Q}_{sum}(\mathbf{w}) - \mathbf{Q}_{NL}\big)\mathbf{w}\ . \qquad (8.46)$$

Stationary points $\mathbf{w}_s$, i.e. filter coefficients $\mathbf{w}$ for which the gradient is 0, satisfy

$$\boxed{\big(\mathbf{Q}_{sum}(\mathbf{w}_s) - \mathbf{Q}_{NL}\big)\mathbf{w}_s = \mathbf{0}} \qquad (8.47)$$

This implies that for a stationary point, either $\mathbf{w}_s = \mathbf{0}$, $\mathbf{Q}_{sum}(\mathbf{w}_s) = \mathbf{Q}_{NL}$ or $\mathbf{w}_s$ lies in the null space of the matrix $\mathbf{Q}_{sum}(\mathbf{w}_s) - \mathbf{Q}_{NL}$. Simulations indicate that the several stationary points exist and that the latter condition is prevalent. Since $J_{sum}(\mathbf{w}) = \mathbf{w}^T\mathbf{Q}_{sum}(\mathbf{w})\mathbf{w}$, the cost function in a stationary point $\mathbf{w}_s$ is equal to

$$\begin{aligned} J_{NL}(\mathbf{w}_s) &= \mathbf{w}_s^T\mathbf{Q}_{sum}(\mathbf{w}_s)\mathbf{w}_s + d_{NL} - 2\mathbf{w}_s^T\mathbf{Q}_{NL}\mathbf{w}_s && (8.48)\\ &= d_{NL} - \mathbf{w}_s^T\mathbf{Q}_{NL}\mathbf{w}_s \leq d_{NL}\ . && (8.49)\end{aligned}$$

Since $J_{NL}(\mathbf{w}_s) \geq 0$, all stationary points are located in the region

$$\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w} \leq d_{NL}\ . \tag{8.50}$$

In order to improve the convergence speed and the numerical robustness of the large-scale algorithms, also the Hessian

$$\mathbf{H}_{NL}(\mathbf{w}) = \frac{\partial^2 J_{NL}(\mathbf{w})}{\partial^2\mathbf{w}} = \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}} - 4\mathbf{Q}_{NL} \tag{8.51}$$

can be provided. In Appendix E.4 the calculation of the Hessian $\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}}$ is discussed. Using (E.88),

$$\mathbf{w}^T\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}}\mathbf{w} = 12\,\mathbf{w}^T\mathbf{Q}_{sum}(\mathbf{w})\mathbf{w} = 12\,J_{sum}(\mathbf{w})\ , \tag{8.52}$$

the quadratic form $\mathbf{w}^T\mathbf{H}_{NL}(\mathbf{w})\mathbf{w}$ can be written as

$$\mathbf{w}^T\mathbf{H}_{NL}(\mathbf{w})\mathbf{w} = 12\mathbf{w}^T\mathbf{Q}_{sum}(\mathbf{w})\mathbf{w} - 4\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w} = 12J_{sum}(\mathbf{w}) - 4\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w}\ ,$$

which can be either positive or negative. Since $J_{NL}(\mathbf{w}) \geq 0$, it follows from (8.37) that

$$J_{sum}(\mathbf{w}) \geq 2\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w} - d_{NL}\ , \tag{8.53}$$

such that

$$\mathbf{w}^T\mathbf{H}_{NL}(\mathbf{w})\mathbf{w} \geq 20\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w} - 12d_{NL}\ . \tag{8.54}$$

Hence, if

$$\mathbf{w}^T\mathbf{Q}_{NL}\mathbf{w} \geq \frac{3}{5}d_{NL}\ , \tag{8.55}$$

then the quadratic form $\mathbf{w}^T\mathbf{H}_{NL}(\mathbf{w})\mathbf{w}$ is positive (leading to a convex optimisation problem). However, no conclusions can be drawn for $\mathbf{w}$ where this condition is not satisfied. In a stationary point $\mathbf{w}_s$, the quadratic form $\mathbf{w}^T\mathbf{H}_{NL}(\mathbf{w})\mathbf{w}$ is equal to

$$\mathbf{w}_s^T\mathbf{H}_{NL}(\mathbf{w}_s)\mathbf{w}_s = 12\,\mathbf{w}_s^T\mathbf{Q}_{sum}(\mathbf{w}_s)\mathbf{w}_s - 4\,\mathbf{w}_s^T\mathbf{Q}_{NL}\mathbf{w}_s = 8\,\mathbf{w}_s^T\mathbf{Q}_{NL}\mathbf{w}_s\ . \tag{8.56}$$

Since in general the matrix $\mathbf{Q}_{NL}$, defined in (8.40), is positive definite (only in very special cases $\mathbf{Q}_{NL}$ is singular and hence positive semi-definite), the quadratic form $\mathbf{w}_s^T\mathbf{H}_{NL}(\mathbf{w}_s)\mathbf{w}_s$ is strictly positive in all stationary points except for $\mathbf{w}_s = \mathbf{0}$, where it is equal to zero. Hence, *all stationary points are either local minima or saddle points.* For $\mathbf{w}_s = \mathbf{0}$, the Hessian $\mathbf{H}_{NL}(\mathbf{0}) = -4\mathbf{Q}_{NL}$ is negative definite, such that $\mathbf{w}_s = \mathbf{0}$ is the only local maximum.

Simulations have indicated that for each design problem a number of local minima exist, related to the symmetry present in the considered problem. E.g. if $\mathbf{w}_m$ is a local minimum, then $-\mathbf{w}_m$ is a local minimum and for a symmetric linear array $\mathbf{J}_M\mathbf{w}_m$ also is a local minimum, with $\mathbf{J}_M$ the $M \times M$ reversal matrix, cf. (A.8). In these local minima the cost function has the same value, since (for a symmetric linear array)

$$d_{NL} - \mathbf{w}_m^T\mathbf{Q}_{NL}\mathbf{w}_m = d_{NL} - (-\mathbf{w}_m^T)\mathbf{Q}_{NL}(-\mathbf{w}_m) = d_{NL} - \mathbf{w}_m^T\mathbf{J}_M\mathbf{Q}_{NL}\mathbf{J}_M\mathbf{w}_m \,.$$

Simulations have also shown that other local minima exist, which appear not to be (easily) related to $\mathbf{w}_m$. However, the cost function in all local minima seems to be approximately equal, such that any of these local minima can be used as the final solution for the broadband beamformer.

Linear constraints

Incorporating linear constraints $\mathbf{Cw} = \mathbf{b}$ can be done by using the MATLAB function `fmincon` [35], which finds the minimum of a constrained nonlinear multi-variable function (we have used the large-scale subspace trust region method, based on the interior-reflective Newton method using preconditioned conjugate gradients [93]).

8.4 Eigenfilter design procedures

In this section we present two novel non-iterative design procedures for broadband beamformers, which are based on eigenfilters. Eigenfilters have been introduced for designing 1-dimensional linear phase FIR filters [253]. Their main advantage is the fact that no matrix inversion is required (as in LS filter design) and that time-domain and frequency-domain constraints are easily incorporated. Eigenfilter techniques have also been applied for designing 2-dimensional FIR and spatial filters [32][208]. In this section, we extend the application domain of eigenfilters to the design of broadband beamformers.

In this section two eigenfilter cost functions will be considered:

- the conventional eigenfilter cost function J_{eig}, minimising the error between the spatial directivity patterns $D(\omega,\theta)\,H(\omega_c,\theta_c)/D(\omega_c,\theta_c)$ and $H(\omega,\theta)$. Note that a reference frequency-angle point (ω_c,θ_c) is required for this technique. Minimising this cost function with/without additional constraints leads to a (generalised) eigenvalue problem (cf. Section 8.4.1);

- the TLS eigenfilter cost function J_{TLS}, minimising the total least-squares (TLS) error between the actual and the desired spatial directivity pattern. This cost function does not require a reference point and also leads to a generalised eigenvalue problem (cf. Section 8.4.2).

8.4.1 Conventional eigenfilter technique

General design

In the conventional eigenfilter technique first a reference frequency-angle point (ω_c, θ_c) is chosen and the filter $\mathbf{w}$ is calculated such that the error between the spatial directivity patterns $H(\omega, \theta)$ and $D(\omega, \theta)\, H(\omega_c, \theta_c)/D(\omega_c, \theta_c)$ is minimised. Note that we do not specify the exact value of $H(\omega_c, \theta_c)$, which can however be done at a later stage by imposing a linear point constraint, cf. Section 8.5. The eigenfilter cost function is defined as

$$J_{eig}(\mathbf{w}) = \int_{\Theta}\int_{\Omega} F(\omega,\theta)\Big|\frac{D(\omega,\theta)}{D(\omega_c,\theta_c)} H(\omega_c,\theta_c) - H(\omega,\theta)\Big|^2 d\omega d\theta\ , \tag{8.57}$$

Using (8.6) it can be shown that $J_{eig}(\mathbf{w})$ is equal to the quadratic form

$$\boxed{J_{eig}(\mathbf{w}) = \mathbf{w}^T \mathbf{Q}_{eig} \mathbf{w}} \tag{8.58}$$

with $\mathbf{Q}_{eig}$ equal to

$$\int_{\Theta}\int_{\Omega} F(\omega,\theta)\, \mathrm{Re}\Big\{\Big[\frac{D(\omega,\theta)}{D(\omega_c,\theta_c)}\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)\Big]\cdot \Big[\frac{D(\omega,\theta)}{D(\omega_c,\theta_c)}\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)\Big]^H\Big\} d\omega d\theta\ . \tag{8.59}$$

When minimising the cost function $J_{eig}(\mathbf{w})$, an additional constraint is required in order to avoid the trivial solution $\mathbf{w} = \mathbf{0}$. Both a quadratic (energy) constraint and a linear constraint are possible.

Specific design case

For the specific design case, assuming that the reference point (ω_c, θ_c) does not belong to the stopband region (Θ_s, Ω_s), the cost function $J_{eig}(\mathbf{w})$ in (8.57) can be written as

$$J_{eig}(\mathbf{w}) = \int_{\Theta_p}\int_{\Omega_p} |H(\omega_c,\theta_c) - H(\omega,\theta)|^2 d\omega d\theta + \alpha \int_{\Theta_s}\int_{\Omega_s} |H(\omega,\theta)|^2 d\omega d\theta\ , \tag{8.60}$$

such that the matrix $\mathbf{Q}_{eig}$ is equal to

$$\mathbf{Q}_{eig} = \underbrace{\int_{\Theta_p}\int_{\Omega_p} \mathrm{Re}\big\{[\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)][\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)]^H\big\} d\omega d\theta}_{\mathbf{Q}_p} +$$

$$\underbrace{\alpha \int_{\Theta_s} \int_{\Omega_s} \mathbf{G}_R(\omega, \theta) d\omega d\theta}_{\mathbf{Q}_e^s} \tag{8.61}$$

The quantity $\mathbf{w}^T \mathbf{Q}_p \mathbf{w}$ is equal to the error in the passband, whereas $\mathbf{w}^T \mathbf{Q}_e^s \mathbf{w}$ is equal to the energy (=error) in the stopband, such that

$$J_{eig}(\mathbf{w}) = \mathbf{w}^T \underbrace{(\mathbf{Q}_p + \alpha \mathbf{Q}_e^s)}_{\mathbf{Q}_{eig}} \mathbf{w} \tag{8.62}$$

is a weighted error function over passband and stopband. The calculation of the integrals in (8.61) is discussed in Appendix E.2 and E.3.

Quadratic energy constraint

The most common constraint on the filter $\mathbf{w}$ is the unit-norm (quadratic) constraint $\mathbf{w}^T \mathbf{w} = 1$, which leads to the following eigenvalue problem,

$$\min_{\mathbf{w}} \mathbf{w}^T \mathbf{Q}_{eig} \mathbf{w}, \quad \text{subject to } \mathbf{w}^T \mathbf{w} = 1 \tag{8.63}$$

of which the solution is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{Q}_{eig}$ (hence the name eigenfilters).

In the 1-dimensional FIR filter design-case [253], this unit-norm constraint corresponds to the total area under the frequency response $|W(\omega)|^2$ being equal to 1, since using Parseval's theorem we can write

$$\int_0^{\pi} |W(\omega)|^2 \frac{d\omega}{\pi} = \mathbf{w}^T \mathbf{w} \ . \tag{8.64}$$

In broadband beamformer design, a unit-norm constraint apparently does not have a physical meaning any more. Hence, we have modified this quadratic constraint by constraining the total area under the spatial directivity spectrum $|H(\omega, \theta)|^2$ to be equal to 1, i.e.

$$\int_0^{\pi} \int_0^{\pi} |H(\omega, \theta)|^2 d\omega d\theta = \mathbf{w}^T \mathbf{Q}_e^{tot} \mathbf{w} = 1 \ , \tag{8.65}$$

with

$$\mathbf{Q}_e^{tot} = \int_0^{\pi} \int_0^{\pi} \mathbf{G}_R(\omega, \theta) d\omega d\theta \ . \tag{8.66}$$

The calculation of this integral is discussed in Appendix E.2. This constraint gives rise to the following constrained optimisation problem,

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^T \mathbf{Q}_{eig} \mathbf{w}, \quad \text{subject to } \mathbf{w}^T \mathbf{Q}_e^{tot} \mathbf{w} = 1} \tag{8.67}$$

of which the solution $\mathbf{w}_{eig}$ is the generalised eigenvector, corresponding to the minimum generalised eigenvalue in the GEVD of $\mathbf{Q}_{eig}$ and $\mathbf{Q}_e^{tot}$.

Linear constraints

Instead of imposing a quadratic constraint, it is also possible to impose linear constraints $\mathbf{Cw} = \mathbf{b}$ in order to avoid the trivial solution $\mathbf{w} = \mathbf{0}$. We then have to solve the constrained optimisation problem

$$\min_{\mathbf{w}} \mathbf{w}^T \mathbf{Q}_{eig} \mathbf{w}, \quad \text{subject to } \mathbf{Cw} = \mathbf{b} \,, \tag{8.68}$$

which is the same optimisation problem as (8.22) with $\mathbf{a} = \mathbf{0}$ and $d_{LS} = 0$, such that the solution (8.23) becomes

$$\mathbf{w}_{eig}^c = \mathbf{Q}_{eig}^{-1} \mathbf{C}^T (\mathbf{C} \mathbf{Q}_{eig}^{-1} \mathbf{C}^T)^{-1} \mathbf{b} \,. \tag{8.69}$$

8.4.2 Eigenfilter based on TLS error

General design

Recently, an eigenfilter, based on a TLS error criterion, has been described in [209] for designing 2-dimensional FIR filters. The advantage of this eigenfilter is that no reference point is required. We have extended this TLS eigenfilter technique to the design of broadband beamformers. Instead of minimising the LS error (cf. Section 8.3.2), the TLS error

$$\frac{|D(\omega,\theta) - H(\omega,\theta)|^2}{\mathbf{w}^T \mathbf{w} + 1} \tag{8.70}$$

is used and the cost function to be minimised can be written as

$$\bar{J}_{TLS}(\mathbf{w}) = \int_{\Theta} \int_{\Omega} F(\omega,\theta) \frac{|D(\omega,\theta) - H(\omega,\theta)|^2}{\mathbf{w}^T \mathbf{w} + 1} d\omega d\theta \,. \tag{8.71}$$

As in the conventional eigenfilter technique (cf. Section 8.4.1), we replace $\mathbf{w}^T\mathbf{w}$ with $\mathbf{w}^T \mathbf{Q}_e^{tot} \mathbf{w}$, which has a physical meaning, and instead minimise the cost function

$$J_{TLS}(\mathbf{w}) = \int_{\Theta} \int_{\Omega} F(\omega,\theta) \frac{|D(\omega,\theta) - H(\omega,\theta)|^2}{\mathbf{w}^T \mathbf{Q}_e^{tot} \mathbf{w} + 1} d\omega d\theta \,, \tag{8.72}$$

which can be written as

$$\boxed{J_{TLS}(\mathbf{w}) = \frac{\hat{\mathbf{w}}^T \hat{\mathbf{Q}}_{TLS} \hat{\mathbf{w}}}{\hat{\mathbf{w}}^T \hat{\mathbf{Q}}_e^{tot} \hat{\mathbf{w}}}} \tag{8.73}$$

with the extended vector $\hat{\mathbf{w}}$ and matrices $\hat{\mathbf{Q}}_{TLS}$ and $\hat{\mathbf{Q}}_e^{tot}$ defined as

$$\hat{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix}, \quad \hat{\mathbf{Q}}_{TLS} = \begin{bmatrix} \mathbf{Q}_{LS} & \mathbf{a} \\ \mathbf{a}^T & d_{LS} \end{bmatrix}, \quad \hat{\mathbf{Q}}_e^{tot} = \begin{bmatrix} \mathbf{Q}_e^{tot} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} . \tag{8.74}$$

The definitions of $\mathbf{Q}_{LS}$, $\mathbf{a}$ and d_{LS} are given in Section 8.3.2, while the definition of $\mathbf{Q}_e^{tot}$ is given in Section 8.4.1. The filter $\hat{\mathbf{w}}_{TLS}$ minimising (8.73) is the generalised eigenvalue of $\hat{\mathbf{Q}}_{TLS}$ and $\hat{\mathbf{Q}}_e^{tot}$, corresponding to the smallest generalised eigenvalue. After scaling the last element of $\hat{\mathbf{w}}_{TLS}$ to -1, the actual solution $\mathbf{w}_{TLS}$ is obtained as the first M elements of $\hat{\mathbf{w}}_{TLS}$.

It will be shown by simulations that the TLS eigenfilter technique has a better performance than the weighted LS, the maximum energy array and the conventional eigenfilter technique and therefore appears to be the preferred non-iterative design procedure.

Linear constraints

In [209] it is shown that linear constraints $\mathbf{Cw} = \mathbf{b}$ can be easily rewritten as

$$\hat{\mathbf{C}}\hat{\mathbf{w}} = \mathbf{0}, \qquad \hat{\mathbf{C}} = \left[\begin{array}{cc} \mathbf{C} & \mathbf{b} \end{array} \right], \tag{8.75}$$

such that the constrained optimisation problem can be rewritten as

$$\boxed{\min_{\hat{\mathbf{w}}} \frac{\hat{\mathbf{w}}^T \hat{\mathbf{Q}}_{TLS} \hat{\mathbf{w}}}{\hat{\mathbf{w}}^T \hat{\mathbf{Q}}_e^{tot} \hat{\mathbf{w}}}, \quad \text{subject to } \hat{\mathbf{C}}\hat{\mathbf{w}} = \mathbf{0}} \tag{8.76}$$

which is similar to (8.27) in Section 8.3.3. The solution $\tilde{\mathbf{w}}_{TLS}$ of the unconstrained optimisation problem is given by the generalised eigenvector corresponding to the minimum generalised eigenvalue of $\mathbf{B}\hat{\mathbf{Q}}_{TLS}\mathbf{B}^T$ and $\mathbf{B}\hat{\mathbf{Q}}_e^{tot}\mathbf{B}^T$ (with $\mathbf{B}$ the null space of $\hat{\mathbf{C}}$), such that the solution $\hat{\mathbf{w}}_{TLS}^c$ of the constrained optimisation problem (8.76) is equal to

$$\hat{\mathbf{w}}_{TLS}^c = \mathbf{B}^T \tilde{\mathbf{w}}_{TLS} \ . \tag{8.77}$$

After scaling the last element of $\hat{\mathbf{w}}_{TLS}^c$ to -1, the actual solution $\mathbf{w}_{TLS}^c$ is obtained as the first M elements of $\hat{\mathbf{w}}_{TLS}^c$.

8.5 Linear constraints

In this section several types of linear constraints are discussed, which can be imposed on the filter $\mathbf{w}$. These linear constraints can always be written in the form

$$\mathbf{Cw} = \mathbf{b} \tag{8.78}$$

with $\mathbf{C}$ a $J \times M$-dimensional matrix (with the number of constraints $J \leq M$) and $\mathbf{b}$ a J-dimensional vector. In this section point constraints, line constraints and derivative constraints are discussed.

8.5.1 Point constraints

Point constraints can be used for constraining the spatial directivity pattern $H(\omega, \theta)$ to some predefined value at a specific frequency-angle point. The absolute point constraint $H(\omega_f, \theta_f) = b$, with $b = b_R + jb_I$ a complex scalar, corresponds to 2 real-valued constraints,

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_f, \theta_f) \\ \mathbf{g}_I^T(\omega_f, \theta_f) \end{bmatrix}}_{\mathbf{C}} \mathbf{w} = \underbrace{\begin{bmatrix} b_R \\ b_I \end{bmatrix}}_{\mathbf{b}} , \tag{8.79}$$

whereas the relative point constraint $H(\omega_{f_1}, \theta_{f_1}) = b\, H(\omega_{f_2}, \theta_{f_2})$ can be written as

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_{f_1}, \theta_{f_1}) - b_R\, \mathbf{g}_R^T(\omega_{f_2}, \theta_{f_2}) + b_I\, \mathbf{g}_I^T(\omega_{f_2}, \theta_{f_2}) \\ \mathbf{g}_I^T(\omega_{f_1}, \theta_{f_1}) - b_I\, \mathbf{g}_R^T(\omega_{f_2}, \theta_{f_2}) - b_R\, \mathbf{g}_I^T(\omega_{f_2}, \theta_{f_2}) \end{bmatrix}}_{\mathbf{C}} \mathbf{w} = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{b}} . \tag{8.80}$$

8.5.2 Line constraint

Constraining the spatial directivity pattern $H(\omega, \theta)$ at the angle θ_f to a predefined frequency response $B(\omega) = \sum_{l=0}^{L-1} b_l\, e^{-jl\omega} = \mathbf{b}^T \mathbf{e}(\omega)$, with $\mathbf{b}$ a real-valued vector defined similarly as in (8.3), corresponds to

$$\begin{aligned} H(\omega, \theta_f) &= \sum_{n=0}^{N-1} W_n(\omega) e^{-j\omega\tau_n(\theta_f)} = \sum_{l=0}^{L-1} \left(\sum_{n=0}^{N-1} w_{n,l}\, e^{-j\omega\tau_n(\theta_f)} \right) e^{-jl\omega} \\ &\triangleq B(\omega) = \sum_{l=0}^{L-1} b_l\, e^{-jl\omega} . \end{aligned} \tag{8.81}$$

Obviously, this can be done by putting

$$\sum_{n=0}^{N-1} w_{n,l}\, e^{-j\omega\tau_n(\theta_f)} = b_l , \quad l = 0 \ldots L-1 , \tag{8.82}$$

which can be written as

$$\begin{bmatrix} e^{-j\omega\tau_0(\theta_f)} \cdot \mathbf{I}_L & e^{-j\omega\tau_1(\theta_f)} \cdot \mathbf{I}_L & \ldots & e^{-j\omega\tau_{N-1}(\theta_f)} \cdot \mathbf{I}_L \end{bmatrix} \mathbf{w} = \mathbf{b} , \tag{8.83}$$

and corresponds to $2L$ real-valued constraints,

$$\begin{bmatrix} \cos\big(\omega\tau_0(\theta_f)\big)\, \mathbf{I}_L & \cos\big(\omega\tau_1(\theta_f)\big)\, \mathbf{I}_L & \ldots & \cos\big(\omega\tau_{N-1}(\theta_f)\big)\, \mathbf{I}_L \\ \sin\big(\omega\tau_0(\theta_f)\big)\, \mathbf{I}_L & \sin\big(\omega\tau_1(\theta_f)\big)\, \mathbf{I}_L & \ldots & \sin\big(\omega\tau_{N-1}(\theta_f)\big)\, \mathbf{I}_L \end{bmatrix} \mathbf{w} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} . \tag{8.84}$$

This equation has to hold *for all* ω. However, since $J = 2L \leq M$, in general these constraints can be satisfied maximally for $N/2$ frequency points. An exception is $\theta_f = \pi/2$ (i.e. broadside direction), since in this case $\tau_n(\theta_f) = 0$, $n = 0 \ldots N-1$, such that (8.84) reduces to

$$\underbrace{\begin{bmatrix} \mathbf{I}_L & \mathbf{I}_L & \ldots & \mathbf{I}_L \end{bmatrix}}_{\mathbf{C}} \mathbf{w} = \mathbf{b} . \tag{8.85}$$

8.5.3 Derivative constraints

In order to smoothen the spatial directivity pattern $H(\omega, \theta)$, we can introduce derivative constraints, e.g. flattening the spatial directivity pattern at certain frequencies and angles by putting the frequency and/or angle derivatives to 0 [87], i.e.

$$\left.\frac{\partial H(\omega,\theta)}{\partial\theta}\right|_{\substack{\omega=\omega_f\\ \theta=\theta_f}} = 0, \qquad \left.\frac{\partial H(\omega,\theta)}{\partial\omega}\right|_{\substack{\omega=\omega_f\\ \theta=\theta_f}} = 0\ . \tag{8.86}$$

Since $H(\omega,\theta) = \mathbf{w}^T\mathbf{g}(\omega,\theta)$, these derivatives are equal to

$$\frac{\partial H(\omega,\theta)}{\partial\theta} = \mathbf{w}^T \underbrace{\frac{\partial \mathbf{g}(\omega,\theta)}{\partial\theta}}_{\mathbf{g}'_\theta(\omega,\theta)}, \qquad \frac{\partial H(\omega,\theta)}{\partial\omega} = \mathbf{w}^T \underbrace{\frac{\partial \mathbf{g}(\omega,\theta)}{\partial\omega}}_{\mathbf{g}'_\omega(\omega,\theta)}\ . \tag{8.87}$$

Using (8.5) and (8.7), the derivative $\mathbf{g}'_\theta(\omega,\theta)$ can be written as

$$\begin{aligned}\mathbf{g}'_\theta(\omega,\theta) &= \frac{\partial}{\partial\theta}\begin{bmatrix}\mathbf{e}(\omega)e^{-j\omega\tau_0(\theta)}\\ \mathbf{e}(\omega)e^{-j\omega\tau_1(\theta)}\\ \vdots\\ \mathbf{e}(\omega)e^{-j\omega\tau_{N-1}(\theta)}\end{bmatrix} = j\omega\frac{f_s}{c}\sin\theta\begin{bmatrix}\mathbf{e}(\omega)e^{-j\omega\tau_0(\theta)}d_0\\ \mathbf{e}(\omega)e^{-j\omega\tau_1(\theta)}d_1\\ \vdots\\ \mathbf{e}(\omega)e^{-j\omega\tau_{N-1}(\theta)}d_{N-1}\end{bmatrix}\\ &= j\omega\frac{f_s}{c}\sin\theta\,\mathbf{\Delta}_\theta\,\mathbf{g}(\omega,\theta)\ ,\end{aligned} \tag{8.88}$$

with $\mathbf{\Delta}_\theta$ an $M \times M$-dimensional diagonal matrix, containing the microphone distances,

$$\mathbf{\Delta}_\theta = \begin{bmatrix} d_0\,\mathbf{I}_L & & & \\ & d_1\,\mathbf{I}_L & & \\ & & \ddots & \\ & & & d_{N-1}\,\mathbf{I}_L\end{bmatrix}. \tag{8.89}$$

If $\sin\theta_f = 0$, i.e. $\theta_f = 0$ or $\theta_f = \pi$, the first-order angle derivative constraint $\mathbf{w}^T\mathbf{g}'_\theta(\omega_f,\theta_f) = 0$ is satisfied for all frequencies. For all other angles, this constraint can be written as

$$\left[\,\mathbf{e}^H(\omega_f)e^{j\omega_f\tau_0(\theta_f)}d_0\ \ \mathbf{e}^H(\omega_f)e^{j\omega_f\tau_1(\theta_f)}d_1\ \ \ldots\ \ \mathbf{e}^H(\omega_f)e^{j\omega_f\tau_{N-1}(\theta_f)}d_{N-1}\,\right]\mathbf{w} = 0,$$

which corresponds to 2 real-valued linear constraints,

$$\underbrace{\begin{bmatrix}\mathbf{g}_R^T(\omega_f,\theta_f)\\ \mathbf{g}_I^T(\omega_f,\theta_f)\end{bmatrix}\mathbf{\Delta}_\theta}_{\mathbf{C}}\,\mathbf{w} = \underbrace{\begin{bmatrix}0\\0\end{bmatrix}}_{\mathbf{b}}\ . \tag{8.90}$$

The derivative $\mathbf{g}'_\omega(\omega,\theta)$ can be written as

$$\mathbf{g}'_\omega(\omega,\theta) = \frac{\partial}{\partial\omega}\begin{bmatrix} \mathbf{e}(\omega)e^{-j\omega\tau_0(\theta)} \\ \mathbf{e}(\omega)e^{-j\omega\tau_1(\theta)} \\ \vdots \\ \mathbf{e}(\omega)e^{-j\omega\tau_{N-1}(\theta)} \end{bmatrix} = \begin{bmatrix} e^{-j\omega\tau_0(\theta)}\Big(\mathbf{e}'(\omega) - \mathbf{e}(\omega)j\tau_0(\theta)\Big) \\ e^{-j\omega\tau_1(\theta)}\Big(\mathbf{e}'(\omega) - \mathbf{e}(\omega)j\tau_1(\theta)\Big) \\ \vdots \\ e^{-j\omega\tau_{N-1}(\theta)}\Big(\mathbf{e}'(\omega) - \mathbf{e}(\omega)j\tau_{N-1}(\theta)\Big) \end{bmatrix},$$

with

$$\mathbf{e}'(\omega) = \begin{bmatrix} 0 \\ -j\,e^{-j\omega} \\ \vdots \\ -j(L-1)\,e^{-j(L-1)\omega} \end{bmatrix} = -j\,\mathbf{D}\,\mathbf{e}(\omega)\;, \tag{8.91}$$

and $\mathbf{D}$ an $L \times L$ diagonal matrix,

$$\mathbf{D} = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & L-1 \end{bmatrix}, \tag{8.92}$$

such that

$$\mathbf{g}'_\omega(\omega,\theta) = -j\boldsymbol{\Delta}_\omega(\theta)\,\mathbf{g}(\omega,\theta)\;, \tag{8.93}$$

with $\boldsymbol{\Delta}_\omega(\theta)$ an $M \times M$ diagonal matrix,

$$\boldsymbol{\Delta}_\omega(\theta) = \begin{bmatrix} \mathbf{D}+\tau_0(\theta)\,\mathbf{I}_L & & & \\ & \mathbf{D}+\tau_1(\theta)\,\mathbf{I}_L & & \\ & & \ddots & \\ & & & \mathbf{D}+\tau_{N-1}(\theta)\,\mathbf{I}_L \end{bmatrix}. \tag{8.94}$$

The first-order frequency derivative constraint $\mathbf{w}^T\mathbf{g}'_\omega(\omega_f,\theta_f) = 0$ now corresponds to 2 linear constraints

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_f,\theta_f) \\ \mathbf{g}_I^T(\omega_f,\theta_f) \end{bmatrix}\boldsymbol{\Delta}_\omega(\theta_f)}_{\mathbf{C}}\,\mathbf{w} = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{b}}. \tag{8.95}$$

In order to further smoothen the spatial directivity pattern, higher-order frequency and/or angle derivatives can be set to zero. Using (8.88), the second-order angle derivative $\mathbf{g}''_\theta(\omega,\theta)$ can be written as

$$\begin{aligned} \mathbf{g}''_\theta(\omega,\theta) &= \frac{\partial\mathbf{g}'_\theta(\omega,\theta)}{\partial\theta} = j\omega\frac{f_s}{c}\boldsymbol{\Delta}_\theta\Big(\cos\theta\,\mathbf{g}(\omega,\theta) + \sin\theta\,\mathbf{g}'_\theta(\omega,\theta)\Big) && (8.96) \\ &= j\omega\frac{f_s}{c}\boldsymbol{\Delta}_\theta\Big(\cos\theta\,\mathbf{I}_M + j\omega\frac{f_s}{c}\sin^2\theta\,\boldsymbol{\Delta}_\theta\Big)\mathbf{g}(\omega,\theta)\;. && (8.97) \end{aligned}$$

Similar expressions can be derived for higher-order angle derivatives. Using (8.93), the rth-order frequency derivative $\mathbf{g}_\omega^{(r)}(\omega, \theta)$ can be written as

$$\mathbf{g}_\omega^{(r)}(\omega, \theta) = \big(-j\boldsymbol{\Delta}_\omega(\theta)\big)^r \, \mathbf{g}(\omega, \theta) \; . \tag{8.98}$$

8.6 Simulations

In this section, simulation results for far-field broadband beamformer design are discussed for the specific design case with $D(\omega, \theta) = 1$ in the passband and $D(\omega, \theta) = 0$ in the stopband. We have performed simulations for all cost functions, which have been discussed in Sections 8.3 and 8.4, using a linear uniform microphone array with $N = 5$ microphones, an inter-microphone distance $d = 4\,\text{cm}$ and sampling frequency $f_s = 8\,\text{kHz}$. Two specifications for the passband and the stopband regions have been considered:

- specification 1: passband $(\Omega_p, \Theta_p) = (300\text{–}4000\,\text{Hz}, 70°\text{–}110°)$ and stopband $(\Omega_s, \Theta_s) = (300\text{–}4000\,\text{Hz}, 0°\text{–}60° + 120°\text{–}180°)$
- specification 2: passband $(\Omega_p, \Theta_p) = (300\text{–}4000\,\text{Hz}, 40°\text{–}80°)$ and stopband $(\Omega_s, \Theta_s) = (300\text{–}4000\,\text{Hz}, 0°\text{–}30° + 90°\text{–}180°)$

For the first specification, we have performed simulations without linear constraints and with a line constraint at 90°, whereas for the second specification, we have only performed simulations without linear constraints. For the conventional eigenfilter technique, the reference point for the first specification is $(\omega_c, \theta_c) = (1500\,\text{Hz}, 90°)$ and the reference point for the second specification is $(\omega_c, \theta_c) = (1500\,\text{Hz}, 60°)$. Both for the conventional eigenfilter technique and for the TLS eigenfilter technique, the matrix $\mathbf{Q}_e^{tot}$ is computed with frequency and angle specifications $(\Omega, \Theta) = (300\text{–}4000\,\text{Hz}, 0°\text{–}180°)$.

All broadband beamformers have been designed using the following parameters: filter length $L = 20$ and stopband weight $\alpha = 0.1, 1, 10$. For all beamformers we have computed the different cost functions J_{LS}, J_{eig}, J_{TLS}, J_{ME} and J_{NL}, which have been defined in Sections 8.3 and 8.4. We will plot the total spatial directivity pattern $H(\omega, \theta)$ in the frequency-angle region $(\Omega, \Theta) = (300\text{–}3500\,\text{Hz}, 0°\text{–}180°)$ and the angular pattern for the specific frequencies $(500, 1000, 1500, 2000, 2500, 3500)\,\text{Hz}$.

8.6.1 Design specification 1

Considering the first design specification *without linear constraints*, the different cost functions for the different beamformer design procedures are summarised in Table 8.1. Obviously, the design procedure optimising a specific cost function gives rise to the best value for this particular cost function (bold values).

		Cost function				
Design	α	J_{LS}	J_{eig}	J_{TLS}	J_{ME}	J_{NL}
LS	0.1	**0.07015**	0.02688	0.01803	3.87628	0.07734
EIG	0.1	0.08169	**0.02179**	0.02008	4.02636	0.06917
TLS	0.1	0.07234	0.02593	**0.01752**	3.51239	0.06759
ME	0.1	2824.61	0.92219	0.92061	**130.189**	$5.10\ 10^7$
NL	0.1	0.63243	0.15624	0.14475	2.97090	**0.02540**
LS	1	**0.32012**	0.12644	0.10712	7.82490	0.24624
EIG	1	0.44332	**0.10786**	0.12097	10.9793	0.29769
TLS	1	0.34927	0.12651	**0.09851**	7.72356	0.18891
ME	1	2844.15	0.92856	0.92698	**130.189**	$5.10\ 10^7$
NL	1	0.84110	0.24517	0.22330	5.24686	**0.10301**
LS	10	**1.00743**	0.58272	0.56422	17.83966	0.97683
EIG	10	2.10339	**0.44667**	0.51747	35.37774	2.52124
TLS	10	1.35343	0.54114	**0.44637**	22.22030	0.37251
ME	10	3039.51	0.99225	0.99065	**130.189**	$5.10\ 10^7$
NL	10	4.08658	1.61600	1.29464	18.66897	**0.21410**

Table 8.1: Different cost functions for design specification 1 without linear constraints ($N = 5$; $L = 20$; $\alpha = 0.1, 1, 10$)

We will now compare the performance of the non-iterative design procedures (LS, EIG, TLS, ME) with the non-linear design procedure (NL) and determine which non-iterative design procedure has the best performance, using the non-linear cost function J_{NL} as a performance criterion. The maximum energy array technique has a quite poor performance (this can also be seen from the spatial directivity pattern in Fig. 8.6). In addition, the TLS eigenfilter technique always has a better performance than the weighted LS technique (this is also true for other filter lengths and number of microphones). For small stopband weights α, the conventional eigenfilter technique also gives rise to a better performance than the weighted LS technique (and even the TLS eigenfilter technique), but this is not true any more for large stopband weights. Therefore, *the TLS eigenfilter technique appears to be the preferred non-iterative design procedure*, best resembling the performance of the non-linear design procedure but having a significantly lower computational complexity.

Figures 8.3, 8.4, 8.5, 8.6 and 8.7 show the spatial directivity patterns for all design procedures with $\alpha = 1$. Figures 8.8 and 8.9 show the spatial directivity pattern for the TLS eigenfilter technique and the non-linear criterion with $\alpha = 0.1$. Figures 8.10 and 8.11 show the spatial directivity pattern for the TLS eigenfilter technique and the non-linear criterion with $\alpha = 10$.

When a *line constraint at* 90° is imposed, one can see by comparing Tables 8.1 and 8.2 that the cost functions with a line constraint are worse than the cost functions without constraints, but that all design procedures now give rise to

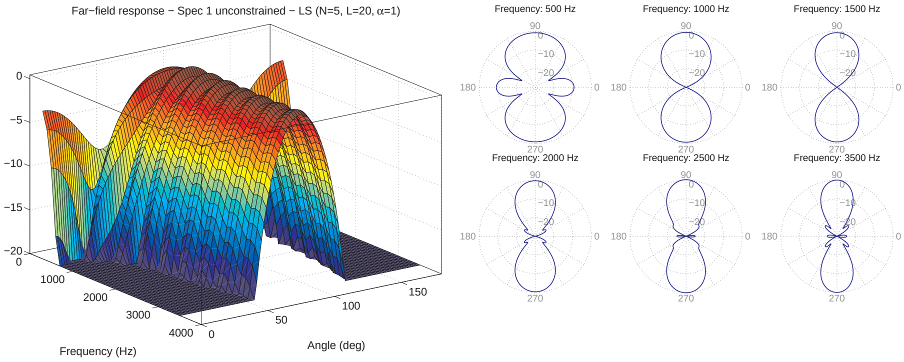


Figure 8.3: Weighted LS technique (design specification 1, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

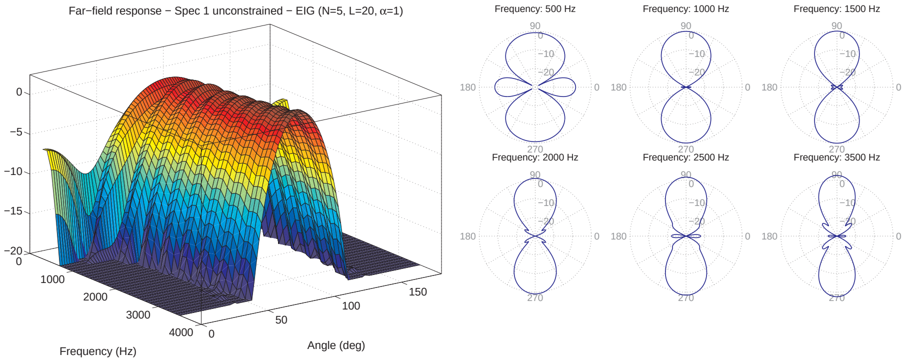


Figure 8.4: Conventional eigenfilter technique (design specification 1, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

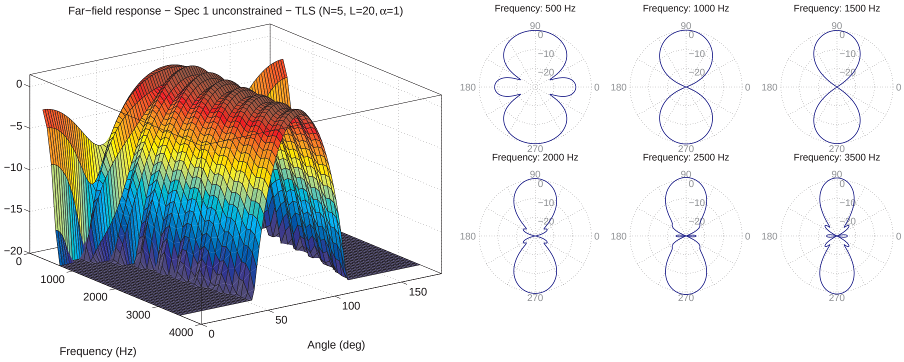


Figure 8.5: TLS eigenfilter technique (design specification 1, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

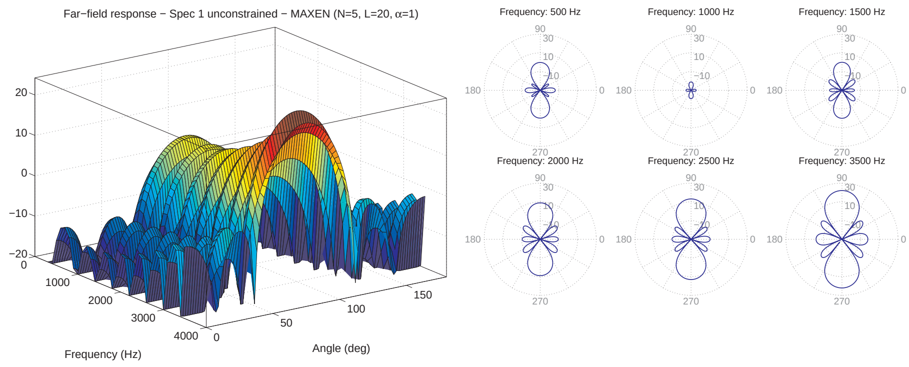


Figure 8.6: Maximum energy array technique (design specification 1, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

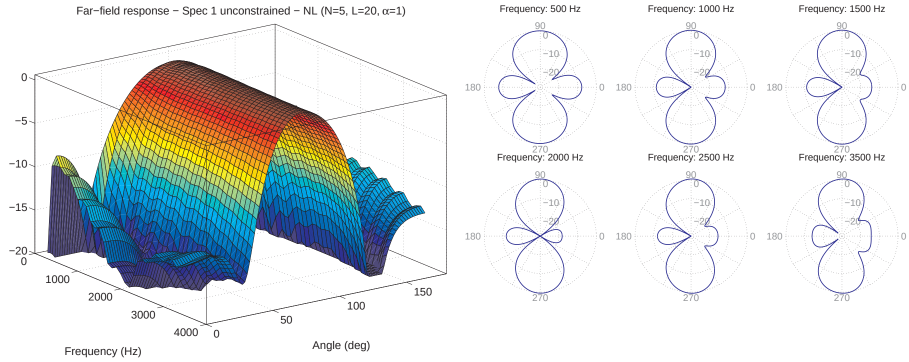


Figure 8.7: Non-linear criterion (design specification 1, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

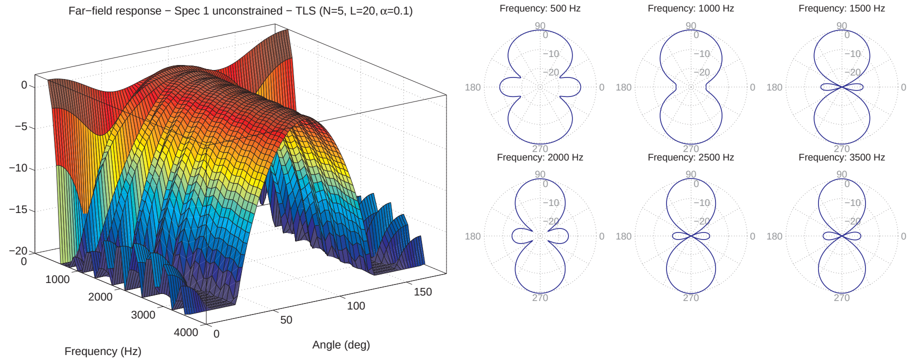


Figure 8.8: TLS eigenfilter technique (design specification 1, no linear constraints, $\alpha = 0.1$, $N = 5$, $L = 20$)

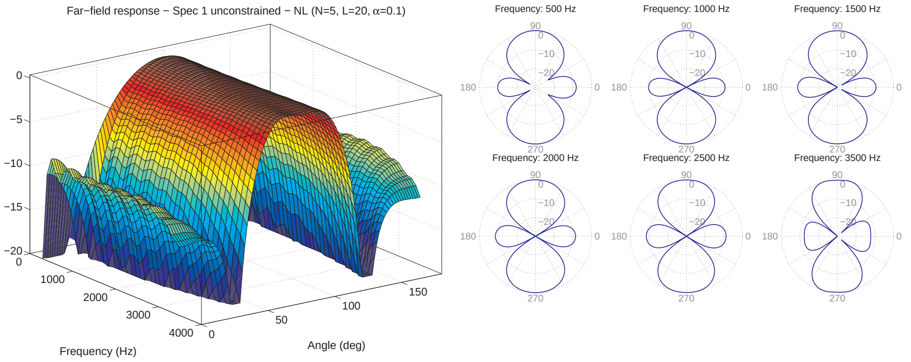


Figure 8.9: Non-linear criterion (design specification 1, no linear constraints, $\alpha = 0.1$, $N = 5$, $L = 20$)

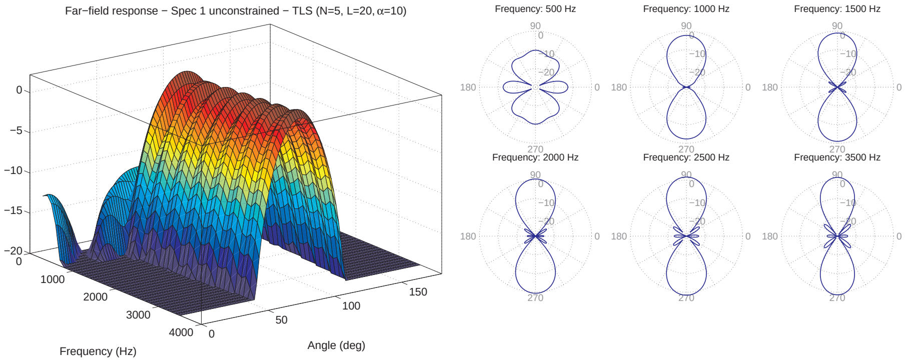


Figure 8.10: TLS eigenfilter technique (design specification 1, no linear constraints, $\alpha = 10$, $N = 5$, $L = 20$)

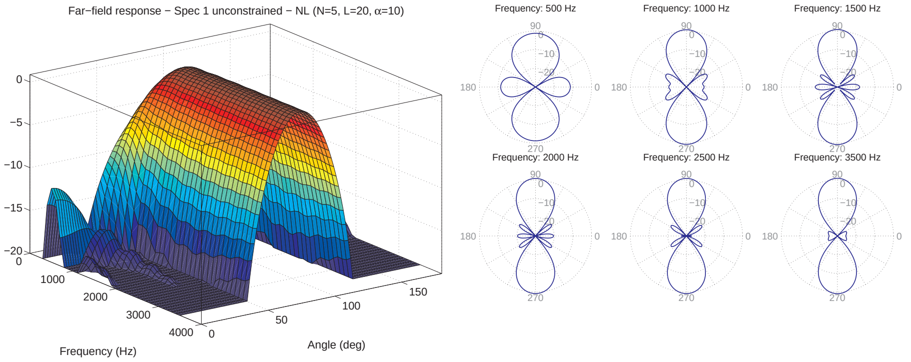


Figure 8.11: Non-linear criterion (design specification 1, no linear constraints, $\alpha = 10$, $N = 5$, $L = 20$)

		Cost function				
Design	α	J_{LS}	J_{eig}	J_{TLS}	J_{ME}	J_{NL}
LS	10	**3.96204**	1.74435	1.21113	4.05166	1.85361
EIG	10	3.96204	**1.74435**	1.21113	4.05166	1.85361
TLS	10	3.99103	1.72361	**1.20375**	4.06720	1.83286
ME	10	3.97901	1.72672	1.20416	**4.06885**	1.84120
NL	10	5.01514	2.09900	1.47970	3.19583	**1.29136**

Table 8.2: Different cost functions for design specification 1 with line constraint ($N = 5$; $L = 20$; $\alpha = 10$)

		Cost function				
Design	α	J_{LS}	J_{eig}	J_{TLS}	J_{ME}	J_{NL}
LS	1	**0.50350**	0.24804	0.18191	4.62621	0.40657
EIG	1	3.54617	**0.15078**	0.94322	8.06733	0.29521
TLS	1	0.58258	0.24821	**0.15872**	4.58828	0.25312
ME	1	287.043	0.89290	0.87877	**38.9523**	252775
NL	1	1.98809	0.86805	0.54727	6.58217	**0.10891**

Table 8.3: Different cost functions for design specification 2 without linear constraints ($N = 5$; $L = 20$; $\alpha = 1$)

quite similar results (also the maximum energy array technique). Again, the TLS eigenfilter technique has a better performance, i.e. non-linear cost function J_{NL}, than the weighted LS, the maximum energy array and the conventional eigenfilter technique, such that it appears to be the preferred non-iterative design procedure. Figure 8.12 shows the spatial directivity pattern of the maximum energy array technique with $\alpha = 10$.

8.6.2 Design specification 2

Considering the second design specification *without linear constraints*, the different cost functions for the different beamformer design procedures are summarised in Table 8.3 ($\alpha = 1$). Again, the maximum energy array technique has a quite poor performance. In addition, the TLS eigenfilter technique again has a better performance, i.e. non-linear cost function J_{NL}, than the weighted LS, the maximum energy array and the conventional eigenfilter technique and therefore appears to be the preferred non-iterative design procedure. Figures 8.13 and 8.14 show the spatial directivity patterns for the TLS eigenfilter technique and the non-linear criterion with $\alpha = 1$. Figures 8.15 and 8.16 show the spatial directivity patterns for the TLS eigenfilter technique and the non-linear criterion with $\alpha = 0.1$.

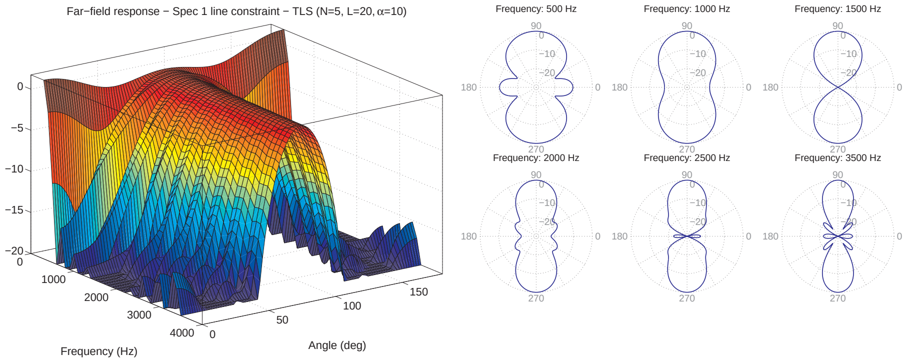


Figure 8.12: Maximum energy array technique (design specification 1, line constraint, $\alpha = 10$, $N = 5$, $L = 20$)

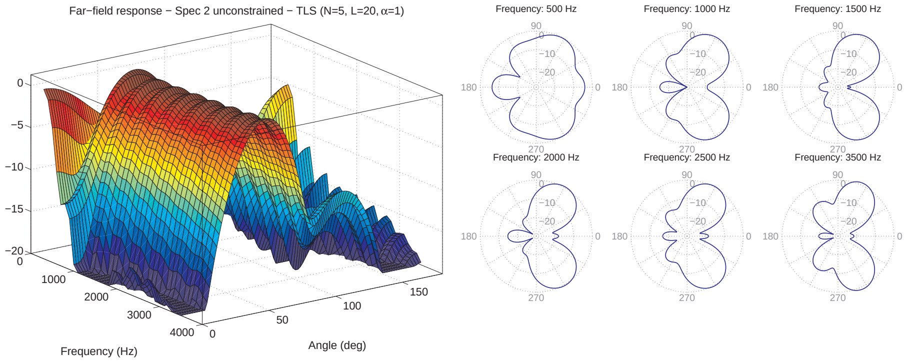


Figure 8.13: TLS eigenfilter technique (design specification 2, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

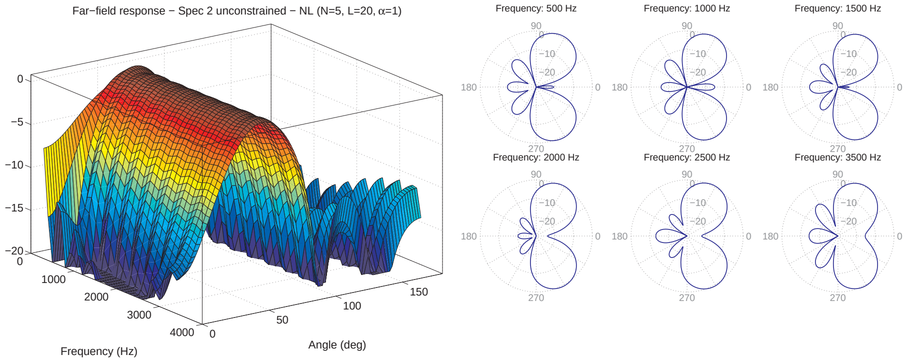


Figure 8.14: Non-linear criterion (design specification 2, no linear constraints, $\alpha = 1$, $N = 5$, $L = 20$)

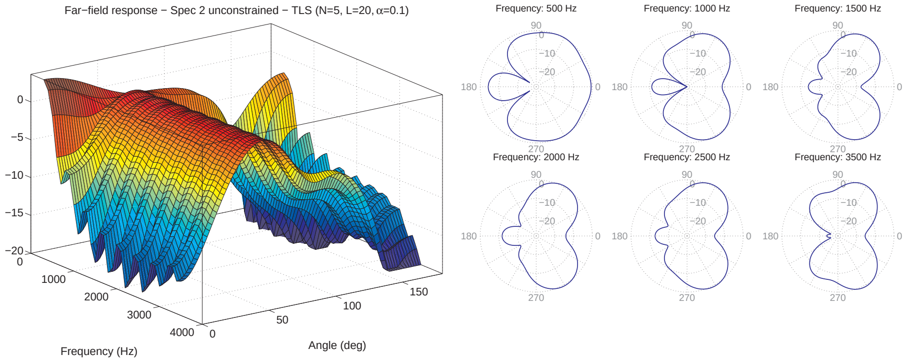


Figure 8.15: TLS eigenfilter technique (design specification 2, no linear constraints, $\alpha = 0.1$, $N = 5$, $L = 20$)

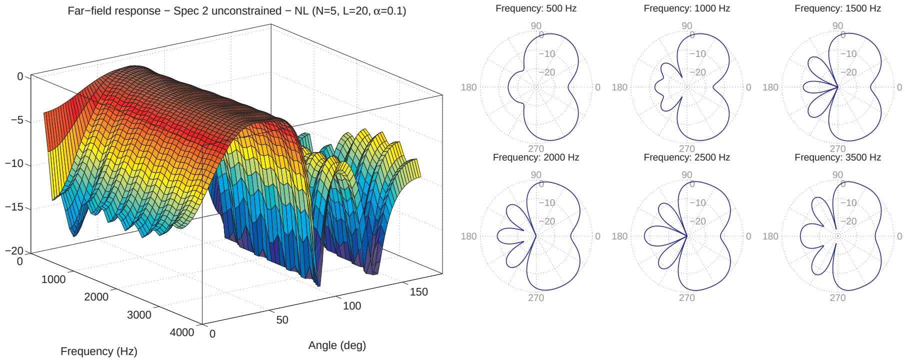


Figure 8.16: Non-linear criterion (design specification 2, no linear constraints, $\alpha = 0.1$, $N = 5$, $L = 20$)

8.7 Conclusion

In this chapter we have described several design procedures for designing far-field broadband beamformers with an arbitrary spatial directivity pattern using an arbitrary microphone configuration and an FIR filter-and-sum structure. In Section 8.3 several cost functions have been defined: a weighted LS cost function, a maximum energy array cost function, and a non-linear criterion, not taking into account the phase of the spatial directivity patterns. In Section 8.3.4 we have proposed a modified non-linear cost function, such that the double integrals arising in the optimisation problem only need to be computed once. However, optimising this non-linear cost function requires an iterative optimisation technique, giving rise to a large computational complexity. In Section 8.4 we have presented two novel non-iterative design procedures, which are based on eigenfilters. In the conventional eigenfilter technique, a reference frequency-angle point is required, whereas this reference point is not required in the TLS eigenfilter technique, which minimises the TLS error between the actual and the desired spatial directivity pattern. In Section 8.5 several linear constraints (point, line and derivative constraints) have been discussed that can be imposed on the filter coefficients. In Section 8.6 different simulations have shown that among all considered non-iterative design procedures the TLS eigenfilter technique has the best performance, i.e. best resembling the performance of the non-linear design procedure but having a significantly lower computational complexity.

Chapter 9

Near-Field Broadband Beamforming

This chapter discusses the design of *near-field broadband beamformers*. The ultimate goal is to design a broadband beamformer whose spatial directivity pattern optimally fits a desired spatial directivity pattern *for all distances* from the microphone array. However, in this chapter we will only consider the design of near-field broadband beamformers which operate at one specific distance or at a limited number of distances from the microphone array.

Section 9.1 describes the near-field configuration. In Section 9.2 it is shown that the design of near-field beamformers for one specific distance is very similar to the design of far-field beamformers. The same design procedures and cost functions can be used; the only difference lies in the calculation of the double integrals involved. This section also discusses design procedures for broadband beamformers which operate at several distances. Although this extension is straightforward for most cost functions, for the TLS eigenfilter and the maximum energy array cost function this extension leads to a significantly different optimisation problem, for which no closed-form solution is available.

Section 9.3 discusses linear constraints for the near-field case. Only point constraints and derivative constraints are discussed, since line constraints can not be defined for the near-field case.

Section 9.4 gives simulations results for near-field broadband beamformers operating at one specific distance and for mixed near-field far-field broadband beamforming. It is shown that the TLS eigenfilter technique again is the preferred non-iterative design procedure and that mixed near-field far-field design provides a trade-off between the near-field and the far-field performance.

9.1 Near-field configuration

When the speech source is close to the microphone array, the far-field assumptions are no longer valid and spherical wavefronts and signal attenuation have to be taken into account. Consider the linear microphone array depicted in Fig. 9.1, where the speech source $S(\omega)$ is located at a distance r from the centre of the microphone array and with the angle θ as defined in the figure. As already mentioned in Section 1.3.4, the typical rule of thumb is that the far-field assumptions are no longer valid when

$$r < \frac{d_{tot}^2 f_s}{c} \ , \tag{9.1}$$

with r the distance of the speech source to the centre of the microphone array and $d_{tot} = d_{N-1} - d_0$ the total length of the (linear) microphone array [169]. Hence, in the near-field of a microphone array, not only the direction θ of the speech source, but also its distance r to the microphone array has to be taken into account. E.g. in [153][225] superdirective beamformers and frequency-invariant beamformers have been designed for the near-field case .

In this chapter we will discuss the design of near-field broadband beamformers with an arbitrary desired spatial directivity pattern using an FIR filter-and-sum structure. It will be shown that *the design of near-field broadband beamformers is very similar to the design of far-field broadband beamformers (which are actually a special case for $r = \infty$). All the cost functions in Sections 8.3 and 8.4 remain valid, whereas only the steering vector $\mathbf{g}(\omega, \theta)$ in (8.7) and all related quantities are defined differently for the near-field case.*

Using simple geometrical relationships, the distance $r_n(\theta, r)$ from the signal source to the nth microphone is equal to

$$r_n(\theta, r) = \sqrt{(r \sin\theta)^2 + (d_n + r\cos\theta)^2} = \sqrt{r^2 + d_n^2 + 2d_n r \cos\theta} \ . \tag{9.2}$$

For convenience this equation can be rewritten as

$$\boxed{r_n(\theta, r) = \sqrt{p_n + q_n \cos\theta}} \tag{9.3}$$

with

$$p_n = r^2 + d_n^2, \qquad q_n = 2d_n r \ . \tag{9.4}$$

Taking into account spherical wavefronts and signal attenuation, the microphone signals $Y_n(\omega, \theta, r)$ are delayed and attenuated versions of the signal $\bar{Y}(\omega, \theta, r)$ at the centre of the microphone array, i.e.

$$Y_n(\omega, \theta, r) = a_n(\theta, r) e^{-j\omega\tau_n(\theta, r)} \bar{Y}(\omega, \theta, r), \ -\pi \leq \omega \leq \pi, \ -\pi \leq \theta \leq \pi \ , \tag{9.5}$$

with the attenuation $a_n(\theta, r)$ and the delay $\tau_n(\theta, r)$ equal to

$$\boxed{a_n(\theta, r) = \frac{r}{r_n(\theta, r)} \qquad \tau_n(\theta, r) = \frac{r_n(\theta, r) - r}{c} f_s} \tag{9.6}$$

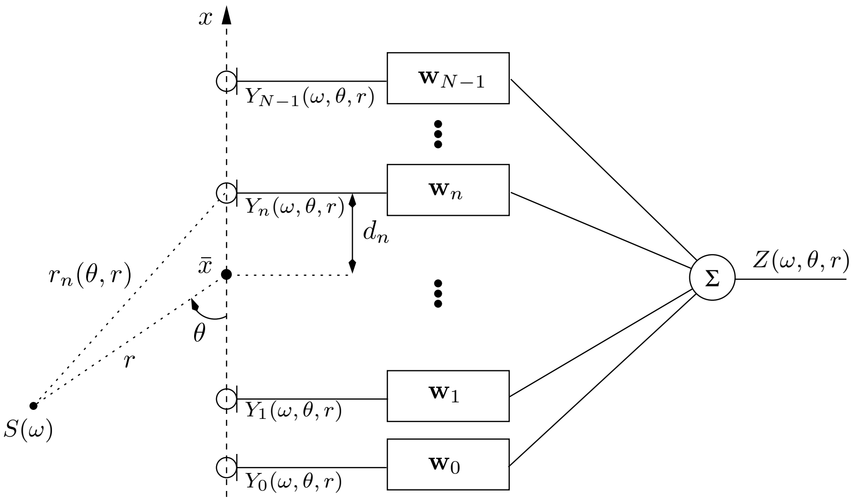


Figure 9.1: Linear microphone array configuration for near-field

Remark 9.1 For $r \to \infty$, i.e. far-field assumptions, it can be proved that

$$\lim_{r\to\infty} a_n(\theta,r) = \lim_{r\to\infty} \frac{r}{\sqrt{r^2+d_n^2+2d_n r\cos\theta}} = 1 \tag{9.7}$$

$$\begin{aligned}\lim_{r\to\infty} \tau_n(\theta,r) &= \lim_{r\to\infty} \frac{\sqrt{r^2+d_n^2+2d_n r\cos\theta}-r}{c} f_s \\ &= \lim_{r\to\infty} \frac{r\left[1+\frac{1}{2}\left(\frac{d_n^2}{r^2}+\frac{2d_n\cos\theta}{r}\right)-1\right]}{c} f_s = \frac{d_n\cos\theta}{c} f_s\ ,\end{aligned} \tag{9.8}$$

which are the far-field expressions. △

The spatial directivity pattern $H(\omega,\theta,r)$ is defined as

$$H(\omega,\theta,r) = \frac{Z(\omega,\theta,r)}{\bar{Y}(\omega,\theta,r)} = \frac{\sum_{n=0}^{N-1} W_n(\omega) Y_n(\omega,\theta,r)}{\bar{Y}(\omega,\theta,r)}\ . \tag{9.9}$$

Using (9.6), the spatial directivity pattern $H(\omega,\theta,r)$ can be written as

$$\boxed{H(\omega,\theta,r) = \sum_{n=0}^{N-1} a_n(\theta,r) W_n(\omega) e^{-j\omega\tau_n(\theta,r)} = \mathbf{w}^T \mathbf{g}(\omega,\theta,r)} \tag{9.10}$$

with the M-dimensional steering vector $\mathbf{g}(\omega,\theta,r)$ now dependent of r,

$$\mathbf{g}(\omega,\theta,r) = \begin{bmatrix} a_0(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_0(\theta,r)} \\ a_1(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_1(\theta,r)} \\ \vdots \\ a_{N-1}(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_{N-1}(\theta,r)} \end{bmatrix}\ . \tag{9.11}$$

As in the far-field case, the steering vector $\mathbf{g}(\omega, \theta, r)$ can be decomposed into a real part $\mathbf{g}_R(\omega, \theta, r)$ and an imaginary part $\mathbf{g}_I(\omega, \theta, r)$. Using (9.10), the spatial directivity spectrum $|H(\omega, \theta, r)|^2$ can be written as

$$|H(\omega, \theta, r)|^2 = H(\omega, \theta, r)H^*(\omega, \theta, r) = \mathbf{w}^T\mathbf{G}(\omega, \theta, r)\mathbf{w} \ , \tag{9.12}$$

with

$$\mathbf{G}(\omega, \theta, r) = \mathbf{g}(\omega, \theta, r)\mathbf{g}^H(\omega, \theta, r) \ , \tag{9.13}$$

which can also be decomposed into a real part $\mathbf{G}_R(\omega, \theta, r)$ and an imaginary part $\mathbf{G}_I(\omega, \theta, r)$. Since $\mathbf{G}_I(\omega, \theta, r)$ is anti-symmetric (cf. Appendix G.2), the spatial directivity spectrum $|H(\omega, \theta, r)|^2$ is equal to (G.13),

$$\boxed{|H(\omega, \theta, r)|^2 = \mathbf{w}^T\mathbf{G}_R(\omega, \theta, r)\mathbf{w}} \tag{9.14}$$

9.2 Near-field beamformer design procedures

The ultimate goal of broadband beamformer design is to design a beamformer such that the spatial directivity pattern $H(\omega, \theta, r)$ optimally fits a desired spatial directivity pattern $D(\omega, \theta, r)$ *for all distances* r, i.e.

$$\boxed{\min_{\mathbf{w}} \int_r \int_\Theta \int_\Omega F(\omega, \theta, r)|H(\omega, \theta, r) - D(\omega, \theta, r)|^2 d\omega d\theta dr} \tag{9.15}$$

However, since this is quite a difficult task, near-field broadband beamformers are generally designed for one or a limited number of predefined distances, i.e. the outer integral in (9.15) is approximated by a finite sum.

9.2.1 Design for one distance

If the near-field broadband beamformer design is performed for *one fixed distance* r, the cost functions and derivations in Sections 8.3 and 8.4 remain valid, but the following substitutions have to be made

$$H(\omega, \theta), \mathbf{g}(\omega, \theta), \mathbf{G}(\omega, \theta) \rightarrow H(\omega, \theta, r), \mathbf{g}(\omega, \theta, r), \mathbf{G}(\omega, \theta, r) \ . \tag{9.16}$$

The only difference lies in the calculation of the double integrals, which is discussed in Appendices G and H.

9.2.2 Mixed near-field far-field beamforming

The spatial directivity pattern of a near-field broadband beamformer designed for one specific distance can be quite unsatisfactory at other distances (cf. simulations in Section 9.4.2). If the broadband beamformer should be able to operate at several distances – possibly having a different desired spatial directivity pattern $D(\omega, \theta, r)$ at these distances – we can define the total cost function

$$\boxed{J_{tot}(\mathbf{w}) = \sum_{r=1}^{R} \alpha_r J_r(\mathbf{w})} \tag{9.17}$$

with α_r a positive weighting factor, assigning more or less importance to the cost function $J_r(\mathbf{w})$. The cost function $J_r(\mathbf{w})$ can be any of the cost functions discussed in Sections 8.3 and 8.4, defined at distance r. If one of the considered distances is $r = \infty$, this is called mixed near-field far-field beamforming.

For most design procedures (weighted LS, non-linear criterion, conventional eigenfilter), this extension is straightforward. E.g. in [273] mixed near-field far-field beamforming has been discussed for the weighted LS cost function. However, for the TLS eigenfilter and the maximum energy array cost functions this extension gives rise to a significantly different optimisation problem, for which no closed-form solution is available.

Weighted least-squares

The weighted LS cost function is equal to (cf. Section 8.3.2)

$$\begin{aligned} J_{LS}^{tot}(\mathbf{w}) &= \sum_{r=1}^{R} \alpha_r J_{LS,r}(\mathbf{w}) = \sum_{r=1}^{R} \alpha_r \big(\mathbf{w}^T \mathbf{Q}_{LS,r} \mathbf{w} - 2\mathbf{w}^T \mathbf{a}_r + d_{LS,r}\big) && (9.18) \\ &= \mathbf{w}^T \Big(\sum_{r=1}^{R} \alpha_r \mathbf{Q}_{LS,r}\Big) \mathbf{w} - 2\mathbf{w}^T \Big(\sum_{r=1}^{R} \alpha_r \mathbf{a}_r\Big) + \sum_{r=1}^{R} \alpha_r d_{LS,r} \ , && (9.19) \end{aligned}$$

with $\mathbf{Q}_{LS,r}$, $\mathbf{a}_r$ and $d_{LS,r}$ defined at the distance r. This equation is equivalent to the cost function in (8.14), with

$$\mathbf{Q}_{LS} = \sum_{r=1}^{R} \alpha_r \mathbf{Q}_{LS,r}, \quad \mathbf{a} = \sum_{r=1}^{R} \alpha_r \mathbf{a}_r, \quad d_{LS} = \sum_{r=1}^{R} \alpha_r d_{LS,r} \ . \tag{9.20}$$

The solution of the constrained and the unconstrained weighted LS cost function has been discussed in Section 8.3.2.

Non-linear criterion

The non-linear cost function is equal to (cf. Section 8.3.4)

$$J_{NL}^{tot}(\mathbf{w}) = \sum_{r=1}^{R} \alpha_r J_{NL,r}(\mathbf{w}) = \sum_{r=1}^{R} \alpha_r \big(J_{sum,r}^{tot}(\mathbf{w}) + d_{NL,r} - 2\mathbf{w}^T \mathbf{Q}_{NL,r} \mathbf{w}\big) \ ,$$

where $J_{sum,r}^{tot}(\mathbf{w})$ can be written as (cf. Appendix E.4)

$$J_{sum,r}^{tot}(\mathbf{w}) = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{l=1}^{M} w_i w_j w_k w_l \, \rho_{ijkl,r} \ , \tag{9.21}$$

with $J^{tot}_{sum,r}(\mathbf{w})$, $\rho_{ijkl,r}$, $\mathbf{Q}_{NL,r}$ and $d_{NL,r}$ defined at the distance r. The non-linear cost function can now be written as

$$\begin{aligned} J^{tot}_{NL}(\mathbf{w}) &= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l \Big(\sum_{r=1}^{R}\alpha_r \rho_{ijkl,r}\Big) + \Big(\sum_{r=1}^{R}\alpha_r d_{NL,r}\Big) - \\ & \quad 2\mathbf{w}^T\Big(\sum_{r=1}^{R}\alpha_r \mathbf{Q}_{NL,r}\Big)\mathbf{w}\ . \end{aligned} \tag{9.22}$$

This equation is equivalent to the cost function in (8.37), with

$$\rho_{ijkl} = \sum_{r=1}^{R}\alpha_r \rho_{ijkl,r}, \quad d_{NL} = \sum_{r=1}^{R}\alpha_r d_{NL,r}, \quad \mathbf{Q}_{NL} = \sum_{r=1}^{R}\alpha_r \mathbf{Q}_{NL,r}\ . \tag{9.23}$$

The solution of the unconstrained and the unconstrained non-linear cost function using iterative optimisation techniques has been discussed in Section 8.3.4.

Conventional eigenfilter technique

The conventional eigenfilter cost function is equal to (cf. Section 8.4.1)

$$J^{tot}_{eig}(\mathbf{w}) = \sum_{r=1}^{R}\alpha_r J_{eig,r}(\mathbf{w}) = \sum_{r=1}^{R}\alpha_r \mathbf{w}^T\mathbf{Q}_{eig,r}\mathbf{w} = \mathbf{w}^T\Big(\sum_{r=1}^{R}\alpha_r \mathbf{Q}_{eig,r}\Big)\mathbf{w}\ ,$$

with $\mathbf{Q}_{eig,r}$ defined at the distance r. This equation is equivalent to the cost function in (8.58), with

$$\mathbf{Q}_{eig} = \sum_{r=1}^{R}\alpha_r \mathbf{Q}_{eig,r}\ . \tag{9.24}$$

The solution of the eigenfilter cost function with a quadratic energy constraint and with linear constraints is discussed in Section 8.4.1 (where only one quadratic energy constraint $\mathbf{w}^T\mathbf{Q}^{tot}_{e,r}\mathbf{w} = 1$ at one distance r is allowed).

TLS eigenfilter technique

The TLS eigenfilter cost function is equal to (cf. Section 8.4.2)

$$J^{tot}_{TLS}(\mathbf{w}) = \sum_{r=1}^{R}\alpha_r J_{TLS,r}(\mathbf{w}) = \sum_{r=1}^{R}\alpha_r \frac{\hat{\mathbf{w}}^T\hat{\mathbf{Q}}_{TLS,r}\hat{\mathbf{w}}}{\hat{\mathbf{w}}^T\hat{\mathbf{Q}}^{tot}_{e,r}\hat{\mathbf{w}}}\ , \tag{9.25}$$

with

$$\hat{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix}, \quad \hat{\mathbf{Q}}_{TLS,r} = \begin{bmatrix} \mathbf{Q}_{LS,r} & \mathbf{a}_r \\ \mathbf{a}_r^T & d_{LS,r} \end{bmatrix}, \quad \hat{\mathbf{Q}}^{tot}_{e,r} = \begin{bmatrix} \mathbf{Q}^{tot}_{e,r} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix},$$

and $\hat{\mathbf{Q}}_{LS,r}$, $\mathbf{a}_r$, $d_{LS,r}$ and $\hat{\mathbf{Q}}^{tot}_{e,r}$ defined at the distance r.

The TLS eigenfilter cost function with linear constraints $\mathbf{Cw} = \mathbf{b}$ can be transformed into the unconstrained cost function (cf. Section 8.4.2)

$$\sum_{r=1}^{R} \alpha_r \frac{\tilde{\mathbf{w}}^T \mathbf{B} \hat{\mathbf{Q}}_{TLS,r} \mathbf{B}^T \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^T \mathbf{B} \hat{\mathbf{Q}}_{e,r}^{tot} \mathbf{B}^T \tilde{\mathbf{w}}} \; . \tag{9.26}$$

Both minimising (9.25) and (9.26) can be considered to be special cases of minimising the cost function

$$\boxed{J_m(\mathbf{w}) = \sum_{r=1}^{R} \frac{\mathbf{w}^T \mathbf{A}_r \mathbf{w}}{\mathbf{w}^T \mathbf{B}_r \mathbf{w}}} \tag{9.27}$$

with $\mathbf{A}_r$ and $\mathbf{B}_r$ symmetric positive-definite matrices. When $\mathbf{B}_r = \mathbf{B}$, $r = 1 \ldots R$, this problem is a generalised eigenvalue problem and the solution is given by the generalised eigenvector, corresponding to the minimum generalised eigenvalue of $\sum_{r=1}^{R} \mathbf{A}_r$ and $\mathbf{B}$. In general however, minimising $J_m(\mathbf{w})$ apparently cannot be transformed into a generalised eigenvalue problem. Hence, we have used an iterative non-linear optimisation technique for minimising this cost function. In order to improve the numerical robustness and the convergence speed of the optimisation technique, both the gradient

$$\frac{\partial J_m(\mathbf{w})}{\partial \mathbf{w}} = 2 \sum_{r=1}^{R} \frac{(\mathbf{w}^T \mathbf{B}_r \mathbf{w}) \mathbf{A}_r - (\mathbf{w}^T \mathbf{A}_r \mathbf{w}) \mathbf{B}_r}{(\mathbf{w}^T \mathbf{B}_r \mathbf{w})^2} \mathbf{w} \tag{9.28}$$

and the Hessian

$$\begin{aligned} \frac{\partial^2 J_m(\mathbf{w})}{\partial^2 \mathbf{w}} = 2 \sum_{r=1}^{R} & \frac{(\mathbf{w}^T \mathbf{B}_r \mathbf{w}) \mathbf{A}_r - (\mathbf{w}^T \mathbf{A}_r \mathbf{w}) \mathbf{B}_r + 2(\mathbf{A}_r \mathbf{w} \mathbf{w}^T \mathbf{B}_r - \mathbf{B}_r \mathbf{w} \mathbf{w}^T \mathbf{A}_r)}{(\mathbf{w}^T \mathbf{B}_r \mathbf{w})^2} \\ & -4 \frac{\left[\mathbf{A}_r \mathbf{w} (\mathbf{w}^T \mathbf{B}_r \mathbf{w}) - \mathbf{B}_r \mathbf{w} (\mathbf{w}^T \mathbf{A}_r \mathbf{w})\right] \mathbf{w}^T \mathbf{B}_r}{(\mathbf{w}^T \mathbf{B}_r \mathbf{w})^3} \end{aligned} \tag{9.29}$$

can be provided analytically. Although we have not been able to prove that this optimisation procedure converges to the global minimum, no problems with local minima have occurred during simulations.

Maximum energy array

The maximum energy array cost function is equal to (cf. Section 8.3.3)

$$J_{ME}^{tot}(\mathbf{w}) = \sum_{r=1}^{R} \alpha_r J_{ME,r}(\mathbf{w}) = \sum_{r=1}^{R} \alpha_r \frac{\mathbf{w}^T \mathbf{Q}_{e,r}^p \mathbf{w}}{\mathbf{w}^T \mathbf{Q}_{e,r}^s \mathbf{w}} \; , \tag{9.30}$$

with $\mathbf{Q}_{e,r}^p$ and $\mathbf{Q}_{e,r}^s$ defined at the distance r. The maximum energy array cost function with linear constraints $\mathbf{Cw} = \mathbf{b}$ can be transformed into the unconstrained cost function (cf. Section 8.3.3)

$$\sum_{r=1}^{R} \alpha_r \frac{\tilde{\mathbf{w}}^T \mathbf{B} \hat{\mathbf{Q}}_{e,r}^{p} \mathbf{B}^T \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^T \mathbf{B} \hat{\mathbf{Q}}_{e,r}^{s} \mathbf{B}^T \tilde{\mathbf{w}}} \,. \tag{9.31}$$

Both maximising (9.30) and (9.31) can be considered to be a special case of maximising the cost function $J_m(\mathbf{w})$ in (9.27).

9.3 Linear constraints

For the far-field case, linear constraints of the form $\mathbf{Cw} = \mathbf{b}$ have been defined in Section 8.5. For the near-field case, point constraints and derivative constraints can be defined similarly as for the far-field case. However, a line constraint of the form (8.85) can not be imposed for the near-field case, since for $\theta_f = \pi/2$ and $r \neq \infty$, the delays $\tau_n(\theta_f, r)$ are not equal to 0.

Linear constraints at a distance r can be written as $\mathbf{C}_r\mathbf{w} = \mathbf{b}_r$, such that the linear constraints for all R distances together can be written as

$$\underbrace{\left[\begin{array}{cccc} \mathbf{C}_1^T & \mathbf{C}_2^T & \dots & \mathbf{C}_R^T \end{array}\right]^T}_{\mathbf{C}} \mathbf{w} = \underbrace{\left[\begin{array}{cccc} \mathbf{b}_1^T & \mathbf{b}_2^T & \dots & \mathbf{b}_R^T \end{array}\right]^T}_{\mathbf{b}} \,. \tag{9.32}$$

9.3.1 Point constraint

Similar to (8.79) and (8.80), the absolute point constraint $H(\omega_f, \theta_f, r) = b$ corresponds to

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_f, \theta_f, r) \\ \mathbf{g}_I^T(\omega_f, \theta_f, r) \end{bmatrix}}_{\mathbf{C}_r} \mathbf{w} = \underbrace{\begin{bmatrix} b_R \\ b_I \end{bmatrix}}_{\mathbf{b}_r} , \tag{9.33}$$

whereas the relative point constraint $H(\omega_{f_1}, \theta_{f_1}, r_1) = b \cdot H(\omega_{f_2}, \theta_{f_2}, r_2)$ corresponds to

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_{f_1}, \theta_{f_1}, r_1) - b_R\, \mathbf{g}_R^T(\omega_{f_2}, \theta_{f_2}, r_2) + b_I\, \mathbf{g}_I^T(\omega_{f_2}, \theta_{f_2}, r_2) \\ \mathbf{g}_I^T(\omega_{f_1}, \theta_{f_1}, r_1) - b_I\, \mathbf{g}_R^T(\omega_{f_2}, \theta_{f_2}, r_2) - b_R\, \mathbf{g}_I^T(\omega_{f_2}, \theta_{f_2}, r_2) \end{bmatrix}}_{\mathbf{C}_r} \mathbf{w} = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{b}_r} \,.$$

9.3.2 Derivative constraint

The derivative constraints for the near-field case are defined similarly as for the far-field case in Section 8.5.3. In Appendix D.2 it is shown that the first-order angle derivative $\mathbf{g}'_\theta(\omega, \theta, r)$ can be written as

$$\mathbf{g}'_\theta(\omega, \theta, r) = \sin\theta\; \mathbf{\Delta}_\theta(\omega, \theta, r)\; \mathbf{g}(\omega, \theta, r) \,, \tag{9.34}$$

with $\mathbf{\Delta}_\theta(\omega,\theta,r)$ a complex-valued $M \times M$-dimensional diagonal matrix,

$$\mathbf{\Delta}_\theta(\omega,\theta,r) = \mathbf{\Delta}_{\theta,R}(\theta,r) + j\mathbf{\Delta}_{\theta,I}(\omega,\theta,r)\ , \tag{9.35}$$

defined in (D.14). If $\sin\theta_f = 0$, i.e. $\theta_f = 0$ or $\theta_f = \pi$, the first-order angle derivative constraint $\mathbf{w}^T\mathbf{g}'_\theta(\omega_f,\theta_f,r) = 0$ is satisfied for all frequencies and distances. For all other angles, this constraint can be written as

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_f,\theta_f,r)\mathbf{\Delta}_{\theta,R}(\theta_f,r) - \mathbf{g}_I^T(\omega_f,\theta_f,r)\mathbf{\Delta}_{\theta,I}(\omega_f,\theta_f,r) \\ \mathbf{g}_R^T(\omega_f,\theta_f,r)\mathbf{\Delta}_{\theta,I}(\omega_f,\theta_f,r) + \mathbf{g}_I^T(\omega_f,\theta_f,r)\mathbf{\Delta}_{\theta,R}(\theta_f,r) \end{bmatrix}}_{\mathbf{C}_r}\mathbf{w} = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{b}_r}\ .$$

The first-order frequency derivative $\mathbf{g}'_\omega(\omega,\theta,r)$ can be written as

$$\mathbf{g}'_\omega(\omega,\theta,r) = -j\mathbf{\Delta}_\omega(\theta,r)\,\mathbf{g}(\omega,\theta,r)\ , \tag{9.36}$$

with $\mathbf{\Delta}_\omega(\theta,r)$ a real-valued $M \times M$-dimensional diagonal matrix,

$$\mathbf{\Delta}_\omega(\theta,r) = \begin{bmatrix} \mathbf{D} + \tau_0(\theta,r)\,\mathbf{I}_L & & & \\ & \mathbf{D} + \tau_1(\theta,r)\,\mathbf{I}_L & & \\ & & \ddots & \\ & & & \mathbf{D} + \tau_{N-1}(\theta,r)\,\mathbf{I}_L \end{bmatrix},$$

with $\mathbf{D}$ defined in (8.92). The frequency derivative constraint $\mathbf{w}^T\mathbf{g}'_\omega(\omega_f,\theta_f,r) = 0$ corresponds to 2 linear constraints,

$$\underbrace{\begin{bmatrix} \mathbf{g}_R^T(\omega_f,\theta_f,r) \\ \mathbf{g}_I^T(\omega_f,\theta_f,r) \end{bmatrix}\mathbf{\Delta}_\omega(\theta_f,r)}_{\mathbf{C}_r}\,\mathbf{w} = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{b}_r}\ . \tag{9.37}$$

9.4 Simulations

In this section simulation results are presented for near-field broadband beamforming at one specific distance and for mixed near-field far-field broadband beamformer design. It will be shown that the TLS eigenfilter technique is the preferred non-iterative design procedure and that mixed near-field far-field design provides a trade-off between near-field and far-field performance.

9.4.1 Near-field broadband beamformer

For the near-field broadband beamformer design, we have used the same design criteria (microphone array, stopband/passband specifications, filter length, stopband weights) as for the far-field design (cf. Section 8.6), but we have designed the beamformer for a distance $r = 0.2\,\text{m}$ from the microphone array. We will only present results for the first specification without linear constraints.

		Cost function				
Design	α	J_{LS}	J_{eig}	J_{TLS}	J_{ME}	J_{NL}
LS	0.1	**0.03991**	0.02244	0.01112	10.5245	0.06909
EIG	0.1	0.04676	**0.01466**	0.01267	9.62192	0.06473
TLS	0.1	0.04055	0.02177	**0.01095**	9.45334	0.06613
ME	0.1	3905.35	0.87992	0.87758	**43.6125**	$5.41\ 10^7$
NL	0.1	0.07700	0.02373	0.01774	2.92942	**0.03369**
LS	1	**0.14284**	0.06894	0.04468	18.11551	0.12597
EIG	1	0.16050	**0.05918**	0.04595	17.91655	0.11590
TLS	1	0.14818	0.06879	**0.04309**	17.99989	0.11903
ME	1	3985.72	0.89799	0.89564	**43.61245**	$5.41\ 10^7$
NL	1	0.23852	0.09870	0.06649	9.13160	**0.08441**
LS	10	**0.73287**	0.45994	0.35987	19.15062	0.72276
EIG	10	1.13534	**0.45116**	0.32303	19.41230	0.25785
TLS	10	0.87317	0.45928	**0.30815**	19.33174	0.26571
ME	10	4789.42	1.07863	1.07624	**43.61245**	$5.44\ 10^7$
NL	10	1.12420	0.51194	0.34444	17.50699	**0.16281**

Table 9.1: Different cost functions for design specification 1 without linear constraints ($N = 5$; $L = 20$; $\alpha = 0.1, 1, 10$)

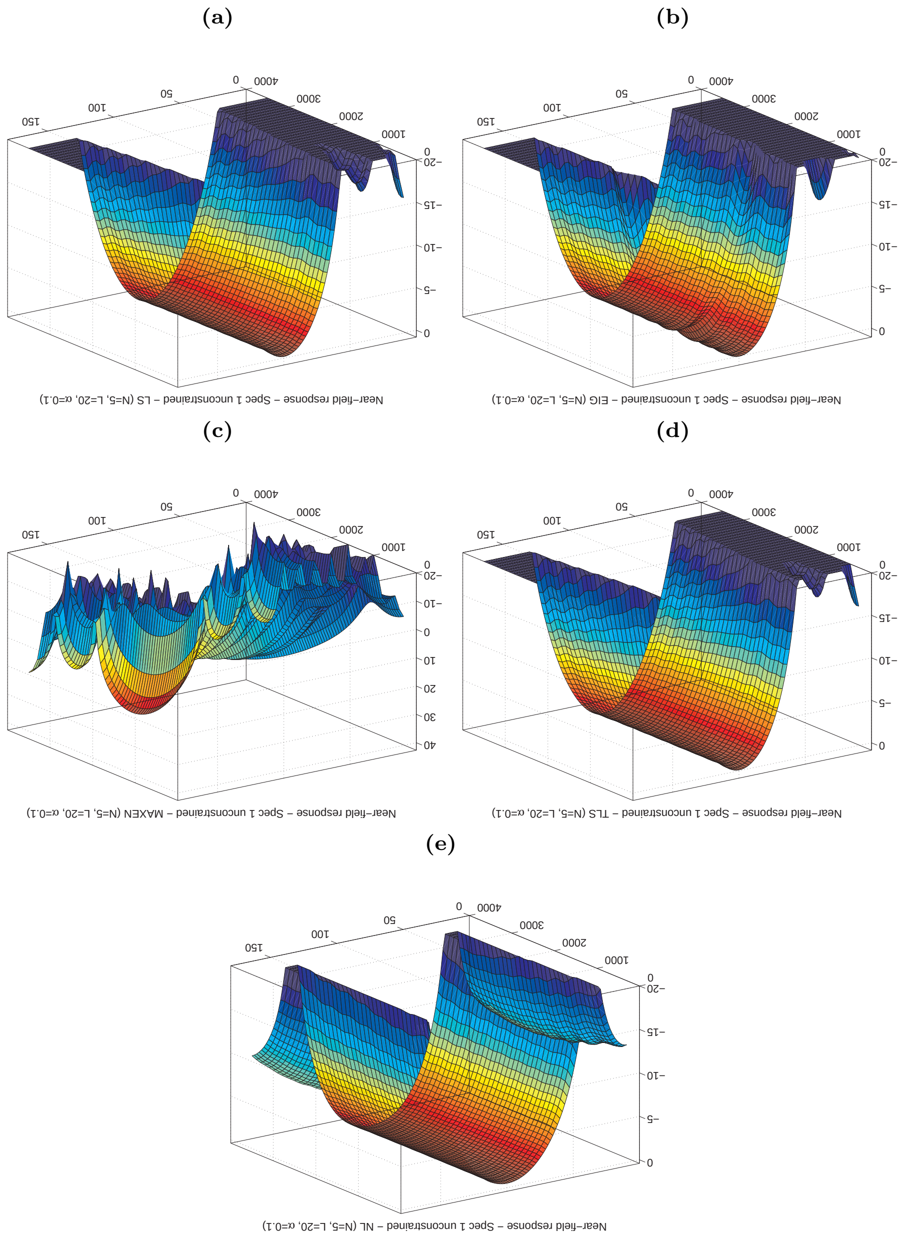
The different cost functions for the different beamformer designs are summarised in Table 9.1. Similar conclusions as for the far-field case hold. Obviously, the design procedure optimising a specific cost function gives rise to the best value for this particular cost function (bold values). The maximum energy array technique has a quite poor performance (this can also be seen from the spatial directivity pattern in Fig. 9.2). In addition, the TLS eigenfilter technique always has a better performance, i.e. non-linear cost function J_{NL}, than the weighted LS technique, and therefore appears to be the preferred non-iterative design procedure[1]. Figures 9.2(a)-(e) show the near-field spatial directivity patterns for all design procedures with $\alpha = 0.1$.

9.4.2 Mixed near-field far-field design

Using the same configuration, we have performed a mixed near-field far-field broadband beamformer design for $r = 0.2\,\mathrm{m}$ (near-field) and $r = \infty$ (far-field) using the weighted LS cost function, the TLS eigenfilter technique and the non-linear criterion. The near-field weighting factor in (9.17) is $\alpha_r = 0.4$. We will only present results for the first design specification without linear constraints.

Table 9.2 summarises the different cost functions (far-field, near-field, total) for the different design procedures (weighted LS, TLS eigenfilter and non-linear

[1] Recall from Chapter 8 that although we would actually like to use the non-linear design procedure, this design procedure gives rise to a high computational complexity and hence we compare the performance of the non-iterative design procedures using the non-linear cost function as a performance criterion.


(a)

(b)

(c)

(d)

(e)

Figure 9.2: Near-field spatial directivity pattern for (**a**) weighted LS, (**b**) conventional eigenfilter, (**c**) maximum energy array, (**d**) TLS eigenfilter and (**e**) non-linear criterion (design specification 1, no linear constraints, $r = 0.2\,\text{m}$, $\alpha = 0.1$, $N = 5$, $L = 20$)

			Cost function	
Design procedure	α	J_∞ (far-field)	J_r (near-field)	J_{tot} (mixed)
LS far-field	1	**0.32012**	1.68710	0.99496
LS near-field	1	0.97135	**0.14284**	1.02849
LS mixed	1	0.42277	0.45489	**0.60472**
TLS far-field	1	**0.09851**	0.40205	0.25933
TLS near-field	1	0.28515	**0.04309**	0.30239
TLS mixed	1	0.12873	0.14564	**0.18698**
NL far-field	1	**0.10301**	3.50694	1.50578
NL near-field	1	0.45379	**0.08441**	0.48756
NL mixed	1	0.15304	0.16557	**0.21926**

Table 9.2: Near-field, far-field and total cost function for different design procedures ($N = 5$; $L = 20$; $\alpha = 1$; $\alpha_r = 0.4$; $r = 0.2\,\mathrm{m}$)

design procedure for far-field, near-field and mixed near-field far-field) and for $\alpha = 1$. As can be seen, the far-field design yields the best far-field cost function, but gives rise to a poor near-field response. On the contrary, the near-field design yields the best near-field cost function, but gives rise to a poor far-field response. The mixed near-field far-field design provides a trade-off between the near-field and the far-field performance. It yields a better far-field cost function than the near-field design but worse than the far-field design, whereas it yields a better near-field cost function than the far-field design but worse than the near-field design.

Figure 9.3 shows the far-field and the near-field spatial directivity patterns for the TLS eigenfilter technique designed for the far-field (with $\alpha = 1$, $N = 5$, $L = 20$). As can be seen from this figure, the near-field response is quite unsatisfactory. Figure 9.4 shows the far-field and the near-field spatial directivity patterns for the TLS eigenfilter technique designed for the near-field (with $\alpha = 1$, $N = 5$, $L = 20$). As can be seen from this figure, the far-field response now is quite unsatisfactory. Providing a trade-off between far-field and near-field performance, Figure 9.5 shows the far-field and the near-field spatial directivity patterns for the TLS eigenfilter technique that has been designed both for far-field and near-field (with $\alpha = 1$, $N = 5$, $L = 20$). Figures 9.6, 9.7 and 9.8 show similar results when the broadband beamformers are designed using the non-linear criterion.

9.5 Conclusion

In this chapter we have shown that the design of near-field broadband beamformers is very similar to the design of far-field broadband beamformers. When designing a near-field broadband beamformer for one specific distance, the same design procedures and cost functions as for the far-field case can be used

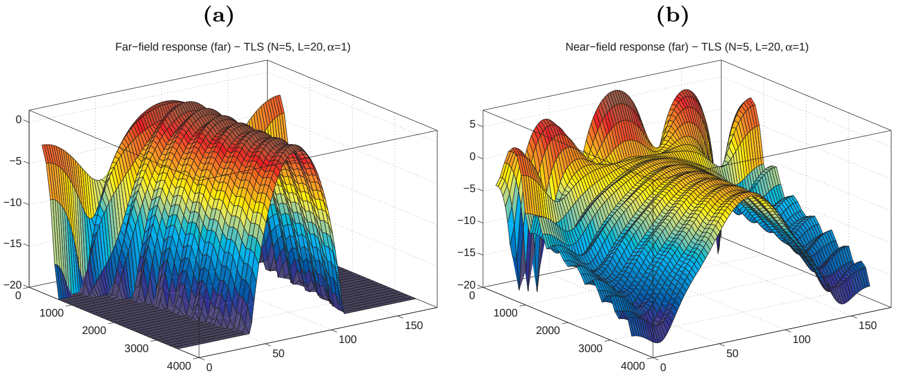


Figure 9.3: (**a**) Far-field and (**b**) near-field spatial directivity pattern for TLS eigenfilter far-field design (design spec 1, $r = 0.2\,\mathrm{m}$, $\alpha = 1$, $N = 5$, $L = 20$)

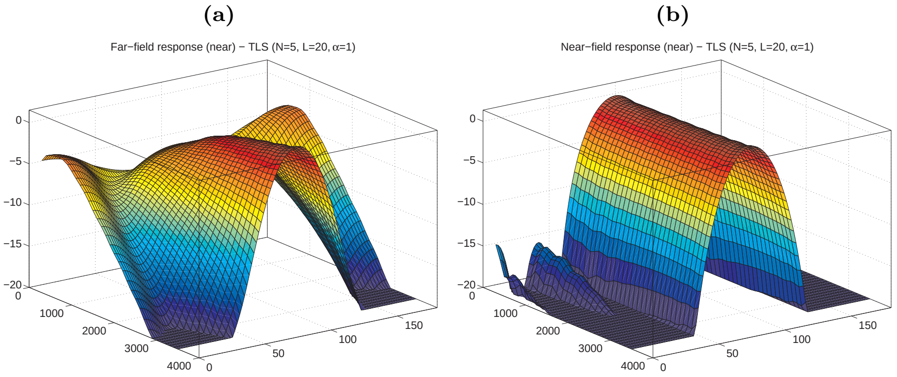


Figure 9.4: (**a**) Far-field and (**b**) near-field spatial directivity pattern for TLS eigenfilter near-field design (design spec 1, $r = 0.2\,\mathrm{m}$, $\alpha = 1$, $N = 5$, $L = 20$)

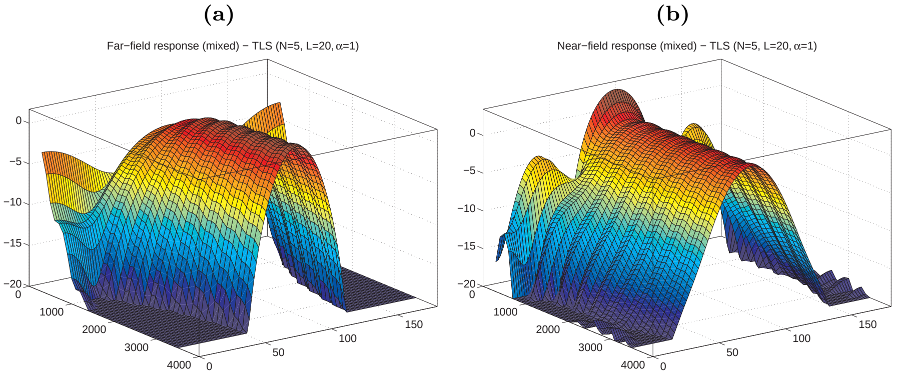


Figure 9.5: (**a**) Far-field and (**b**) near-field spatial directivity pattern for TLS eigenfilter mixed near-field far-field design (design spec 1, $r = 0.2\,\mathrm{m}$, $\alpha = 1$, $N = 5$, $L = 20$)

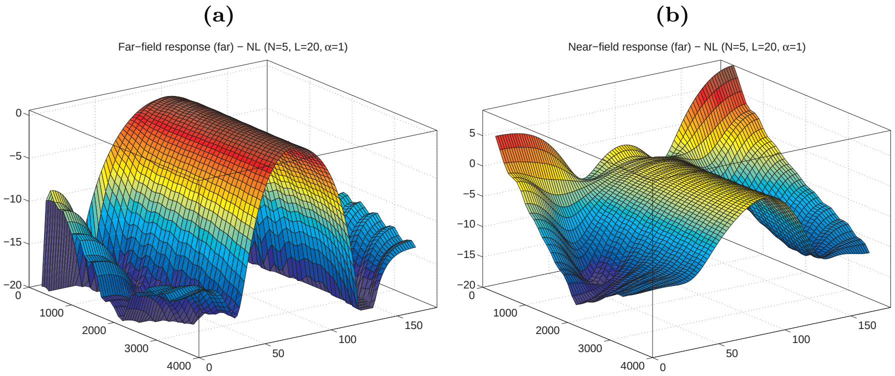


Figure 9.6: (**a**) Far-field and (**b**) near-field spatial directivity pattern for nonlinear far-field design (design spec 1, $r = 0.2\,\mathrm{m}$, $\alpha = 1$, $N = 5$, $L = 20$)

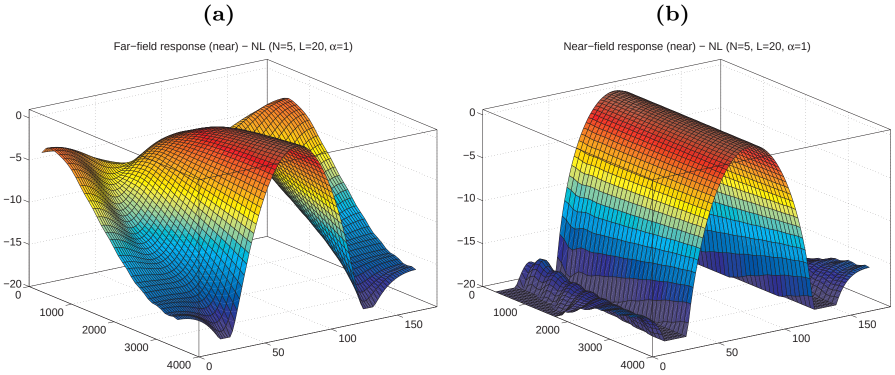


Figure 9.7: (**a**) Far-field and (**b**) near-field spatial directivity pattern for nonlinear near-field design (design spec 1, $r = 0.2\,\mathrm{m}$, $\alpha = 1$, $N = 5$, $L = 20$)

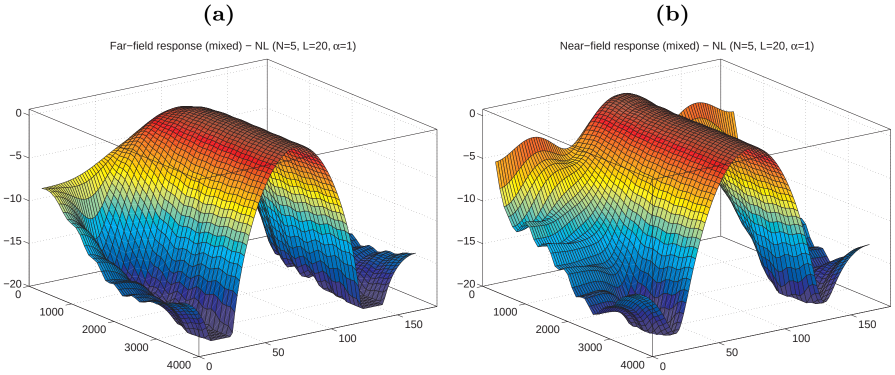


Figure 9.8: (**a**) Far-field and (**b**) near-field spatial directivity pattern for nonlinear mixed near-field far-field design (design spec 1, $r = 0.2\,\mathrm{m}$, $\alpha = 1$, $N = 5$, $L = 20$)

and the only difference lies in the calculation of the double integrals involved. However, the spatial directivity pattern of a near-field broadband beamformer designed for one specific distance can be quite unsatisfactory at other distances. Hence, design procedures have been discussed for designing broadband beamformers that operate at several distances, e.g. mixed near-field far-field beamformers. Although for most cost functions this extension is straightforward, for the TLS eigenfilter and the maximum energy array cost functions this extension gives rise to a significantly different optimisation problem, requiring the use of an iterative non-linear optimisation technique. The simulations in Section 9.4 have shown that the TLS eigenfilter technique again is the preferred non-iterative design procedure for near-field broadband beamformer design and that mixed near-field far-field design provides a trade-off between the near-field and the far-field performance.

Chapter 10

Robust Broadband Beamforming for gain and phase errors

In the previous chapters we have assumed that the microphones are (perfect) omni-directional microphones with a flat frequency response equal to 1. This chapter discusses the design of broadband beamformers that are robust against unknown gain and phase errors in the microphone array characteristics.

In Section 10.2 the broadband beamformer expressions and cost functions are redefined, taking into account the microphone characteristics. In general, the microphone characteristics consist of a gain and a phase, which can both be frequency and angle dependent. Using these redefined expressions, it is possible to design broadband beamformers when the microphone characteristics are exactly known. We will also simplify all expressions for microphone characteristics which are independent of frequency and angle.

However, in many applications the microphone characteristics are not exactly known and can even change over time. Section 10.3 describes two procedures for designing broadband beamformers that are robust against (unknown) gain and phase errors in the microphone array characteristics. The first design procedure optimises the mean performance of the broadband beamformer for all feasible microphone characteristics, whereas the second design procedure optimises the worst-case performance, leading to a minimax problem.

In Section 10.4 simulation results for the different design procedures and cost functions are presented. It is shown that robust beamformer design gives rise to a significant performance improvement when gain and phase errors occur.

10.1 Introduction

It is well known that fixed and adaptive beamformers can be highly *sensitive to errors in the microphone array characteristics* (gain, phase, microphone position), cf. Section 5.4.3 [24][36][137][240]. Small deviations from the assumed microphone characteristics can lead to large deviations in the spatial directivity pattern, especially when using small-size microphone arrays, e.g. in hearing aids and cochlear implants (cf. Section 10.4). Since in practice it is difficult to manufacture microphones having exactly the same characteristics, it is practically impossible to exactly know the microphone array characteristics without a measurement or a calibration procedure [24][248]. However, a measurement or calibration procedure will only limit the error sensitivity for the specific microphone array used and the cost of such a procedure for every individual microphone array clearly is objectionable. Moreover, after calibration the microphone characteristics can still drift over time [137].

For superdirective beamformers, robustness against random errors has been improved by limiting the white noise gain (WNG) of the beamformer, i.e. imposing a norm constraint or a general quadratic constraint on the filter coefficients [16][36][86][146]. Limiting the WNG has also been applied in order to enhance the robustness of adaptive beamformers [37]. In this chapter, we specifically consider (random) gain and phase errors in the microphone characteristics, and we discuss design procedures for designing broadband beamformers with an arbitrary desired spatial directivity pattern that are robust against these specific errors. Since we consider small-size microphone arrays in this chapter, we will assume that the far-field assumptions are valid. However, all derived expressions can be easily extended to the near-field case.

10.2 Known microphone characteristics

In Section 10.2.1 we redefine the beamformer expressions, taking into account the microphone characteristics. Using these expressions, it is possible to design broadband beamformers when the microphone characteristics are exactly known. In Section 10.2.2 the redefined cost functions for the weighted LS, the TLS eigenfilter and the non-linear criterion are discussed. In addition, these expressions are simplified for omni-directional, frequency-flat microphones.

10.2.1 Configuration

When the microphones perform a spatial and a spectral filtering operation on the received signals, their microphone characteristics have to be taken into account in the design of broadband beamformers. The microphone characteristics of the nth microphone are described by the function

$$A_n(\omega,\theta) = a_n(\omega,\theta)e^{-j\psi_n(\omega,\theta)} = a_n(\omega,\theta)\cos\psi_n(\omega,\theta) - ja_n(\omega,\theta)\sin\psi_n(\omega,\theta)\ ,$$

where both the gain $a_n(\omega,\theta)$ and the phase $\psi_n(\omega,\theta)$ can be frequency and angle-dependent. The function $A_n(\omega,\theta)$ is symmetric in ω, i.e.

$$A_n(-\omega,\theta) = A_n(\omega,\theta), \quad a_n(-\omega,\theta) = a_n(\omega,\theta), \quad \psi_n(-\omega,\theta) = \psi_n(\omega,\theta) \; .$$

In Chapters 8 and 9 we have considered perfect microphones with equal microphone characteristics, i.e. $A_n(\omega,\theta) = 1$, $n = 0 \ldots N-1$. Most expressions discussed in these chapters will remain valid when taking into account the microphone characteristics. However, some expressions need to be redefined.

Since the microphone signals $Y_n(\omega,\theta)$ in (8.4) now are equal to

$$Y_n(\omega,\theta) = A_n(\omega,\theta)e^{-j\omega\tau_n(\theta)}\bar{Y}(\omega,\theta), \; -\pi \leq \omega \leq \pi, \; -\pi \leq \theta \leq \pi \; , \tag{10.1}$$

the spatial directivity pattern $H(\omega,\theta)$ in (8.6) can be written as

$$H(\omega,\theta) = \sum_{n=0}^{N-1} W_n(\omega)A_n(\omega,\theta)e^{-j\omega\tau_n(\theta)} = \mathbf{w}^T\bar{\mathbf{g}}(\omega,\theta) \; , \tag{10.2}$$

with the M-dimensional steering vector $\bar{\mathbf{g}}(\omega,\theta)$ now equal to

$$\bar{\mathbf{g}}(\omega,\theta) = \begin{bmatrix} \mathbf{e}(\omega)A_0(\omega,\theta)\,e^{-j\omega\tau_0(\theta)} \\ \mathbf{e}(\omega)A_1(\omega,\theta)\,e^{-j\omega\tau_1(\theta)} \\ \vdots \\ \mathbf{e}(\omega)A_{N-1}(\omega,\theta)\,e^{-j\omega\tau_{N-1}(\theta)} \end{bmatrix} . \tag{10.3}$$

The steering vector $\bar{\mathbf{g}}(\omega,\theta)$ can be written as

$$\boxed{\bar{\mathbf{g}}(\omega,\theta) = \mathbf{A}(\omega,\theta)\cdot\mathbf{g}(\omega,\theta) \; ,} \tag{10.4}$$

with $\mathbf{g}(\omega,\theta)$ the steering vector defined in (8.7), assuming omni-directional microphones with a flat frequency response equal to 1, and $\mathbf{A}(\omega,\theta)$ an $M \times M$-dimensional diagonal matrix consisting of the microphone characteristics,

$$\mathbf{A}(\omega,\theta) = \begin{bmatrix} A_0(\omega,\theta)\,\mathbf{I}_L & & & \\ & A_1(\omega,\theta)\,\mathbf{I}_L & & \\ & & \ddots & \\ & & & A_{N-1}(\omega,\theta)\,\mathbf{I}_L \end{bmatrix} . \tag{10.5}$$

The steering vector $\bar{\mathbf{g}}(\omega,\theta)$ can be decomposed into a real and an imaginary part, $\bar{\mathbf{g}}(\omega,\theta) = \bar{\mathbf{g}}_R(\omega,\theta) + j\bar{\mathbf{g}}_I(\omega,\theta)$. The real part $\bar{\mathbf{g}}_R(\omega,\theta)$ is equal to

$$\bar{\mathbf{g}}_R(\omega,\theta) = \mathbf{A}_R(\omega,\theta)\mathbf{g}_R(\omega,\theta) - \mathbf{A}_I(\omega,\theta)\mathbf{g}_I(\omega,\theta) \; , \tag{10.6}$$

with $\mathbf{A}_R(\omega,\theta)$ and $\mathbf{A}_I(\omega,\theta)$ the real and the imaginary part of $\mathbf{A}(\omega,\theta)$. Using (10.2), the spatial directivity spectrum $|H(\omega,\theta)|^2$ can be written as

$$\boxed{|H(\omega,\theta)|^2 = H(\omega,\theta)H^*(\omega,\theta) = \mathbf{w}^T\bar{\mathbf{G}}(\omega,\theta)\mathbf{w}} \tag{10.7}$$

with $\bar{\mathbf{G}}(\omega,\theta)=\bar{\mathbf{g}}(\omega,\theta)\bar{\mathbf{g}}^H(\omega,\theta)$, which can be written, using (10.4), as

$$\bar{\mathbf{G}}(\omega,\theta)=\mathbf{A}(\omega,\theta)\cdot\mathbf{G}(\omega,\theta)\cdot\mathbf{A}^H(\omega,\theta)\ , \tag{10.8}$$

with $\mathbf{G}(\omega,\theta)=\mathbf{g}(\omega,\theta)\mathbf{g}^H(\omega,\theta)$. Since the imaginary part $\bar{\mathbf{G}}_I(\omega,\theta)$ again is anti-symmetric, the spatial directivity spectrum $|H(\omega,\theta)|^2$ can be written as

$$\boxed{|H(\omega,\theta)|^2=\mathbf{w}^T\bar{\mathbf{G}}_R(\omega,\theta)\mathbf{w}} \tag{10.9}$$

with the real part $\bar{\mathbf{G}}_R(\omega,\theta)$ equal to

$$\begin{aligned}\bar{\mathbf{G}}_R(\omega,\theta)\ =\ &\mathbf{A}_R(\omega,\theta)\mathbf{G}_R(\omega,\theta)\mathbf{A}_R(\omega,\theta)+\mathbf{A}_I(\omega,\theta)\mathbf{G}_R(\omega,\theta)\mathbf{A}_I(\omega,\theta)-\\&\mathbf{A}_I(\omega,\theta)\mathbf{G}_I(\omega,\theta)\mathbf{A}_R(\omega,\theta)+\mathbf{A}_R(\omega,\theta)\mathbf{G}_I(\omega,\theta)\mathbf{A}_I(\omega,\theta)\ .\end{aligned} \tag{10.10}$$

10.2.2 Cost functions

Considering the redefined expressions for the steering vector and the spatial directivity pattern, it is now possible to design broadband beamformers using the cost functions discussed in Chapter 8, when the microphone characteristics $\mathbf{A}(\omega,\theta)$ are exactly known. E.g. in [60] a design example using first-order differential microphones has been given. The only difference lies in the calculation of the double integrals involved. We will briefly discuss the redefined cost functions for the weighted LS, the TLS eigenfilter and the non-linear criterion. In addition, all expressions can be significantly simplified – certainly for the robust broadband beamformer design discussed in Section 10.3 – when we assume that the microphone characteristics are independent of frequency and angle, i.e. for omni-directional, frequency-flat microphones, such that $\mathbf{A}(\omega,\theta)=\mathbf{A}=\mathbf{A}_R+j\mathbf{A}_I$. Even if this assumption is not exactly satisfied in practice, it is generally possible to split up the considered frequency-angle region into several (small) frequency-angle regions where this assumption does hold.

Weighted LS cost function

When taking into account the microphone characteristics, the weighted LS cost function $J_{LS}(\mathbf{w})$ in (8.14) is equal to

$$J_{LS}(\mathbf{w})=\mathbf{w}^T\bar{\mathbf{Q}}_{LS}\mathbf{w}-2\mathbf{w}^T\bar{\mathbf{a}}+d_{LS}\ , \tag{10.11}$$

with

$$\bar{\mathbf{Q}}_{LS}\ =\ \int_\Theta\int_\Omega F(\omega,\theta)\bar{\mathbf{G}}_R(\omega,\theta)d\omega d\theta \tag{10.12}$$

$$\bar{\mathbf{a}}\ =\ \int_\Theta\int_\Omega F(\omega,\theta)\left[D_R(\omega,\theta)\bar{\mathbf{g}}_R(\omega,\theta)+D_I(\omega,\theta)\bar{\mathbf{g}}_I(\omega,\theta)\right]d\omega d\theta\ , \tag{10.13}$$

and d_{LS} defined in (8.17).

Using (10.6) and (10.10), for *omni-directional, frequency-flat microphones* these expressions can be simplified to (assuming $D(\omega,\theta)$ to be real-valued)

$$\boxed{\begin{array}{lcl}\bar{\mathbf{a}} & = & \mathbf{A}_R\,\mathbf{a}-\mathbf{A}_I\,\mathbf{a}^\circ \\ \bar{\mathbf{Q}}_{LS} & = & \mathbf{A}_R\mathbf{Q}_{LS}\mathbf{A}_R+\mathbf{A}_I\mathbf{Q}_{LS}\mathbf{A}_I-\mathbf{A}_I\mathbf{Q}_{LS}^\circ\mathbf{A}_R+\mathbf{A}_R\mathbf{Q}_{LS}^\circ\mathbf{A}_I\end{array}} \tag{10.14}$$

with $\mathbf{Q}_{LS}$ defined in (8.15) and

$$\mathbf{Q}_{LS}^\circ = \int_\Theta\int_\Omega F(\omega,\theta)\mathbf{G}_I(\omega,\theta)d\omega d\theta \tag{10.15}$$

$$\mathbf{a} = \int_\Theta\int_\Omega F(\omega,\theta)D(\omega,\theta)\mathbf{g}_R(\omega,\theta)d\omega d\theta \tag{10.16}$$

$$\mathbf{a}^\circ = \int_\Theta\int_\Omega F(\omega,\theta)D(\omega,\theta)\mathbf{g}_I(\omega,\theta)d\omega d\theta\ . \tag{10.17}$$

The ith element of $\bar{\mathbf{a}}$ and the (i,j)-th element of $\bar{\mathbf{Q}}_{LS}$ are equal to

$$\bar{\mathbf{a}}^i = a_n\big(\cos\psi_n\,\mathbf{a}^i+\sin\psi_n\,\mathbf{a}^{\circ,i}\big) \tag{10.18}$$

$$\bar{\mathbf{Q}}_{LS}^{ij} = a_na_m\Big(\cos\big(\psi_n-\psi_m\big)\,\mathbf{Q}_{LS}^{ij}+\sin\big(\psi_n-\psi_m\big)\,\mathbf{Q}_{LS}^{\circ,ij}\Big)\ , \tag{10.19}$$

with $n=\lfloor\frac{i-1}{L}\rfloor$ and $m=\lfloor\frac{j-1}{L}\rfloor$.

TLS eigenfilter technique

When taking into account the microphone characteristics, the TLS eigenfilter cost function $J_{TLS}(\mathbf{w})$ in (8.72) is equal to

$$J_{TLS}(\mathbf{w}) = \int_\Theta\int_\Omega F(\omega,\theta)\frac{|D(\omega,\theta)-H(\omega,\theta)|^2}{\mathbf{w}^T\bar{\mathbf{Q}}_e^{tot}\mathbf{w}+1}d\omega d\theta\ , \tag{10.20}$$

with

$$\bar{\mathbf{Q}}_e^{tot} = \int_\Theta\int_\Omega \bar{\mathbf{G}}_R(\omega,\theta)d\omega d\theta\ . \tag{10.21}$$

This cost function can be written as

$$J_{TLS}(\mathbf{w}) = \frac{\hat{\mathbf{w}}^T\hat{\bar{\mathbf{Q}}}_{TLS}\hat{\mathbf{w}}}{\hat{\mathbf{w}}^T\hat{\bar{\mathbf{Q}}}_e^{tot}\hat{\mathbf{w}}}\ , \tag{10.22}$$

with

$$\hat{\mathbf{w}} = \begin{bmatrix}\mathbf{w}\\ -1\end{bmatrix},\quad \hat{\bar{\mathbf{Q}}}_{TLS} = \begin{bmatrix}\bar{\mathbf{Q}}_{LS} & \bar{\mathbf{a}}\\ \bar{\mathbf{a}}^T & d_{LS}\end{bmatrix},\quad \hat{\bar{\mathbf{Q}}}_e^{tot} = \begin{bmatrix}\bar{\mathbf{Q}}_e^{tot} & \mathbf{0}\\ \mathbf{0}^T & 1\end{bmatrix}\ . \tag{10.23}$$

Non-linear criterion

When taking into account the microphone characteristics, the non-linear criterion $J_{NL}(\mathbf{w})$ in (8.37) is equal to

$$J_{NL}(\mathbf{w}) = \bar{J}_{sum}(\mathbf{w}) + d_{NL} - 2\mathbf{w}^T \bar{\mathbf{Q}}_{NL}\mathbf{w} \;, \tag{10.24}$$

with

$$\begin{aligned} \bar{J}_{sum}(\mathbf{w}) &= \int_\Theta \int_\Omega F(\omega,\theta)|H(\omega,\theta)|^4 d\omega d\theta = \int_\Theta \int_\Omega F(\omega,\theta)\big(\mathbf{w}^T \bar{\mathbf{G}}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta \\ \bar{\mathbf{Q}}_{NL} &= \int_\Theta \int_\Omega F(\omega,\theta)|D(\omega,\theta)|^2 \bar{\mathbf{G}}_R(\omega,\theta) d\omega d\theta \;, \end{aligned} \tag{10.25}$$

and d_{NL} defined in (8.39).

For *omni-directional, frequency-flat microphones* the matrix $\bar{\mathbf{Q}}_{NL}$ can be computed similarly as $\bar{\mathbf{Q}}_{LS}$ in (10.14) as

$$\bar{\mathbf{Q}}_{NL} = \mathbf{A}_R \mathbf{Q}_{NL} \mathbf{A}_R + \mathbf{A}_I \mathbf{Q}_{NL} \mathbf{A}_I - \mathbf{A}_I \mathbf{Q}^\circ_{NL} \mathbf{A}_R + \mathbf{A}_R \mathbf{Q}^\circ_{NL} \mathbf{A}_I \;, \tag{10.26}$$

with $\mathbf{Q}_{NL}$ defined in (8.40) and

$$\mathbf{Q}^\circ_{NL} = \int_\Theta \int_\Omega F(\omega,\theta)|D(\omega,\theta)|^2 \mathbf{G}_I(\omega,\theta) d\omega d\theta \;. \tag{10.27}$$

In Appendix E.4 it is shown that $\bar{J}_{sum}(\mathbf{w})$ can be written as

$$\boxed{\bar{J}_{sum}(\mathbf{w}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l \; \underbrace{a_{ijkl}\Big(\cos\psi_{ijkl}\cdot\rho_{ijkl} - \sin\psi_{ijkl}\cdot\rho^\circ_{ijkl}\Big)}_{\bar{\rho}_{ijkl}}} \tag{10.28}$$

with a_{ijkl} and ψ_{ijkl} defined in (E.61) and ρ_{ijkl} and ρ°_{ijkl} defined in (E.64) and (E.65).

10.3 Robust broadband beamforming

Using the cost functions defined in Section 10.2.2, it is possible to design broadband beamformers with an arbitrary desired spatial directivity pattern $D(\omega,\theta)$, when the microphone characteristics $A_n(\omega,\theta)$, $n = 0 \ldots N-1$, are exactly known (and fixed). However, these fixed broadband beamformers are known to be highly sensitive to errors in the microphone array characteristics (gain, phase, microphone position) [24][36][137]. Small deviations from the assumed microphone array characteristics can lead to large deviations from the desired spatial directivity pattern, especially when using small-size microphone arrays, e.g. in hearing aids and cochlear implants (cf. Section 10.4). Since in practice it is difficult to manufacture microphones having exactly the same characteristics, it is practically impossible to exactly know the microphone array characteristics without a measurement or a calibration procedure. Obviously, the cost of such a calibration procedure for every individual microphone array is objectionable. Moreover, after calibration the characteristics can still drift over time [137].

Instead of measuring or calibrating every individual microphone array, it is better to consider all feasible microphone characteristics (in this chapter we only consider gain and phase[1]) and to either optimise:

- the *mean performance*, i.e. the weighted sum of the cost functions for all feasible microphone characteristics, using the probability of the microphone characteristics as weights (cf. Section 10.3.1). This procedure requires knowledge of the gain and the phase probability density functions (pdf). It will be shown that for gain errors only the moments of the gain pdf are required, whereas for phase errors in general complete knowledge of the phase pdf is required. We will apply this mean performance criterion to the weighted LS and the non-linear cost function, whereas it is not straightforward to apply this criterion to the TLS eigenfilter cost function. When optimising this mean performance criterion, it is however still possible that for some specific gain/phase combination (typically with a low probability), the cost function is quite high.
- the *worst-case performance*, i.e. the maximum cost function for all feasible microphone characteristics, leading to a minimax problem (cf. Section 10.3.2). This is a stronger criterion, since the cost for the worst-case scenario is minimised. We will apply this criterion to all cost functions.

The same problem of gain and phase errors has been studied in [86]. However, in [86] only the narrowband case for a specific directivity pattern, with a uniform pdf and a LS cost function has been considered. The approach presented here is more general in the sense that we consider broadband beamformers with an arbitrary spatial directivity pattern, arbitrary probability density functions and several cost functions.

10.3.1 Weighted sum using probability density functions

The mean cost function J_{mean} is defined as the weighted sum of the cost functions for all feasible microphone characteristics, using the probability of the microphone characteristics as weights, i.e.

$$\boxed{J_{mean}(\mathbf{w}) = \int_{A_0} \dots \int_{A_{N-1}} J(\mathbf{w}, \mathbf{A})\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1})\, dA_0 \dots dA_{N-1}} \tag{10.29}$$

[1] A microphone position error can be considered as a frequency- and angle-dependent phase error [63].

with $J(\mathbf{w}, \mathbf{A})$ the cost function for a specific characteristic $\{A_0, \ldots, A_{N-1}\}$ and $f_{\mathcal{A}}(A)$ the probability density function (pdf) of the stochastic variable $A = ae^{-j\psi}$, i.e. the joint pdf of the stochastic variables a (gain) and ψ (phase), $f_{\mathcal{A}}(A) = f_{\alpha,\Psi}(a, \psi)$. We assume that $f_{\mathcal{A}}(A)$ is independent of frequency and angle, or that $f_{\mathcal{A}}(A)$ is available for several frequency-angle regions. Without loss of generality, we also assume that all microphone characteristics $A_n, n = 0 \ldots N-1$, are described by the same pdf $f_{\mathcal{A}}(A)$. Furthermore, we assume that a and ψ are independent stochastic variables, such that the joint pdf is separable, i.e.

$$f_{\mathcal{A}}(A) = f_{\alpha}(a) f_{\Psi}(\psi) \ , \tag{10.30}$$

with $f_{\alpha}(a)$ the pdf of the gain a and $f_{\Psi}(\psi)$ the pdf of the phase ψ. For these pdfs the relation

$$\int_a f_{\alpha}(a)\, da = 1, \quad \int_{\psi} f_{\Psi}(\psi)\, d\psi = 1 \tag{10.31}$$

holds. We will consider 2 cost functions from Section 10.2.2: the weighted LS and the non-linear cost function (it is not straightforward to apply this criterion to the TLS eigenfilter cost function). *Remarkably, the same design procedures as for the non-robust design in Section 8.3 can be used, and we only require some additional parameters which can be easily calculated from the gain and the phase pdf.*

Weighted LS cost function

The mean performance weighted LS cost function can be written as

$$J_{LS}^{mean}(\mathbf{w}) = \int_{A_0} \ldots \int_{A_{N-1}} J_{LS}(\mathbf{w}, \mathbf{A})\, f_{\mathcal{A}}(A_0) \ldots f_{\mathcal{A}}(A_{N-1})\, dA_0 \ldots dA_{N-1} \ . \tag{10.32}$$

The cost function $J_{LS}(\mathbf{w}, \mathbf{A})$ for a specific microphone characteristic is equal to (10.11), i.e.

$$J_{LS}(\mathbf{w}, \mathbf{A}) = \mathbf{w}^T \bar{\mathbf{Q}}_{LS} \mathbf{w} - 2\mathbf{w}^T \bar{\mathbf{a}} + d_{LS} \ . \tag{10.33}$$

By combining (10.32) and (10.33), the mean performance weighted LS cost function can be written as

$$\begin{aligned} J_{LS}^{mean}(\mathbf{w}) \ = \ & \mathbf{w}^T \int_{A_0} \ldots \int_{A_{N-1}} \bar{\mathbf{Q}}_{LS}\, f_{\mathcal{A}}(A_0) \ldots f_{\mathcal{A}}(A_{N-1})\, dA_0 \ldots dA_{N-1}\, \mathbf{w} - \\ & 2\mathbf{w}^T \int_{A_0} \ldots \int_{A_{N-1}} \bar{\mathbf{a}}\, f_{\mathcal{A}}(A_0) \ldots f_{\mathcal{A}}(A_{N-1})\, dA_0 \ldots dA_{N-1} + \\ & \int_{A_0} \ldots \int_{A_{N-1}} d_{LS} f_{\mathcal{A}}(A_0) \ldots f_{\mathcal{A}}(A_{N-1})\, dA_0 \ldots dA_{N-1} \qquad (10.34) \\ = \ & \mathbf{w}^T \bar{\mathbf{Q}}_{mean} \mathbf{w} - 2\mathbf{w}^T \bar{\mathbf{a}}_{mean} + d_{LS} \ , \qquad (10.35) \end{aligned}$$

which has the same form as (10.11) and (8.14). Using (10.18), the ith element of $\bar{\mathbf{a}}_{mean}$ is equal to

$$\begin{aligned}
\bar{\mathbf{a}}^{i}_{mean} &= \int_{A_0} \dots \int_{A_{N-1}} a_n \big(\cos\psi_n \, \mathbf{a}^{i} + \sin\psi_n \, \mathbf{a}^{\circ,i} \big) f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1}) \\
&\qquad\qquad dA_0 \dots dA_{N-1} \\
&= \int_{A_n} a_n \big(\cos\psi_n \, \mathbf{a}^{i} + \sin\psi_n \, \mathbf{a}^{\circ,i} \big) f_{\mathcal{A}}(A_n) dA_n \qquad (10.36) \\
&= \int_{a_n} a_n f_\alpha(a_n) da_n \left[\int_{\psi_n} \cos\psi_n f_\Psi(\psi_n) d\psi_n \, \mathbf{a}^{i} + \right. \\
&\qquad \left. \int_{\psi_n} \sin\psi_n f_\Psi(\psi_n) d\psi_n \, \mathbf{a}^{\circ,i} \right] \qquad (10.37) \\
&= \mu_a \left[\mu^c_\psi \, \mathbf{a}^{i} + \mu^s_\psi \, \mathbf{a}^{\circ,i} \right] \ , \qquad (10.38)
\end{aligned}$$

with

$$\boxed{\mu_a = \int_a a f_\alpha(a) da, \qquad \mu^c_\psi = \int_\psi \cos\psi f_\Psi(\psi) d\psi, \qquad \mu^s_\psi = \int_\psi \sin\psi f_\Psi(\psi) d\psi} \tag{10.39}$$

such that

$$\boxed{\bar{\mathbf{a}}_{mean} = \mu_a \mu^c_\psi \, \mathbf{a} + \mu_a \mu^s_\psi \, \mathbf{a}^\circ} \tag{10.40}$$

Using (10.19), the (i,j)-th element of $\bar{\mathbf{Q}}_{mean}$ is equal to

$$\begin{aligned}
\bar{\mathbf{Q}}^{ij}_{mean} &= \int_{A_0} \dots \int_{A_{N-1}} a_n a_m \Big(\cos\big(\psi_n - \psi_m\big) \, \mathbf{Q}^{ij}_{LS} + \sin\big(\psi_n - \psi_m\big) \, \mathbf{Q}^{\circ,ij}_{LS} \Big) \\
&\qquad\qquad f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1}) \, dA_0 \dots dA_{N-1} \qquad (10.41) \\
&= \int_{a_n} \int_{a_m} a_n a_m f_\alpha(a_n) f_\alpha(a_m) da_n da_m \cdot \left[\int_{\psi_n} \int_{\psi_m} \cos\big(\psi_n - \psi_m\big) \cdot \right. \\
&\qquad f_\Psi(\psi_n) f_\Psi(\psi_m) d\psi_n d\psi_m \, \mathbf{Q}^{ij}_{LS} + \int_{\psi_n} \int_{\psi_m} \sin\big(\psi_n - \psi_m\big) \cdot \\
&\qquad \left. f_\Psi(\psi_n) f_\Psi(\psi_m) d\psi_n d\psi_m \, \mathbf{Q}^{\circ,ij}_{LS} \right] \ . \qquad (10.42)
\end{aligned}$$

If $n = m$, $\bar{\mathbf{Q}}^{ij}_{mean}$ is equal to

$$\bar{\mathbf{Q}}^{ij}_{mean} = \int_{a_n} a_n^2 f_\alpha(a_n) da_n \, \mathbf{Q}^{ij}_{LS} = \sigma_a^2 \, \mathbf{Q}^{ij}_{LS} \ , \tag{10.43}$$

with σ_a^2 the variance of the gain pdf,

$$\boxed{\sigma_a^2 = \int_a a^2 f_\alpha(a) da} \tag{10.44}$$

whereas, if $n \neq m$, $\bar{\mathbf{Q}}^{ij}_{mean}$ is equal to

$$\bar{\mathbf{Q}}^{ij}_{mean} = \mu_a^2 \left[\sigma_\psi^c \mathbf{Q}^{ij}_{LS} + \sigma_\psi^s \mathbf{Q}^{\circ,ij}_{LS} \right] \ , \tag{10.45}$$

with μ_a the mean of the gain pdf and

$$\begin{aligned} \sigma_\psi^c &= \int_{\psi_1} \int_{\psi_2} \cos\big(\psi_1 - \psi_2\big) f_\Psi(\psi_1) f_\Psi(\psi_2) d\psi_1 d\psi_2 && (10.46) \\ &= \int_{\psi_1} \int_{\psi_2} \big(\cos\psi_1 \cos\psi_2 + \sin\psi_1 \sin\psi_2\big)\, f_\Psi(\psi_1) f_\Psi(\psi_2) d\psi_1 d\psi_2 \\ \sigma_\psi^s &= \int_{\psi_1} \int_{\psi_2} \sin\big(\psi_1 - \psi_2\big) f_\Psi(\psi_1) f_\Psi(\psi_2) d\psi_1 d\psi_2 && (10.47) \\ &= \int_{\psi_1} \int_{\psi_2} \big(\sin\psi_1 \cos\psi_2 - \cos\psi_1 \sin\psi_2\big)\, f_\Psi(\psi_1) f_\Psi(\psi_2) d\psi_1 d\psi_2 \ , \end{aligned}$$

such that

$$\boxed{\sigma_\psi^c = \big(\mu_\psi^c\big)^2 + \big(\mu_\psi^s\big)^2, \qquad \sigma_\psi^s = \mu_\psi^s \mu_\psi^c - \mu_\psi^c \mu_\psi^s = 0} \tag{10.48}$$

The matrix $\bar{\mathbf{Q}}_{mean}$ can now be easily computed as

$$\boxed{\bar{\mathbf{Q}}_{mean} = \begin{bmatrix} \sigma_a^2\, \mathbf{1}_L & \mu_a^2 \sigma_\psi^c\, \mathbf{1}_L & \dots & \mu_a^2 \sigma_\psi^c\, \mathbf{1}_L \\ \mu_a^2 \sigma_\psi^c\, \mathbf{1}_L & \sigma_a^2 \mathbf{1}_L & \dots & \mu_a^2 \sigma_\psi^c\, \mathbf{1}_L \\ \vdots & \vdots & & \vdots \\ \mu_a^2 \sigma_\psi^c\, \mathbf{1}_L & \mu_a^2 \sigma_\psi^c & \dots & \sigma_a^2\, \mathbf{1}_L \end{bmatrix} \odot \mathbf{Q}_{LS}} \tag{10.49}$$

with $\mathbf{1}_L$ an $L \times L$-dimensional matrix with all elements equal to 1 and $\odot$ denoting element-wise multiplication. As can be seen, we only need the mean and the variance of the gain pdf $f_\alpha(a)$, whereas in general complete knowledge of the phase pdf $f_\Psi(\psi)$ is required.

Frequently used probability density functions are a uniform distribution (with boundary values a_{min} and a_{max}),

$$\left\{ \begin{array}{lcll} f_\alpha(a) & = & \dfrac{1}{a_{max} - a_{min}} & , \quad a_{min} \leq a \leq a_{max} \\ & = & 0 & , \quad a < a_{min}, a > a_{max} \ , \end{array} \right. \tag{10.50}$$

and a Gaussian distribution (with mean μ_a and standard deviation s_a),

$$f_\alpha(a) = \frac{1}{\sqrt{2\pi s_a^2}} e^{-\frac{(a-\mu_a)^2}{2s_a^2}} \ . \tag{10.51}$$

For a uniform distribution the gain and phase parameters are equal to

$$\begin{aligned}
\mu_a &= \frac{a_{min} + a_{max}}{2} & \sigma_a^2 &= \frac{a_{min}^2 + a_{min}a_{max} + a_{max}^2}{3} \\
\mu_\psi^c &= \frac{\sin\psi_{max} - \sin\psi_{min}}{\psi_{max} - \psi_{min}} & \mu_\psi^s &= \frac{\cos\psi_{min} - \cos\psi_{max}}{\psi_{max} - \psi_{min}} \\
\sigma_\psi^c &= \frac{2 - 2\cos\big(\psi_{max} - \psi_{min}\big)}{\big(\psi_{max} - \psi_{min}\big)^2} & \sigma_\psi^s &= 0 \quad .
\end{aligned}$$

For a Gaussian distribution with mean μ_a and standard deviation s_a, the variance is equal to $\sigma_a^2 = \mu_a^2 + s_a^2$, whereas the phase parameters μ_ψ^c, μ_ψ^s and σ_ψ^c have to be calculated numerically.

Non-linear criterion

The mean performance non-linear cost function can be written as

$$J_{NL}^{mean}(\mathbf{w}) = \int_{A_0} \dots \int_{A_{N-1}} J_{NL}(\mathbf{w}, \mathbf{A})\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1})\, dA_0 \dots dA_{N-1} \; . \tag{10.52}$$

The cost function $J_{NL}(\mathbf{w}, \mathbf{A})$ for a specific microphone characteristic is equal to (10.24),

$$J_{NL}(\mathbf{w}, \mathbf{A}) = \bar{J}_{sum}(\mathbf{w}) + d_{NL} - 2\mathbf{w}^T \bar{\mathbf{Q}}_{NL}\mathbf{w} \; . \tag{10.53}$$

By combining (10.52) and (10.53), the mean performance non-linear cost function can be written as

$$\begin{aligned}
J_{NL}^{mean}(\mathbf{w}) = &\int_{A_0} \dots \int_{A_{N-1}} \bar{J}_{sum}(\mathbf{w})\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1})\, dA_0 \dots dA_{N-1} - \\
&2\mathbf{w}^T \int_{A_0} \dots \int_{A_{N-1}} \bar{\mathbf{Q}}_{NL}\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1})\, dA_0 \dots dA_{N-1}\, \mathbf{w} + \\
&\int_{A_0} \dots \int_{A_{N-1}} d_{NL}\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1})\, dA_0 \dots dA_{N-1} \qquad (10.54) \\
= &\bar{J}_{sum}^{mean}(\mathbf{w}) - 2\mathbf{w}^T \bar{\mathbf{Q}}_{NL}^{mean}\mathbf{w} + d_{NL} \; . \qquad (10.55)
\end{aligned}$$

Similar to (10.49), the matrix $\bar{\mathbf{Q}}_{NL}^{mean}$ is equal to

$$\boxed{\bar{\mathbf{Q}}_{NL}^{mean} = \begin{bmatrix} \sigma_a^2\,\mathbf{1}_L & \mu_a^2\sigma_\psi^c\,\mathbf{1}_L & \dots & \mu_a^2\sigma_\psi^c\,\mathbf{1}_L \\ \mu_a^2\sigma_\psi^c\,\mathbf{1}_L & \sigma_a^2\mathbf{1}_L & \dots & \mu_a^2\sigma_\psi^c\,\mathbf{1}_L \\ \vdots & \vdots & & \vdots \\ \mu_a^2\sigma_\psi^c\,\mathbf{1}_L & \mu_a^2\sigma_\psi^c & \dots & \sigma_a^2\,\mathbf{1}_L \end{bmatrix} \odot \mathbf{Q}_{NL}} \tag{10.56}$$

Using (10.28), $\bar{J}^{mean}_{sum}(\mathbf{w})$ can be written as

$$\begin{aligned}\bar{J}^{mean}_{sum}(\mathbf{w}) &= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w(i)w(j)w(k)w(l) \int_{A_0} \dots \int_{A_{N-1}} a_{ijkl} \Big(\cos\psi_{ijkl} \cdot \\ &\qquad \rho_{ijkl} - \sin\psi_{ijkl} \cdot \rho^{\circ}_{ijkl} \Big) f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{N-1})\, dA_0 \dots dA_{N-1} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w(i)w(j)w(k)w(l)\, \delta^{a}_{ijkl} \left(\delta^{c}_{\psi,ijkl} \cdot \rho_{ijkl} - \right. \\ &\qquad\qquad \left. \delta^{s}_{\psi,ijkl} \cdot \rho^{\circ}_{ijkl} \right) , \end{aligned} \tag{10.57}$$

with ρ_{ijkl} and ρ°_{ijkl} defined in (E.64) and (E.65) and

$$\begin{aligned} \delta^{a}_{ijkl} &= \int_{a_0} \dots \int_{a_{N-1}} a_{ijkl}\, f_{\alpha}(a_0) \dots f_{\alpha}(a_{N-1})\, da_0 \dots da_{N-1} \\ \delta^{c}_{\psi,ijkl} &= \int_{\psi_0} \dots \int_{\psi_{N-1}} \cos\psi_{ijkl}\, f_{\Psi}(\psi_0) \dots f_{\Psi}(\psi_{N-1})\, d\psi_0 \dots d\psi_{N-1} \\ \delta^{s}_{\psi,ijkl} &= \int_{\psi_0} \dots \int_{\psi_{N-1}} \sin\psi_{ijkl}\, f_{\Psi}(\psi_0) \dots f_{\Psi}(\psi_{N-1})\, d\psi_0 \dots d\psi_{N-1} \ , \end{aligned}$$

with a_{ijkl} and ψ_{ijkl} defined in (E.61). The expression $\bar{J}^{mean}_{sum}(\mathbf{w})$ in (10.57) has the same form as (10.28) and (E.52), such that the same non-linear optimisation techniques as described in Section 8.3.4 can be used for minimising $J^{mean}_{NL}(\mathbf{w})$. The calculation of the parameters δ^{a}_{ijkl}, $\delta^{c}_{\psi,ijkl}$ and $\delta^{s}_{\psi,ijkl}$ is discussed in Appendix D.3. For the calculation of δ^{a}_{ijkl}, we only need the (higher order) moments of the gain pdf $f_{\alpha}(a)$, whereas for the calculation of $\delta^{c}_{\psi,ijkl}$ and $\delta^{s}_{\psi,ijkl}$, in general complete knowledge of the phase pdf $f_{\Psi}(\psi)$ is required. In Appendix D.3 it is shown that for a symmetric phase pdf $\delta^{s}_{\psi,ijkl} = 0$, such that

$$\boxed{\bar{J}^{mean}_{sum}(\mathbf{w}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w(i)w(j)w(k)w(l)\, \delta^{a}_{ijkl} \cdot \delta^{c}_{\psi,ijkl} \cdot \rho_{ijkl}} \tag{10.58}$$

10.3.2 Minimax criterion

For the minimax criterion, which optimises the worst-case performance, we first have to define a (finite) set of microphone characteristics (K_a gain values and K_{ψ} phase values),

$$\{a_{min} = a_1, a_2, \dots, a_{K_a} = a_{max}\}, \quad \{\psi_{min} = \psi_1, \psi_2, \dots, \psi_{K_{\psi}} = \psi_{max}\} \ ,$$

as an approximation for the continuum of feasible microphone characteristics, and use this set to construct the $(K_a K_\psi)^N$-dimensional vector $\mathbf{F}(\mathbf{w})$, i.e.

$$\mathbf{F}(\mathbf{w}) = \begin{bmatrix} F_1(\mathbf{w}, \mathbf{A}) \\ F_2(\mathbf{w}, \mathbf{A}) \\ \vdots \\ F_{(K_a K_\psi)^N}(\mathbf{w}, \mathbf{A}) \end{bmatrix}, \tag{10.59}$$

which consists of the used cost function (weighted LS, TLS eigenfilter, non-linear, or any other cost function, e.g. defined in [65][157][159][192]) at each possible combination of gain and phase values. The goal then is to minimise the L_∞-norm of $\mathbf{F}(\mathbf{w})$, i.e. the maximum value of the elements $F_k(\mathbf{w})$,

$$\boxed{\min_{\mathbf{w}} \|\mathbf{F}(\mathbf{w})\|_\infty = \min_{\mathbf{w}} \max_k F_k(\mathbf{w})} \tag{10.60}$$

We have used the MATLAB-function `fminimax` [35], which uses a sequential quadratic programming (SQP) method [93]. For the maximum-energy array cost function, the minimum value of $\mathbf{F}(\mathbf{w})$ has to be maximised, or alternatively, the vector $-\mathbf{F}(\mathbf{w})$ can be used. In order to improve the numerical robustness and the convergence speed, the gradient

$$\begin{bmatrix} \frac{\partial F_1(\mathbf{w},\mathbf{A})}{\partial \mathbf{w}} & \frac{\partial F_2(\mathbf{w},\mathbf{A})}{\partial \mathbf{w}} & \dots & \frac{\partial F_{(K_a K_\psi)^N}(\mathbf{w},\mathbf{A})}{\partial \mathbf{w}} \end{bmatrix}, \tag{10.61}$$

which is an $M \times (K_a K_\psi)^N$-dimensional matrix, can be supplied analytically. As can be seen, the larger the values K_a and K_ψ, the denser the grid of feasible microphone characteristics, and the higher the computational complexity for solving the minimax problem. However, when only considering gain errors and using the weighted LS cost function, the number of grid points can be drastically reduced.

Theorem 10.1 *When considering only* **gain errors** *and using the* **weighted LS cost function**, *the maximum value of* $\mathbf{F}(\mathbf{w})$, *for any* $\mathbf{w}$, *occurs on a boundary point (of an N-dimensional hypercube), i.e.* $a_n = a_{min}$ *or* $a_n = a_{max}$, $n = 0 \dots N-1$. *This implies that* $K_a = 2$ *suffices and* $\mathbf{F}(\mathbf{w})$ *only consists of* 2^N *elements. This is not necessarily the case for the TLS eigenfilter and the non-linear cost function.*

Proof : Appendix D.4 □

10.4 Simulations

This section discusses the simulation results of robust broadband beamformer design for gain and phase errors in the microphone characteristics. Since the effect of gain and phase errors is more profound for small-size microphone arrays, we have performed simulations for a small-size linear non-uniform microphone array consisting of $N = 3$ microphones at positions $[\ -0.01 \quad 0 \quad 0.015\ ]$ m, corresponding to a typical configuration for a next-generation multi-microphone BTE hearing aid. The nominal gains and phases of the microphones are $a_n = 1$ and $\psi_n = 0°$, $n = 0 \ldots N-1$. We have designed an end-fire broadband beamformer for a sampling frequency $f_s = 8\,\text{kHz}$ with passband specifications $(\Omega_p, \Theta_p) = (300\text{–}4000\,\text{Hz}, 0°\text{–}60°)$ and stopband specifications $(\Omega_s, \Theta_s) = (300\text{–}4000\,\text{Hz}, 80°\text{–}180°)$. For the TLS eigenfilter, the matrix $\bar{\mathbf{Q}}_e^{tot}$ is computed with frequency and angle specifications $(\Omega, \Theta) = (300\text{–}4000\,\text{Hz}, 0°\text{–}180°)$, cf. Section 10.2.2. The used filter length $L = 20$ and the stopband weight $\alpha = 1$.

We have designed several types of beamformers using the weighted LS cost function and the non-linear criterion:

1. a non-robust broadband beamformer (not taking into account errors, i.e. assuming $a_n = 1$, $\psi_n = 0°$)
2. a robust broadband beamformer using a uniform gain pdf ($a_{min} = 0.85$, $a_{max} = 1.15$)
3. a robust broadband beamformer using a uniform phase pdf ($\psi_{min} = -5°$, $\psi_{max} = 10°$)
4. a robust broadband beamformer using a uniform gain/phase pdf ($a_{min} = 0.85$, $a_{max} = 1.15$, $\psi_{min} = -5°$, $\psi_{max} = 10°$)
5. a robust broadband beamformer using the minimax criterion (only gain errors are taken into account, $a_{min} = 0.85$, $a_{max} = 1.15$, $K_a = 5$)

Using the TLS eigenfilter cost function, we have designed a non-robust beamformer and a robust beamformer using the minimax criterion. For all beamformer designs, we have computed the following cost functions:

1. the cost function J without phase and gain errors ($a_n = 1$, $\psi_n = 0°$)
2. the cost function J_{dev} for microphone gains $[\ 0.9 \quad 1.1 \quad 1.05\ ]$
3. the mean cost function J_a^{tot} for the uniform gain pdf
4. the mean cost function J_ψ^{tot} for the uniform phase pdf
5. the mean cost function J_A^{tot} for the uniform gain/phase pdf
6. the maximum cost function J_{max} when the gain varies between $a_{min} = 0.85$ and $a_{max} = 1.15$

We will plot the spatial directivity pattern in the frequency-angle region $(300\text{–}3500\,\text{Hz}, 0°\text{–}180°)$ and the angular pattern for the specific frequencies $(500, 1000, 1500, 2000, 2500, 3500)$ Hz.

Table 10.1 summarises the different cost functions for the weighted LS, the non-linear and the TLS eigenfilter non-robust and robust broadband beamformer design procedures. Obviously, the design procedure optimising a specific

		Cost function					
Design procedure		J	J_{dev}	J_a^{tot}	J_ψ^{tot}	J_A^{tot}	J_{max}
LS	Non-robust	**0.313**	220.1	123.3	62.67	185.7	961.3
LS	Gain	0.474	0.685	**0.642**	0.576	0.744	1.441
LS	Phase	0.431	0.700	0.666	**0.557**	0.791	1.749
LS	Gain/phase	0.518	**0.652**	0.653	0.596	**0.732**	1.368
LS	Minimax	0.747	0.843	0.804	0.792	0.849	**1.035**
NL	Non-robust	**0.159**	87.13	124.6	70.19	275.4	3624
NL	Gain	0.176	**0.188**	**0.218**	0.339	0.393	0.505
NL	Phase	0.207	0.236	0.259	**0.300**	0.357	0.502
NL	Gain/phase	0.219	0.222	0.248	0.304	**0.337**	0.499
NL	Minimax	0.171	0.199	0.230	0.340	0.411	**0.417**
TLS	Non-robust	**0.075**	0.840				0.936
TLS	Minimax	0.167	**0.196**				**0.246**

Table 10.1: Different cost functions for weighted LS, non-linear and TLS eigenfilter robust broadband beamformer design ($\alpha = 1$; $N = 3$; $L = 20$)

cost function leads to the best value for this cost function (bold values). This implies that when no gain and phase errors occur, the robust design procedures lead to a higher cost function J than the non-robust design procedure. However, the non-robust design procedure leads to very poor results whenever gain and/or phase errors occur (e.g. compare J_{max} for the non-robust and the robust design procedures and see figures). All robust design procedures (using pdf and minimax criterion) yield satisfactory results when gain and/or phase errors occur. For the cost function J_{dev}, the gain/phase-robust beamformer produces the best result for the weighted LS cost function, whereas the gain-robust beamformer produces the best result for the non-linear cost function. This can be explained by the fact that none of the beamformer designs is actually optimised for these specific microphone gains.

Figure 10.1 shows the spatial directivity pattern of the non-robust beamformer, designed with the non-linear cost function, when no gain and phase errors occur. Figure 10.2 shows the spatial directivity pattern for microphone gains $[\ 0.9 \quad 1.1 \quad 1.05\ ]$. As can be seen from this figure, the beamformer performance dramatically degrades, especially for the lower frequencies. Figure 10.3 shows the spatial directivity pattern for microphone gains $[\ 0.9 \quad 1.1 \quad 1.05\ ]$ and phases $[\ 5^\circ \quad -2^\circ \quad 5^\circ\ ]$, i.e. small deviations from the nominal gains and phases. As can be seen from this figure, the beamformer performance dramatically degrades, especially for the lower frequencies, where the spatial directivity pattern is almost omni-directional and the amplification is very high.

Figures 10.4, 10.5 and 10.6 show the spatial directivity pattern of the gain/phase-robust beamformer, designed with the non-linear cost function, when no errors occur, when gain errors occur and when gain and phase errors occur. As can be seen from Figure 10.4, the performance of this beamformer is worse than the performance of the non-robust beamformer when no errors occur. However, as can be clearly seen from Figures 10.5 and 10.6, when gain and/or phase errors occur, the performance of the gain/phase-robust beamformer is much better than the performance of the non-robust beamformer.

Figures 10.7, 10.8 and 10.9 show the spatial directivity pattern of the minimax beamformer, designed with the non-linear cost function, when no errors occur, when gain errors occur and and when gain and phase errors occur. Similar conclusions can be drawn for the minimax beamformer as for the gain/phase-robust beamformer.

10.5 Conclusions

In this chapter two design procedures have been presented for designing fixed broadband beamformers that are robust against unknown gain and phase errors in the microphone array characteristics. The first design procedure optimises the mean performance, by minimising a weighted sum using the gain and the phase probability density functions. When assuming omni-directional, frequency-flat microphones, similar design procedures as for the non-robust design can be used, where we only require some additional parameters which are easily calculated from the gain and the phase pdf (e.g. higher-order moments of the gain pdf). The second design procedure optimises the worst-case performance, by minimising the maximum cost function over a finite set of feasible microphone characteristics. The denser the grid of feasible microphone characteristics, the higher the computational complexity for solving the minimax problem. However, it has been shown that when considering only gain errors and using the weighted LS cost function, the number of grid points can be drastically reduced to 2^N. We have used the weighted LS, the TLS eigenfilter and the non-linear cost function for designing broadband beamformers with an arbitrary spatial directivity pattern. Simulation results for the different design procedures and cost functions have shown that robust broadband beamformer design for a small-size microphone array indeed leads to a significant performance improvement when gain and/or phase errors occur.

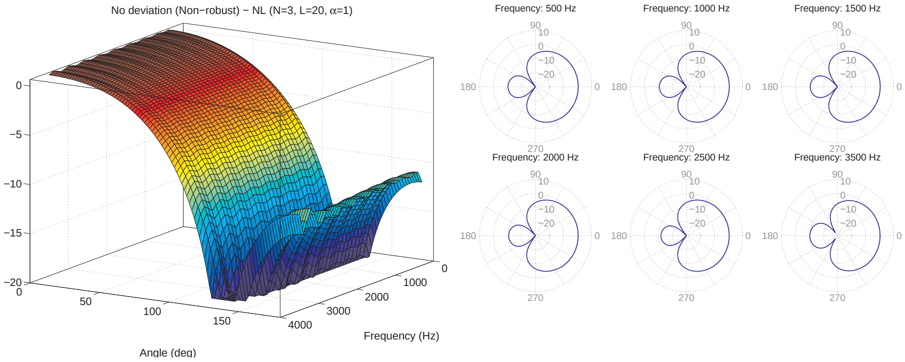


Figure 10.1: Spatial directivity pattern of non-linear non-robust design for no gain and phase errors ($\alpha = 1,\ N = 3,\ L = 20$)

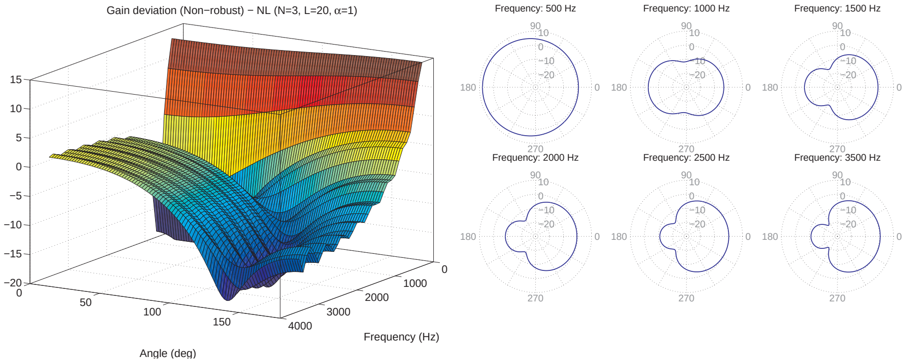


Figure 10.2: Spatial directivity pattern of non-linear non-robust design for gain errors ($\alpha = 1,\ N = 3,\ L = 20$)

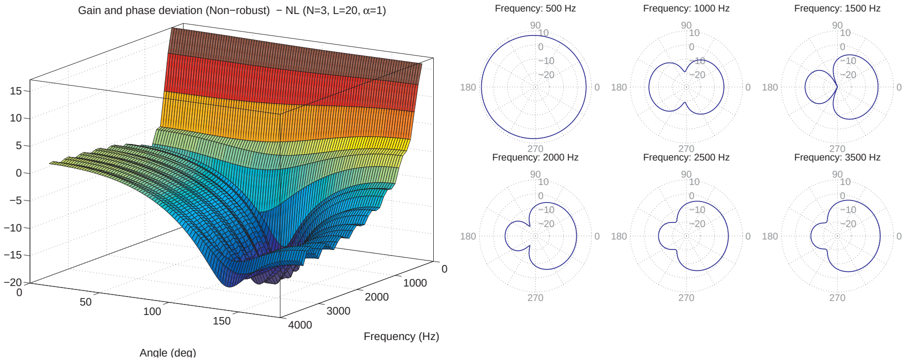


Figure 10.3: Spatial directivity pattern of non-linear non-robust design for gain and phase errors ($\alpha = 1,\ N = 3,\ L = 20$)

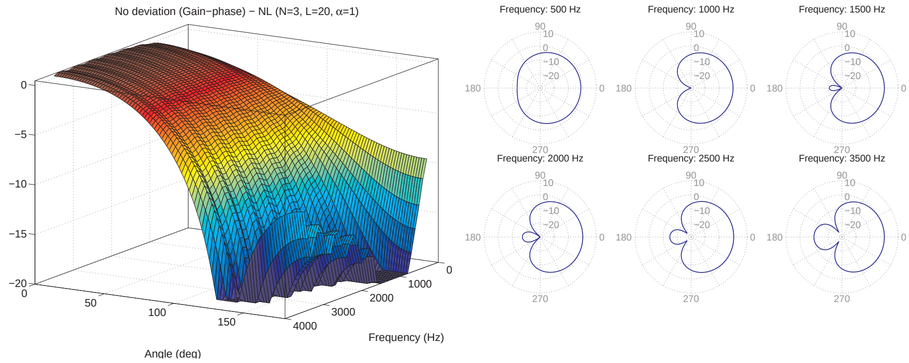


Figure 10.4: Spatial directivity pattern of non-linear gain/phase-robust design for no gain and phase errors ($\alpha = 1,\ N = 3,\ L = 20$)

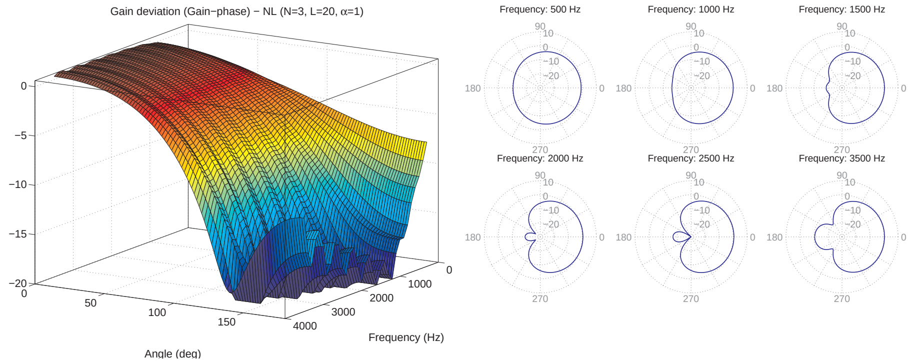


Figure 10.5: Spatial directivity pattern of non-linear gain/phase-robust design for gain errors ($\alpha = 1,\ N = 3,\ L = 20$)

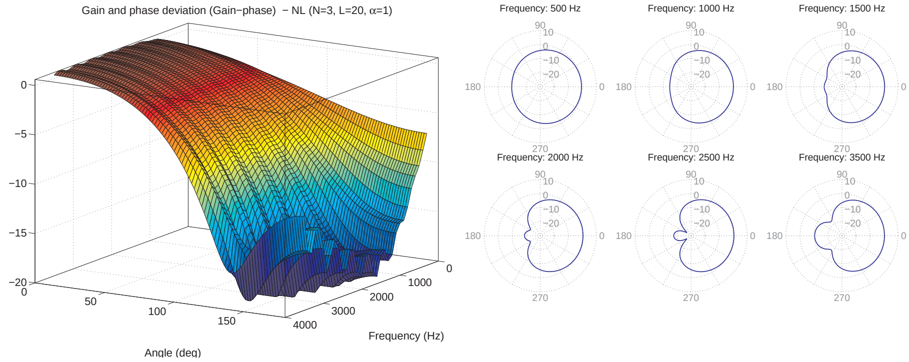


Figure 10.6: Spatial directivity pattern of non-linear gain/phase-robust design for gain and phase errors ($\alpha = 1,\ N = 3,\ L = 20$)

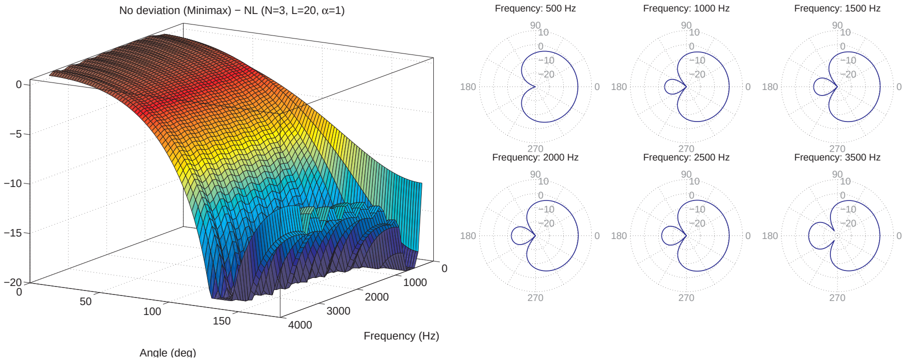


Figure 10.7: Spatial directivity pattern of non-linear minimax design for no gain and phase errors ($\alpha = 1,\ N = 3,\ L = 20$)

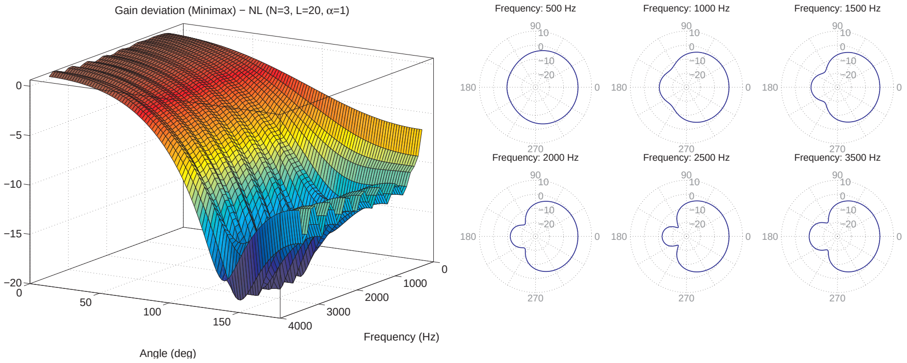


Figure 10.8: Spatial directivity pattern of non-linear minimax design for gain errors ($\alpha = 1,\ N = 3,\ L = 20$)

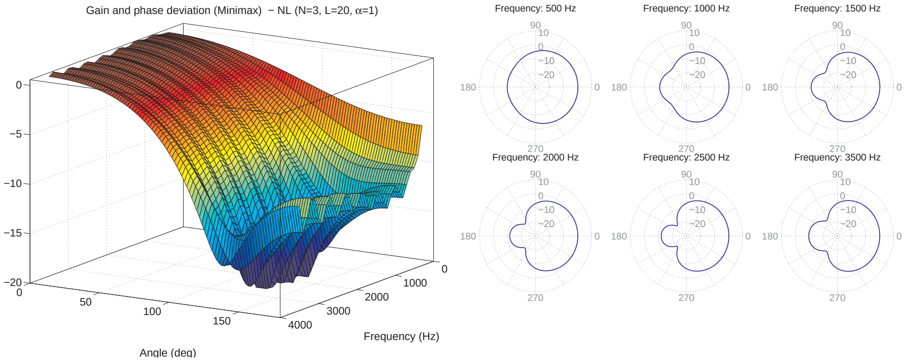


Figure 10.9: Spatial directivity pattern of non-linear minimax design for gain and phase errors ($\alpha = 1,\ N = 3,\ L = 20$)

Chapter 11

Conclusions and Further Research

In this chapter we summarise the main conclusions of this thesis and we list some suggestions for further research.

11.1 Conclusion

In many speech communication applications, the recorded microphone signals are corrupted by background noise, room reverberation and far-end echo signals. This signal degradation may lead to total unintelligibility of the speech signal and decreases the performance of automatic speech recognition systems. Hence, high-performance signal enhancement procedures are called for.

In this thesis we have developed several noise reduction and dereverberation techniques. All presented algorithms are *multi-microphone* signal enhancement algorithms, exploiting both spectral and spatial characteristics. In addition, most algorithms are *adaptive*, enabling these algorithms to deal with different noise situations and with changing acoustic environments. Generally, we have assumed that the noise sources and the acoustic impulse responses are *unknown*, requiring 'blind' estimation techniques. Where possible, we have also incorporated *robustness* against errors in the microphone array characteristics (gain, phase, position) and against other deviations from the assumed signal model (e.g. look direction error, speech detection errors).

In **Part I** we have presented a class of unconstrained optimal filtering techniques for multi-microphone speech enhancement. The optimal filter in the MSE sense is the multi-channel Wiener filter, which produces an MMSE estimate of the speech components in the microphone signals. Using a more general class of estimators, it is possible to trade off speech distortion and noise reduction. Although different possibilities exist to implement the multi-channel Wiener filter, we have considered a GEVD-based implementation, which enables us to easily incorporate the low-rank speech signal model. We have shown that hence the described class of optimal filtering techniques can be considered a multi-microphone extension of the single-microphone subspace-based techniques. An empirical estimate of the optimal filter matrix can be computed using the GSVD of a speech and a noise data matrix. These data matrices are constructed based on the output of a VAD algorithm, which is the only a-priori information the GSVD-based optimal filtering technique relies on. We have derived a number of symmetry properties for the optimal filter, which are valid for the white noise case as well as for the coloured noise case and for any weighting function. When analysing the multi-channel Wiener filter in the frequency-domain, it can be decomposed into a spectral and a spatial filtering term. Furthermore, we have shown that the unconstrained optimal filtering technique can also be used for combined noise and echo reduction.

Both for the batch and the recursive version of the GSVD-based optimal filtering technique, the computational complexity is quite high. Therefore, several techniques have been developed for reducing the overall complexity (recursive GSVD-updating, square root-free implementation, sub-sampling). When considering realistic parameter values, the recursive GSVD-based optimal filtering technique indeed becomes suitable for real-time implementation. In addition, we have shown that the GSVD-based optimal filtering technique can be incorporated into a GSC-type structure with an ANC postprocessing stage.

The performance (unbiased SNR improvement, speech distortion and robustness) of the GSVD-based implementation of the multi-channel optimal filtering technique has been analysed for several acoustic environments (including a real-life recording) and has been compared with standard fixed and adaptive beamforming techniques. For higher filter lengths and for lower reverberation times, the unbiased SNR improvement increases and the speech distortion decreases. The ANC postprocessing stage can either be used for increasing the noise reduction performance or for complexity reduction without decreasing the performance. The ANC postprocessing stage however also gives rise to a slight increase in speech distortion, which can be limited by using longer filter lengths. Since the GSVD-based optimal filtering technique uses no other a-priori information than the output of a VAD algorithm, it is expected to be quite sensitive to speech detection errors. However, it has been shown that the unbiased SNR improvement is not degraded by speech detection errors, but that speech distortion increases with increasing error rate (for error rates smaller than 0.2, speech distortion however remains limited). When comparing the performance of the GSVD-based optimal filtering technique with standard beamforming techniques, the SNR improvement of the GSVD-based optimal filtering technique with ANC postprocessing stage outperforms the SNR improvement of the GSC for all considered scenarios. In addition, the robustness of the GSC and the GSVD-based optimal filter have been analysed for several deviations from the assumed nominal signal model. It has been shown that the performance of the GSVD-based optimal filter is independent of a deviation in the microphone gain and phase and that the GSVD-based optimal filter is more robust than the GSC for microphone mismatch, microphone displacement and look direction error.

In **Part II** we have discussed multi-microphone algorithms for time-delay estimation (TDE), dereverberation, and combined noise reduction and dereverberation. Since the presented algorithms require a (partial) estimate of the acoustic impulse responses, we have also developed batch and adaptive algorithms for (partially) estimating the acoustic impulse responses, both in the time-domain and in the frequency-domain. We have extended a recently developed adaptive EVD algorithm for TDE to noisy environments, by using an adaptive GEVD or by pre-whitening the microphone signals. For the adaptive GEVD, we have derived a stochastic gradient algorithm which iteratively estimates the generalised eigenvector corresponding to the smallest generalised eigenvalue. In addition, we have extended all TDE algorithms to the case of more than two microphones. Simulations show that the time-delays can be estimated more robustly using the adaptive GEVD algorithm than using the adaptive EVD algorithm and the adaptive pre-whitening algorithm.

We have presented a frequency-domain technique for estimating the acoustic transfer functions when the microphone signals are corrupted by spatially coloured noise. The acoustic transfer function vector can be calculated from the generalised eigenvector, corresponding to the largest generalised eigenvalue of the speech and the noise correlation matrices. However, unlike time-domain subspace-based techniques, this frequency-domain technique requires some prior knowledge, i.e. the norm of the transfer function vector, limiting its practical use to rather time-invariant acoustic environments. Using the estimated transfer function vector, dereverberation can be performed with a normalised matched filtering approach. We have shown that the MMSE estimate of the clean dereverberated speech signal can be obtained by matched filtering of the MMSE estimates of the speech components in the microphone signals. Hence, by integrating the normalised matched filter with the multi-channel Wiener filter, we have developed a combined noise reduction and dereverberation technique. Simulations show that this combined technique provides a trade-off between SNR improvement and dereverberation.

In **Part III** we have discussed design procedures for fixed broadband beamformers, which can be used both for noise reduction and for dereverberation. We have presented several cost functions for designing far-field and near-field broadband beamformers with an arbitrary spatial directivity pattern using an arbitrary microphone configuration and an FIR filter-and-sum structure. We have discussed the weighted least-squares, the maximum energy array and a modified non-linear cost function and we have developed two novel cost functions, which are based on eigenfilters. In the conventional eigenfilter technique a reference frequency-angle point is required, whereas in the eigenfilter technique based on a TLS error criterion, this reference point is not required. Although in general we would like to use the non-linear design procedure, this procedure gives rise to a high computational complexity, since it requires an iterative optimisation technique. Hence, we have compared the performance of the non-iterative design procedures, having a lower computational complexity, using the non-linear cost function as a performance criterion. Using simulations with different passband and stopband specifications, we have shown that the TLS eigenfilter technique is the preferred non-iterative design procedure.

We have also presented design procedures for broadband beamformers which operate at several distances from the microphone array. Although this extension is straightforward for most cost functions, for the TLS eigenfilter and the maximum energy array cost function this extension leads to a significantly different optimisation problem, for which no closed-form solution is available. Using simulations we have shown that mixed near-field far-field design provides a trade-off between the near-field and the far-field performance.

Since in many applications the microphone characteristics are not exactly known and can even change over time, we have developed two design procedures for designing broadband beamformers that are robust against (unknown) gain and phase errors in the microphone characteristics. The first design procedure optimises the mean performance, requiring knowledge about the gain and the phase pdfs. When assuming omni-directional, frequency-flat microphones, similar design procedures as for the non-robust design can be used, where we only require some additional parameters which are easily calculated from the gain and the phase pdf. The second design procedure optimises the worst-case performance, by minimising the maximum cost function over a finite set of feasible microphone characteristics. The denser the grid of feasible microphone characteristics, the higher the computational complexity for solving the minimax problem. However, it has been shown that when considering only gain errors and using the weighted LS cost function, the number of grid points can be drastically reduced. Simulation results have shown that robust broadband beamformer design for a small-size microphone array indeed leads to a significant performance improvement when gain and/or phase errors occur.

11.2 Suggestions for further research

In Part I the GSVD-based optimal filtering technique for multi-microphone noise reduction has been discussed. It has been shown that the noise reduction performance of this technique is better than the noise reduction performance of standard fixed and adaptive beamforming techniques and that it is more robust against deviations from the assumed signal model. Although several techniques have been discussed for reducing the overall computational complexity (cf. Chapter 4), the complexity remains quite high – in fact much higher than the complexity of standard beamforming techniques. In addition to the presented techniques, it would therefore be interesting to investigate other techniques for *reducing the computational complexity* (using e.g. subband QR-decomposition-based techniques or stochastic gradient, i.e. LMS-type, algorithms) without severely reducing the performance and the robustness.

Furthermore, when the SNR of the microphone signals is very low or when highly non-stationary noise sources are present, it is possible that the VAD-algorithm completely fails. In this case, the performance of the multi-channel Wiener filter becomes unreliable, i.e. resulting in an unacceptably large speech distortion or slow convergence. Hence, it is necessary to incorporate a larger robustness for these scenarios. On the other hand, fixed broadband beamforming techniques, cf. Part III, do not rely on the output of a VAD-algorithm and hence are very robust in scenarios where the VAD fails. Therefore, an interesting research topic would be *the combination of multi-channel Wiener filtering and fixed broadband beamforming techniques*. We expect the combined technique to be more robust than the multi-channel Wiener filter in scenarios where the VAD fails, and the performance of the combined technique to be better than the performance of fixed broadband beamforming techniques in other scenarios. Combining multi-channel Wiener filtering and fixed broadband beamforming may be possible by adding a (regularisation) term, which is related to fixed broadband beamforming, to the MSE cost function in (3.7).

As already mentioned in Chapter 6, the *computational complexity of the adaptive GEVD and the adaptive pre-whitening algorithm for time-delay estimation* is much higher than the complexity of the adaptive EVD algorithm, since in each iteration step two additional matrix-vector multiplications need to be performed. Reducing the computational complexity of these algorithms is a topic of further research. One could e.g. replace the empirical noise correlation matrix $\mathbf{R}_{vv}[k]$ in the adaptive GEVD algorithm by an instantaneous estimate $\mathbf{v}[k']\mathbf{v}^T[k']$, where $\mathbf{v}[k']$ is a noise data vector which is stored in a buffer during noise-only periods and which is used in the update equations during subsequent speech-and-noise periods.

In Section 6.3.2 we have developed a stochastic gradient algorithm which estimates and tracks the generalised singular vector corresponding to the *smallest* generalised singular value. In Section 7.2.3 a stochastic gradient algorithm is required which estimates and tracks the generalised singular vector corresponding to the *largest* generalised singular value. Such subspace tracking algorithms do exist for the SVD [41][199][281], but it remains a topic of further research to *extend these subspace tracking algorithms to the GSVD*, while still maintaining the $\mathcal{O}(N)$ computational complexity.

We believe that *acoustic impulse response estimation and dereverberation* still require much research attention, since some fundamental issues still need to be resolved. Both time-domain and frequency-domain techniques experience major problems in accurately estimating the complete acoustic impulse responses, cf. Chapter 6 and 7. The main problem for the time-domain subspace-based techniques is the length of the acoustic impulse responses in combination with the low-rank model of the speech signal and the presence of background noise. Moreover, time-domain subspace-based techniques appear to be very sensitive to underestimating the length of the acoustic impulse responses, while the underlying reason for this sensitivity is not well understood. The effect of some of these problems can be reduced by using frequency-domain techniques. However, frequency-domain techniques require some prior knowledge about the acoustic transfer functions in order to resolve a scaling problem which occurs in each frequency bin (cf. Section 7.2). Solving this scaling problem remains a topic of further research. A similar scaling problem however also occurs in frequency-domain blind source separation (BSS) techniques. Recently, using relations between BSS and (adaptive) beamforming techniques, algorithms have been developed which (partially) solve the scaling and the permutation problem occurring in BSS [6][207]. It would be interesting to *investigate if these BSS-algorithms can also be used for solving the scaling problem occurring in frequency-domain acoustic transfer function estimation.* Also other blind system identification techniques (e.g. based on multi-channel linear prediction [160][166][282] or non-linear Kalman filtering [98][141][270]) should be further investigated, since these techniques have already proved their usefulness in other applications (e.g. digital communications).

Bibliography

[1] K. Abed-Meraim, W. Qiu, and Y. Hua. Blind System Identification. *Proc. IEEE*, 85(8):1310–1322, August 1997.

[2] S. Affes and Y. Grenier. A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech. *IEEE Trans. Speech and Audio Processing*, 5(5):425–437, September 1997.

[3] M. Ali. Stereophonic echo cancellation system using time-varying all-pass filtering for signal decorrelation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 3689–3692, Seattle WA, USA, May 1998.

[4] J. Allen and D. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65:943–950, April 1979.

[5] J. B. Allen, D. A. Berkley, and J. Blauert. Multimicrophone signal processing technique to remove room reverberation of speech signals. *Journal of the Acoustical Society of America*, 62(4):912–915, October 1977.

[6] S. Araki, S. Makino, R. Mukai, Y. Hinamoto, T. Nishikawa, and H. Saruwatari. Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1785–1788, Orlando FL, USA, May 2002.

[7] C. Avendano, J. Benesty, and D. R. Morgan. A Least Squares Component Normalization Approach to Blind Channel Identification. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1797–1800, Phoenix AZ, USA, May 1999.

[8] D. Bees, M. Blostein, and P. Kabal. Reverberant Speech Enhancement Using Cepstral Processing. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 977–980, Toronto, Ontario, Canada, May 1991.

[9] J. Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *Journal of the Acoustical Society of America*, 107(1):384–391, January 2000.

[10] J. Benesty and T. Gänsler. A robust fast recursive least squares adaptive algorithm. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3785–3788, Salt Lake City UT, USA, May 2001.

[11] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, editors. *Advances in Network and Acoustic Echo Cancellation.* Springer-Verlag, 2001.

[12] J. Benesty and Y. Huang, editors. *Adaptive Signal Processing : Applications to Real-World Problems.* Springer-Verlag, 2003.

[13] J. Benesty and D. R. Morgan. Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 789–792, Istanbul, Turkey, May 2000.

[14] J. Benesty, D. R Morgan, J. L. Hall, and M. M. Sondhi. Stereophonic acoustic echo cancellation using nonlinear transformations and comb filtering. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 3673–3676, Seattle WA, USA, May 1998.

[15] J. Benesty, D. R. Morgan, and M. M. Sondhi. A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation. *IEEE Trans. Speech and Audio Processing*, 6:156–165, March 1998.

[16] J. Bitzer and K. U. Simmer. *Superdirective Microphone Arrays*, chapter 2 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 19–38. Springer-Verlag, May 2001.

[17] 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 Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 2965–2968, Phoenix, Arizona, USA, May 1999.

[18] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. Multi-microphone noise reduction techniques as front-end devices for speech recognition. *Speech Communication*, 34(1-2):3–12, April 2001.

[19] S. Black. Minnesota no longer hearing-aid mecca. The Minneapolis St. Paul Business Journal. November 2002.

`http://twincities.bizjournals.com/twincities/stories/2002/11/18/story4.html`.

[20] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localisation.* The MIT Press, 1983.

[21] S. F. Boll. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Trans. Acoust., Speech, Signal Processing*, 27(2):113–120, April 1979.

[22] M. Brandstein and D. Ward, editors. *Microphone Arrays: Signal Processing Techniques and Applications.* Digital Signal Processing. Springer-Verlag, 2001.

[23] M. S. Brandstein, J. E. Adcock, and H. F. Silverman. A closed-form location estimator for use with room environment microphone arrays. *IEEE Trans. Speech and Audio Processing*, 5(1):45–50, January 1997.

[24] M. Buck. Aspects of first-order differential microphone arrays in the presence of sensor imperfections. *European Transactions on Telecommunications, special issue on Acoustic Echo and Noise Control*, 13(2):115–122, Mar-Apr 2002.

[25] K. M. Buckley. Broad-Band Beamforming and the Generalized Sidelobe Canceller. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(5):1322–1323, October 1986.

[26] P. Butler and A. Cantoni. Eigenvalues and eigenvectors of symmetric centrosymmetric matrices. *Linear Algebra and its Applications*, 13:275–288, March 1976.

[27] J. A. Cadzow. Signal Enhancement – A Composite Property Mapping Algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, 36(1):49–62, January 1988.

[28] O. Cappé. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. Speech and Audio Processing*, 2(2):345–349, April 1994.

[29] B. Champagne, S. Bédard, and A. Stéphenne. Performance of time-delay estimation in the presence of room reverberation. *IEEE Trans. Speech and Audio Processing*, 4(2):148–152, March 1996.

[30] Y. T. Chan and K. C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Trans. Signal Processing*, 42(8):1905–1919, August 1994.

[31] J. P. Charlier, M. Vanbegin, and P. Van Dooren. On efficient implementations of Kogbetliantz's algorithm for computing the singular value decomposition. *Numerische Mathematik*, 52:279–300, 1988.

[32] T. Chen. Unified eigenfilter approach: with applications to spectral/spatial filtering. In *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, pages 331–334, Chicago, USA, March 1993.

[33] P. C. Ching, Y. T. Chan, and K. C. Ho. Constrained adaptation for time delay estimation with multipath propagation. *IEE Proceedings-F*, 138(5):453–458, October 1991.

[34] I. Claesson and S. Nordholm. A Spatial Filtering Approach to Robust Adaptive Beaming. *IEEE Trans. Antennas Propagat.*, 40(9):1093–1096, September 1992.

[35] T. Coleman, M. A. Branch, and A. Grace. *MATLAB Optimization Toolbox User's Guide.* The Mathworks Inc., Natick MA, USA, January 1999.

[36] H. Cox, R. Zeskind, and T. Kooij. Practical supergain. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(3):393–398, June 1986.

[37] H. Cox, R. M. Zeskind, and M. M. Owen. Robust Adaptive Beamforming. *IEEE Trans. Acoust., Speech, Signal Processing*, 35(10):1365–1376, October 1987.

[38] J. Davies. Wireless carriers await fallout from N.Y. ban. The San-Diego Union Tribune. June 2001. `http://www.jabra.com/news/viewarticle.cfm?id=61`.

[39] T. Dawn. Soon machines will understand every word you say. *Scientific Computing World*, pages 25–30, January 1995.

[40] B. De Moor. The singular value decomposition and long and short spaces of noisy matrices. *IEEE Trans. Signal Processing*, 41(9):2826–2839, September 1993.

[41] J. Dehaene, M. Moonen, and J. Vandewalle. An improved stochastic gradient algorithm for principal component analysis and subspace tracking. *IEEE Trans. Signal Processing*, 45(10):2582–2586, October 1997.

[42] J. R. Deller, J. G. Proakis, and J. H. L. Hansen. *Discrete-Time Processing of Speech Signals.* Macmillan Publishing Company, Englewood Cliffs, New Jersey, 1993.

[43] M. Dendrinos, S. Bakamidis, and G. Carayannis. Speech enhancement from noise : A regenerative approach. *Speech Communication*, 10(2):45–57, February 1991.

[44] R. M. M. Derkx, G. P. M. Egelmeers, and P. C. W. Sommen. New constraining method for partitioned block frequency-domain adaptive filters. *IEEE Trans. Signal Processing*, 50(9):2177–2186, September 2002.

[45] E. J. Diethorn. *Subband Noise Reduction Methods for Speech Enhancement*, chapter 9 in "Acoustic Signal Processing for Telecommunication" (Gay, S. L. and Benesty, J., Eds.), pages 155–178. Kluwer Academic Publishers, Boston, 2000.

[46] S. Doclo and E. De Clippel. Verbetering van spraakverstaan bij hoortoestellen via adaptieve ruisonderdrukking in reële tijd. Master's thesis, Katholieke Universiteit Leuven, Belgium, 1997. UDC : 681.5.017(043).

[47] S. Doclo, E. De Clippel, and M. Moonen. Combined Acoustic Echo and Noise Reduction using GSVD-based Optimal Filtering. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 1061–1064, Istanbul, Turkey, June 2000.

[48] S. Doclo, E. De Clippel, and M. Moonen. Multi-microphone noise reduction using GSVD-based optimal filtering with ANC postprocessing stage. In *Proc. of DSP2000 Workshop*, pages 383–388, Hunt TX, USA, October 2000.

[49] S. Doclo, I. Dologlou, and M. Moonen. A novel iterative signal enhancement algorithm for noise reduction in speech. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, pages 1435–1438, Sydney, Australia, December 1998.

[50] S. Doclo, K. Eneman, and M. Moonen. Voice activity detection. Technical Report ESAT-SISTA/TR 2001-62, ESAT, Katholieke Universiteit Leuven, Belgium, January 2002.

[51] S. Doclo and M. Moonen. Robustness of SVD-based Optimal Filtering for Noise Reduction in Multi-Microphone Speech Signals. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 80–83, Pocono Manor, Pennsylvania, USA, September 1999.

[52] S. Doclo and M. Moonen. SVD-based optimal filtering with applications to noise reduction in speech signals. Technical Report ESAT-SISTA/TR 1999-33, ESAT, Katholieke Universiteit Leuven, Belgium, April 1999.

[53] S. Doclo and M. Moonen. SVD-based optimal filtering with applications to noise reduction in speech signals. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 143–146, New Paltz, New York, USA, October 1999.

[54] S. Doclo and M. Moonen. Noise Reduction in Multi-Microphone Speech Signals using Recursive and Approximate GSVD-based Optimal Filtering. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS2000)*, Hilvarenbeek, The Netherlands, March 2000.

[55] S. Doclo and M. Moonen. Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 31–34, Darmstadt, Germany, September 2001.

[56] S. Doclo and M. Moonen. *GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement*, chapter 6 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 111–132. Springer-Verlag, May 2001.

[57] S. Doclo and M. Moonen. Robust time-delay estimation in highly adverse acoustic environments. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 59–62, New Paltz NY, USA, October 2001.

[58] S. Doclo and M. Moonen. Comparison of least-squares and eigenfilter techniques for broadband beamforming. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS2002)*, pages 73–76, Leuven, Belgium, March 2002.

[59] S. Doclo and M. Moonen. Design of far-field broadband beamformers using eigenfilters. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages III 237–240, Toulouse, France, September 2002.

[60] S. Doclo and M. Moonen. Far-field and near-field broadband beamformer design. Technical Report ESAT-SISTA/TR 2002-109, ESAT, Katholieke Universiteit Leuven, Belgium, July 2002.

[61] S. Doclo and M. Moonen. GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Trans. Signal Processing*, 50(9):2230–2244, September 2002.

[62] S. Doclo and M. Moonen. Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. *IEEE Trans. Signal Processing*, 51(10):2511–2526, October 2003.

[63] S. Doclo and M. Moonen. Design of broadband beamformers robust against microphone position errors. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 267–270, Kyoto, Japan, September 2003.

[64] S. Doclo and M. Moonen. Design of broadband speech beamformers robust against errors in the microphone array characteristics. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages V 473–476, Hong Kong, China, April 2003.

[65] S. Doclo and M. Moonen. Design of far-field and near-field broadband beamformers using eigenfilters. *Signal Processing*, 83(12):2641–2673, December 2003.

[66] S. Doclo and M. Moonen. Robust adaptive time delay estimation for speaker localisation in noisy and reverberant acoustic environments. *EURASIP Journal on Applied Signal Processing*, 2003(11):1110–1124, October 2003.

[67] S. Doclo and M. Moonen. Multi-Microphone Noise Reduction Using Recursive GSVD-Based Optimal Filtering with ANC Postprocessing Stage. *IEEE Trans. Speech and Audio Processing*, 13(1), January 2005.

[68] I. Dologlou and G. Carayannis. Physical Representation of Signal Reconstruction from Reduced Rank Matrices. *IEEE Trans. Signal Processing*, 39(7):1682–1684, July 1991.

[69] I. Dologlou, J.-C. Pesquet, and J. Skowronski. Projection-based rank reduction algorithms for multichannel modelling and image compression. *Signal Processing*, 48(2):97–109, January 1996.

[70] I. Dologlou, S. Van Huffel, and D. Van Ormondt. Improved Signal Enhancement Procedures Applied to Exponential Data Modeling. *IEEE Trans. Signal Processing*, 45(3):799–803, March 1997.

[71] C. L. Dolph. A current distribution for broadside arrays which optimizes the relationship between beam width and sidelobe level. *Proc. IRE*, 34:335–348, 1946.

[72] Y. Duk, K. Al-Naimi, and A. Kondoz. Improved Voice Activity Detection Based on a Smoothed Statistical Likelihood Ratio. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City UT, USA, May 2001.

[73] G. P. M. Egelmeers and P. C. W. Sommen. A new method for efficient convolution in frequency domain by nonuniform partitioning for adaptive filtering. *IEEE Trans. Signal Processing*, 42(12):3123–3129, December 1994.

[74] N. Elkin. Look ma - no wires!. eMarketer. November 2002. `http://www.emarketer.com/news/article.php?1001838`.

[75] G. Elko. *Superdirectional Microphone Arrays*, chapter 10 in “Acoustic Signal Processing for Telecommunication” (Gay, S. L. and Benesty, J., Eds.), pages 181–237. Kluwer Academic Publishers, Boston, 2000.

[76] G. W. Elko. Microphone array systems for hands-free telecommunication. *Speech Communication*, 20(3-4):229–240, December 1996.

[77] EMC. EMC announces 1 billion cellular phone users in the world! `http://www.emc-database.com/website.nsf/index/pr020319`, March 2002.

[78] K. Eneman. *Subband and Frequency-Domain Adaptive Filtering Techniques for Speech Enhancement in Hands-free Communication.* PhD thesis, Katholieke Universiteit Leuven, Belgium, March 2002.

[79] K. Eneman and M. Moonen. A Relation between Subband and Frequency-Domain Adaptive Filtering. In *Proc. 13th International Conference on Digital Signal Processing*, Santorini, Greece, July 1997.

[80] K. Eneman and M. Moonen. Hybrid subband/frequency-domain adaptive systems. *Signal Processing*, 81(1):117–136, January 2001.

[81] K. Eneman and M. Moonen. Iterated partitioned block frequency-domain adaptive filtering for acoustic echo cancellation. *IEEE Trans. Speech and Audio Processing*, 11(2):143–158, March 2003.

[82] P. Eneroth, S. Gay, T. Gänsler, and J. Benesty. A Hybrid FRLS/NLMS Stereo Acoustic Echo Canceller. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 20–23, Pocono Manor PA, USA, September 1999.

[83] Y. Ephraim and D. Malah. Speech Enhancement Using a Minimun Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(6):1109–1121, December 1984.

[84] Y. Ephraim and D. Malah. Speech Enhancement Using a Minimun Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 33(2):443–445, April 1985.

[85] Y. Ephraim and H. L. Van Trees. A Signal Subspace Approach for Speech Enhancement. *IEEE Trans. Speech and Audio Processing*, 3(4):251–266, July 1995.

[86] M. H. Er. A robust formulation for an optimum beamformer subject to amplitude and phase perturbations. *Signal Processing*, 19(1):17–26, 1990.

[87] M. H. Er and A. Cantoni. Derivative Constraints for Broad-Band Element Space Antenna Array Processors. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(6):1378–1393, December 1983.

[88] F. A. Everest. *The Master Handbook of Acoustics.* McGraw-Hill, 4th edition, 2001.

[89] C. F. Eyring. Reverberation Time in “Dead” Rooms. *Journal of the Acoustical Society of America*, 1:217–241, 1930.

[90] L. Faiget, R. Ruiz, and C. Legros. The True Duration of the Impulse Response Used to Estimate Reverberation Time. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 913–916, Atlanta GA, USA, May 1996.

[91] J. L. Flanagan, A. C. Surendran, and E. E. Jan. Spatially selective sound capture for speech and audio processing. *Speech Communication*, 13:207–222, 1993.

[92] J.L. Flanagan. Parametric coding of speech spectra. *Journal of the Acoustical Society of America*, 68(2):412–419, August 1980.

[93] R. Fletcher. *Practical Methods of Optimization.* Wiley, New York, 1987.

[94] D. A. Florêncio and H. S. Malvar. Multichannel filtering for optimum noise reduction in microphone arrays. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 197–200, Salt Lake City UT, USA, May 2001.

[95] O. L. Frost III. An Algorithm for Linearly Constrained Adaptive Array Processing. *Proc. IEEE*, 60:926–935, August 1972.

[96] S. Furui and M. M. Sondhi. *Advances in Speech Signal Processing.* Marcel Dekker, 1991.

[97] M. Gabrea, E. Grivel, and M. Najun. A single microphone Kalman filter-based noise canceller. *IEEE Signal Processing Lett.*, 6(3):55–57, March 1999.

[98] S. Gannot. Sequential-joint estimation of signal and parameters using the unscented kalman filter with application to single- and multi-microphone speech enhancement. Technical report, ESAT, Katholieke Universiteit Leuven, Belgium, September 2001.

[99] S. Gannot, D. Burshtein, and E. Weinstein. Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms. *IEEE Trans. Speech and Audio Processing*, 6(4):373–385, July 1998.

[100] S. Gannot, D. Burshtein, and E. Weinstein. Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech. *IEEE Trans. Signal Processing*, 49(8):1614–1626, August 2001.

[101] S. Gannot and M. Moonen. Subspace methods for multi-microphone speech dereverberation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 47–50, Darmstadt, Germany, September 2001.

[102] T. Gänsler and J. Benesty. Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview. *International Journal on Adaptive Control*, 14:565–586, February 2000.

[103] S. Gay. *Fast Projection Algorithms with Application to Voice Echo Cancellation.* PhD thesis, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA, October 1994.

[104] S. L. Gay and J. Benesty, editors. *Acoustic Signal Processing for Telecommunication*. Kluwer Academic Publishers, Boston, 2000.

[105] W. M. Gentleman. Least squares computation by Givens transformations without square roots. *Journal of the Institute of Mathematics and its Applications*, 12:329–336, 1973.

[106] P. G. Georgiou, C. Kyriakakis, and P. Tsakalides. Robust time delay estimation for sound source localization in noisy environments. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 237–240, New Paltz NY, USA, October 1997.

[107] J. D. Gibson, B. Koo, and S. D. Gray. Filtering of colored noise for speech enhancement and coding. *IEEE Trans. Signal Processing*, 39(8):1732–1742, August 1991.

[108] A. Gilloire and V. Turbin. Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 3681–3684, Seattle WA, USA, May 1998.

[109] Z. Goh, K.-C. Tan, and B. T. G. Tan. Postprocessing Method for Suppressing Musical Tones Generated by Spectral Subtraction. *IEEE Trans. Speech and Audio Processing*, 6(3):287–292, May 1998.

[110] G. H. Golub and C. F. Van Loan. *Matrix Computations*. MD : John Hopkins University Press, Baltimore, 3rd edition, 1996.

[111] R. M. Gray. Toeplitz and circulant matices: A review. `http://ee.stanford.edu/~gray/toeplitz.pdf`. Technical report, Dept. of Electrical Engineering, Stanford University, Stanford CA, USA, August 2002.

[112] J. E. Greenberg, P. M. Peterson, and P. M. Zurek. Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *Journal of the Acoustical Society of America*, 94(5):3009–3010, November 1993.

[113] J. E. Greenberg and P. M. Zurek. Evaluation of an adaptive beamforming method for hearing aids. *Journal of the Acoustical Society of America*, 91(3):1662–1676, March 1992.

[114] J. E. Greenberg, P. M. Zurek, and M. Brantley. Evaluation of feedback-reduction algorithms for hearing aids. *Journal of the Acoustical Society of America*, 108(5):2366–2376, November 2000.

[115] Y. Grenier. A microphone array for car environments. *Speech Communication*, 12:25–39, 1993.

[116] L. J. Griffiths and C. W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagat.*, 30(1):27–34, January 1982.

[117] I. Gürelli and C. L. Nikias. EVAM: An Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals. *IEEE Trans. Signal Processing*, 43(1):134–149, January 1995.

[118] Y. Haneda. Common Acoustical Pole and Zero Modeling of Room Transfer Functions. *IEEE Trans. Speech and Audio Processing*, 2(2):320–328, April 1994.

[119] J. H. L. Hansen and M. A. Clements. Constrained iterative speech enhancement with application to speech recognition. *IEEE Trans. Signal Processing*, 39(4):795–805, April 1991.

[120] P. C. Hansen and S. H. Jensen. FIR Filter Representations of Reduced-Rank Noise Reduction. *IEEE Trans. Signal Processing*, 46(6):1737–1741, June 1998.

[121] P. S. K. Hansen. *Signal Subspace Methods for Speech Enhancement.* PhD thesis, Technical University of Denmark, Lyngby, Denmark, September 1997.

[122] E. Hänsler. The hands-free telephone problem – an annotated bibliography. *Signal Processing*, 27:259–271, June 1992.

[123] S. Haykin. *Adaptive Filter Theory.* Information and system sciences series. Prentice Hall, Englewood Cliffs, New Jersey, 4th edition, 2001.

[124] P. Heitkämper. An Adaptation Control for Acoustic Echo Cancellers. *IEEE Signal Processing Lett.*, 4:170–172, June 1997.

[125] W. Herbordt and W. Kellermann. GSAEC - Acoustic Echo Cancellation embedded into the Generalized Sidelobe Canceller. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 1843–1846, Tampere, Finland, September 2000.

[126] W. Herbordt and W. Kellermann. Frequency-Domain Integration of Acoustic Echo Cancellation with a Generalized Sidelobe Canceller. *European Transactions on Telecommunications, special issue on Acoustic Echo and Noise Control*, 13(2):123–132, March-April 2002.

[127] M. W. Hoffman, T. D. Trine, K. M. Buckley, and D. J. Van Tasell. Robust adaptive microphone array processing for hearing aids: realistic speech enhancement predictions. *Journal of the Acoustical Society of America*, 96:759–770, 1994.

[128] 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*, 47(10):2677–2684, October 1999.

[129] T. Houtgast and H. J. M. Steeneken. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America*, 77(3):1069–1077, March 1985.

[130] J. Huang and Y. Zhao. Energy-Constrained Signal Subspace Method for Speech Enhancement and Recognition. *IEEE Signal Processing Lett.*, 4(10):283–285, October 1997.

[131] L.-S. Huang and C.-H. Yang. A Novel Approach to Robust Speech Endpoint Detection in Car Environments. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1751–1754, Istanbul, Turkey, June 2000.

[132] Y. Huang, J. Benesty, and G. W. Elko. *Microphone Arrays for Video Camera Steering*, chapter 11 in "Acoustic Signal Processing for Telecommunication" (Gay, S. L. and Benesty, J., Eds.), pages 239–259. Kluwer Academic Publishers, Boston, 2000.

[133] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau. Real-Time Passive Source Localization: A Practical Linear-Correction Least-Squares Approach. *IEEE Trans. Speech and Audio Processing*, 9(8):943–956, November 2001.

[134] N. K. Jablon. Adaptive beamforming with the Generalized Sidelobe Canceller in the presence of array imperfections. *IEEE Trans. Antennas Propagat.*, 34(8):996–1012, August 1986.

[135] F. Jabloun and B. Champagne. A multi-microphone signal subspace approach for speech enhancement. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 205–208, Salt Lake City UT, USA, May 2001.

[136] C. P. Janse. Audio Conferencing. In *Signaalbewerking voor communicatie : Nieuwe algebraïsche signaalbewerkingsmethoden voor mobiele communicatie en audiotoepassingen*, chapter 8, pages 173–198. cursus PATO, Delft, September 1996.

[137] L. B. Jensen. Hearing aid with adaptive matching of input transducers. United Stated Patent No. 6,741,714, May 25, 2004.

[138] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen. Reduction of Broad-Band Noise in Speech by Truncated QSVD. *IEEE Trans. Speech and Audio Processing*, 3(6):439–448, November 1995.

[139] C. W. Jim. A comparison of two LMS constrained optimal array structures. *Proc. IEEE*, 65(12):1730–1731, December 1977.

[140] D. H Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques.* Prentice Hall, Englewood Cliffs, New Jersey, 1st edition, 1993.

[141] S. Julier, J. Uhlmann, and H.F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. on Automatic Control*, 45(3):477–482, March 2000.

[142] J. C. Junqua, B. Mak, and B. Reaves. A Robust Algorithm for Word Boundary Detection in the Presence of Noise. *IEEE Trans. Speech and Audio Processing*, 2(3):406–412, April 1994.

[143] J. C. Junqua, B. Reaves, and B. Mak. A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognizer. In *Proc. EUROSPEECH*, pages 1371–1374, 1991.

[144] M. Kajala and M. Hämäläinen. Broadband beamforming optimization for speech enhancement in noisy environments. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 19–22, New Paltz NY, USA, October 1999.

[145] J. M. Kates. Feedback Cancellation in Hearing Aids: Results form a Computer Simulation. *IEEE Trans. Signal Processing*, 39(3):553–562, March 1991.

[146] J. M. Kates. Superdirective arrays for hearing aids. *Journal of the Acoustical Society of America*, 94(4):1930–1933, October 1993.

[147] J. M. Kates. Classification of background noises for hearing-aid applications. *Journal of the Acoustical Society of America*, 97(1):461–470, January 1995.

[148] J. M. Kates. Constrained adaptation for feedback cancellation in hearing aids. *Journal of the Acoustical Society of America*, 106(2):1010–1019, August 1999.

[149] J. M. Kates and M. R. Weiss. A comparison of hearing-aid array-processing techniques. *Journal of the Acoustical Society of America*, 99(5):3138–3148, May 1996.

[150] W. Kellermann. A Self-Steering Digital Microphone Array. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3581–3584, Toronto, Canada, May 1991.

[151] W. Kellermann. Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 219–222, München, Germany, April 1997.

[152] W. Kellermann. *Acoustic Echo Cancellation for Beamforming Microphone Arrays*, chapter 13 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 281–306. Springer-Verlag, May 2001.

[153] R. A. Kennedy, T. Abhayapala, D. B. Ward, and R. C. Williamson. Nearfield broadband frequency invariant beamforming. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 905–908, Atlanta GA, USA, May 1996.

[154] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust., Speech, Signal Processing*, 24(4):320–327, August 1976.

[155] D. Korompis, K. Yao, and F. Lorenzelli. Broadband Maximum Energy Array with User Imposed Spatial and Frequency Constraints. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 529–532, Adelaide, Australia, April 1994.

[156] H. Kuttruff. *Room Acoustics.* Spon Press, 4th edition, 2000.

[157] B. K. Lau, Y. H. Leung, K. L. Teo, and V. Sreeram. Minimax filters for microphone arrays. *IEEE Trans. Circuits Syst. II*, 46(12):1522–1525, December 1999.

[158] R. Le Bouquin-Jeannes, P. Scalart, G. Faucon, and C. Beaugeant. Combined noise and echo reduction in hands-free systems: A survey. *IEEE Trans. Speech and Audio Processing*, 9(8):808–820, November 2001.

[159] H. Lebret and S. Boyd. Antenna array pattern synthesis via convex optimization. *IEEE Trans. Signal Processing*, 45(3):526–532, March 1997.

[160] A.P. Liavas, P.A. Regalia, and J.-P. Delmas. On the robustness of the linear prediction method for blind channel identification with respect to effective channel undermodeling/overmodeling. *IEEE Trans. Signal Processing*, 48(5):1477–1481, May 2000.

[161] J. S. Lim, editor. *Speech Enhancement.* Prentice Hall, Englewood Cliffs, New Jersey, 1983.

[162] J. S. Lim and A. V. Oppenheim. All-pole modeling of degraded speech. *IEEE Trans. Acoust., Speech, Signal Processing*, 26(3):197–210, June 1978.

[163] J. S. Lim and A. V. Oppenheim. Enhancement and bandwidth compression of noisy speech. *Proc. IEEE*, 67, December 1979.

[164] Q.-G. Liu, B. Champagne, and P. Kabal. A microphone array processing technique for speech enhancement in a reverberant space. *Speech Communication*, 18:317–334, 1996.

[165] P. Lockwood and J. Boudy. Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars. *Speech Communication*, 11:215–228, 1992.

[166] R. Lopez-Valcarce and S. Dasgupta. Blind channel equalization with colored sources based on second-order statistics: a linear prediction approach. *IEEE Trans. Signal Processing*, 49(9):2050–2059, September 2001.

[167] W.-S. Lu and A. Antoniou. Design of digital filters and filter banks by optimization: A state of the art review. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 351–354, Tampere, Finland, September 2000.

[168] F. T. Luk. A parallel method for computing the generalized singular value decomposition. *Journal of Parallel and Distributed Computing*, 2:250–260, 1985.

[169] R. J. Mailloux. *Phased Array Antenna Handbook*. Artech House, Boston, 1994.

[170] J. Makhoul. On the Eigenvectors of Symmetric Toeplitz Matrices. *IEEE Trans. Acoust., Speech, Signal Processing*, 29(4):868–872, August 1981.

[171] D. Mansour and A. Gray. Uconstrained Frequency-Domain Adaptive Filter. *IEEE Trans. Acoust., Speech, Signal Processing*, 30(5):726–734, October 1982.

[172] R. Martin. Spectral Subtraction Based on Minimum Statistics. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 1182–1185, Edinburgh, Scotland, September 1994.

[173] R. Martin and P. Vary. Combined acoustic echo cancellation, dereverberation and noise reduction : a two microphone approach. *Annales des Télécommunications*, pages 429–438, 1994.

[174] R. Martin and P. Vary. Combined Acoustic Echo Control and Noise Reduction for Hands-Free Telephony - State of the Art and Perspectives. In *Proc. European Signal Processing Conf. (EUSIPCO)*, volume 2, pages 1107–1110, Trieste, Italy, September 1996.

[175] J. A. Maxwell and P. M. Zurek. Reducing Acoustic Feedback in Hearing Aids. *IEEE Trans. Speech and Audio Processing*, 3(4):304–313, July 1995.

[176] M. Mboup and M. Bonnet. On the Adequateness of IIR Adaptive Filtering for Acoustic Echo Cancellation. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI: Theories and Applications*, pages 111–114. Elsevier Science Publishers B.V., 1992.

[177] R. J. McAulay and M. L. Malpass. Speech Enhancement Using a Soft-Decision Noise Suppression Filter. *IEEE Trans. Acoust., Speech, Signal Processing*, 28(2):137–145, April 1980.

[178] R. J. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(4):744–754, August 1986.

[179] U. Mittal and N. Phamdo. Signal/Noise KLT Based Approach for Enhancing Speech Degraded by Colored Noise. *IEEE Trans. Speech and Audio Processing*, 8(2):159–167, March 2000.

[180] M. Miyoshi and Y. Kaneda. Inverse Filtering of Room Acoustics. *IEEE Trans. Acoust., Speech, Signal Processing*, 36(2):145–152, February 1988.

[181] M. Moonen. *Jacobi-type updating algorithms for signal processing, systems identification and control.* PhD thesis, ESAT, Katholieke Universiteit Leuven, Belgium, 1990.

[182] M. Moonen and I. Proudler. Generating "Fast QR" algorithms using signal flow graph techniques. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove CA, USA, November 1996.

[183] M. Moonen, P. Van Dooren, and J. Vandewalle. A systolic algorithm for QSVD updating. *Signal Processing*, 25:203–213, 1991.

[184] M. Moonen, P. Van Dooren, and J. Vandewalle. A Singular Value Decomposition Updating Algorithm for Subspace Tracking. *SIAM Journal on Matrix Analysis and Applications*, 13(4):1015–1038, October 1992.

[185] M. Moonen, P. Van Dooren, and J. Vandewalle. A systolic array for SVD updating. *SIAM Journal on Matrix Analysis and Applications*, 14(2):353–371, 1993.

[186] D. R. Morgan and S. G. Kratzer. On a Class of Computationally Efficient, Rapidly Converging, Generalized NLMS Algorithms. *IEEE Signal Processing Lett.*, 3(8):245–247, August 1996.

[187] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue. Subspace Methods for the Blind Identification of Multichannel FIR Filters. *IEEE Trans. Signal Processing*, 43(2):516–525, February 1995.

[188] S. T. Neely and J. B. Allen. Invertibility of a room impulse response. *Journal of the Acoustical Society of America*, 66:165–169, 1979.

[189] R. O. Neubauer. Existing Reverberation Time Formulae - A Comparison with Computer Simulated Reverberation Times. In *8th International Congress on Sound and Vibration*, pages 805–812, Hong Kong, China, July 2001.

[190] M. Nilsson, S. D. Soli, and A. Sullivan. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95(2):1085–1099, February 1994.

[191] S. Nordebo, I. Claesson, and S. Nordholm. Adaptive beamforming: Spatial filter designed blocking matrix. *IEEE Journal of Oceanic Engineering*, 19(4):583–590, October 1994.

[192] S. Nordebo, I. Claesson, and S. Nordholm. Weighted Chebyshev approximation for the design of broadband beamformers using quadratic programming. *IEEE Signal Processing Lett.*, 1(7):103–105, July 1994.

[193] S. Nordebo and S. Nordholm. Noise reduction using an adaptive microphone array in a car -a speech recognition evaluation. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 17–20, New Paltz NY, USA, October 1993.

[194] S. Nordholm, I. Claesson, and B. Bengtsson. Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars. *IEEE Trans. Veh. Technol.*, 42(4):514–518, November 1993.

[195] S. Nordholm, I. Claesson, and M. Dahl. Adaptive Microphone Array Employing Calibration Signals: An Analytical Evaluation. *IEEE Trans. Speech and Audio Processing*, 7(3):241–252, May 1999.

[196] S. Nordholm, I. Claesson, and N. Grbić. *Optimal and Adaptive Microphone Arrays for Speech Input in Automobiles*, chapter 14 in “Microphone Arrays: Signal Processing Techniques and Applications” (Brandstein, M. S. and Ward, D. B., Eds.), pages 307–330. Springer-Verlag, May 2001.

[197] D. O’ Shaughnessy. *Speech Communication, Human and Machine.* Addison Wesley Publishing Company, 1987.

[198] Acoustical Society of America. ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index, June 1997.

[199] E. Oja. A simplified neuron model as a principal component analyzer. *Journal Math. Biol.*, 15:267–273, 1982.

[200] M. Omologo and P. Svaizer. Use of the crosspower-spectrum phase in acoustic event location. *IEEE Trans. Speech and Audio Processing*, 5(3):288–29, May 1997.

[201] M. Omologo, P. Svaizer, and M. Matassoni. Environmental conditions and acoustic transduction in hands-free speech recognition. *Speech Communication*, 25(1-3):75–95, August 1998.

[202] A. Oppenheim, R. Schafer, and T. Stockham. Non-linear filtering of multiplied and convolved signals. In *IEEE Trans. Audio and Electroacoustics*, volume 16, pages 437–466, September 1968.

[203] K. Ozeki and T. Umeda. An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. *Electron. Commun. Japan*, 67-A:19–27, 1984.

[204] H. Özer and S. G. Tanyer. A geometric algorithm for voice activity detection in nonstationary gaussian noise. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Rhodes, Greece, September 1998.

[205] C. C. Paige. Computing the generalized singular value decomposition. *SIAM Journal on Scientific and Statistical Computing*, 7:1126–1146, 1986.

[206] C. C. Paige and M. A. Saunders. Towards a Generalized Singular Value Decomposition. *SIAM Journal on Numerical Analysis*, 18(3):398–405, 1981.

[207] L. C. Parra and C. V. Alvino. Geometric source separation: merging convolutive source separation with geometric beamforming. *IEEE Trans. Speech and Audio Processing*, 10(6):352–362, September 2002.

[208] S.-C. Pei and J.-J. Shyu. 2-D FIR eigenfilters: A least-squares approach. *IEEE Trans. Circuits Syst.*, 37(1):24–34, January 1990.

[209] S.-C. Pei and C.-C. Tseng. A new eigenfilter based on total least squares error criterion. *IEEE Trans. Circuits Syst. I*, 48(6):699–709, June 2001.

[210] P. M. Peterson. Simulating the response of multiple microphones to a single acoustic source in a reverberant room. *Journal of the Acoustical Society of America*, 80(5):1527–1529, 1986.

[211] V. M. A. Peutz. Articulation Loss of Consonants as a Criterion for Speech Transmission Index in a Room. *Journal Audio Engineering Society*, 19(11):915–919, 1971.

[212] R. Plomp. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research*, 29:146–154, 1986.

[213] J. Proakis and D. Manolakis. *Digital Signal Processing : Principles, Algorithms and Applications.* Prentice Hall, Englewood Cliffs, New Jersey, 1996.

[214] J. Proakis, C. M. Rader, F. Ling, M. Moonen, I. K. Proudler, and C. L. Nikias. *Algorithms for Statistical Signal Processing.* Prentice Hall, Englewood Cliffs, New Jersey, 2002.

[215] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals.* Prentice Hall, Englewood Cliffs, New Jersey, 1978.

[216] D. Rabinkin, R. Renomeron, J. Flanagan, and D. F. Macomber. Optimal truncation time for matched filter array processing. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3629–3632, Seattle WA, USA, May 1998.

[217] D. V. Rabinkin, R. J. Renomeron, A. Dahl, J. C. French, J. L. Flanagan, and M. H. Bianchi. A DSP implementation of source location using microphone arrays. *Proc. SPIE*, 2846:88–99, 1996.

[218] P. A. Regalia. An Adaptive Unit Norm Filter with Applications to Signal Analysis and Karhunen-Loève Transformations. *IEEE Trans. Circuits Syst.*, 37(5):646–649, May 1990.

[219] Wainhouse Research. New Research Reports from Wainhouse Research Says Market for Audio, Video, and Web Conferencing Services to Reach $9.8 billion by 2006, up from $2.8 billion in 2000. `http://www.wainhouse.com/prcms01v3.html`, January 2002.

[220] A. Rezayee and S. Gazor. An Adaptive KLT Approach for Speech Enhancement. *IEEE Trans. Speech and Audio Processing*, 9(2):87–95, February 2001.

[221] G. Rombouts and M. Moonen. Fast QRD-lattice-based unconstrained optimal filtering for acoustic noise reduction. *IEEE Trans. Speech and Audio Processing,* in press.

[222] G. Rombouts and M. Moonen. A sparse block exact affine projection algorithm. *IEEE Trans. Speech and Audio Processing*, 10(2):100–108, February 2002.

[223] G. Rombouts and M. Moonen. An integrated approach to acoustic echo and noise suppression. *Submitted to Signal Processing*, 2003.

[224] G. Rombouts and M. Moonen. QRD-based unconstrained optimal filtering for acoustic noise reduction. *Signal Processing*, 83(9):1889–1904, September 2003.

[225] J. G. Ryan and R. A. Goubran. Array optimization applied in the near field of a microphone array. *IEEE Trans. Speech and Audio Processing*, 8(2):173–176, March 2000.

[226] M. Savoji. A Robust Algorithm for Accurate Endpointing of Speech Signals. *Speech Communication*, pages 45–60, March 1989.

[227] L. L. Scharf. *Statistical Signal Processing : Detection, Estimation and Time Series Analysis.* Addison Wesley, 1st edition, July 1991.

[228] L. L. Scharf. The SVD and Reduced Rank Signal Processing. *Signal Processing*, 25:113–133, 1991.

[229] J.-L. Shen, J.-W. Hung, and L.-S. Lee. Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy environments. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, pages 1015–1018, Sydney, Australia, December 1998.

[230] J. Shynk. Adaptive IIR Filtering. *IEEE ASSP Magazine*, pages 4–21, April 1989.

[231] J. J. Shynk. Frequency-Domain and Multirate Adaptive Filtering. *IEEE Signal Processing Magazine*, 9(1):14–37, January 1992.

[232] K. U. Simmer, J. Bitzer, and C. Marro. *Post-Filtering Techniques*, chapter 3 in “Microphone Arrays: Signal Processing Techniques and Applications” (Brandstein, M. S. and Ward, D. B., Eds.), pages 39–60. Springer-Verlag, May 2001.

[233] M. G. Siqueira and A. Alwan. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Trans. Speech and Audio Processing*, 8(4):443–453, July 2000.

[234] W. Soede, A. J. Berkhout, and F. A. Bilsen. Development of a directional hearing instrument based on array technology. *Journal of the Acoustical Society of America*, 94(2):785–798, August 1993.

[235] J. Sohn, N. S. Kim, and W. Sung. A Statistical Model-Based Voice Activity Detection. *IEEE Signal Processing Lett.*, 6(1):1–3, January 1999.

[236] J. Sohn and W. Sung. A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 365–368, Seattle WA, USA, May 1998.

[237] P.C.W. Sommen. *Adaptive filtering methods. On methods to use a priori information in order to reduce complexity while maintaining convergence properties.* PhD thesis, T.U. Eindhoven, The Netherlands, 1992.

[238] M. M. Sondhi, D. R. Morgan, and J. L. Hall. Stereophonic acoustic echo cancellation - an overview of the fundamental problem. *IEEE Signal Processing Lett.*, 2:148–151, August 1995.

[239] A. Spriet, M. Moonen, and J. Wouters. The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation. *Signal Processing,* in press.

[240] A. Spriet, M. Moonen, and J. Wouters. Robustness Analysis of Multichannel Wiener Filtering and Generalized Sidelobe Cancellation for Multi-microphone Noise Reduction in Hearing Aid Applications. *IEEE Trans. on Speech and Audio Processing,* in press.

[241] A. Spriet, M. Moonen, and J. Wouters. Robustness analysis of GSVD based optimal filtering and Generalized Sidelobe Canceller for hearing aid applications. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 31–34, New Paltz NY, USA, October 2001.

[242] A. Spriet, M. Moonen, and J. Wouters. A Multi-Channel Subband Generalized Singular Value Decomposition Approach to Speech Enhancement. *European Transactions on Telecommunications, special issue on Acoustic Echo and Noise Control*, 13(2):149–158, Mar.-Apr. 2002.

[243] A. Spriet, M. Moonen, and J. Wouters. Feedback cancellation in hearing aids : an unbiased modeling approach. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 531–534, Toulouse, France, September 2002.

[244] A. Spriet and K. Vanbleu. Ruisonderdrukking met behulp van iteratieve Wiener-filters. Master's thesis, Katholieke Universiteit Leuven, Belgium, 1999.

[245] R. W. Stadler and W. M. Rabinowitz. On the potential of fixed arrays for hearing aids. *Journal of the Acoustical Society of America*, 94(3):1332–1342, September 1993.

[246] A. Stéphenne and B. Champagne. A new cepstral prefiltering technique for estimating time delay under reverberant conditions. *Signal Processing*, 59:253–266, 1997.

[247] A. Swinnen, S. Van Huffel, K. Van Loven, and R. Jacobs. Detection and multichannel SVD-based filtering of trigeminal somatosensory evoked potentials. *Medical & Biological Engineering & Computing*, 38(3):297–305, May 2000.

[248] C. Sydow. Broadband beamforming for a microphone array. *Journal of the Acoustical Society of America*, 96(2):845–849, August 1994.

[249] M. Tanaka and S. Makino. A Block Exact Fast Affine Projection Algorithm. *IEEE Trans. Speech and Audio Processing*, 7(1):79–86, January 1999.

[250] S. G. Tanyer and H. Özer. Voice activity detection in nonstationary noise. *IEEE Trans. Speech and Audio Processing*, 8(4):478–482, July 2000.

[251] M. Tohyama, R. H. Lyon, and T. Koike. Source Waveform Recovery in a Reverberant Space by Cepstrum Dereverberation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 157–160, Minneapolis, Minnesota, USA, April 1993.

[252] L. Tong, G. Xu, and T. Kailath. Fast blind equalization via antenna arrays. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 272–275, Minneapolis, USA, April 1993.

[253] P. P. Vaidyanathan and T. Q. Nguyen. Eigenfilters: a new approach to least-squares FIR filter design and applications including Nyquist filters. *IEEE Trans. Circuits Syst.*, 34(1):11–23, January 1987.

[254] D. Van Compernolle. Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 833–836, Albuquerque NM, USA, April 1990.

[255] D. Van Compernolle. Speech detection and speech-noise discrimination. Technical Report MI2-SPCH-93-4, ESAT, Katholieke Universiteit Leuven, Belgium, March 1993.

[256] D. Van Compernolle, T. Claes, F. Xie, and J. Smolders. Measuring the signal-to-noise ratio of noisy data. Technical Report MI2-SPCH-94-1, ESAT, Katholieke Universiteit Leuven, Belgium, February 1994.

[257] D. Van Compernolle, J. Smolders, and F. Xie. Noise suppression for hands free mobile telepohone communication. In *Proc. INTERNOISE*, pages 1657–1660, Leuven, Belgium, August 1993.

[258] D. Van Compernolle and S. Van Gerven. Beamforming with Microphone Arrays. In V. Cappellini and A. Figueiras-Vidal, editors, *COST 229 : Applications of Digital Signal Processing to Telecommunications*, pages 107–131. 1995.

[259] S. Van Gerven, D. Van Compernolle, P. Wauters, W. Verstraeten, K. Eneman, and K. Delaet. Multiple beam broadband beamforming: Filter design and real-time implementation. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 173–176, New Paltz NY, USA, October 1995.

[260] S. Van Gerven and F. Xie. A Comparative Study of Speech Detection Methods. In *Proc. EUROSPEECH*, volume 3, pages 1095–1098, Rhodos, Greece, September 1997.

[261] S. Van Huffel. Enhanced resolution based on minimum variance estimation and exponential data modelling. *Signal Processing*, 33(3):333–355, September 1993.

[262] C. Van Loan. Computing the CS and the Generalized Singular Value Decomposition. *Numerische Mathematik*, (46):479–491, 1985.

[263] C. F. Van Loan. Generalizing the Singular Value Decomposition. *SIAM Journal on Numerical Analysis*, 13(1):76–83, March 1976.

[264] B. D. Van Veen and K. M. Buckley. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Magazine*, 5(2):4–24, April 1988.

[265] P. Vandaele and M. Moonen. Two deterministic blind channel estimation algorithms based on oblique projections. *Signal Processing*, 80:481–495, 2000.

[266] J. Vanden Berghe and J. Wouters. An adaptive noise canceller for hearing aids using two nearby microphones. *Journal of the Acoustical Society of America*, 103(6):3621–3626, June 1998.

[267] 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*, 12(3):247–251, 1993.

[268] N. Virag. Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System. *IEEE Trans. Speech and Audio Processing*, 7(2):126–137, March 1999.

[269] Belgisch Instituut voor de Verkeersveiligheid. Voorstelling van de campagne 'Straf, wat je allemaal kan met een GSM'. `http://www.bivv.be/nl/pdf/0202cGsmNL.pdf`, February 2002.

[270] E. A. Wan and R. van der Merwe. *Kalman Filtering and Neural Networks*, chapter The Unscented Kalman Filter. 2001.

[271] H. Wang and P. Chu. Voice source localization for automatic camera pointing system in videoconferencing. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 187–190, Munich, Germany, April 1997.

[272] H. Wang and F. Itakura. An approach of dereverberation using multimicrophone sub-band envelope estimation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 953–956, Toronto, Canada, 1991.

[273] D. B. Ward and G. W. Elko. Mixed nearfield/farfield beamforming: A new technique for speech acquisition in a reverberant environment. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 213–216, New Paltz NY, USA, 1997.

[274] D. B. Ward, R. A. Kennedy, and R. C. Williamson. Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns. *Journal of the Acoustical Society of America*, 97(2):91–95, February 1995.

[275] D. B. Ward, R. A. Kennedy, and R. C. Williamson. *Constant Directivity Beamforming*, chapter 1 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 3–17. Springer-Verlag, May 2001.

[276] M. Wax and T. Kailath. Optimum localization of multiple sources by passive arrays. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(5):1210–1218, October 1983.

[277] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Ziedler, E.J. Dong, and R.C. Goodlin. Adaptive noise cancellation: principles and applications. *Proc. IEEE*, 63(12):1692–1716, December 1975.

[278] GSM World. Subscriber statistics and forecasts. October 2002. `http://www.gsmworld.com/news/statistics/substats.shtml` and `http://www.gsmworld.com/news/statistics/subforecasts.shtml`.

[279] F. Xie and D. Van Compernolle. A Family of MLP based Nonlinear Spectral Estimators for Noise Reduction. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 53–56, Adelaide, Australia, April 1994.

[280] F. Xie and D. Van Compernolle. Speech Enhancement by Spectral Magnitude Estimation - A Unifying Approach. *Speech Communication*, 19(2):89–104, August 1996.

[281] B. Yang. Projection approximation subspace tracking. *IEEE Trans. Signal Processing*, 43(1):95–107, January 1995.

[282] Y. Zhou, H. Leung, and P. Yip. Blind identification of multichannel FIR systems based on linear prediction. *IEEE Trans. Signal Processing*, 48(9):2674–2678, September 2000.

Appendices

A Linear algebra definitions

In this appendix some linear algebra definitions and properties are briefly reviewed. A good reference text on linear algebra and matrix decompositions that focuses on numerical aspects and implementation issues is [110].

A.1 Structured real matrices

Consider the $N \times N$-dimensional real matrix $\mathbf{A}$, the N-dimensional real vector $\mathbf{v}$ and the $NL \times NL$-dimensional real block-matrix $\mathbf{B}$,

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix} , \tag{A.1}$$

$$\mathbf{v} = [v_i] = \begin{bmatrix} v_1 & v_2 & \dots & v_N \end{bmatrix}^T , \tag{A.2}$$

$$\mathbf{B} = [\mathbf{B}_{ij}] = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \dots & \mathbf{B}_{1N} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \dots & \mathbf{B}_{2N} \\ \vdots & \vdots & & \vdots \\ \mathbf{B}_{N1} & \mathbf{B}_{N2} & \dots & \mathbf{B}_{NN} \end{bmatrix} , \tag{A.3}$$

with $\mathbf{B}_{ij}$ $L \times L$-dimensional matrices and ^{T} denoting the transpose operation.

Definition A.1 The matrix $\mathbf{A}$ is *symmetric* iff (if and only if) $\mathbf{A}$ is symmetric along its main diagonal ($a_{ij} = a_{ji}$),

$$\mathbf{A} \text{ is symmetric} \Leftrightarrow \mathbf{A} = \mathbf{A}^T . \tag{A.4}$$

Definition A.2 The matrix $\mathbf{A}$ is *orthogonal* iff

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_N , \tag{A.5}$$

with $\mathbf{I}_N$ the $N \times N$ identity matrix. Hence, the inverse of an orthogonal matrix is equal to its transpose, $\mathbf{A}^{-1} = \mathbf{A}^T$.

Definition A.3 The matrix $\mathbf{A}$ is *positive definite* iff all its eigenvalues are strictly positive, i.e.

$$\mathbf{A} \text{ is positive definite} \Leftrightarrow \mathbf{v}^T\mathbf{A}\mathbf{v} > 0, \quad \forall \mathbf{v} \in \mathbb{R}^N . \tag{A.6}$$

The matrix $\mathbf{A}$ is *positive semi-definite* iff all its eigenvalues are positive, i.e.

$$\mathbf{A} \text{ is positive semi-definite} \Leftrightarrow \mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0, \quad \forall \mathbf{v} \in \mathbb{R}^N \ . \tag{A.7}$$

Definition A.4 The $N \times N$-dimensional *reversal matrix* $\mathbf{J}_N$ is a matrix with ones along the secondary diagonal and zeros everywhere else,

$$\mathbf{J}_N = \begin{bmatrix} 0 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 \end{bmatrix} \ . \tag{A.8}$$

$\mathbf{J}_N\mathbf{A}$ reverses the rows of $\mathbf{A}$, $\mathbf{A}\mathbf{J}_N$ reverses the columns of $\mathbf{A}$ and $\mathbf{J}_N\mathbf{A}\mathbf{J}_N$ reverses both the rows and the columns of $\mathbf{A}$. $\mathbf{J}_N$ is an orthogonal and symmetric matrix, hence $\mathbf{J}_N = \mathbf{J}_N^T$, $\mathbf{J}_N\mathbf{J}_N = \mathbf{I}_N$ and $\mathbf{J}_N^{-1} = \mathbf{J}_N$.

Definition A.5 The matrix $\mathbf{A}$ is *centro-symmetric* iff $\mathbf{A}$ is symmetric along its secondary diagonal ($a_{ij} = a_{N-j+1,N-i+1}$),

$$\mathbf{A} \text{ is centro-symmetric} \Leftrightarrow \mathbf{J}_N\mathbf{A} = \mathbf{A}^T\mathbf{J}_N \ . \tag{A.9}$$

Definition A.6 The matrix $\mathbf{A}$ is *double-symmetric* (or symmetric centro-symmetric) iff $\mathbf{A}$ is symmetric and centro-symmetric ($a_{ij} = a_{ji} = a_{N-i+1,N-j+1}$),

$$\mathbf{A} \text{ is double-symmetric} \quad \Leftrightarrow \quad \left\{ \begin{array}{c} \mathbf{A} = \mathbf{A}^T \\ \mathbf{J}_N\mathbf{A} = \mathbf{A}^T\mathbf{J}_N \end{array} \right. \quad \Rightarrow \quad \mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A} \quad . \tag{A.10}$$

Remark A.7 From the property $\mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A}$, nothing can be concluded about the symmetry nor the centro-symmetry of the matrix $\mathbf{A}$. E.g. consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix} \ .$$

If $\mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A}$, this simply means that the ith row/column of $\mathbf{A}$ is equal to the $(N - i + 1)$th row/column of $\mathbf{A}$ in reverse. For N odd, this implies that the middle row/column of $\mathbf{A}$ is symmetric. △

Definition A.8 The vector $\mathbf{v}$ is called *symmetric* iff $\mathbf{J}_N\mathbf{v} = \mathbf{v}$ and *skew-symmetric* iff $\mathbf{J}_N\mathbf{v} = -\mathbf{v}$.

Definition A.9 The matrix $\mathbf{A}$ is a *Toeplitz matrix* iff the elements along the diagonals of $\mathbf{A}$ are equal,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1N} \\ a_{21} & a_{11} & a_{12} & \dots & a_{1,N-1} \\ a_{31} & a_{21} & a_{11} & \dots & a_{2,N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{N1} & a_{N-1,1} & a_{N-2,1} & \dots & a_{11} \end{bmatrix}. \tag{A.11}$$

As can be readily verified, all Toeplitz matrices are centro-symmetric.

Definition A.10 The matrix $\mathbf{A}$ is *symmetric Toeplitz* iff it is both symmetric and Toeplitz,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1N} \\ a_{12} & a_{11} & a_{12} & \dots & a_{1,N-1} \\ a_{13} & a_{12} & a_{11} & \dots & a_{2,N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{1N} & a_{1,N-1} & a_{1,N-2} & \dots & a_{11} \end{bmatrix}. \tag{A.12}$$

As can be readily verified, all symmetric Toeplitz matrices are double-symmetric.

Definition A.11 The $NL \times NL$-dimensional *block-reversal matrix* $\mathbf{S}_{NL}$ is a matrix with identity matrices $\mathbf{I}_L$ along its secondary diagonal and zeros everywhere else.

$$\mathbf{S}_{NL} = \begin{bmatrix} 0 & 0 & \dots & \mathbf{I}_L \\ \vdots & \vdots & & \vdots \\ 0 & \mathbf{I}_L & \dots & 0 \\ \mathbf{I}_L & 0 & \dots & 0 \end{bmatrix} \tag{A.13}$$

$\mathbf{S}_{NL}$ is an orthogonal and a symmetric matrix, hence $\mathbf{S}_{NL} = \mathbf{S}_{NL}^T$, $\mathbf{S}_{NL}^{-1} = \mathbf{S}_{NL}$ and $\mathbf{S}_{NL}\mathbf{S}_{NL} = \mathbf{I}_{NL}$.

Definition A.12 The block-matrix $\mathbf{B}$ is *block-symmetric* iff the $L \times L$ matrices $\mathbf{B}_{ij}$ are symmetric along the main diagonal of the block-matrix $\mathbf{B}$ ($\mathbf{B}_{ij} = \mathbf{B}_{ji}$),

$$\mathbf{B} = [\mathbf{B}_{ij}] = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} & \dots & \mathbf{B}_{1N} \\ \mathbf{B}_{12} & \mathbf{B}_{22} & \mathbf{B}_{23} & \dots & \mathbf{B}_{2N} \\ \mathbf{B}_{13} & \mathbf{B}_{23} & \mathbf{B}_{33} & \dots & \mathbf{B}_{3N} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{B}_{1N} & \mathbf{B}_{2N} & \mathbf{B}_{3N} & \dots & \mathbf{B}_{NN} \end{bmatrix}. \tag{A.14}$$

In general, block-symmetric matrices are not symmetric. If all block-matrices $\mathbf{B}_{ij}$ are symmetric ($\mathbf{B}_{ij} = \mathbf{B}_{ij}^T$), the block-symmetric matrix $\mathbf{B}$ is symmetric.

Definition A.13 The block-matrix $\mathbf{B}$ is called *block-Toeplitz* iff the $L \times L$ matrices $\mathbf{B}_{ij}$ along the diagonals of $\mathbf{B}$ are equal,

$$\mathbf{B} = [\mathbf{B}_{ij}] = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} & \dots & \mathbf{B}_{1N} \\ \mathbf{B}_{21} & \mathbf{B}_{11} & \mathbf{B}_{12} & \dots & \mathbf{B}_{1,N-1} \\ \mathbf{B}_{31} & \mathbf{B}_{21} & \mathbf{B}_{11} & \dots & \mathbf{B}_{1,N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{B}_{N1} & \mathbf{B}_{N-1,1} & \mathbf{B}_{N-2,1} & \dots & \mathbf{B}_{11} \end{bmatrix} . \tag{A.15}$$

In general, block-Toeplitz matrices are not Toeplitz.

A.2 Matrix decompositions

Definition A.14 Given the $N \times N$-dimensional real matrix $\mathbf{A}$, all (complex) N-dimensional vectors $\mathbf{v}_i \neq \mathbf{0}$ and (complex) scalars λ_i that satisfy

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{A.16}$$

are called *eigenvectors* and *eigenvalues* of the matrix $\mathbf{A}$.

Definition A.15 If the $N \times N$-dimensional real matrix $\mathbf{A}$ has N linearly independent eigenvectors, then the *eigenvalue decomposition* (EVD) of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{-1} , \tag{A.17}$$

with $\mathbf{V}$ an $N \times N$-dimensional matrix containing the eigenvectors $\mathbf{v}_i$ as columns and $\boldsymbol{\Delta}$ a diagonal matrix containing the eigenvalues λ_i. Normally, the eigenvectors are normalised such that $||\mathbf{v}_i||_2 = 1$, $i = 1 \dots N$. It can be shown that the EVD of a symmetric matrix $\mathbf{A}$ is equal to

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^T , \tag{A.18}$$

with $\mathbf{V}$ an orthogonal matrix and all eigenvalues real scalars.

Definition A.16 Given the $P \times N$-dimensional real matrices $\mathbf{A}$ and $\mathbf{B}$, all (complex) N-dimensional vectors $\mathbf{x}_i \neq \mathbf{0}$ and (complex) scalars λ_i that satisfy

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{B}\mathbf{x}_i \tag{A.19}$$

are called *generalised eigenvectors* and *generalised eigenvalues* of the matrices $\mathbf{A}$ and $\mathbf{B}$.

Definition A.17 If the $P \times N$-dimensional real matrices $\mathbf{A}$ and $\mathbf{B}$ have N linearly independent generalised eigenvectors, then the *generalised eigenvalue decomposition* (GEVD) of $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\mathbf{A}\mathbf{X} = \mathbf{B}\mathbf{X}\mathbf{\Lambda}\ , \tag{A.20}$$

with $\mathbf{X}$ an $N \times N$-dimensional (invertible) matrix containing the generalised eigenvectors $\mathbf{x}_i$ as columns and $\mathbf{\Lambda}$ a diagonal matrix containing the generalised eigenvalues λ_i. Alternatively, the GEVD of $\mathbf{A}$ and $\mathbf{B}$ can be written as

$$\left\{\begin{array}{lcl} \mathbf{A} & = & \mathbf{Q}\mathbf{\Lambda}_A\mathbf{X}^{-1} \\ \mathbf{B} & = & \mathbf{Q}\mathbf{\Lambda}_B\mathbf{X}^{-1}\ , \end{array}\right. \tag{A.21}$$

with $\mathbf{Q}$ a $P \times N$ matrix and $\mathbf{\Lambda} = \mathbf{\Lambda}_A\mathbf{\Lambda}_B^{-1}$. The GEVD of two symmetric $N \times N$-dimensional matrices $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\left\{\begin{array}{lcl} \mathbf{A} & = & \mathbf{Q}\mathbf{\Lambda}_A\mathbf{Q}^T \\ \mathbf{B} & = & \mathbf{Q}\mathbf{\Lambda}_B\mathbf{Q}^T\ , \end{array}\right. \tag{A.22}$$

with $\mathbf{Q}$ an invertible, but not necessarily orthogonal, matrix.

Definition A.18 The *singular value decomposition* (SVD) of the $P \times N$-dimensional real matrix $\mathbf{A}$ (with $P \geq N$) is given by

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\ , \tag{A.23}$$

with $\mathbf{U}$ a $P \times N$-dimensional orthogonal matrix containing the (left) singular vectors $\mathbf{u}_i$, $\mathbf{V}$ an $N \times N$-dimensional orthogonal matrix containing the (right) singular vectors $\mathbf{v}_i$, and $\mathbf{\Sigma}$ an $N \times N$-dimensional diagonal matrix containing the singular values σ_i, with $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_N \geq 0$. The matrix $\mathbf{A}$ can be written as the dyadic decomposition

$$\mathbf{A} = \sum_{i=1}^{N} \sigma_i \mathbf{u}_i \mathbf{v}_i^T\ , \tag{A.24}$$

such that

$$\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i, \quad \mathbf{A}^T\mathbf{u}_i = \sigma_i\mathbf{v}_i\ . \tag{A.25}$$

If the matrix $\mathbf{A}$ has rank $R < N$, then $N - R$ singular values are equal to zero, such that the SVD of $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T\ , \tag{A.26}$$

with $\mathbf{U}_1$ a $P \times R$-dimensional and $\mathbf{V}_1$ an $R \times R$-dimensional orthogonal matrix, and $\mathbf{\Sigma}_1 = \text{diag}\{\sigma_1, \ldots, \sigma_R\}$ an $R \times R$-dimensional diagonal matrix.

Definition A.19 The (Moore-Penrose) *pseudo-inverse* of the $P \times N$-dimensional real matrix $\mathbf{A}$ with rank R is defined using (A.26) as

$$\mathbf{A}^{\dagger} = \mathbf{V}_1 \mathbf{\Sigma}_1^{-1} \mathbf{U}_1^T \ . \tag{A.27}$$

Definition A.20 The *generalised singular value decomposition* (GSVD) of the $P \times N$-dimensional real matrix $\mathbf{A}$ and the $Q \times N$-dimensional real matrix $\mathbf{B}$ (with $P \geq N$ and $Q \geq N$) is given by

$$\left\{ \begin{array}{lcl} \mathbf{A} & = & \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{Q}^T \\ \mathbf{B} & = & \mathbf{U}_B \mathbf{\Sigma}_B \mathbf{Q}^T \ , \end{array} \right. \tag{A.28}$$

with $\mathbf{U}_A$ a $P \times N$-dimensional orthogonal matrix, $\mathbf{U}_B$ a $Q \times N$-dimensional orthogonal matrix, $\mathbf{Q}$ an $N \times N$-dimensional invertible, but not necessarily orthogonal, matrix containing the (right) generalised singular vectors $\mathbf{q}_i$ and $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$ $N \times N$-dimensional diagonal matrices containing σ_{Ai} and σ_{Bi}, with $\sigma_{A1} \geq \sigma_{A2} \ldots \geq \sigma_{AN} \geq 0$ and $0 \leq \sigma_{B1} \leq \sigma_{B2} \ldots \leq \sigma_{BN}$. The generalised singular values are equal to σ_{Ai}/σ_{Bi}. The matrices $\mathbf{A}$ and $\mathbf{B}$ can be written as the dyadic decomposition

$$\left\{ \begin{array}{lcl} \mathbf{A} & = & \displaystyle\sum_{i=1}^{N} \sigma_{Ai} \mathbf{u}_{Ai} \mathbf{q}_i^T \\ \mathbf{B} & = & \displaystyle\sum_{i=1}^{N} \sigma_{Bi} \mathbf{u}_{Bi} \mathbf{q}_i^T \ . \end{array} \right. \tag{A.29}$$

If we define the matrix $\mathbf{T} = \mathbf{Q}^{-T}$, containing the vectors $\mathbf{t}_i$, such that $\mathbf{q_i}^T \mathbf{t}_j = \delta_{ij}$, with δ_{ij} the Kronecker-delta, then

$$\left\{ \begin{array}{lcll} \mathbf{A}\mathbf{t}_i & = & \sigma_{Ai} \mathbf{u}_{Ai}, & \mathbf{A}^T \mathbf{u}_{Ai} = \sigma_{Ai} \mathbf{q}_i \\ \mathbf{B}\mathbf{t}_i & = & \sigma_{Bi} \mathbf{u}_{Bi}, & \mathbf{B}^T \mathbf{u}_{Bi} = \sigma_{Bi} \mathbf{q}_i \ . \end{array} \right. \tag{A.30}$$

Definition A.21 The QR-decomposition of the $P \times N$ real matrix $\mathbf{A}$ (with $P \geq N$) is defined

$$\mathbf{A} = \mathbf{Q}_A \mathbf{R}_A \ , \tag{A.31}$$

with $\mathbf{Q}_A$ a $P \times N$ orthogonal matrix and $\mathbf{R}_A$ an $N \times N$ upper-triangular matrix.

A.3 Matrix and vector norms

Definition A.22 The *2-norm* of an N-dimensional vector $\mathbf{v}$ is equal to

$$||\mathbf{v}||_2 = \sqrt{\sum_{j=1}^{N} v_j^2} \ . \tag{A.32}$$

Generally, we will write $||\mathbf{v}||$ instead of $||\mathbf{v}||_2$.

Definition A.23 The *2-norm* of an $M \times N$-dimensional matrix $\mathbf{A}$ is defined as

$$||\mathbf{A}||_2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} \; . \tag{A.33}$$

It can be shown that the 2-norm of $\mathbf{A}$ is equal to its largest singular value, i.e.

$$||\mathbf{A}||_2 = \sigma_1(\mathbf{A}) \; . \tag{A.34}$$

Definition A.24 The *Frobenius-norm* of an $M \times N$-dimensional matrix $\mathbf{A}$ is defined as

$$||\mathbf{A}||_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij}^2} \; . \tag{A.35}$$

It can be shown that the Frobenius-norm of $\mathbf{A}$ is related to its singular values as

$$||\mathbf{A}||_F = \sqrt{\sum_{j=1}^{N} \sigma_j^2(\mathbf{A})} \; . \tag{A.36}$$

A.4 Matrix inversion lemma

Lemma A.25 *The inverse of the $N \times N$-dimensional matrix $\mathbf{A} + \mathbf{BDC}$, with $\mathbf{A}$ a full-rank $N \times N$-dimensional matrix, $\mathbf{B}$ an $N \times R$-dimensional matrix, $\mathbf{C}$ an $R \times N$-dimensional matrix and $\mathbf{D}$ a full-rank $R \times R$-dimensional matrix, is equal to*

$$(\mathbf{A} + \mathbf{BDC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{CA}^{-1}\mathbf{B} + \mathbf{D}^{-1})^{-1}\mathbf{CA}^{-1} \; . \tag{A.37}$$

Corollary A.26 *Using (A.37), the inverse of the rank-1 update $\mathbf{A} + \mathbf{uv}^H$, with $\mathbf{u}$ and $\mathbf{v}$ complex N-dimensional vectors, is equal to*

$$(\mathbf{A} + \mathbf{uv}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{uv}^H\mathbf{A}^{-1}}{1 + \mathbf{v}^H\mathbf{A}^{-1}\mathbf{u}} \; . \tag{A.38}$$

Corollary A.27 *It can be proved that*

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^H & c \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \frac{\mathbf{bb}^H}{c}\right)^{-1} & -\frac{\mathbf{A}^{-1}\mathbf{b}}{c - \mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} \\ -\frac{\mathbf{b}^H\mathbf{A}^{-1}}{c - \mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} & \frac{1}{c - \mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} \end{bmatrix} , \tag{A.39}$$

with $\mathbf{A}$ an $N \times N$-dimensional matrix, $\mathbf{b}$ an N-dimensional vector and c a scalar.

Proof: Using (A.38),

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & c \end{bmatrix}^{-1} = \left(\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & c \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{0} & 1 \end{bmatrix} \right)^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\frac{\mathbf{A}^{-1}\mathbf{b}}{c} \\ \mathbf{0} & \frac{1}{c} \end{bmatrix}, \tag{A.40}$$

such that

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^H & c \end{bmatrix}^{-1} = \left(\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & c \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{b}^H & 0 \end{bmatrix} \right)^{-1} \tag{A.41}$$

$$= \begin{bmatrix} \mathbf{A}^{-1} & -\frac{\mathbf{A}^{-1}\mathbf{b}}{c} \\ \mathbf{0} & \frac{1}{c} \end{bmatrix} - \frac{\begin{bmatrix} -\mathbf{A}^{-1}\mathbf{b}\mathbf{b}^H\mathbf{A}^{-1} & \frac{\mathbf{A}^{-1}\mathbf{b}\,(\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b})}{c} \\ \mathbf{b}^H\mathbf{A}^{-1} & -\frac{\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}}{c} \end{bmatrix}}{c - \mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}}$$

$$= \begin{bmatrix} \left(\mathbf{A} - \frac{\mathbf{b}\mathbf{b}^H}{c}\right)^{-1} & -\frac{\mathbf{A}^{-1}\mathbf{b}}{c-\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} \\ -\frac{\mathbf{b}^H\mathbf{A}^{-1}}{c-\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} & \frac{1}{c-\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} \end{bmatrix}. \tag{A.42}$$

□

A.5 Symmetry properties of eigenvectors

Theorem A.28 *If the $N \times N$-dimensional matrix $\mathbf{A}$ satisfies $\mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A}$ and has N distinct eigenvalues, then $\mathbf{A}$ has $\lceil N/2 \rceil$ symmetric eigenvectors and $\lfloor N/2 \rfloor$ skew-symmetric eigenvectors which span the eigenspace of $\mathbf{A}$, where $\lceil x \rceil$ represents the smallest integer greater than or equal to x and $\lfloor x \rfloor$ represents the largest integer smaller than or equal to x.*

Proof [26]: The matrix $\mathbf{A}$ has N orthonormal eigenvectors $\mathbf{v}_i$ which are unique apart from their sign. Therefore, for any eigenvector $\mathbf{v}_i$, $i = 1 \dots N$,

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i \Rightarrow \mathbf{J}_N\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{J}_N\mathbf{v}_i \Rightarrow \mathbf{A}\mathbf{J}_N\mathbf{v}_i = \lambda_i\mathbf{J}_N\mathbf{v}_i \tag{A.43}$$

holds. Hence $\mathbf{J}_N\mathbf{v}_i$ is an eigenvector of $\mathbf{A}$ corresponding to λ_i. Since the N eigenvalues of $\mathbf{A}$ are distinct and $\mathbf{J}_N\mathbf{v}_i$ has the same norm as $\mathbf{v}_i$, then $\mathbf{J}_N\mathbf{v}_i = \pm\mathbf{v}_i$, such that $\mathbf{v}_i$ is either symmetric or skew-symmetric. The only possible way for the eigenspace to consist of N mutually orthogonal, symmetric (skew-symmetric) nonzero eigenvectors, is that is consists of $\lceil N/2 \rceil$ symmetric eigenvectors and $\lfloor N/2 \rfloor$ skew-symmetric eigenvectors. □

In [26] it has been proved that when the multiplicity of some eigenvalues is larger than 1, the matrix $\mathbf{A}$ can have eigenvectors which are a linear combination of symmetric and skew-symmetric vectors, and hence, are neither symmetric nor skew-symmetric.

Corollary A.29 *If* $\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{-1}$ *is the eigenvalue decomposition of* $\mathbf{A}$*, then*

$$\mathbf{J}_N\mathbf{V} = \mathbf{V}\,diag\{\pm 1\}\ . \tag{A.44}$$

Corollary A.30 *Since* $\mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A}$ *is true for all double-symmetric and symmetric Toeplitz matrices* $\mathbf{A}$*, all eigenvectors of double-symmetric and symmetric Toeplitz matrices generally are symmetric or skew-symmetric.*

Lemma A.31 *If* $\mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A}$*, then* $\mathbf{J}_N\mathbf{A}^T\mathbf{J}_N = \mathbf{A}^T$ *and* $\mathbf{J}_N\mathbf{A}^{-1}\mathbf{J}_N = \mathbf{A}^{-1}$ *(if* $\mathbf{A}$ *is invertible).*

Proof :

$$\begin{aligned} \mathbf{A}^T &= (\mathbf{J}_N\mathbf{A}\mathbf{J}_N)^T = \mathbf{J}_N^T\mathbf{A}^T\mathbf{J}_N^T = \mathbf{J}_N\mathbf{A}^T\mathbf{J}_N & \text{(A.45)}\\ \mathbf{A}^{-1} &= (\mathbf{J}_N\mathbf{A}\mathbf{J}_N)^{-1} = \mathbf{J}_N^{-1}\mathbf{A}^{-1}\mathbf{J}_N^{-1} = \mathbf{J}_N\mathbf{A}^{-1}\mathbf{J}_N\ . & \text{(A.46)} \end{aligned}$$

□

Lemma A.32 *The inverse of a nonsingular double-symmetric matrix is also double-symmetric [170]. The inverse of a nonsingular symmetric Toeplitz matrix is double symmetric, but not necessarily Toeplitz.*

Lemma A.33 *Consider* $\mathbf{A} \in \mathbb{R}^{N\times N}$ *and* $\mathbf{B} \in \mathbb{R}^{N\times N}$*. If* $\mathbf{J}_N\mathbf{A}\mathbf{J}_N = \mathbf{A}$ *and* $\mathbf{J}_N\mathbf{B}\mathbf{J}_N = \mathbf{B}$*, then* $\mathbf{J}_N(\mathbf{A}+\mathbf{B})\mathbf{J}_N = \mathbf{A}+\mathbf{B}$ *and* $\mathbf{J}_N(\mathbf{A}\mathbf{B})\mathbf{J}_N = \mathbf{A}\mathbf{B}$.

Lemma A.34 *The sum of two double-symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is also double-symmetric. The sum of two symmetric Toeplitz matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is also symmetric Toeplitz. The product of two double-symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is double-symmetric, only if* $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$*. The product of two symmetric Toeplitz matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is double-symmetric, only if* $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$*, and is not necessarily Toeplitz.*

Theorem A.35 *If the block-matrix* $\mathbf{B} \in \mathbb{R}^{NL\times NL}$ *satisfies* $\mathbf{S}_{NL}\mathbf{B}\mathbf{S}_{NL} = \mathbf{B}$ *and has* NL *distinct eigenvalues, then all* NL *eigenvectors* $\mathbf{v}_i$ *of* $\mathbf{B}$ *satisfy the property* $\mathbf{S}_{NL}\mathbf{v}_i = \pm\mathbf{v}_i$*.*

Proof : $\mathbf{B}$ has NL orthonormal eigenvectors $\mathbf{v}_i$ which are unique apart from their sign. For any eigenvector $\mathbf{v}_i$, $i = 1\ldots NL$,

$$\mathbf{B}\mathbf{v}_i = \lambda_i\mathbf{v}_i \Rightarrow \mathbf{S}_{NL}\mathbf{B}\mathbf{v}_i = \lambda_i\mathbf{S}_{NL}\mathbf{v}_i \Rightarrow \mathbf{B}\mathbf{S}_{NL}\mathbf{v}_i = \lambda_i\mathbf{S}_{NL}\mathbf{v}_i \tag{A.47}$$

holds. Hence $\mathbf{S}_{NL}\mathbf{v}_i$ is an eigenvector of $\mathbf{B}$ corresponding to λ_i. The NL eigenvalues of $\mathbf{B}$ are distinct and $\mathbf{S}_{NL}\mathbf{v}_i$ has the same norm as $\mathbf{v}_i$, such that $\mathbf{S}_{NL}\mathbf{v}_i = \pm\mathbf{v}_i$. □

Corollary A.36 *If* $\mathbf{B} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{-1}$ *is the eigenvalue decomposition of* $\mathbf{B}$*, then*

$$\mathbf{S}_{NL}\mathbf{V} = \mathbf{V}\,diag\{\pm 1\}\ . \tag{A.48}$$

Corollary A.37 *Since* $\mathbf{S}_{NL}\mathbf{B}\mathbf{S}_{NL} = \mathbf{B}$ *is true for all matrices* $\mathbf{B}$ *which are both block-symmetric and block-Toeplitz, all eigenvectors* $\mathbf{v}_i$ *of these matrices generally satisfy* $\mathbf{S}_{NL}\mathbf{v}_i = \pm\mathbf{v}_i$*.*

Lemma A.38 *If* $\mathbf{S}_{NL}\mathbf{B}\mathbf{S}_{NL} = \mathbf{B}$*, then* $\mathbf{S}_{NL}\mathbf{B}^T\mathbf{S}_{NL} = \mathbf{B}^T$ *and* $\mathbf{S}_{NL}\mathbf{B}^{-1}\mathbf{S}_{NL} = \mathbf{B}^{-1}$ *(if* $\mathbf{B}$ *is invertible).*

Proof :

$$\begin{aligned} \mathbf{B}^T &= (\mathbf{S}_{NL}\mathbf{B}\mathbf{S}_{NL})^T = \mathbf{S}_{NL}^T\mathbf{B}^T\mathbf{S}_{NL}^T = \mathbf{S}_{NL}\mathbf{B}^T\mathbf{S}_{NL} && \text{(A.49)} \\ \mathbf{B}^{-1} &= (\mathbf{S}_{NL}\mathbf{B}\mathbf{S}_{NL})^{-1} = \mathbf{S}_{NL}^{-1}\mathbf{B}^{-1}\mathbf{S}_{NL}^{-1} = \mathbf{S}_{NL}\mathbf{B}^{-1}\mathbf{S}_{NL}\ . && \text{(A.50)} \end{aligned}$$

□

Lemma A.39 *Consider* $\mathbf{B} \in \mathbb{R}^{NL\times NL}$ *and* $\mathbf{C} \in \mathbb{R}^{NL\times NL}$*. If* $\mathbf{S}_{NL}\mathbf{B}\mathbf{S}_{NL} = \mathbf{B}$ *and* $\mathbf{S}_{NL}\mathbf{C}\mathbf{S}_{NL} = \mathbf{C}$*, then* $\mathbf{S}_{NL}(\mathbf{B}+\mathbf{C})\mathbf{S}_{NL} = \mathbf{B}+\mathbf{C}$ *and* $\mathbf{S}_{NL}(\mathbf{B}\mathbf{C})\mathbf{S}_{NL} = \mathbf{B}\mathbf{C}$*.*

Lemma A.40 *The sum of two block-symmetric matrices* $\mathbf{B}$ *and* $\mathbf{C}$ *is also block-symmetric. The sum of two block-Toeplitz matrices* $\mathbf{B}$ *and* $\mathbf{C}$ *is also block-Toeplitz.*

The properties proved in theorems A.28 and A.35 and lemmas A.31, A.33, A.38 and A.39 hold for any transformation matrix $\mathbf{T}$ and matrix $\mathbf{A}$ which satisfy

$$\begin{cases} \mathbf{T}\mathbf{A}\mathbf{T} = \mathbf{A} \\ \mathbf{T} = \mathbf{T}^T \\ \mathbf{T} = \mathbf{T}^{-1} \end{cases} \tag{A.51}$$

A.6 Derivative to vectors and matrices

Consider the vectors $\mathbf{u} \in \mathbb{R}^N$, $\mathbf{d} \in \mathbb{R}^N$ and $\mathbf{w} \in \mathbb{R}^N$, and the matrices $\mathbf{A} \in \mathbb{R}^{N\times N}$ and $\mathbf{W} \in \mathbb{R}^{N\times N}$,

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \vdots & \vdots & & \vdots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{bmatrix} .$$

We can prove the following properties (these properties are easily extendible to the complex-valued case).

Property A.41 $\boxed{\mathbf{J} = \mathbf{w}^T\mathbf{u} = \mathbf{u}^T\mathbf{w}, \quad \frac{\partial \mathbf{J}}{\partial \mathbf{w}} = \mathbf{u}}$

Proof:

$$\begin{aligned} \mathbf{J} &= \mathbf{w}^T\mathbf{u} = \sum_{i=1}^{N} u_i w_i \Rightarrow \frac{\partial \mathbf{J}}{\partial w_k} = u_k \\ \frac{\partial \mathbf{J}}{\partial \mathbf{w}} &= \begin{bmatrix} \frac{\partial \mathbf{J}}{\partial w_1} \\ \frac{\partial \mathbf{J}}{\partial w_2} \\ \vdots \\ \frac{\partial \mathbf{J}}{\partial w_N} \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} = \mathbf{u} . \end{aligned}$$

□

Property A.42 $\boxed{\mathbf{J} = \mathbf{w}^T\mathbf{A}\mathbf{w}, \quad \frac{\partial \mathbf{J}}{\partial \mathbf{w}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{w}}$

$\boxed{\mathbf{J} = \mathbf{w}^T\mathbf{u}\mathbf{u}^T\mathbf{w}, \quad \frac{\partial \mathbf{J}}{\partial \mathbf{w}} = 2\mathbf{u}\mathbf{u}^T\mathbf{w}}$

Proof:

$$\begin{aligned} \mathbf{J} &= \mathbf{w}^T\mathbf{A}\mathbf{w} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i a_{ij} w_j \\ \frac{\partial \mathbf{J}}{\partial w_k} &= \frac{\partial}{\partial w_k}\left(w_k a_{kk} w_k + \sum_{j=1, j\neq k}^{N} w_k a_{kj} w_j + \sum_{i=1, i\neq k}^{N} w_i a_{ik} w_k \right) \\ &= \sum_{j=1}^{N} a_{kj} w_j + \sum_{i=1}^{N} w_i a_{ik} = \mathbf{A}(1,:)\mathbf{w} + \mathbf{A}(:,1)^T\mathbf{w} \\ \frac{\partial \mathbf{J}}{\partial \mathbf{w}} &= \begin{bmatrix} \frac{\partial \mathbf{J}}{\partial w_1} \\ \frac{\partial \mathbf{J}}{\partial w_2} \\ \vdots \\ \frac{\partial \mathbf{J}}{\partial w_N} \end{bmatrix} = \begin{bmatrix} \mathbf{A}(1,:) \\ \mathbf{A}(2,:) \\ \vdots \\ \mathbf{A}(N,:) \end{bmatrix} \mathbf{w} + \begin{bmatrix} \mathbf{A}(:,1)^T \\ \mathbf{A}(:,2)^T \\ \vdots \\ \mathbf{A}(:,N)^T \end{bmatrix} \mathbf{w} \\ &= (\mathbf{A} + \mathbf{A}^T)\mathbf{w} . \end{aligned}$$

For symmetric $\mathbf{A}$: $\frac{\partial \mathbf{J}}{\partial \mathbf{w}} = 2\mathbf{A}\mathbf{w}$. □

Property A.43 $\boxed{\mathbf{J} = \mathbf{u}^T\mathbf{W}\mathbf{d}, \quad \frac{\partial \mathbf{J}}{\partial \mathbf{W}} = \mathbf{u}\mathbf{d}^T}$

Proof:

$$\begin{aligned}
\mathbf{J} &= \mathbf{u}^T\mathbf{W}\mathbf{d} = \sum_{i=1}^{N}\sum_{j=1}^{N} u_i w_{ij} d_j \Rightarrow \frac{\partial \mathbf{J}}{\partial w_{kl}} = u_k d_l \\
\frac{\partial \mathbf{J}}{\partial \mathbf{W}} &= \begin{bmatrix} \frac{\partial \mathbf{J}}{\partial w_{11}} & \frac{\partial \mathbf{J}}{\partial w_{12}} & \cdots & \frac{\partial \mathbf{J}}{\partial w_{1N}} \\ \frac{\partial \mathbf{J}}{\partial w_{21}} & \frac{\partial \mathbf{J}}{\partial w_{22}} & \cdots & \frac{\partial \mathbf{J}}{\partial w_{2N}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \mathbf{J}}{\partial w_{N1}} & \frac{\partial \mathbf{J}}{\partial w_{N2}} & \cdots & \frac{\partial \mathbf{J}}{\partial w_{NN}} \end{bmatrix} = \begin{bmatrix} u_1d_1 & u_1d_2 & \ldots & u_1d_N \\ u_2d_1 & u_2d_2 & \ldots & u_2d_N \\ \vdots & \vdots & & \vdots \\ u_Nd_1 & u_Nd_2 & \ldots & u_Nd_N \end{bmatrix} \\
&= \mathbf{u}\mathbf{d}^T \ .
\end{aligned}$$

□

Property A.44 $\boxed{\mathbf{J} = \mathbf{u}^T\mathbf{W}\mathbf{W}^T\mathbf{u}, \quad \frac{\partial \mathbf{J}}{\partial \mathbf{W}} = 2\mathbf{u}\mathbf{u}^T\mathbf{W}}$

Proof:

$$\begin{aligned}
\mathbf{J} &= \mathbf{u}^T\mathbf{W}\mathbf{W}^T\mathbf{u} = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} u_j w_{ji}\right)\left(\sum_{k=1}^{N} w_{ki} u_k\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} u_j w_{ji} w_{ki} u_k \\
\frac{\partial \mathbf{J}}{\partial w_{pq}} &= \frac{\partial}{\partial w_{pq}}\left(u_p w_{pq} w_{pq} u_p + \sum_{j=1, j\neq p}^{N} u_j w_{jq} w_{pq} u_p + \sum_{k=1, k\neq p}^{N} u_p w_{pq} w_{kq} u_k \right) \\
&= \sum_{j=1}^{N} u_j w_{jq} u_p + \sum_{k=1}^{N} u_p w_{kq} u_k = 2u_p \sum_{j=1}^{N} u_j w_{jq} = 2u_p \mathbf{u}\mathbf{W}(:,q) \\
\frac{\partial \mathbf{J}}{\partial \mathbf{W}} &= \begin{bmatrix} \frac{\partial \mathbf{J}}{\partial w_{11}} & \frac{\partial \mathbf{J}}{\partial w_{12}} & \cdots & \frac{\partial \mathbf{J}}{\partial w_{1N}} \\ \frac{\partial \mathbf{J}}{\partial w_{21}} & \frac{\partial \mathbf{J}}{\partial w_{22}} & \cdots & \frac{\partial \mathbf{J}}{\partial w_{2N}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \mathbf{J}}{\partial w_{N1}} & \frac{\partial \mathbf{J}}{\partial w_{N2}} & \cdots & \frac{\partial \mathbf{J}}{\partial w_{NN}} \end{bmatrix} \\
&= 2\begin{bmatrix} u_1\mathbf{u}^T\mathbf{W}(:,1) & u_1\mathbf{u}^T\mathbf{W}(:,2) & \ldots & u_1\mathbf{u}^T\mathbf{W}(:,N) \\ u_2\mathbf{u}^T\mathbf{W}(:,1) & u_2\mathbf{u}^T\mathbf{W}(:,2) & \ldots & u_2\mathbf{u}^T\mathbf{W}(:,N) \\ \vdots & \vdots & & \vdots \\ u_N\mathbf{u}^T\mathbf{W}(:,1) & u_N\mathbf{u}^T\mathbf{W}(:,2) & \ldots & u_N\mathbf{u}^T\mathbf{W}(:,N) \end{bmatrix} \\
&= 2\left[\ \mathbf{u}\mathbf{u}^T\mathbf{W}(:,1) \quad \mathbf{u}\mathbf{u}^T\mathbf{W}(:,2) \quad \ldots \quad \mathbf{u}\mathbf{u}^T\mathbf{W}(:,N)\ \right] \\
&= 2\mathbf{u}\mathbf{u}^T\mathbf{W}
\end{aligned}$$

□

B Appendix to Chapter 2

B.1 Orthogonality of $\mathbf{Q}_V^T\,\mathbf{U}_V$

Using (2.62), the matrix $\mathbf{R}_V^T\mathbf{R}_V$ can be both written as

$$\mathbf{R}_V^T\mathbf{R}_V = \mathbf{V}_0[k]^T\mathbf{Q}_V\mathbf{Q}_V^T\mathbf{V}_0[k] = \mathbf{Q}^T\boldsymbol{\Sigma}_V\,\underbrace{\mathbf{U}_V^T\mathbf{Q}_V\mathbf{Q}_V^T\mathbf{U}_V}\,\boldsymbol{\Sigma}_V\mathbf{Q} \tag{B.1}$$

and

$$\mathbf{R}_V^T\mathbf{R}_V = \mathbf{V}_0[k]^T\mathbf{V}_0[k] = \mathbf{Q}^T\boldsymbol{\Sigma}_V^2\mathbf{Q}\ , \tag{B.2}$$

such that $\mathbf{U}_V^T\mathbf{Q}_V\mathbf{Q}_V^T\mathbf{U}_V = \left(\mathbf{Q}_V^T\,\mathbf{U}_V\right)^T\left(\mathbf{Q}_V^T\,\mathbf{U}_V\right) = \mathbf{I}_L$. Hence, $\mathbf{Q}_V^T\,\mathbf{U}_V$ is an orthogonal matrix.

B.2 Minimisation of $||\mathbf{Y}_0[k]\mathbf{W} - \mathbf{X}_0[k]||_F^2$

The cost function $J_{MV}(\mathbf{W})$ in (2.77) can be written as

$$\begin{aligned}
J_{MV}(\mathbf{W}) &= ||\mathbf{Y}_0[k]\mathbf{W} - \mathbf{X}_0[k]||_F^2 && \text{(B.3)}\\
&= \sum_{i=0}^{P-1} ||\,\mathbf{y}_0^T[k-i]\mathbf{W} - \mathbf{x}_0^T[k-i]\,||_2^2 && \text{(B.4)}\\
&= \sum_{i=0}^{P-1} \mathbf{y}_0^T[k-i]\mathbf{W}\mathbf{W}^T\mathbf{y}_0[k-i] - 2\mathbf{y}_0^T[k-i]\mathbf{W}\mathbf{x}_0[k-i] && \text{(B.5)}\\
&\qquad +\mathbf{x}_0^T[k-i]\mathbf{x}_0[k-i]\ , && \text{(B.6)}
\end{aligned}$$

which can be minimised by putting the derivative (cf. Appendix A.6)

$$\begin{aligned}
\frac{\partial J_{MV}(\mathbf{W})}{\partial \mathbf{W}} &= \sum_{i=0}^{P-1} 2\mathbf{y}_0[k-i]\mathbf{y}_0^T[k-i]\mathbf{W} - 2\mathbf{y}_0[k-i]\mathbf{x}_0^T[k-i] && \text{(B.7)}\\
&= 2\mathbf{Y}_0^T[k]\mathbf{Y}_0[k]\mathbf{W} - 2\mathbf{Y}_0^T[k]\mathbf{X}_0[k] && \text{(B.8)}
\end{aligned}$$

to zero, such that

$$\mathbf{W} = \left(\mathbf{Y}_0^T[k]\mathbf{Y}_0[k]\right)^{-1}\mathbf{Y}_0^T[k]\mathbf{X}_0[k]\ . \tag{B.9}$$

B.3 Solution of optimisation problem (2.129)

The optimisation problem

$$\min_{\mathbf{W}(\omega)} \mathbf{W}^H(\omega)\bar{\mathbf{R}}_{yy}(\omega)\mathbf{W}(\omega), \quad \text{subject to } \mathbf{W}^H(\omega)\mathbf{d}(\omega,\theta_x) = 1\ , \tag{B.10}$$

can be solved by introducing the Lagrange multiplier λ and considering the cost function[1]

$$J(\mathbf{W}) = \mathbf{W}^H\bar{\mathbf{R}}_{yy}\mathbf{W} + \lambda(\mathbf{W}^H\mathbf{d} - 1)^H(\mathbf{W}^H\mathbf{d} - 1)\ . \tag{B.11}$$

[1] For the sake of conciseness the frequency parameter ω and the angle parameter θ_x will be frequently omitted in the following equations.

Putting the derivative of $J(\mathbf{W})$ with respect to $\mathbf{W}$, i.e.

$$\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} = 2\bar{\mathbf{R}}_{yy}\mathbf{W} + 2\lambda(\mathbf{d}\mathbf{d}^H\mathbf{W} - \mathbf{d}) \tag{B.12}$$

to zero, yields the solution

$$\mathbf{W} = (\bar{\mathbf{R}}_{yy} + \lambda\mathbf{d}\mathbf{d}^H)^{-1}\mathbf{d}\ , \tag{B.13}$$

which can be rewritten, using the matrix inversion lemma (A.38), as

$$\mathbf{W} = \left[\bar{\mathbf{R}}_{yy}^{-1} - \frac{\lambda\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}}{1+\lambda\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}}\right]\mathbf{d}\ . \tag{B.14}$$

The parameter λ can be calculated by satisfying the constraint $\mathbf{W}^H\mathbf{d} = 1$, i.e.

$$\mathbf{W}^H\mathbf{d} = \mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d} - \frac{\lambda(\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d})^2}{1+\lambda\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}} = \frac{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}}{1+\lambda\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}} = 1\ , \tag{B.15}$$

such that

$$\lambda = \frac{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d} - 1}{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}}\ . \tag{B.16}$$

The filter $\mathbf{W}$ in (B.14) can now be written as

$$\begin{aligned}
\mathbf{W} &= \left[\bar{\mathbf{R}}_{yy}^{-1} - \frac{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d} - 1}{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}} \cdot \frac{\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}}{1+(\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}-1)}\right]\mathbf{d} && \text{(B.17)}\\
&= \bar{\mathbf{R}}_{yy}^{-1}\mathbf{d} - \frac{(\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}-1)\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}}{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}} && \text{(B.18)}\\
&= \frac{\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}}{\mathbf{d}^H\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}}\ . && \text{(B.19)}
\end{aligned}$$

Using (2.109), the signal vector $\mathbf{Y}(\omega)$ can be written as

$$\mathbf{Y}(\omega) = S(\omega)e^{-j\omega\bar{\tau}(\theta_x)}\mathbf{d}(\omega,\theta_x) + \mathbf{V}(\omega)\ , \tag{B.20}$$

such that the correlation matrix $\bar{\mathbf{R}}_{yy}$ is equal to

$$\bar{\mathbf{R}}_{yy} = \mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\} = P_s\mathbf{d}\mathbf{d}^H + \bar{\mathbf{R}}_{vv}\ . \tag{B.21}$$

Using the matrix inversion lemma (A.38), $\bar{\mathbf{R}}_{yy}^{-1}$ can be written as

$$\bar{\mathbf{R}}_{yy}^{-1} = \bar{\mathbf{R}}_{vv}^{-1} - \frac{P_s\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}\mathbf{d}^H\bar{\mathbf{R}}_{vv}^{-1}}{1+P_s\mathbf{d}^H\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}\ , \tag{B.22}$$

such that $\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d}$ is equal to

$$\bar{\mathbf{R}}_{yy}^{-1}\mathbf{d} = \bar{\mathbf{R}}_{vv}^{-1}\mathbf{d} - \frac{P_s\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}\mathbf{d}^H\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}{1+P_s\mathbf{d}^H\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}} = \frac{\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}{1+P_s\mathbf{d}^H\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}\ , \tag{B.23}$$

and the filter $\mathbf{W}$ in (B.14) can be written as

$$\mathbf{W} = \frac{\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}{1 + P_s \mathbf{d}^H \bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}} \cdot \frac{1 + P_s \mathbf{d}^H \bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}{\mathbf{d}^H \bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}} = \frac{\bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}}{\mathbf{d}^H \bar{\mathbf{R}}_{vv}^{-1}\mathbf{d}} \; . \tag{B.24}$$

If we assume a homogeneous noise field, such that $\bar{\mathbf{R}}_{vv} = P_v \mathbf{\Gamma}_v$, cf. (2.37), then $\mathbf{W}$ can be written as

$$\mathbf{W} = \frac{\mathbf{\Gamma}_v^{-1}\mathbf{d}}{\mathbf{d}^H \mathbf{\Gamma}_v^{-1}\mathbf{d}} \; . \tag{B.25}$$

B.4 Solution of optimisation problem (2.134)

The optimisation problem

$$\min_{\mathbf{w}[k]} \mathbf{w}^T[k] \bar{\mathbf{R}}_{yy}[k] \mathbf{w} \quad \text{subject to } \; \mathbf{C}\mathbf{w}[k] = \mathbf{b} \; , \tag{B.26}$$

can be solved by introducing the J-dimensional vector of Lagrange multipliers $\boldsymbol{\lambda}$ and considering the cost function

$$J(\mathbf{w}[k]) = \mathbf{w}^T[k] \bar{\mathbf{R}}_{yy}[k] \mathbf{w}[k] + \boldsymbol{\lambda}^T (\mathbf{C}\mathbf{w}[k] - \mathbf{b}) \; . \tag{B.27}$$

Putting the derivative of $J(\mathbf{w}[k])$ with respect to $\mathbf{w}[k]$, i.e.

$$\frac{\partial J(\mathbf{w}[k])}{\partial \mathbf{w}[k]} = 2\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k] + 2\mathbf{C}^T \boldsymbol{\lambda} \tag{B.28}$$

to zero, yields the solution

$$\mathbf{w}[k] = -\bar{\mathbf{R}}_{yy}^{-1}[k] \mathbf{C}^T \boldsymbol{\lambda} \; . \tag{B.29}$$

The parameter vector $\boldsymbol{\lambda}$ can be calculated by satisfying the constraint $\mathbf{C}\mathbf{w}[k] = \mathbf{b}$, i.e.

$$\mathbf{C}\mathbf{w}[k] = -\mathbf{C}\bar{\mathbf{R}}_{yy}^{-1}[k] \mathbf{C}^T \boldsymbol{\lambda} = \mathbf{b} \; , \tag{B.30}$$

such that

$$\boldsymbol{\lambda} = -(\mathbf{C}\bar{\mathbf{R}}_{yy}^{-1}[k]\mathbf{C}^T)^{-1}\mathbf{b} \; , \tag{B.31}$$

and hence the solution $\mathbf{w}[k]$ is equal to

$$\mathbf{w}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k]\mathbf{C}^T (\mathbf{C}\bar{\mathbf{R}}_{yy}^{-1}[k]\mathbf{C}^T)^{-1}\mathbf{b} \; . \tag{B.32}$$

B.5 Constrained gradient-descent procedure (2.137)

The constrained optimisation problem

$$\min_{\mathbf{w}[k]} \mathbf{w}^T[k] \bar{\mathbf{R}}_{yy}[k] \mathbf{w} \quad \text{subject to } \; \mathbf{C}\mathbf{w}[k] = \mathbf{b} \; , \tag{B.33}$$

can be adaptively solved using a gradient-descent optimisation technique, where in each iteration step the filters are updated in the direction of the constrained gradient, i.e. using (B.28),

$$\mathbf{w}[k+1] = \mathbf{w}[k] - \frac{\mu}{2}\frac{\partial J(\mathbf{w}[k])}{\partial \mathbf{w}[k]} = \mathbf{w}[k] - \mu(\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k] + \mathbf{C}^T\boldsymbol{\lambda}[k]) \ , \quad \text{(B.34)}$$

with μ the step size of the adaptive algorithm. Since the filter $\mathbf{w}[k+1]$ also has to satisfy the constraint

$$\mathbf{C}\mathbf{w}[k+1] = \mathbf{C}\mathbf{w}[k] - \mu\mathbf{C}(\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k] + \mathbf{C}^T\boldsymbol{\lambda}[k]) = \mathbf{b} \ , \quad \text{(B.35)}$$

the parameter vector $\boldsymbol{\lambda}[k]$ should be equal to

$$\boldsymbol{\lambda}[k] = (\mu\mathbf{C}\mathbf{C}^T)^{-1}(\underbrace{\mathbf{C}\mathbf{w}[k] - \mathbf{b}} - \mu\mathbf{C}\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k]) \quad \text{(B.36)}$$

Note that we do not assume the term $\mathbf{C}\mathbf{w}[k] - \mathbf{b}$ to be exactly equal to zero, in order to prevent error accumulation. Using (B.36), the filter update can now be written as

$$\begin{aligned}\mathbf{w}[k+1] = \mathbf{w}[k] - \mu &\left[\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k] + \mathbf{C}^T(\mu\mathbf{C}\mathbf{C}^T)^{-1}(\underbrace{\mathbf{C}\mathbf{w}[k] - \mathbf{b}} - \mu\mathbf{C}\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k])\right] \\ &= \left[\mathbf{I}_M - \mu\bar{\mathbf{R}}_{yy}[k] - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C} + \mu\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C}\bar{\mathbf{R}}_{yy}[k]\right]\mathbf{w}[k] \\ &\quad + \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b} \\ &= \mathbf{P}_C(\mathbf{w}[k] - \mu\bar{\mathbf{R}}_{yy}[k]\mathbf{w}[k]) + \mathbf{b}_C \ ,\end{aligned}$$

with

$$\begin{aligned}\mathbf{P}_C &= \mathbf{I}_M - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C} \ , & \text{(B.37)} \\ \mathbf{b}_C &= \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b} \ . & \text{(B.38)}\end{aligned}$$

By using the instantaneous gradient, i.e. approximating $\bar{\mathbf{R}}_{yy}[k] \approx \mathbf{y}\mathbf{y}^T[k]$, the constrained LMS algorithm can be written as

$$\mathbf{w}[k+1] = \mathbf{P}_C(\mathbf{w}[k] - \mu z[k]\mathbf{y}[k]) + \mathbf{b}_C \ . \quad \text{(B.39)}$$

In fact, a geometrical interpretation can be given (see Fig. B.1a). The J linear constraints restrict the filter $\mathbf{w}[k]$ to lie in an $(M-J)$-dimensional hyperplane, i.e. the constraint hyperplane $\Lambda = \{\mathbf{w} : \mathbf{C}\mathbf{w} = \mathbf{b}\}$, which is orthogonal to the subspace spanned by the rows of $\mathbf{C}$. The matrix $\mathbf{P}_C$ is the projection matrix on the constraint subspace $\Sigma = \{\mathbf{w} : \mathbf{C}\mathbf{w} = \mathbf{0}\}$, while the vector $\mathbf{b}_C$ is the shortest vector orthogonal to the constraint subspace Σ which terminates on the constraint hyperplane Λ. The constrained LMS algorithm is geometrically depicted in Fig. B.1b. It can be easily seen that by using this approach, no error accumulation occurs and all vectors $\mathbf{w}[k]$ lie in the constraint hyperplane.

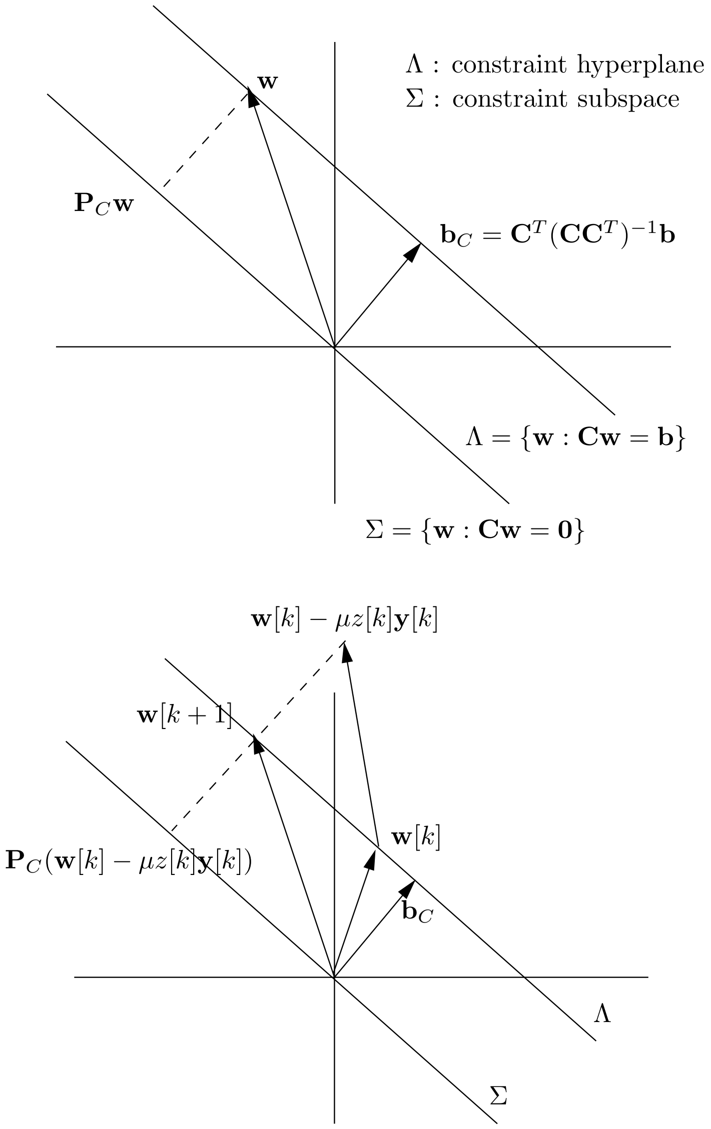

$\mathbf{w}$
Λ : constraint hyperplane
Σ : constraint subspace
$\mathbf{P}_C\mathbf{w}$
$\mathbf{b}_C = \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{b}$
$\Lambda = \{\mathbf{w} : \mathbf{C}\mathbf{w} = \mathbf{b}\}$
$\Sigma = \{\mathbf{w} : \mathbf{C}\mathbf{w} = \mathbf{0}\}$
$\mathbf{w}[k] - \mu z[k]\mathbf{y}[k]$
$\mathbf{w}[k+1]$
$\mathbf{w}[k]$
$\mathbf{P}_C(\mathbf{w}[k] - \mu z[k]\mathbf{y}[k])$
$\mathbf{b}_C$
Λ
Σ

Figure B.1: Constrained gradient-descent procedure (Frost beamformer)

C Appendix to Part I

C.1 Signal distortion $\epsilon_y^2[k]$ versus residual noise $\epsilon_v^2[k]$

Using the filter matrix $\bar{\mathbf{W}}[k]$ in (3.33),

$$\bar{\mathbf{W}}[k] = \left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)^{-1} \bar{\mathbf{R}}_{xx}[k] \;, \tag{C.1}$$

with the Lagrange-multiplier $\lambda \geq 0$, it can be easily proved that

$$\begin{aligned} \mathbf{I}_M - \bar{\mathbf{W}}^T[k] &= \left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)\left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)^{-1} - \\ &\quad \bar{\mathbf{R}}_{xx}[k]\ \left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)^{-1} &\text{(C.2)} \\ &= \lambda \bar{\mathbf{R}}_{vv}[k]\ \left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)^{-1} \;. &\text{(C.3)} \end{aligned}$$

Using (3.15) and (C.3), the signal distortion energy $\epsilon_y^2[k]$ can be written in function of λ as

$$\begin{aligned} \epsilon_y^2[k] &= \mathcal{E}\Big\{\mathbf{x}^T[k]\left(\mathbf{I}_M - \bar{\mathbf{W}}^T[k]\right)^T\left(\mathbf{I}_M - \bar{\mathbf{W}}^T[k]\right)\mathbf{x}[k]\Big\} &\text{(C.4)} \\ &= \mathrm{tr}\Big\{\left(\mathbf{I}_M - \bar{\mathbf{W}}^T[k]\right)\bar{\mathbf{R}}_{xx}[k]\left(\mathbf{I}_M - \bar{\mathbf{W}}^T[k]\right)^T\Big\} &\text{(C.5)} \\ &= \mathrm{tr}\Big\{\lambda^2 \bar{\mathbf{R}}_{vv}[k]\left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)^{-1}\bar{\mathbf{R}}_{xx}[k]\left(\bar{\mathbf{R}}_{xx}[k] + \lambda \bar{\mathbf{R}}_{vv}[k]\right)^{-1}\bar{\mathbf{R}}_{vv}[k]\Big\} \\ &= \mathrm{tr}\left\{\bar{\mathbf{Q}}[k]\ \mathrm{diag}\left\{\frac{\left(\bar{\sigma}_i^2[k] - \bar{\eta}_i^2[k]\right)\left(\lambda \bar{\eta}_i^2[k]\right)^2}{\left(\bar{\sigma}_i^2[k] + (\lambda-1)\,\bar{\eta}_i^2[k]\right)^2}\right\}\bar{\mathbf{Q}}^T[k]\right\} \;, &\text{(C.6)} \end{aligned}$$

whereas the residual noise energy $\epsilon_v^2[k]$ can be written as in (3.37),

$$\epsilon_v^2[k] = \mathrm{tr}\left\{\bar{\mathbf{Q}}[k]\ \mathrm{diag}\left\{\frac{\left(\bar{\sigma}_i^2[k] - \bar{\eta}_i^2[k]\right)^2 \bar{\eta}_i^2[k]}{\left(\bar{\sigma}_i^2[k] + (\lambda-1)\,\bar{\eta}_i^2[k]\right)^2}\right\}\bar{\mathbf{Q}}^T[k]\right\} \;. \tag{C.7}$$

Since we want to prove that the smaller the signal distortion $\epsilon_y^2[k]$ is, the larger the residual noise $\epsilon_v^2[k]$ is, this is equivalent to proving that the derivative $\partial\epsilon_y^2[k]/\partial\epsilon_v^2[k]$ is always negative. This derivative can be computed as

$$\frac{\partial \epsilon_y^2[k]}{\partial \epsilon_v^2[k]} = \frac{\partial \epsilon_y^2[k]}{\partial \lambda}\left(\frac{\partial \epsilon_v^2[k]}{\partial \lambda}\right)^{-1} \;. \tag{C.8}$$

Using (C.6), the derivative $\partial\epsilon_y^2[k]/\partial\lambda$ is equal to

$$\begin{aligned} \frac{\partial \epsilon_y^2[k]}{\partial \lambda} &= \mathrm{tr}\left\{\bar{\mathbf{Q}}[k]\ \mathrm{diag}\left\{\frac{\partial}{\partial \lambda}\frac{\left(\bar{\sigma}_i^2[k] - \bar{\eta}_i^2[k]\right)\left(\lambda \bar{\eta}_i^2[k]\right)^2}{\left(\bar{\sigma}_i^2[k] + (\lambda-1)\,\bar{\eta}_i^2[k]\right)^2}\right\}\bar{\mathbf{Q}}^T[k]\right\} &\text{(C.9)} \\ &= \mathrm{tr}\left\{\bar{\mathbf{Q}}[k]\ \mathrm{diag}\left\{\frac{-2\lambda\left(\bar{\sigma}_i^2[k] - \bar{\eta}_i^2[k]\right)^2\left(\bar{\eta}_i^2[k]\right)^2}{\left(\bar{\sigma}_i^2[k] + (\lambda-1)\,\bar{\eta}_i^2[k]\right)^3}\right\}\bar{\mathbf{Q}}^T[k]\right\} \;, &\text{(C.10)} \end{aligned}$$

whereas using (C.7), the derivative $\partial\epsilon_v^2[k]/\partial\lambda$ is equal to

$$\frac{\partial\epsilon_v^2[k]}{\partial\lambda} = \operatorname{tr}\left\{\bar{\mathbf{Q}}[k]\ \operatorname{diag}\left\{\frac{\partial}{\partial\lambda}\ \frac{\left(\bar{\sigma}_i^2[k]-\bar{\eta}_i^2[k]\right)^2\bar{\eta}_i^2[k]}{\left(\bar{\sigma}_i^2[k]+(\lambda-1)\,\bar{\eta}_i^2[k]\right)^2}\right\}\bar{\mathbf{Q}}^T[k]\right\} \tag{C.11}$$

$$= \operatorname{tr}\left\{\bar{\mathbf{Q}}[k]\ \operatorname{diag}\left\{\frac{2\left(\bar{\sigma}_i^2[k]-\bar{\eta}_i^2[k]\right)^2\left(\bar{\eta}_i^2[k]\right)^2}{\left(\bar{\sigma}_i^2[k]+(\lambda-1)\,\bar{\eta}_i^2[k]\right)^3}\right\}\bar{\mathbf{Q}}^T[k]\right\}, \tag{C.12}$$

such that using (C.8)

$$\frac{\partial\epsilon_y^2[k]}{\partial\epsilon_v^2[k]} = -\lambda\ , \tag{C.13}$$

which is always negative. For a specific speech-noise example, we have plotted the signal distortion energy $\epsilon_y^2[k]$ versus the residual noise energy $\epsilon_v^2[k]$ in Fig. 3.2, where it can be seen that this function is monotonically decreasing.

When $\lambda = 0$, i.e. $\bar{\mathbf{W}}[k] = \mathbf{I}_M$, it can be seen from (C.6) and (C.7) that the signal distortion energy $\epsilon_y^2[k] = 0$ and that the residual noise energy reaches its maximum value, i.e.

$$\epsilon_{v,max}^2[k] = \operatorname{tr}\left\{\bar{\mathbf{Q}}[k]\ \operatorname{diag}\{\bar{\eta}_i^2[k]\}\ \bar{\mathbf{Q}}^T[k]\right\} = \operatorname{tr}\left\{\bar{\mathbf{R}}_{vv}[k]\right\}\ . \tag{C.14}$$

When $\lambda = \infty$, i.e. $\bar{\mathbf{W}}[k] = \mathbf{0}$, it can be seen from (C.6) and (C.7) that the residual noise energy $\epsilon_v^2[k] = 0$ and that the signal distortion energy reaches its maximum value, i.e.

$$\epsilon_{y,max}^2[k] = \operatorname{tr}\left\{\bar{\mathbf{Q}}[k]\ \operatorname{diag}\{(\bar{\sigma}_i^2[k]-\bar{\eta}_i^2[k])\}\ \bar{\mathbf{Q}}^T[k]\right\} = \operatorname{tr}\left\{\bar{\mathbf{R}}_{xx}[k]\right\}\ . \tag{C.15}$$

C.2 Wiener filter for combined noise and echo reduction

Using (3.113) and (3.115), the matrix $\bar{\mathbf{R}}_{y_ty_t}(\omega)$ can be written as

$$\bar{\mathbf{R}}_{y_ty_t}(\omega) = \mathcal{E}\{\mathbf{Y}_t(\omega)\mathbf{Y}_t^H(\omega)\} = \mathcal{E}\left\{\begin{bmatrix}\mathbf{Y}(\omega)\\ F(\omega)\end{bmatrix}\begin{bmatrix}\mathbf{Y}^H(\omega) & F^*(\omega)\end{bmatrix}\right\} \tag{C.16}$$

$$= \begin{bmatrix}\bar{\mathbf{R}}_{yy}(\omega) & \bar{\mathbf{r}}_{yf}(\omega)\\ \bar{\mathbf{r}}_{yf}^H(\omega) & P_f(\omega)\end{bmatrix}, \tag{C.17}$$

with $P_f(\omega) = \mathcal{E}\{|F(\omega)|^2\}$ and

$$\bar{\mathbf{r}}_{yf}(\omega) = \mathcal{E}\{\mathbf{Y}(\omega)F^*(\omega)\} = \mathcal{E}\{\mathbf{V}_f(\omega)F^*(\omega)\} = \mathbf{H}_f(\omega)P_f(\omega)\ , \tag{C.18}$$

since the speech components $\mathbf{X}(\omega)$ and the unknown noise components $\mathbf{V}_u(\omega)$ are assumed to be uncorrelated with the far-end echo signal $F(\omega)$. Using (3.113) and (3.116), the matrix $\bar{\mathbf{R}}_{x_tx_t}(\omega)$ can be written as

$$\bar{\mathbf{R}}_{x_tx_t}(\omega) = \mathcal{E}\{\mathbf{X}_t(\omega)\mathbf{X}_t^H(\omega)\} = \begin{bmatrix}\bar{\mathbf{R}}_{xx}(\omega) & \mathbf{0}\\ \mathbf{0} & 0\end{bmatrix}. \tag{C.19}$$

Using (A.39), the matrix $\bar{\mathbf{R}}_{y_t y_t}^{-1}(\omega)$ can be written as

$$\bar{\mathbf{R}}_{y_t y_t}^{-1}(\omega) = \begin{bmatrix} \left(\bar{\mathbf{R}}_{yy}(\omega) - \frac{\bar{\mathbf{r}}_{yf}(\omega)\bar{\mathbf{r}}_{yf}^H(\omega)}{P_f(\omega)}\right)^{-1} & -\frac{\bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{r}}_{yf}(\omega)}{P_f(\omega)-\bar{\mathbf{r}}_{yf}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{r}}_{yf}(\omega)} \\ -\frac{\bar{\mathbf{r}}_{yf}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)}{P_f(\omega)-\bar{\mathbf{r}}_{yf}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{r}}_{yf}(\omega)} & \frac{1}{P_f(\omega)-\bar{\mathbf{r}}_{yf}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{r}}_{yf}(\omega)} \end{bmatrix}, \tag{C.20}$$

such that the Wiener filter $\mathbf{W}_{WF}^t(\omega)$ in (3.114) can be written as

$$\mathbf{W}_{WF}^t(\omega) = \begin{bmatrix} \mathbf{W}_{WF}(\omega) \\ W_{WF}^f(\omega) \end{bmatrix} = \begin{bmatrix} \left(\bar{\mathbf{R}}_{yy}(\omega) - \frac{\bar{\mathbf{r}}_{yf}(\omega)\bar{\mathbf{r}}_{yf}^H(\omega)}{P_f(\omega)}\right)^{-1} \\ -\frac{\bar{\mathbf{r}}_{yf}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)}{P_f(\omega)-\bar{\mathbf{r}}_{yf}^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\bar{\mathbf{r}}_{yf}(\omega)} \end{bmatrix} \bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1\,. \tag{C.21}$$

Using (C.18) and (3.111), the filter $\mathbf{W}_{WF}(\omega)$ can be written as

$$\begin{aligned} \mathbf{W}_{WF}(\omega) &= \left[\bar{\mathbf{R}}_{yy}(\omega) - P_f(\omega)\mathbf{H}_f(\omega)\mathbf{H}_f^H(\omega)\right]^{-1}\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1 & \text{(C.22)} \\ &= \left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1\,, & \text{(C.23)} \end{aligned}$$

which is the same formula for the multi-channel Wiener filter as if no echo source were present, cf. (3.81), implying that the echo source has no influence on $\mathbf{W}_{WF}(\omega)$. Using (C.18), the filter $W_{WF}^f(\omega)$ can be written as

$$W_{WF}^f(\omega) = -\frac{\mathbf{H}_f^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\,\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1}{1 - P_f(\omega)\,\mathbf{H}_f^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega)\mathbf{H}_f(\omega)}\,. \tag{C.24}$$

Using (3.111) and the matrix inversion lemma in (A.38), the matrix $\bar{\mathbf{R}}_{yy}^{-1}(\omega)$ is equal to

$$\begin{aligned} \bar{\mathbf{R}}_{yy}^{-1}(\omega) &= \left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega) + P_f(\omega)\mathbf{H}_f(\omega)\mathbf{H}_f^H(\omega)\right]^{-1} & \text{(C.25)} \\ &= \left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1} - \\ &\quad \frac{P_f(\omega)\,\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\mathbf{H}_f(\omega)\mathbf{H}_f^H(\omega)\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}}{1 + P_f(\omega)\,\mathbf{H}_f^H(\omega)\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\mathbf{H}_f(\omega)}\,, \end{aligned}$$

such that

$$\mathbf{H}_f^H(\omega)\bar{\mathbf{R}}_{yy}^{-1}(\omega) = \frac{\mathbf{H}_f^H(\omega)\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}}{1 + P_f(\omega)\,\mathbf{H}_f^H(\omega)\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\mathbf{H}_f(\omega)}\,. \tag{C.26}$$

Inserting (C.26) into (C.24) yields

$$W_{WF}^f(\omega) = -\mathbf{H}_f^H(\omega)\left[\bar{\mathbf{R}}_{xx}(\omega) + \bar{\mathbf{R}}_{vv}^u(\omega)\right]^{-1}\bar{\mathbf{R}}_{xx}(\omega)\,\mathbf{e}_1 = -\mathbf{H}_f^H(\omega)\mathbf{W}_{WF}(\omega)\,. \tag{C.27}$$

D Appendix to Part III

D.1 Weighted LS criterion with linear constraint

The constrained minimisation problem

$$\min_{\mathbf{w}} \mathbf{w}^T \mathbf{Q}_{LS} \mathbf{w} - 2\mathbf{w}^T \mathbf{a} + d_{LS}, \quad \text{subject to } \mathbf{C}\mathbf{w} = \mathbf{b} \ , \tag{D.1}$$

with $\mathbf{C}$ a $J \times M$-dimensional constraint matrix and $\mathbf{b}$ a J-dimensional constraint vector, can easily be transformed into an unconstrained minimisation problem. Using (2.146), the filter $\mathbf{w}$ can be parametrised as

$$\mathbf{w} = \mathbf{w}_q - \mathbf{C}_a^T \mathbf{w}_a \ , \tag{D.2}$$

with $\mathbf{w}_q$ the fixed (quiescent) part, equal to

$$\mathbf{w}_q = \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1} \mathbf{b} \ , \tag{D.3}$$

$\mathbf{w}_a$ the $(M-J)$-dimensional variable part, and $\mathbf{C}_a$ the $(M-J) \times M$-dimensional null-space of $\mathbf{C}$. The cost function $J_{LS}(\mathbf{w}) = \mathbf{w}^T \mathbf{Q}_{LS} \mathbf{w} - 2\mathbf{w}^T \mathbf{a} + d_{LS}$ can now be written as

$$\begin{aligned} J_{LS}(\mathbf{w}) &= (\mathbf{w}_q - \mathbf{C}_a^T \mathbf{w}_a)^T \mathbf{Q}_{LS} (\mathbf{w}_q - \mathbf{C}_a^T \mathbf{w}_a) - 2(\mathbf{w}_q - \mathbf{C}_a^T \mathbf{w}_a)^T \mathbf{a} + d_{LS} \\ &= \mathbf{w}_q^T \mathbf{Q}_{LS} \mathbf{w}_q - 2\mathbf{w}_a^T \mathbf{C}_a \mathbf{Q}_{LS} \mathbf{w}_q + \mathbf{w}_a^T \mathbf{C}_a \mathbf{Q}_{LS} \mathbf{C}_a^T \mathbf{w}_a - 2\mathbf{w}_q^T \mathbf{a} + \\ &\quad 2\mathbf{w}_a^T \mathbf{C}_a \mathbf{a} + d_{LS} \end{aligned}$$

and can be minimised by setting the derivate

$$\frac{\partial J_{LS}(\mathbf{w})}{\partial \mathbf{w}_a} = -2\mathbf{C}_a \mathbf{Q}_{LS} \mathbf{w}_q + 2\mathbf{C}_a \mathbf{Q}_{LS} \mathbf{C}_a^T \mathbf{w}_a + 2\mathbf{C}_a \mathbf{a} \tag{D.4}$$

equal to 0, yielding the solution

$$\mathbf{w}_a^{min} = (\mathbf{C}_a \mathbf{Q}_{LS} \mathbf{C}_a^T)^{-1} \mathbf{C}_a (\mathbf{Q}_{LS} \mathbf{w}_q - \mathbf{a}) \ , \tag{D.5}$$

such that the solution $\mathbf{w}_{LS}^c$ of the constrained minimisation problem is equal to

$$\mathbf{w}_{LS}^c = \mathbf{w}_q - \mathbf{C}_a^T \mathbf{w}_a^{min} = \mathbf{w}_q - \mathbf{C}_a^T (\mathbf{C}_a \mathbf{Q}_{LS} \mathbf{C}_a^T)^{-1} \mathbf{C}_a (\mathbf{Q}_{LS} \mathbf{w}_q - \mathbf{a}) \tag{D.6}$$

$$= \left[\mathbf{I}_M - \mathbf{C}_a^T (\mathbf{C}_a \mathbf{Q}_{LS} \mathbf{C}_a^T)^{-1} \mathbf{C}_a \mathbf{Q}_{LS}\right] (\mathbf{w}_q - \mathbf{Q}_{LS}^{-1} \mathbf{a}) + \mathbf{Q}_{LS}^{-1} \mathbf{a} \ . \tag{D.7}$$

In [139] it has been proved that

$$\mathbf{I}_M - \mathbf{C}_a^T (\mathbf{C}_a \mathbf{A} \mathbf{C}_a^T)^{-1} \mathbf{C}_a \mathbf{A} = \mathbf{A}^{-1} \mathbf{C}^T (\mathbf{C} \mathbf{A}^{-1} \mathbf{C}^T)^{-1} \mathbf{C} \ , \tag{D.8}$$

such that using (D.2) $\mathbf{w}_{LS}^c$ can be rewritten as

$$\mathbf{w}_{LS}^c = \mathbf{Q}_{LS}^{-1} \mathbf{C}^T (\mathbf{C} \mathbf{Q}_{LS}^{-1} \mathbf{C}^T)^{-1} \mathbf{C} \left[\mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1} \mathbf{b} - \mathbf{Q}_{LS}^{-1} \mathbf{a}\right] + \mathbf{Q}_{LS}^{-1} \mathbf{a} \ , \tag{D.9}$$

such that

$$\boxed{\mathbf{w}_{LS}^c = \mathbf{Q}_{LS}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{Q}_{LS}^{-1}\mathbf{C}^T)^{-1}(\mathbf{b} - \mathbf{C}\mathbf{Q}_{LS}^{-1}\mathbf{a}) + \mathbf{Q}_{LS}^{-1}\mathbf{a}} \tag{D.10}$$

which can be written in function of the unconstrained solution as

$$\mathbf{w}_{LS}^c = \mathbf{w}_{LS} + \mathbf{Q}_{LS}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{Q}_{LS}^{-1}\mathbf{C}^T)^{-1}(\mathbf{b} - \mathbf{C}\mathbf{w}_{LS}) \ . \tag{D.11}$$

D.2 Derivative constraints for near-field case

The first-order angle derivative $\mathbf{g}'_\theta(\omega, \theta, r)$ is equal to

$$\mathbf{g}'_\theta(\omega,\theta,r) = \frac{\partial \mathbf{g}(\omega,\theta,r)}{\partial\theta} = \frac{\partial}{\partial\theta}\begin{bmatrix} a_0(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_0(\theta,r)} \\ a_1(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_1(\theta,r)} \\ \vdots \\ a_{N-1}(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_{N-1}(\theta,r)} \end{bmatrix} , \tag{D.12}$$

with

$$\begin{aligned} \frac{\partial a_n(\theta,r)}{\partial\theta} &= \frac{\partial}{\partial\theta}\frac{r}{\sqrt{p_n + q_n\cos\theta}} = \frac{rq_n\sin\theta}{2(p_n+q_n\cos\theta)^{3/2}} = \frac{q_n\sin\theta}{2r_n^2(\theta,r)}a_n(\theta,r) \ , \\ \frac{\partial \tau_n(\theta,r)}{\partial\theta} &= \frac{\partial}{\partial\theta}\frac{\sqrt{p_n+q_n\cos\theta}}{c}f_s = \frac{-q_n\sin\theta}{2c\sqrt{p_n+q_n\cos\theta}}f_s = \frac{-q_n\sin\theta}{2r_n(\theta,r)c}f_s \ , \end{aligned}$$

and

$$\frac{\partial}{\partial\theta}a_n(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_n(\theta,r)} = a_n(\theta,r)\mathbf{e}(\omega)e^{-j\omega\tau_n(\theta,r)}\frac{q_n\sin\theta}{2r_n(\theta,r)}\left(\frac{1}{r_n(\theta,r)}+j\frac{\omega f_s}{c}\right),$$

such that the first-order angle derivative $\mathbf{g}'_\theta(\omega,\theta,r)$ can be written as

$$\mathbf{g}'_\theta(\omega,\theta,r) = \sin\theta\ \mathbf{\Delta}_\theta(\omega,\theta,r)\ \mathbf{g}(\omega,\theta,r) \ , \tag{D.13}$$

with $\mathbf{\Delta}_\theta(\omega,\theta,r)$ a complex-valued $M \times M$-dimensional diagonal matrix,

$$\begin{aligned} \mathbf{\Delta}_\theta(\omega,\theta,r) &= \begin{bmatrix} \frac{q_0}{2r_0^2(\theta,r)}\mathbf{I}_L & & & \\ & \frac{q_1}{2r_1^2(\theta,r)}\mathbf{I}_L & & \\ & & \ddots & \\ & & & \frac{q_{N-1}}{2r_{N-1}^2(\theta,r)}\mathbf{I}_L \end{bmatrix} + \\ & \quad j\frac{\omega f_s}{c}\begin{bmatrix} \frac{q_0}{2r_0(\theta,r)}\mathbf{I}_L & & & \\ & \frac{q_1}{2r_1(\theta,r)}\mathbf{I}_L & & \\ & & \ddots & \\ & & & \frac{q_{N-1}}{2r_{N-1}(\theta,r)}\mathbf{I}_L \end{bmatrix} \\ &= \mathbf{\Delta}_{\theta,R}(\theta,r) + j\mathbf{\Delta}_{\theta,I}(\omega,\theta,r) \ . \end{aligned} \tag{D.14}$$

Hence, $\mathbf{g}'_\theta(\omega,\theta,r)$ can be written as

$$\begin{aligned}\mathbf{g}'_\theta(\omega,\theta,r) &= \sin\theta\Big[\Big(\mathbf{\Delta}_{\theta,R}(\theta,r)\mathbf{g}_R(\omega,\theta,r)-\mathbf{\Delta}_{\theta,I}(\omega,\theta,r)\mathbf{g}_I(\omega,\theta,r)\Big)+\\ &\qquad j\Big(\mathbf{\Delta}_{\theta,I}(\omega,\theta,r)\mathbf{g}_R(\omega,\theta,r)+\mathbf{\Delta}_{\theta,R}(\theta,r)\mathbf{g}_I(\omega,\theta,r)\Big)\Big]\ . \quad \text{(D.15)}\end{aligned}$$

Remark D.1 For $r\to\infty$, i.e. far-field assumptions,

$$\lim_{r\to\infty}\frac{q_n}{2r_n(\theta,r)} = \lim_{r\to\infty}\frac{2rd_n}{2\sqrt{r^2+d_n^2+2rd_n\cos\theta}} = d_n \quad \text{(D.16)}$$

$$\lim_{r\to\infty}\frac{q_n}{2r_n^2(\theta,r)} = \lim_{r\to\infty}\frac{2rd_n}{2(r^2+d_n^2+2rd_n\cos\theta)} = 0\ , \quad \text{(D.17)}$$

such that $\mathbf{g}'_\theta(\omega,\theta,r)$ in (D.13) for the near-field case reduces to $\mathbf{g}'_\theta(\omega,\theta)$ in (8.88) for the far-field case. △

D.3 Expressions for robust non-linear criterion

Depending on the values of $\lfloor\frac{i-1}{L}\rfloor$, $\lfloor\frac{j-1}{L}\rfloor$, $\lfloor\frac{k-1}{L}\rfloor$ and $\lfloor\frac{l-1}{L}\rfloor$, different cases have to be considered:

- **4 equal values**: $\gamma_{ijkl}=0$

$$\delta^a_{ijkl}=\int_a a^4\, f_\alpha(a)\,da, \quad \delta^c_{\gamma,ijkl}=1, \quad \delta^s_{\gamma,ijkl}=0 \quad \text{(D.18)}$$

- **3 equal values, 1 different value**: $a_{ijkl}=a_1^3a_2$, $\gamma_{ijkl}=\pm(\gamma_1-\gamma_2)$

$$\delta^a_{ijkl} = \int_{a_1}\int_{a_2} a_1^3a_2\, f_\alpha(a_1)f_\alpha(a_2)\,da_1da_2 = \mu_a\int_a a^3\, f_\alpha(a)\,da \quad \text{(D.19)}$$

$$\delta^c_{\gamma,ijkl} = \int_{\gamma_1}\int_{\gamma_2}\cos\big[\pm\big(\gamma_1-\gamma_2\big)\big]\, f_{\mathcal{G}}(\gamma_1)f_{\mathcal{G}}(\gamma_2)d\gamma_1d\gamma_2 = \sigma^c_\gamma \quad \text{(D.20)}$$

$$\delta^s_{\gamma,ijkl} = \int_{\gamma_1}\int_{\gamma_2}\sin\big[\pm\big(\gamma_1-\gamma_2\big)\big]\, f_{\mathcal{G}}(\gamma_1)f_{\mathcal{G}}(\gamma_2)d\gamma_1d\gamma_2 = 0 \quad \text{(D.21)}$$

- **2 equal values, 2 equal values**: $a_{ijkl}=a_1^2a_2^2$

$$\delta^a_{ijkl}=\int_{a_1}\int_{a_2} a_1^2a_2^2\, f_\alpha(a_1)f_\alpha(a_2)\,da_1da_2=\sigma_a^4 \quad \text{(D.22)}$$

$\boxed{\lfloor\frac{i-1}{L}\rfloor=\lfloor\frac{k-1}{L}\rfloor\neq\lfloor\frac{j-1}{L}\rfloor=\lfloor\frac{l-1}{L}\rfloor}$: $\gamma_{ijkl}=2(\gamma_1-\gamma_2)$

$$\delta^c_{\gamma,ijkl} = \int_{\gamma_1}\int_{\gamma_2}\cos 2\big(\gamma_1-\gamma_2\big)\, f_{\mathcal{G}}(\gamma_1)f_{\mathcal{G}}(\gamma_2)d\gamma_1d\gamma_2$$

$$\begin{aligned} &= \int_{\gamma_1}\int_{\gamma_2} \big(\cos 2\gamma_1 \cos 2\gamma_2 + \sin 2\gamma_1 \sin 2\gamma_2\big)\, f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) d\gamma_1 d\gamma_2 \\ &= \big(\mu^c_{2\gamma}\big)^2 + \big(\mu^s_{2\gamma}\big)^2 \end{aligned} \tag{D.23}$$

$$\delta^s_{\gamma,ijkl} = \int_{\gamma_1}\int_{\gamma_2} \sin 2\big(\gamma_1 - \gamma_2\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) d\gamma_1 d\gamma_2 = 0 \tag{D.24}$$

with

$$\boxed{\mu^c_{2\gamma} = \int_{\gamma} \cos 2\gamma\, f_{\mathcal{G}}(\gamma) d\gamma, \qquad \mu^s_{2\gamma} = \int_{\gamma} \sin 2\gamma\, f_{\mathcal{G}}(\gamma) d\gamma} \tag{D.25}$$

$\boxed{\lfloor\frac{i-1}{L}\rfloor = \lfloor\frac{j-1}{L}\rfloor \neq \lfloor\frac{k-1}{L}\rfloor = \lfloor\frac{l-1}{L}\rfloor,\ \lfloor\frac{i-1}{L}\rfloor = \lfloor\frac{l-1}{L}\rfloor \neq \lfloor\frac{j-1}{L}\rfloor = \lfloor\frac{k-1}{L}\rfloor}$: $\gamma_{ijkl} = 0$

$$\delta^c_{\gamma,ijkl} = 1, \quad \delta^s_{\gamma,ijkl} = 0 \tag{D.26}$$

- **2 equal values, 2 different values**: $a_{ijkl} = a_1^2 a_2 a_3$

$$\delta^a_{ijkl} = \int_{a_1}\int_{a_2}\int_{a_3} a_1^2 a_2 a_3\, f_\alpha(a_1) f_\alpha(a_2) f_\alpha(a_3)\, da_1 da_2 da_3 = \sigma_a^2 \mu_a^2 \tag{D.27}$$

$\boxed{\lfloor\frac{i-1}{L}\rfloor = \lfloor\frac{k-1}{L}\rfloor \neq \lfloor\frac{j-1}{L}\rfloor \neq \lfloor\frac{l-1}{L}\rfloor}$: $\gamma_{ijkl} = 2\gamma_1 - \gamma_2 - \gamma_3$

$$\begin{aligned} \delta^c_{\gamma,ijkl} &= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3} \cos\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 \\ &= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3} \cos 2\gamma_1 \big(\cos\gamma_2 \cos\gamma_3 - \sin\gamma_2 \sin\gamma_3\big) + \sin 2\gamma_1 \cdot \\ &\quad \big(\sin\gamma_2 \cos\gamma_3 + \cos\gamma_2 \sin\gamma_3\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 \\ &= \mu^c_{2\gamma}\left[\big(\mu^c_\gamma\big)^2 - \big(\mu^s_\gamma\big)^2\right] + 2\mu^s_{2\gamma}\mu^c_\gamma\mu^s_\gamma = \bar{\delta}^c_\gamma \end{aligned} \tag{D.28}$$

$$\begin{aligned} \delta^s_{\gamma,ijkl} &= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3} \sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 \\ &= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3} \sin 2\gamma_1 \big(\cos\gamma_2 \cos\gamma_3 - \sin\gamma_2 \sin\gamma_3\big) - \cos 2\gamma_1 \cdot \\ &\quad \big(\sin\gamma_2 \cos\gamma_3 + \cos\gamma_2 \sin\gamma_3\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 \\ &= \mu^s_{2\gamma}\left[\big(\mu^c_\gamma\big)^2 - \big(\mu^s_\gamma\big)^2\right] - 2\mu^c_{2\gamma}\mu^c_\gamma\mu^s_\gamma = \bar{\delta}^s_\gamma \end{aligned} \tag{D.29}$$

$\boxed{\lfloor\frac{j-1}{L}\rfloor = \lfloor\frac{l-1}{L}\rfloor \neq \lfloor\frac{i-1}{L}\rfloor \neq \lfloor\frac{k-1}{L}\rfloor}$: $\gamma_{ijkl} = -2\gamma_1 + \gamma_2 + \gamma_3$

$$\delta^c_{\gamma,ijkl} = \bar{\delta}^c_\gamma, \quad \delta^s_{\gamma,ijkl} = -\bar{\delta}^s_\gamma \tag{D.30}$$

$\boxed{\text{all other cases}}$: $\gamma_{ijkl} = \pm(\gamma_1 - \gamma_2)$

$$\delta^c_{\gamma,ijkl} = \sigma^c_\gamma, \quad \delta^s_{\gamma,ijkl} = 0 \tag{D.31}$$

- **4 different values**: $a_{ijkl} = a_1 a_2 a_3 a_4$, $\gamma_{ijkl} = \gamma_1 - \gamma_2 + \gamma_3 - \gamma_4$

$$\begin{aligned}
\delta^a_{ijkl} &= \int_{a_1}\int_{a_2}\int_{a_3}\int_{a_4} a_1 a_2 a_3 a_4\, f_\alpha(a_1) f_\alpha(a_2) f_\alpha(a_3) f_\alpha(a_4)\, da_1 da_2 da_3 da_4 \\
&= \mu_a^4 \qquad\qquad \text{(D.32)}\\
\delta^c_{\gamma,ijkl} &= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3}\int_{\gamma_4} \cos\big(\gamma_1 - \gamma_2 + \gamma_3 - \gamma_4\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) f_{\mathcal{G}}(\gamma_4)\cdot \\
&\qquad d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4 \\
&= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3}\int_{\gamma_4} \big[\cos\big(\gamma_1-\gamma_2\big)\cos\big(\gamma_3-\gamma_4\big) - \sin\big(\gamma_1-\gamma_2\big)\cdot \\
&\qquad \sin\big(\gamma_3-\gamma_4\big)\big]\ f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) f_{\mathcal{G}}(\gamma_4) d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4 \\
&= \big(\sigma^c_\gamma\big)^2 \qquad\qquad \text{(D.33)}\\
\delta^s_{\gamma,ijkl} &= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3}\int_{\gamma_4} \sin\big(\gamma_1 - \gamma_2 + \gamma_3 - \gamma_4\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) f_{\mathcal{G}}(\gamma_4)\cdot \\
&\qquad d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4 \\
&= \int_{\gamma_1}\int_{\gamma_2}\int_{\gamma_3}\int_{\gamma_4} \big[\sin\big(\gamma_1-\gamma_2\big)\cos\big(\gamma_3-\gamma_4\big) + \cos\big(\gamma_1-\gamma_2\big)\cdot \\
&\qquad \sin\big(\gamma_3-\gamma_4\big)\big]\ f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) f_{\mathcal{G}}(\gamma_4) d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4 = 0 \qquad \text{(D.34)}
\end{aligned}$$

For a *symmetric phase pdf* $f_{\mathcal{G}}(\gamma)$, i.e. a function $f_{\mathcal{G}}(\gamma_c + \gamma) = f_{\mathcal{G}}(\gamma_c - \gamma)$, $\forall\gamma$, for a certain γ_c, it can easily be proved that $\bar{\delta}^s_\gamma = 0$, since

$$\begin{aligned}
\bar{\delta}^s_\gamma &= \int\int\int_{\gamma_c-\gamma_I}^{\gamma_c+\gamma_I} \sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 \\
&= \int\int\int_{-\gamma_I}^{0} \sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_c+\gamma_1) f_{\mathcal{G}}(\gamma_c+\gamma_2) f_{\mathcal{G}}(\gamma_c+\gamma_3)\cdot \\
&\qquad d\gamma_1 d\gamma_2 d\gamma_3 + \int\int\int_0^{\gamma_I} \sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_c+\gamma_1) f_{\mathcal{G}}(\gamma_c+\gamma_2)\cdot \\
&\qquad f_{\mathcal{G}}(\gamma_c+\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 \\
&= \int\int\int_{\gamma_I}^{0} -\sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_c-\gamma_1) f_{\mathcal{G}}(\gamma_c-\gamma_2) f_{\mathcal{G}}(\gamma_c-\gamma_3)\cdot \\
&\qquad (-d\gamma_1)(-d\gamma_2)(-d\gamma_3) + \int\int\int_0^{\gamma_I} \sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_c+\gamma_1)\cdot \\
&\qquad f_{\mathcal{G}}(\gamma_c+\gamma_2) f_{\mathcal{G}}(\gamma_c+\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 = 0\ ,
\end{aligned}$$

such that for $\gamma_I = \infty$ we obtain

$$\bar{\delta}^s_\gamma = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sin\big(2\gamma_1 - \gamma_2 - \gamma_3\big) f_{\mathcal{G}}(\gamma_1) f_{\mathcal{G}}(\gamma_2) f_{\mathcal{G}}(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3 = 0\ .$$

For a uniform distribution the phase parameters $\mu^c_{2\gamma}$ and $\mu^s_{2\gamma}$ are equal to

$$\mu^c_{2\gamma} = \frac{\sin 2\gamma_{max} - \sin 2\gamma_{min}}{2(\gamma_{max} - \gamma_{min})}, \qquad \mu^s_{2\gamma} = \frac{\cos 2\gamma_{min} - \cos 2\gamma_{max}}{2(\gamma_{max} - \gamma_{min})}, \tag{D.35}$$

whereas for a Gaussian distribution these parameters have to be calculated numerically.

D.4 Proof of Theorem 10.1

When considering only gain errors, the weighted LS cost function in (10.33) can be written as

$$J_{LS}(\mathbf{w}, \mathbf{A}) = \mathbf{w}^T \bar{\mathbf{Q}}_{LS} \mathbf{w} - 2\mathbf{w}^T \bar{\mathbf{a}} + d_{LS} , \tag{D.36}$$

with $\bar{\mathbf{a}} = \mathbf{A}_R\, \mathbf{a}$ and $\bar{\mathbf{Q}}_{LS} = \mathbf{A}_R \mathbf{Q}_{LS} \mathbf{A}_R$, and $\mathbf{A}_R$ equal to

$$\mathbf{A}_R = \begin{bmatrix} a_0\, \mathbf{I}_L & & & \\ & a_1\, \mathbf{I}_L & & \\ & & \ddots & \\ & & & a_{N-1}\, \mathbf{I}_L \end{bmatrix} . \tag{D.37}$$

The expression $\mathbf{w}^T \bar{\mathbf{Q}}_{LS} \mathbf{w}$ can be rewritten as

$$\mathbf{w}^T \bar{\mathbf{Q}}_{LS} \mathbf{w} = \mathbf{w}^T \mathbf{A}_R \mathbf{Q}_{LS} \mathbf{A}_R \mathbf{w} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a_m a_n \mathbf{w}_m^T \left[\mathbf{Q}_{LS}\right]_{mn} \mathbf{w}_n , \tag{D.38}$$

with $\left[\mathbf{Q}_{LS}\right]_{mn}$ an $L \times L$-dimensional sub-matrix of $\mathbf{Q}_{LS}$,

$$\left[\mathbf{Q}_{LS}\right]_{mn} = \mathbf{Q}_{LS}^{mL+1:(m+1)L\, ,\, nL+1:(n+1)L}, \quad m = 0 \ldots N-1, \;\; n = 0 \ldots N-1 .$$

If we substitute $\mathbf{w}_m^T \left[\mathbf{Q}_{LS}\right]_{mn} \mathbf{w}_n$ by $b_{mn}(\mathbf{w})$, then $\mathbf{w}^T \bar{\mathbf{Q}}_{LS} \mathbf{w}$ in (D.38) can be rewritten as

$$\mathbf{w}^T \bar{\mathbf{Q}}_{LS} \mathbf{w} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a_m a_n b_{mn}(\mathbf{w}) = \boldsymbol{\alpha}^T \mathbf{B}_{LS}(\mathbf{w}) \boldsymbol{\alpha} , \tag{D.39}$$

with $\boldsymbol{\alpha}$ an N-dimensional vector, consisting of the microphone gains,

$$\boldsymbol{\alpha} = \begin{bmatrix} a_0 & a_1 & \ldots & a_{N-1} \end{bmatrix}^T . \tag{D.40}$$

Similarly, if we define $c_n(\mathbf{w})$ by $\mathbf{w}_n^T \left[\mathbf{a}\right]_n$, with $\left[\mathbf{a}\right]_n$ an L-dimensional sub-vector of $\mathbf{a}$,

$$\left[\mathbf{a}\right]_n = \mathbf{a}^{nL+1:(n+1)L}, \; n = 0 \ldots N-1 , \tag{D.41}$$

then the weighted LS cost function can be written as

$$J_{LS}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \mathbf{B}_{LS}(\mathbf{w}) \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^T \mathbf{c} + d_{LS} . \tag{D.42}$$

Since $\mathbf{Q}_{LS}$ is a positive-(semi)definite matrix, $\mathbf{w}^T\mathbf{Q}_{LS}\mathbf{w} \geq 0$, $\forall\mathbf{w}$, such that

$$\mathbf{w}^T\bar{\mathbf{Q}}_{LS}\mathbf{w} = \mathbf{w}^T\mathbf{A}_R\mathbf{Q}_{LS}\mathbf{A}_R\mathbf{w} = \boldsymbol{\alpha}^T\mathbf{B}_{LS}(\mathbf{w})\boldsymbol{\alpha} \geq 0, \quad \forall\mathbf{w}, \forall\boldsymbol{\alpha} \tag{D.43}$$

and hence $\mathbf{B}_{LS}(\mathbf{w})$ is a positive-(semi)definite matrix for every $\mathbf{w}$. Therefore the weighted LS cost function $J_{LS}(\boldsymbol{\alpha})$ is a quadratic function (with a single minimum), such that the maximum value of $J_{LS}(\boldsymbol{\alpha})$ for all points inside an N-dimensional hypercube, defined by $a_{min} \leq a_n \leq a_{max}, n = 0 \dots N-1$, occurs on one of the 2^N boundary points of the hypercube.

Considering only gain errors, the TLS eigenfilter cost function in (10.22) can be written as

$$J_{TLS}(\mathbf{w}, \mathbf{A}) = \frac{\hat{\mathbf{w}}^T\hat{\bar{\mathbf{Q}}}_{TLS}\hat{\mathbf{w}}}{\hat{\mathbf{w}}^T\hat{\bar{\mathbf{Q}}}_e^{tot}\hat{\mathbf{w}}} = \frac{\hat{\mathbf{w}}^T\hat{\mathbf{A}}_R\hat{\mathbf{Q}}_{TLS}\hat{\mathbf{A}}_R\hat{\mathbf{w}}}{\hat{\mathbf{w}}^T\hat{\mathbf{A}}_R\hat{\mathbf{Q}}_e^{tot}\hat{\mathbf{A}}_R\hat{\mathbf{w}}} = \frac{\hat{\boldsymbol{\alpha}}^T\hat{\mathbf{B}}_{TLS}(\mathbf{w})\hat{\boldsymbol{\alpha}}}{\hat{\boldsymbol{\alpha}}^T\hat{\mathbf{B}}_e^{tot}(\mathbf{w})\hat{\boldsymbol{\alpha}}}\,, \tag{D.44}$$

with $\hat{\mathbf{Q}}_{TLS}$ and $\hat{\mathbf{Q}}_e^{tot}$ defined in (8.74) and

$$\hat{\mathbf{A}}_R = \begin{bmatrix} \mathbf{A}_R & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\alpha} \\ 1 \end{bmatrix}. \tag{D.45}$$

For this cost function, the maximum value of $\mathbf{F}(\mathbf{w})$ does *not* necessarily occur on one of the boundary points of the hypercube. This is also not necessarily the case for the non-linear cost function.

E Calculation of expressions for far-field broadband beamforming

In this appendix, the calculation of the following expressions is discussed:

- Appendix E.1 (WLS criterion) : weighted LS, TLS eigenfilter

$$\boxed{\int_{\Theta_p}\int_{\Omega_p} \mathrm{Re}\big\{H(\omega,\theta)\big\}d\omega d\theta = \mathbf{w}^T \cdot \int_{\Theta_p}\int_{\Omega_p} \mathbf{g}_R(\omega,\theta)d\omega d\theta = \mathbf{w}^T\mathbf{a}}$$

- Appendix E.2 (Energy criterion): weighted LS, conventional eigenfilter, TLS eigenfilter, maximum energy array, non-linear criterion

$$\boxed{\int_{\Theta}\int_{\Omega} |H(\omega,\theta)|^2 d\omega d\theta = \mathbf{w}^T \cdot \int_{\Theta}\int_{\Omega} \mathbf{G}_R(\omega,\theta)d\omega d\theta \cdot \mathbf{w} = \mathbf{w}^T\mathbf{Q}_e\mathbf{w}}$$

- Appendix E.3 (Passband error) : conventional eigenfilter

$$\boxed{\begin{array}{l} \displaystyle\int_{\Theta_p}\int_{\Omega_p} |H(\omega_c,\theta_c) - H(\omega,\theta)|^2 d\omega d\theta = \mathbf{w}^T\mathbf{Q}_p\mathbf{w} = \\ \displaystyle\mathbf{w}^T \cdot \int_{\Theta_p}\int_{\Omega_p} \mathrm{Re}\big\{[\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)][\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)]^H\big\}d\omega d\theta \cdot \mathbf{w} \end{array}}$$

- Appendix E.4 (Non-linear criterion) : non-linear criterion

$$\boxed{J_{sum}(\mathbf{w}) = \int_{\Theta}\int_{\Omega} |H(\omega,\theta)|^4 d\omega d\theta = \int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta}$$

E.1 WLS criterion

The vector $\mathbf{a}$ which needs to be calculated in the weighted LS cost function and the TLS eigenfilter is equal to

$$\mathbf{a} = \int_{\Theta_p}\int_{\Omega_p} \mathbf{g}_R(\omega,\theta)d\omega d\theta\ . \tag{E.1}$$

Using (8.7), the ith element of $\mathbf{g}_R(\omega,\theta)$ is equal to

$$\mathbf{g}_R^i(\omega,\theta) = \cos\left[\omega\Big(k + \frac{d_n\cos\theta}{c}f_s\Big)\right], \quad i = 1\dots M\ , \tag{E.2}$$

with

$$k = \mathrm{mod}(i-1,L) \qquad n = \lfloor\frac{i-1}{L}\rfloor\ . \tag{E.3}$$

The ith element of $\mathbf{a}$ therefore is equal to

$$\mathbf{a}^i = \int_{\Theta_p}\int_{\Omega_p} \mathbf{g}_R^i(\omega,\theta) d\omega d\theta = \int_{\Theta_p}\int_{\Omega_p} \cos\left[\omega\Big(k + \frac{d_n \cos\theta}{c} f_s\Big)\right] d\omega d\theta \,. \tag{E.4}$$

This integral can be considered to be a special case of the integral

$$\int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2} \cos\left[\omega\big(\alpha_i + \beta_i \cos\theta\big) + \gamma_i\right] d\omega d\theta \,, \tag{E.5}$$

with

$$\boxed{\alpha_i = k \qquad \beta_i = \frac{d_n}{c} f_s \qquad \gamma_i = 0} \tag{E.6}$$

Solving integrals of the type (E.5) is discussed in Appendix F. Since α_i can take on L distinct values and β_i can take on N distinct values, $LN = M$ different integrals need to be calculated.

E.2 Energy criterion

The energy (both in the stopband as in the passband) is defined as

$$\int_{\Theta}\int_{\Omega} |H(\omega,\theta)|^2 d\omega d\theta = \mathbf{w}^T \cdot \int_{\Theta}\int_{\Omega} \mathbf{G}(\omega,\theta) d\omega d\theta \cdot \mathbf{w} \,. \tag{E.7}$$

Using (8.7) and (8.9), the (i,j)-th element of $\mathbf{G}(\omega,\theta)$ is equal to

$$\mathbf{G}^{ij}(\omega,\theta) = \mathbf{g}^i(\omega,\theta)\mathbf{g}^j(\omega,\theta)^* = e^{-j\omega\left((k-l)+\frac{(d_n-d_m)\cos\theta}{c} f_s\right)}, \quad i,j = 1\ldots M \,,$$

with

$$k = \mathrm{mod}(i-1, L) \qquad n = \lfloor \frac{i-1}{L} \rfloor \tag{E.8}$$

$$l = \mathrm{mod}(j-1, L) \qquad m = \lfloor \frac{j-1}{L} \rfloor \,. \tag{E.9}$$

The (i,j)-th element of the real part $\mathbf{G}_R(\omega,\theta)$ and the imaginary part $\mathbf{G}_I(\omega,\theta)$ then are equal to

$$\mathbf{G}_R^{ij}(\omega,\theta) = [\mathbf{G}_R]_{nm}^{kl}(\omega,\theta) = \cos\left[\omega\Big((k-l) + \frac{(d_n-d_m)\cos\theta}{c} f_s\Big)\right] \tag{E.10}$$

$$\mathbf{G}_I^{ij}(\omega,\theta) = [\mathbf{G}_I]_{nm}^{kl}(\omega,\theta) = -\sin\left[\omega\Big((k-l) + \frac{(d_n-d_m)\cos\theta}{c} f_s\Big)\right] \,. \tag{E.11}$$

The real part $\mathbf{G}_R(\omega,\theta)$ is symmetric, since

$$\mathbf{G}_R^{ji}(\omega,\theta) = [\mathbf{G}_R]_{mn}^{lk}(\omega,\theta) = [\mathbf{G}_R]_{nm}^{kl}(\omega,\theta) = \mathbf{G}_R^{ij}(\omega,\theta) \,, \tag{E.12}$$

whereas the imaginary part $\mathbf{G}_I(\omega,\theta)$ is anti-symmetric, since

$$\mathbf{G}_I^{ji}(\omega,\theta) = \left[\mathbf{G}_I\right]_{mn}^{lk}(\omega,\theta) = -\left[\mathbf{G}_I\right]_{nm}^{kl}(\omega,\theta) = -\mathbf{G}_I^{ij}(\omega,\theta)\ . \tag{E.13}$$

The spatial directivity spectrum $|H(\omega,\theta)|^2$ can be written as

$$|H(\omega,\theta)|^2 = \mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w} = \mathbf{w}^T\mathbf{G}_R(\omega,\theta)\mathbf{w} + j\mathbf{w}^T\mathbf{G}_I(\omega,\theta)\mathbf{w} \tag{E.14}$$

Since $\mathbf{G}_I(\omega,\theta)$ is anti-symmetric, $\mathbf{w}^T\mathbf{G}_I(\omega,\theta)\mathbf{w} = 0$, such that the spatial directivity spectrum can be written as

$$\boxed{|H(\omega,\theta)|^2 = \mathbf{w}^T\mathbf{G}_R(\omega,\theta)\mathbf{w}} \tag{E.15}$$

which implies that $|H(\omega,\theta)|^2$ is symmetric in both variables ω and θ, i.e.

$$|H(\omega,\theta)|^2 = |H(-\omega,\theta)|^2 = |H(\omega,-\theta)|^2 = |H(-\omega,-\theta)|^2\ . \tag{E.16}$$

The energy criterion (E.7) can now be written as

$$\int_\Theta\int_\Omega |H(\omega,\theta)|^2 d\omega d\theta = \mathbf{w}^T \cdot \int_\Theta\int_\Omega \mathbf{G}_R(\omega,\theta)d\omega d\theta \cdot \mathbf{w} = \mathbf{w}^T\mathbf{Q}_e\mathbf{w}\ , \tag{E.17}$$

with the (i,j)-th element of $\mathbf{Q}_e$ is equal to

$$\mathbf{Q}_e^{ij} = \int_\Theta\int_\Omega \mathbf{G}_R^{ij}(\omega,\theta)d\omega d\theta = \int_\Theta\int_\Omega \cos\left[\omega\Big((k-l) + \frac{(d_n-d_m)\cos\theta}{c}f_s\Big)\right]d\omega d\theta. \tag{E.18}$$

This integral can be considered to be a special case of the integral

$$\int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2} \cos\left[\omega\big(\alpha_{ij} + \beta_{ij}\cos\theta\big) + \gamma_{ij}\right] d\omega d\theta\ , \tag{E.19}$$

with

$$\boxed{\alpha_{ij} = k-l \qquad \beta_{ij} = \frac{(d_n-d_m)}{c}f_s \qquad \gamma_{ij} = 0} \tag{E.20}$$

Solving integrals of the type (E.19) is discussed in Appendix F. Since α_{ij} can take on $2L-1$ distinct values and β_{ij} can take on N^2-N+1 distinct values (for the most general microphone configuration), $(2L-1)(N^2-N+1)$ different integrals need to be calculated. For a symmetric microphone array, β_{ij} can only take on $\frac{N^2}{2}+1$ (for even N) or $\frac{N^2-1}{2}+1$ (for odd N) distinct values, while for a uniform microphone array, β_{ij} can only take on $2N-1$ distinct values.

Quadratic energy constraint

A special case of the energy criterion is the quadratic energy constraint in the conventional eigenfilter technique in Section 8.4.1, where the matrix $\mathbf{Q}_e^{tot}$ is defined as

$$\mathbf{Q}_e^{tot} = \int_0^\pi\int_0^\pi \mathbf{G}_R(\omega,\theta)d\omega d\theta\ . \tag{E.21}$$

The (i,j)-th element of $\mathbf{Q}_e^{tot}$ is equal to

$$\mathbf{Q}_e^{tot,ij} = \int_0^{\pi}\int_0^{\pi} \cos\left[\omega\big(\alpha_{ij} + \beta_{ij}\cos\theta\big)\right] d\omega d\theta\ , \tag{E.22}$$

which can be further simplified. Since

$$\left\{\begin{array}{rcl} \displaystyle\int_0^{\pi} \sin(p\cos\theta)d\theta & = & 0 \\ \displaystyle\int_0^{\pi} \cos(p\cos\theta)d\theta & = & \pi J_0(p)\ , \end{array}\right. \tag{E.23}$$

with $J_r(\cdot)$ the Bessel function of the first kind of order r, we can write

$$\int_0^{\pi} \cos\left[\omega\big(\alpha_{ij} + \beta_{ij}\cos\theta\big)\right] d\theta = \int_0^{\pi} \cos(\beta_{ij}\omega\cos\theta + \alpha_{ij}\omega)d\theta \tag{E.24}$$

$$= \int_0^{\pi} \cos(\beta_{ij}\omega\cos\theta)\cos(\alpha_{ij}\omega)d\theta - \underbrace{\int_0^{\pi} \sin(\beta_{ij}\omega\cos\theta)\sin(\alpha_{ij}\omega)d\theta}_{0} \tag{E.25}$$

$$= \pi J_0(\beta_{ij}\omega)\cos(\alpha_{ij}\omega)\ , \tag{E.26}$$

such that

$$\mathbf{Q}_e^{tot,ij} = \pi\int_0^{\pi} J_0(\beta_{ij}\omega)\cos(\alpha_{ij}\omega)d\omega\ , \tag{E.27}$$

which has to be integrated numerically.

E.3 Passband error

The passband error, used in the conventional eigenfilter technique, is defined as

$$\int_{\Theta_p}\int_{\Omega_p} |H(\omega_c,\theta_c) - H(\omega,\theta)|^2 d\omega d\theta\ . \tag{E.28}$$

The integrand $|H(\omega_c,\theta_c) - H(\omega,\theta)|^2$ of (E.28) can be written as

$$\begin{aligned} |H(\omega_c,\theta_c) - H(\omega,\theta)|^2 &= |\mathbf{w}^T\mathbf{g}(\omega_c,\theta_c) - \mathbf{w}^T\mathbf{g}(\omega,\theta)|^2 && \text{(E.29)} \\ &= \mathbf{w}^T\left[\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)\right]\left[\mathbf{g}(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)\right]^H \mathbf{w} \\ &= \mathbf{w}^T\big[\mathbf{g}(\omega_c,\theta_c)\mathbf{g}^H(\omega_c,\theta_c) - \mathbf{g}(\omega,\theta)\mathbf{g}^H(\omega_c,\theta_c) \\ &\quad - \mathbf{g}(\omega_c,\theta_c)\mathbf{g}^H(\omega,\theta) + \mathbf{g}(\omega,\theta)\mathbf{g}^H(\omega,\theta)\big]\,\mathbf{w}\ . && \text{(E.30)} \end{aligned}$$

If we define $\tilde{\mathbf{G}}(\omega_1,\theta_1,\omega_2,\theta_2)$ as

$$\tilde{\mathbf{G}}(\omega_1,\theta_1,\omega_2,\theta_2) = \mathbf{g}(\omega_1,\theta_1)\mathbf{g}^H(\omega_2,\theta_2)\ , \tag{E.31}$$

then we can write (E.30) as

$$\mathbf{w}^T \underbrace{\left[\mathbf{G}(\omega_c,\theta_c) - \tilde{\mathbf{G}}(\omega,\theta,\omega_c,\theta_c) - \tilde{\mathbf{G}}(\omega_c,\theta_c,\omega,\theta) + \mathbf{G}(\omega,\theta)\right]}_{\hat{\mathbf{G}}(\omega_c,\theta_c,\omega,\theta)=\hat{\mathbf{G}}(\omega,\theta,\omega_c,\theta_c)} \mathbf{w}\ . \tag{E.32}$$

The (i,j)-th element of $\tilde{\mathbf{G}}(\omega_1,\theta_1,\omega_2,\theta_2)$ is equal to

$$\tilde{\mathbf{G}}^{ij}(\omega_1,\theta_1,\omega_2,\theta_2) = e^{-j\omega_1\left(k+\frac{d_n\cos\theta_1}{c}f_s\right)} \cdot e^{j\omega_2\left(l+\frac{d_m\cos\theta_2}{c}f_s\right)} \ , \tag{E.33}$$

with

$$k = \mathrm{mod}(i-1,L) \qquad n = \lfloor\frac{i-1}{L}\rfloor \tag{E.34}$$

$$l = \mathrm{mod}(j-1,L) \qquad m = \lfloor\frac{j-1}{L}\rfloor \ . \tag{E.35}$$

The matrix $\hat{\mathbf{G}}(\omega_c,\theta_c,\omega,\theta)$ is complex Hermitian, since $\hat{\mathbf{G}}^{ji}(\omega_c,\theta_c,\omega,\theta)^* = \hat{\mathbf{G}}^{ij}(\omega_c,\theta_c,\omega,\theta)$, because

$$\begin{cases} \mathbf{G}^{ji}(\omega_c,\theta_c)^* = e^{j\omega_c\left(l+\frac{d_m\cos\theta_c}{c}f_s\right)} \cdot e^{-j\omega_c\left(k+\frac{d_n\cos\theta_c}{c}f_s\right)} = \mathbf{G}^{ij}(\omega_c,\theta_c) \\ \tilde{\mathbf{G}}^{ji}(\omega,\theta,\omega_c,\theta_c)^* = e^{j\omega\left(l+\frac{d_m\cos\theta}{c}f_s\right)} \cdot e^{-j\omega_c\left(k+\frac{d_n\cos\theta_c}{c}f_s\right)} = \tilde{\mathbf{G}}^{ij}(\omega_c,\theta_c,\omega,\theta) \\ \tilde{\mathbf{G}}^{ji}(\omega_c,\theta_c,\omega,\theta)^* = e^{j\omega_c\left(l+\frac{d_m\cos\theta_c}{c}f_s\right)} \cdot e^{-j\omega\left(k+\frac{d_n\cos\theta}{c}f_s\right)} = \tilde{\mathbf{G}}^{ij}(\omega,\theta,\omega_c,\theta_c) \\ \mathbf{G}^{ji}(\omega,\theta)^* = e^{j\omega\left(l+\frac{d_m\cos\theta}{c}f_s\right)} \cdot e^{-j\omega\left(k+\frac{d_n\cos\theta}{c}f_s\right)} = \mathbf{G}^{ij}(\omega,\theta) \ . \end{cases}$$

This implies that the real part $\hat{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta)$ is symmetric and the imaginary part $\hat{\mathbf{G}}_I(\omega_c,\theta_c,\omega,\theta)$ is anti-symmetric, such that $\mathbf{w}^T\hat{\mathbf{G}}_I(\omega_c,\theta_c,\omega,\theta)\mathbf{w} = 0$, and (E.30) can be written as

$$\boxed{\begin{aligned} |H(\omega_c,\theta_c) - H(\omega,\theta)|^2 &= \mathbf{w}^T\hat{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta)\mathbf{w} \\ &= \mathbf{w}^T\Big[\mathbf{G}_R(\omega_c,\theta_c) - \tilde{\mathbf{G}}_R(\omega,\theta,\omega_c,\theta_c) \\ &\quad - \tilde{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta) + \mathbf{G}_R(\omega,\theta)\Big]\mathbf{w} \ . \end{aligned}} \tag{E.36}$$

The (i,j)-th element of $\tilde{\mathbf{G}}_R(\omega_1,\theta_1,\omega_2,\theta_2)$ is equal to

$$\tilde{\mathbf{G}}_R^{ij}(\omega_1,\theta_1,\omega_2,\theta_2) = \cos\left[\omega_1\Big(k+\frac{d_n\cos\theta_1}{c}f_s\Big) - \omega_2\Big(l+\frac{d_m\cos\theta_2}{c}f_s\Big)\right] \ , \tag{E.37}$$

such that the following symmetry properties hold,

$$\begin{aligned} \tilde{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta) &= \tilde{\mathbf{G}}_R(-\omega_c,\theta_c,-\omega,\theta) \\ &= \tilde{\mathbf{G}}_R(\omega_c,\pm\theta_c,\omega,\pm\theta) = \tilde{\mathbf{G}}_R(-\omega_c,\pm\theta_c,-\omega,\pm\theta) \ . \end{aligned}$$

The passband error (E.28) can now be written as

$$\begin{aligned} \int_{\Theta_p}\int_{\Omega_p} |H(\omega_c,\theta_c) - H(\omega,\theta)|^2 d\omega d\theta &= \mathbf{w}^T \cdot \int_{\Theta_p}\int_{\Omega_p} \hat{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta) d\omega d\theta \cdot \mathbf{w} \\ &= \mathbf{w}^T\mathbf{Q}_p\mathbf{w} \ . \end{aligned} \tag{E.38}$$

The (i, j)-th element of $\mathbf{Q}_p$ is equal to

$$\begin{aligned}
\mathbf{Q}_p^{ij} &= \int_{\Theta_p}\int_{\Omega_p} \hat{\mathbf{G}}_R^{ij}(\omega_c, \theta_c, \omega, \theta)d\omega d\theta \qquad \text{(E.39)}\\
&= \int_{\Theta_p}\int_{\Omega_p} \cos\left[\omega_c\Big((k-l) + \frac{(d_n - d_m)\cos\theta_c}{c}f_s\Big)\right] d\omega d\theta\\
&\quad - \int_{\Theta_p}\int_{\Omega_p} \cos\left[\omega\Big(k + \frac{d_n\cos\theta}{c}f_s\Big) - \omega_c\Big(l + \frac{d_m\cos\theta_c}{c}f_s\Big)\right] d\omega d\theta\\
&\quad - \int_{\Theta_p}\int_{\Omega_p} \cos\left[\omega\Big(l + \frac{d_m\cos\theta}{c}f_s\Big) - \omega_c\Big(k + \frac{d_n\cos\theta_c}{c}f_s\Big)\right] d\omega d\theta\\
&\quad + \int_{\Theta_p}\int_{\Omega_p} \cos\left[\omega\Big((k-l) + \frac{(d_n - d_m)\cos\theta}{c}f_s\Big)\right] d\omega d\theta\ . \qquad \text{(E.40)}
\end{aligned}$$

All these integrals can again be considered to be special cases of the integral

$$\int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2} \cos\big[\omega\big(\alpha_{ij} + \beta_{ij}\cos\theta\big) + \gamma_{ij}\big]\, d\omega d\theta\ , \qquad \text{(E.41)}$$

with

$\alpha_{ij} = 0$ $\beta_{ij} = 0$ $\gamma_{ij} = \omega_c\Big((k-l) + \frac{(d_n - d_m)\cos\theta_c}{c}f_s\Big)$	$\alpha_{ij} = k$ $\beta_{ij} = \frac{d_n}{c}f_s$ $\gamma_{ij} = -\omega_c\Big(l + \frac{d_m\cos\theta_c}{c}f_s\Big)$
$\alpha_{ij} = l$ $\beta_{ij} = \frac{d_m}{c}f_s$ $\gamma_{ij} = -\omega_c\Big(k + \frac{d_n\cos\theta_c}{c}f_s\Big)$	$\alpha_{ij} = k - l$ $\beta_{ij} = \frac{d_n - d_m}{c}f_s$ $\gamma_{ij} = 0$

(E.42)

Solving integrals of the type (E.41) is discussed in Appendix F. For computing the passband error $2(2L-1)(N^2 - N + 1) + 2(LN)^2$ different integrals need to be calculated.

E.4 Non-linear criterion

Calculation of the cost function $J_{abs}(\mathbf{w})$

For calculating the non-linear cost function $\bar{J}_{NL}(\mathbf{w})$ in (8.31), the criterion

$$J_{abs}(\mathbf{w}) = 2\int_{\Theta}\int_{\Omega} |H(\omega, \theta)| d\omega d\theta \qquad \text{(E.43)}$$

needs to be computed (assuming $F(\omega, \theta) = 1$ and $|D(\omega, \theta)| = 1$ without loss of generality). Using (E.15), the integrand $|H(\omega, \theta)|$ can be written as

$$|H(\omega, \theta)| = \sqrt{\mathbf{w}^T\mathbf{G}_R(\omega, \theta)\mathbf{w}} = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \mathbf{G}_R^{ij}(\omega, \theta)}\ , \qquad \text{(E.44)}$$

with w_i the ith element of $\mathbf{w}$. Using (E.10) and (E.20), $J_{abs}(\mathbf{w})$ can be written as

$$J_{abs}(\mathbf{w}) = 2\int_{\Theta}\int_{\Omega}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \cos\left[\omega\big(\alpha_{ij} + \beta_{ij}\cos\theta\big)\right]}d\omega d\theta\ . \tag{E.45}$$

Because of the square root, the filter coefficients can not be extracted from the double integral, and *for every* $\mathbf{w}$ *the double integrals need to be recomputed numerically.*

Calculation of cost function $J_{sum}(\mathbf{w})$

For calculating the non-linear cost function $J_{NL}(\mathbf{w})$ in (8.35), the criterion

$$J_{sum}(\mathbf{w}) = \int_{\Theta}\int_{\Omega}|H(\omega,\theta)|^4 d\omega d\theta = \int_{\Theta}\int_{\Omega}\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta \tag{E.46}$$

needs to be calculated (both for the passband and for the stopband). In this section we will show that is possible to calculate $J_{sum}(\mathbf{w})$ *without having to recalculate double integrals for every* $\mathbf{w}$.

Using (8.8) and (8.9), the integrand $|H(\omega,\theta)|^4$ can be written as

$$\begin{aligned}
|H(\omega,\theta)|^4 &= \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big) && \text{(E.47)}\\
&= \Big(\sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \mathbf{G}^{ij}(\omega,\theta)\Big)\Big(\sum_{k=1}^{M}\sum_{l=1}^{M} w_k w_l \mathbf{G}^{kl}(\omega,\theta)\Big) && \text{(E.48)}\\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l\, \mathbf{g}^i(\omega,\theta)\mathbf{g}^j(\omega,\theta)^*\, \mathbf{g}^k(\omega,\theta)\mathbf{g}^l(\omega,\theta)^* \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l\, e^{-j\omega\big(\alpha_{ijkl}+\beta_{ijkl}\cos\theta\big)}\ , && \text{(E.49)}
\end{aligned}$$

with

$$\boxed{\begin{aligned}
\alpha_{ijkl} &= \mathrm{mod}(i-1,L) - \mathrm{mod}(j-1,L) + \mathrm{mod}(k-1,L) - \mathrm{mod}(l-1,L)\\
\beta_{ijkl} &= \frac{f_s}{c}\Big(d_{\lfloor\frac{i-1}{L}\rfloor} - d_{\lfloor\frac{j-1}{L}\rfloor} + d_{\lfloor\frac{k-1}{L}\rfloor} - d_{\lfloor\frac{l-1}{L}\rfloor}\Big)
\end{aligned}} \tag{E.50}$$

Since $|H(\omega,\theta)|^4$ is real (and the filter coefficients are real), only the real part of the exponential function has to be considered, such that

$$|H(\omega,\theta)|^4 = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l\ \cos\left[\omega\big(\alpha_{ijkl} + \beta_{ijkl}\cos\theta\big)\right]\ , \tag{E.51}$$

and $J_{sum}(\mathbf{w})$ can be written as

$$\boxed{J_{sum}(\mathbf{w}) = \int_{\Theta}\int_{\Omega} |H(\omega,\theta)|^4 d\omega d\theta = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l \rho_{ijkl}} \tag{E.52}$$

with

$$\rho_{ijkl} = \int_{\Theta}\int_{\Omega} \cos\left[\omega\big(\alpha_{ijkl} + \beta_{ijkl}\cos\theta\big)\right] d\omega d\theta \; . \tag{E.53}$$

These integrals are discussed in Appendix F and only need to be computed once (since ρ_{ijkl} is independent of $\mathbf{w}$).

As can be seen, α_{ijkl} can take on $4L-3$ distinct values $(-4L+2\ldots 4L-2)$ and since

$$\beta_{ijkl} = \beta_{kjil} = \beta_{ilkj} = \beta_{klij} \; , \tag{E.54}$$

β_{ijkl} can only take on $\frac{N^4-2N^3+7N^2-6N}{4}+1$ distinct values for the most general microphone configuration. Moreover, since $\alpha_{jilk} = -\alpha_{ijkl}$ and $\beta_{jilk} = -\beta_{ijkl}$, the following symmetry properties hold for ρ_{ijkl},

$$\rho_{ijkl} = \rho_{kjil} = \rho_{ilkj} = \rho_{klij} = \rho_{jilk} = \rho_{lijk} = \rho_{jkli} = \rho_{lkji} \; . \tag{E.55}$$

Therefore $(2L-2)\big(\frac{N^4-2N^3+7N^2-6N}{4}+1\big)$ different integrals ρ_{ijkl} need to be computed.

Remark E.1 Although $|H(\omega,\theta)|^4$ is equal to $(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w})(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w})$, it is not possible to represent $J_{sum}(\mathbf{w})$ in the form $(\mathbf{w}^T\mathbf{A}\mathbf{w})(\mathbf{w}^T\mathbf{B}\mathbf{w})$, since it is not possible to write ρ_{ijkl} as $a_{ij}\cdot b_{kl}$. △

Calculation of $\bar{J}_{sum}(\mathbf{w})$ for omni-directional, frequency-flat microphones

When taking into account the microphone characteristics (cf. Section 10.2.2), the criterion

$$\bar{J}_{sum}(\mathbf{w}) = \int_{\Theta}\int_{\Omega} F(\omega,\theta)|H(\omega,\theta)|^4 d\omega d\theta = \int_{\Theta}\int_{\Omega} F(\omega,\theta)\big(\mathbf{w}^T\bar{\mathbf{G}}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta$$

needs to be calculated. Using (10.8), the (i,j)-th element of $\bar{\mathbf{G}}(\omega,\theta)$ for omni-directional, frequency-flat microphones is equal to

$$\begin{aligned} \bar{\mathbf{G}}^{ij}(\omega,\theta) &= a_n a_m e^{-j(\psi_n-\psi_m)}\mathbf{G}^{ij}(\omega,\theta) && \text{(E.56)} \\ &= a_n a_m e^{-j\left[\omega\left((k-l)+\frac{(d_n-d_m)\cos\theta}{c}f_s\right)+(\psi_n-\psi_m)\right]} \; . && \text{(E.57)} \end{aligned}$$

Hence, the expression $|H(\omega,\theta)|^4$ can be written as

$$|H(\omega,\theta)|^4 = \big(\mathbf{w}^T\bar{\mathbf{G}}(\omega,\theta)\mathbf{w}\big)\big(\mathbf{w}^T\bar{\mathbf{G}}(\omega,\theta)\mathbf{w}\big) \tag{E.58}$$

$$= \Bigg(\sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \bar{\mathbf{G}}^{ij}(\omega,\theta)\Bigg)\Bigg(\sum_{k=1}^{M}\sum_{l=1}^{M} w_k w_l \bar{\mathbf{G}}^{kl}(\omega,\theta)\Bigg) \tag{E.59}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l\, a_{ijkl}\, e^{-j\left[\omega\left(\alpha_{ijkl}+\beta_{ijkl}\cos\theta\right)+\psi_{ijkl}\right]}, \tag{E.60}$$

with α_{ijkl} and β_{ijkl} defined in (E.50) and

$$\boxed{\begin{aligned} a_{ijkl} &= a_{\lfloor\frac{i-1}{L}\rfloor}\cdot a_{\lfloor\frac{j-1}{L}\rfloor}\cdot a_{\lfloor\frac{k-1}{L}\rfloor}\cdot a_{\lfloor\frac{l-1}{L}\rfloor} \\ \psi_{ijkl} &= \psi_{\lfloor\frac{i-1}{L}\rfloor} - \psi_{\lfloor\frac{j-1}{L}\rfloor} + \psi_{\lfloor\frac{k-1}{L}\rfloor} - \psi_{\lfloor\frac{l-1}{L}\rfloor} \end{aligned}} \tag{E.61}$$

Since $|H(\omega,\theta)|^4$ is real (and the filter coefficients are real), only the real part of the exponential function in (E.60) has to be considered, i.e.

$$\begin{aligned} |H(\omega,\theta)|^4 \;=\; & \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l\, a_{ijkl}\Big(\cos\big[\omega\big(\alpha_{ijkl}+\beta_{ijkl}\cos\theta\big)\big]\cdot \\ & \cos\psi_{ijkl} - \sin\big[\omega\big(\alpha_{ijkl}+\beta_{ijkl}\cos\theta\big)\big]\cdot\sin\psi_{ijkl}\Big)\,. \end{aligned} \tag{E.62}$$

Hence, $\bar{J}_{sum}(\mathbf{w})$ can be written as

$$\boxed{\bar{J}_{sum}(\mathbf{w}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l\, \underbrace{a_{ijkl}\Big(\cos\psi_{ijkl}\cdot\rho_{ijkl} - \sin\psi_{ijkl}\cdot\rho^{\circ}_{ijkl}\Big)}_{\bar{\rho}_{ijkl}}} \tag{E.63}$$

with

$$\rho_{ijkl} \;=\; \int_{\Theta}\int_{\Omega} F(\omega,\theta)\cos\big[\omega\big(\alpha_{ijkl}+\beta_{ijkl}\cos\theta\big)\big]\,d\omega d\theta \tag{E.64}$$

$$\rho^{\circ}_{ijkl} \;=\; \int_{\Theta}\int_{\Omega} F(\omega,\theta)\sin\big[\omega\big(\alpha_{ijkl}+\beta_{ijkl}\cos\theta\big)\big]\,d\omega d\theta\,. \tag{E.65}$$

The calculation of these integrals is discussed in Appendix F.

Calculation of gradient $\frac{\partial J_{sum}(\mathbf{w})}{\partial \mathbf{w}}$

In many optimisation techniques the gradient of the cost function is required. This gradient can either be approximated numerically or can be supplied in analytical form, which is more robust. The gradient $\frac{\partial J_{sum}(\mathbf{w})}{\partial \mathbf{w}}$ is equal to

$$\frac{\partial J_{sum}(\mathbf{w})}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial J_{sum}(\mathbf{w})}{\partial w_1} \\ \frac{\partial J_{sum}(\mathbf{w})}{\partial w_2} \\ \vdots \\ \frac{\partial J_{sum}(\mathbf{w})}{\partial w_M} \end{bmatrix}, \tag{E.66}$$

and can be calculated by taking the derivative of $J_{sum}(\mathbf{w})$ in (E.46), i.e.

$$\begin{aligned}\frac{\partial J_{sum}(\mathbf{w})}{\partial \mathbf{w}} &= \int_{\Theta}\int_{\Omega} \frac{\partial}{\partial \mathbf{w}} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)^2 d\omega d\theta && \text{(E.67)}\\ &= 2\int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\big(\mathbf{G}(\omega,\theta)+\mathbf{G}^T(\omega,\theta)\big)\mathbf{w} d\omega d\theta && \text{(E.68)}\\ &= 4\int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}_R(\omega,\theta) d\omega d\theta \cdot \mathbf{w}\ , && \text{(E.69)}\end{aligned}$$

such that

$$\boxed{\frac{\partial J_{sum}(\mathbf{w})}{\partial \mathbf{w}} = 4\mathbf{Q}_{sum}(\mathbf{w})\cdot\mathbf{w}} \tag{E.70}$$

with

$$\mathbf{Q}_{sum}(\mathbf{w}) = \mathrm{Re}\Big\{\int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}(\omega,\theta) d\omega d\theta\Big\}\ . \tag{E.71}$$

The (m,n)-th element of $\mathbf{Q}_{sum}(\mathbf{w})$ is equal to

$$\begin{aligned}\mathbf{Q}_{sum}^{mn}(\mathbf{w}) &= \mathrm{Re}\Big\{\int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}^{mn}(\omega,\theta) d\omega d\theta\Big\} && \text{(E.72)}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \int_{\Theta}\int_{\Omega} \mathrm{Re}\big\{\mathbf{g}^i(\omega,\theta)\mathbf{g}^j(\omega,\theta)^*\mathbf{g}^m(\omega,\theta)\mathbf{g}^n(\omega,\theta)^*\big\} d\omega d\theta \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \rho_{ijmn}\ , && \text{(E.73)}\end{aligned}$$

such that the nth element of the gradient $\frac{\partial J_{sum}(\mathbf{w})}{\partial \mathbf{w}}$ can be computed as

$$\frac{\partial J_{sum}(\mathbf{w})}{\partial w_n} = 4\sum_{k=1}^{M} \mathbf{Q}_{sum}^{kn}(\mathbf{w})\, w_k\ , \tag{E.74}$$

which can eventually be written as

$$\boxed{\frac{\partial J_{sum}(\mathbf{w})}{\partial w_n} = 4\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} w_i w_j w_k \rho_{ijkn}} \tag{E.75}$$

The matrix $\mathbf{Q}_{sum}(\mathbf{w})$ is a symmetric positive-definite matrix, since

$$\mathbf{Q}_{sum}^{nm}(\mathbf{w}) = \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \rho_{ijnm} = \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \rho_{ijmn} = \mathbf{Q}_{sum}^{mn}(\mathbf{w})\ , \tag{E.76}$$

and since

$$\mathbf{Q}_{sum}(\mathbf{w}) = \int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}_R(\omega,\theta) d\omega d\theta \tag{E.77}$$

is the integral, i.e. the infinite summation of the positive-definite matrices $\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}_R(\omega,\theta)$. Only for $\mathbf{w}=\mathbf{0}$, the matrix $\mathbf{Q}_{sum}(\mathbf{w})=\mathbf{0}$ becomes positive semi-definite.

Calculation of Hessian $\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}}$

For large-scale (constrained) optimisation methods, it is advisable also to provide the Hessian of the cost function, which is a symmetric matrix defined as

$$\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}} = \begin{bmatrix} \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 w_1} & \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_1 \partial w_M} \\ \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_2 \partial w_1} & \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 w_2} & \cdots & \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_2 \partial w_M} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_M \partial w_1} & \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_M \partial w_2} & \cdots & \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 w_M} \end{bmatrix} . \tag{E.78}$$

The Hessian can be calculated by taking the derivative of $\frac{\partial J_{sum}(\mathbf{w})}{\partial\mathbf{w}}$ in (E.68),

$$\begin{aligned} \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}} &= 2\int_\Theta\int_\Omega \frac{\partial}{\partial\mathbf{w}}\big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\big(\mathbf{G}(\omega,\theta)+\mathbf{G}^T(\omega,\theta)\big)\mathbf{w}d\omega d\theta \\ &= 2\int_\Theta\int_\Omega \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\big(\mathbf{G}(\omega,\theta)+\mathbf{G}^T(\omega,\theta)\big)d\omega d\theta + \\ &\quad 2\int_\Theta\int_\Omega \big(\mathbf{G}(\omega,\theta)+\mathbf{G}^T(\omega,\theta)\big)\mathbf{w}\mathbf{w}^T\big(\mathbf{G}(\omega,\theta)+\mathbf{G}^T(\omega,\theta)\big)d\omega d\theta \\ &= 4\,\mathrm{Re}\Big\{\int_\Theta\int_\Omega \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}(\omega,\theta)+\mathbf{G}(\omega,\theta)\mathbf{w}\mathbf{w}^T\mathbf{G}(\omega,\theta)+ \\ &\quad \mathbf{G}^T(\omega,\theta)\mathbf{w}\mathbf{w}^T\mathbf{G}(\omega,\theta)d\omega d\theta\Big\}, \end{aligned} \tag{E.79}$$

such that the (m,n)-th element of $\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2\mathbf{w}}$ is equal to

$$\begin{aligned} \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_m \partial w_n} &= 4\,\mathrm{Re}\Big\{\int_\Theta\int_\Omega \Big(\sum_{i=1}^M\sum_{j=1}^M w_i w_j \mathbf{G}^{ij}(\omega,\theta)\Big)\mathbf{G}^{mn}(\omega,\theta)d\omega d\theta\Big\} + \\ &\quad 4\,\mathrm{Re}\Big\{\int_\Theta\int_\Omega \Big(\sum_{i=1}^M w_i \mathbf{G}^{mi}(\omega,\theta)\Big)\Big(\sum_{j=1}^M w_j \mathbf{G}^{jn}(\omega,\theta)\Big)d\omega d\theta\Big\} + \\ &\quad 4\,\mathrm{Re}\Big\{\int_\Theta\int_\Omega \Big(\sum_{i=1}^M w_i \mathbf{G}^{im}(\omega,\theta)\Big)\Big(\sum_{j=1}^M w_j \mathbf{G}^{jn}(\omega,\theta)\Big)d\omega d\theta\Big\} \\ &= 4\sum_{i=1}^M\sum_{j=1}^M w_i w_j\big(\rho_{ijmn}+\rho_{mijn}+\rho_{imjn}\big), \end{aligned} \tag{E.80}$$

which can eventually be written, using (E.55), as

$$\boxed{\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_m \partial w_n} = 4\sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \big(2\rho_{ijmn} + \rho_{imjn}\big)} \tag{E.81}$$

The Hessian $\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 \mathbf{w}}$ is a positive-definite matrix, since

$$\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 \mathbf{w}} = 4\int_\Theta\int_\Omega \big(\mathbf{w}^T\mathbf{G}(\omega,\theta)\mathbf{w}\big)\mathbf{G}_R(\omega,\theta) + 2\mathbf{G}_R(\omega,\theta)\mathbf{w}\mathbf{w}^T\mathbf{G}_R(\omega,\theta)d\omega d\theta \tag{E.82}$$

is the integral, i.e. infinite summation, of positive-definite matrices. It can easily be shown that $\mathbf{G}_R(\omega,\theta)\mathbf{w}\mathbf{w}^T\mathbf{G}_R(\omega,\theta)$ is a positive-definite matrix, since

$$\bar{\mathbf{w}}^T\mathbf{G}_R(\omega,\theta)\mathbf{w}\mathbf{w}^T\mathbf{G}_R(\omega,\theta)\bar{\mathbf{w}} = \big(\bar{\mathbf{w}}^T\mathbf{G}_R(\omega,\theta)\mathbf{w}\big)^2 > 0, \quad \forall \mathbf{w}, \bar{\mathbf{w}} \neq \mathbf{0}\,. \tag{E.83}$$

In order to perform a local stability analysis, the quadratic form $\mathbf{w}^T\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 \mathbf{w}}\mathbf{w}$ can be calculated as

$$\mathbf{w}^T\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 \mathbf{w}}\mathbf{w} = \sum_{m=1}^{M}\sum_{n=1}^{M} w_m w_n \frac{\partial^2 J_{sum}(\mathbf{w})}{\partial w_m \partial w_n} \tag{E.84}$$

$$= 4\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{m=1}^{M}\sum_{n=1}^{M} w_i w_j w_m w_n \big(2\rho_{ijmn} + \rho_{imjn}\big) \tag{E.85}$$

$$= 12\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{m=1}^{M}\sum_{n=1}^{M} w_i w_j w_m w_n \rho_{ijmn}\,, \tag{E.86}$$

using the expression

$$\sum_{i=1}^{M}\sum_{j=1}^{M} w(i)w(j)\rho_{ijkn} = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i)w(j)\rho_{jikn} = \sum_{i=1}^{M}\sum_{j=1}^{M} w(i)w(j)\rho_{ijnk}\,. \tag{E.87}$$

Hence, the quadratic form can be written as

$$\boxed{\mathbf{w}^T\frac{\partial^2 J_{sum}(\mathbf{w})}{\partial^2 \mathbf{w}}\mathbf{w} = 12\,\mathbf{w}^T\mathbf{Q}_{sum}(\mathbf{w})\mathbf{w} = 12\,J_{sum}(\mathbf{w})} \tag{E.88}$$

This quantity is always positive (or equal to zero), since

$$J_{sum}(\mathbf{w}) = \int_\Theta\int_\Omega |H(\omega,\theta)|^4 d\omega d\theta \tag{E.89}$$

is an integral, i.e. infinite summation, of positive values.

F Solving integrals for far-field assumption

The integral

$$I = \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \cos\left[\omega\big(\alpha + \beta\cos\theta\big) + \gamma\right] d\omega d\theta \qquad \text{(F.1)}$$

is equal to

$$\int_{\theta_1}^{\theta_2} \frac{\sin\left[\omega_2\big(\alpha + \beta\cos\theta\big) + \gamma\right]}{\alpha + \beta\cos\theta} d\theta - \int_{\theta_1}^{\theta_2} \frac{\sin\left[\omega_1\big(\alpha + \beta\cos\theta\big) + \gamma\right]}{\alpha + \beta\cos\theta} d\theta\ , \qquad \text{(F.2)}$$

such that in fact we need to solve integrals of the type (F.2),

$$\boxed{I_\theta(\omega) = \int_{\theta_1}^{\theta_2} \frac{\sin\left[\omega\big(\alpha + \beta\cos\theta\big) + \gamma\right]}{\alpha + \beta\cos\theta} d\theta} \qquad \text{(F.3)}$$

Normally this integral can be computed numerically without any problem (e.g. using the MATLAB commands `quad` or `quad8`), but some special cases occur.

In robust broadband beamformer design, also integrals of the type

$$I = \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \sin\left[\omega\big(\alpha + \beta\cos\theta\big) + \gamma\right] d\omega d\theta\ , \qquad \text{(F.4)}$$

arise. However, these can be considered to be special case of (F.1), where γ has been replaced by $\gamma - \pi/2$.

CASE 1: $\beta = 0,\ \alpha \neq 0$

The integral I now reduces to

$$I = \int_{\omega_1}^{\omega_2} \cos(\omega\alpha + \gamma) d\omega \cdot (\theta_2 - \theta_1) = \frac{\sin(\omega_2\alpha + \gamma) - \sin(\omega_1\alpha + \gamma)}{\alpha} \cdot (\theta_2 - \theta_1)\ . \qquad \text{(F.5)}$$

CASE 2: $\beta = 0,\ \alpha = 0$

The integral I now reduces to

$$I = \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \cos\gamma\, d\omega d\theta = \cos\gamma \cdot (\theta_2 - \theta_1) \cdot (\omega_2 - \omega_1)\ . \qquad \text{(F.6)}$$

CASE 3: $\exists\ \theta_n \in]\theta 1, \theta 2[, \alpha + \beta\cos\theta_n = 0$

The singularity θ_n in the denominator will occur at

$$\cos\theta_n = -\frac{\alpha}{\beta}\ , \qquad \text{(F.7)}$$

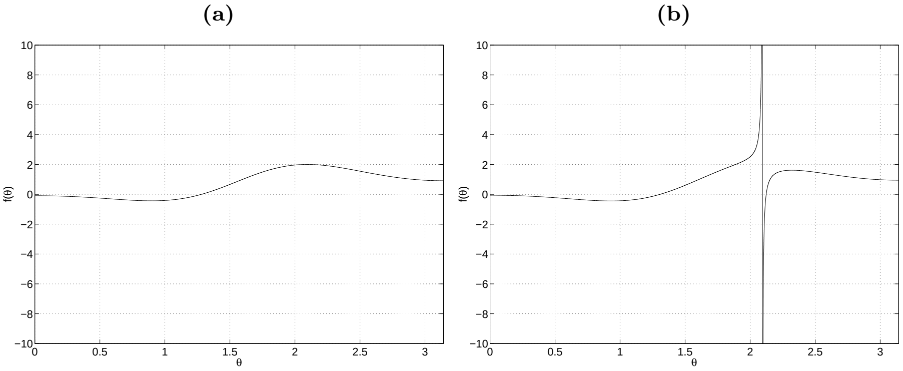


Figure F.1: Function $f(\omega, \theta)$ with singularity at $\theta_n = 2.09$ ($\alpha = 1$, $\beta = 2$, $\omega = 2$) with (**a**) $\gamma = 0$ and (**b**) $\gamma = 0.1$

and hence will only occur if $|\alpha| \leq |\beta|$. Because all considered functions are symmetric in θ, we do not need to consider negative angles θ and can assume that $0 \leq \theta \leq \pi$. Hence, $\sin\theta_n$ will always be positive and can be written as

$$\sin\theta_n = \sqrt{1 - \frac{\alpha^2}{\beta^2}}, \qquad \beta\sin\theta_n = \beta\sqrt{1 - \frac{\alpha^2}{\beta^2}} \ . \tag{F.8}$$

Because of the singularity θ_n in the denominator, numerically calculating the integral $I_\theta(\omega)$ can give rise to numerical problems. Figure F.1a depicts the function

$$\boxed{f(\omega, \theta) = \frac{\sin\left[\omega\left(\alpha + \beta\cos\theta\right) + \gamma\right]}{\alpha + \beta\cos\theta}} \tag{F.9}$$

for $\alpha = 1$, $\beta = 2$, $\omega = 2$ and $\gamma = 0$ in the range $[0, \pi]$, while figure F.1b depicts $f(\omega, \theta)$ for $\gamma = 0.1$. As can be seen, if $\gamma = 0$, $\lim\limits_{\theta \to \theta_n} f(\omega, \theta) = \omega$, whereas if $\gamma \neq 0$, $\lim\limits_{\theta \to \theta_n} f(\omega, \theta) = \pm\infty$. Numerically, this implies that there only exists a problem in calculating $I_\theta(\omega)$ if $\gamma \neq 0$, so from now on we will assume that $\gamma \neq 0$.

If the singularity θ_n exists in the integration interval $]\theta 1, \theta 2[$, the integral $I_\theta(\omega)$ can be split up as

$$I_\theta(\omega) = \underbrace{\int_{\theta_1}^{\theta_n - \epsilon} f(\omega, \theta)\, d\theta}_{I_{\theta,1}(\omega)} + \underbrace{\int_{\theta_n - \epsilon}^{\theta_n + \epsilon} f(\omega, \theta)\, d\theta}_{I_{\theta,2}(\omega)} + \underbrace{\int_{\theta_n + \epsilon}^{\theta_2} f(\omega, \theta)\, d\theta}_{I_{\theta,3}(\omega)} \ . \tag{F.10}$$

The integrals $I_{\theta,1}(\omega)$ and $I_{\theta,3}(\omega)$ can be calculated numerically without any problem, such that we will concentrate on the calculation of the integral $I_{\theta,2}(\omega)$, containing the singularity θ_n.

First, we derive a function $g(\theta)$ which is a good *approximation for* $f(\omega,\theta)$ *around* θ_n. The Taylor-expansions of $\cos\theta$ and $\sin\theta$ around 0 are

$$\cos\theta = \sum_{k=0}^{\infty}(-1)^k\frac{\theta^{2k}}{(2k)!} = 1+\mathcal{O}(\theta^2) \tag{F.11}$$

$$\sin\theta = \sum_{k=0}^{\infty}(-1)^k\frac{\theta^{2k+1}}{(2k+1)!} = \theta+\mathcal{O}(\theta^3)\ , \tag{F.12}$$

such that the Taylor-expansion of $\cos\theta$ around θ_n is

$$\begin{aligned}\cos\theta &= \cos\left[\theta_n+(\theta-\theta_n)\right] &\text{(F.13)}\\ &= \cos\theta_n\cdot\cos(\theta-\theta_n)-\sin\theta_n\cdot\sin(\theta-\theta_n) &\text{(F.14)}\\ &= -\frac{\alpha}{\beta}\left[1+\mathcal{O}\big((\theta-\theta_n)^2\big)\right]-\sqrt{1-\frac{\alpha^2}{\beta^2}}\left[(\theta-\theta_n)+\mathcal{O}\big((\theta-\theta_n)^3\big)\right]\ ,\end{aligned}$$

such that

$$\alpha+\beta\cos\theta = -\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)+\mathcal{O}\big((\theta-\theta_n)^2\big)\ . \tag{F.15}$$

In a first approximation the function $f(\omega,\theta)$ can be approximated by

$$\frac{\sin\left[\omega\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)-\gamma\right]}{\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)}\ , \tag{F.16}$$

which can be further simplified (since we assumed $\gamma\neq 0$) to

$$\boxed{g(\theta) = -\frac{\sin\gamma}{\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)}} \tag{F.17}$$

Note that $g(\theta)$ is independent of ω. We now define the function

$$\bar{f}(\omega,\theta) = f(\omega,\theta)-g(\theta)\ . \tag{F.18}$$

Figure F.2a depicts the function $\bar{f}(\omega,\theta)$ for $\alpha=1$, $\beta=2$, $\omega=2$ and $\gamma=0$ in the range $[0,\pi]$, while figure F.1b depicts $\bar{f}(\omega,\theta)$ for $\gamma=0.1$. As can be seen, for any γ, $\lim\limits_{\theta\to\theta_n}\bar{f}(\omega,\theta)$ is finite.

The integral $I_{\theta,2}(\omega)$ can now be written as

$$I_{\theta,2}(\omega) = \int_{\theta_n-\epsilon}^{\theta_n+\epsilon} f(\omega,\theta)\,d\theta = \int_{\theta_n-\epsilon}^{\theta_n+\epsilon}\bar{f}(\omega,\theta)\,d\theta+\int_{\theta_n-\epsilon}^{\theta_n+\epsilon}g(\theta)\,d\theta \tag{F.19}$$

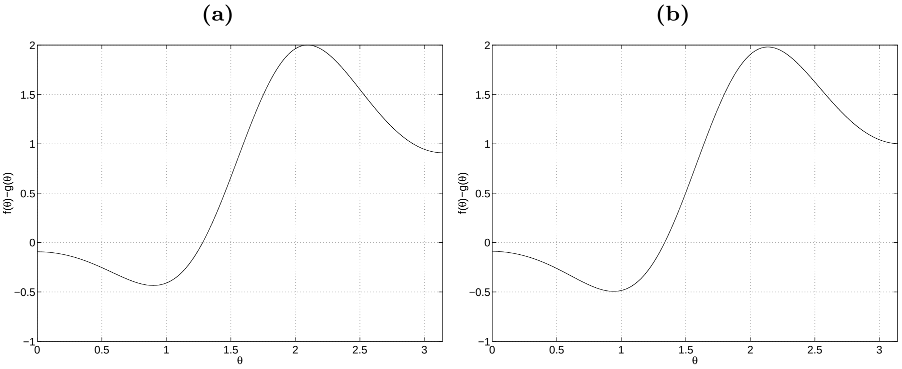


Figure F.2: Function $\bar{f}(\omega,\theta)$ with singularity at $\theta_n = 2.09$ ($\alpha = 1$, $\beta = 2$, $\omega = 2$) with (**a**) $\gamma = 0$ and (**b**) $\gamma = 0.1$

The integral $\int_{\theta_n-\epsilon}^{\theta_n+\epsilon} g(\theta)\,d\theta$ is equal to 0, since

$$\int_{\theta_n-\epsilon}^{\theta_n+\epsilon} g(\theta)\,d\theta = -\int_{\theta_n-\epsilon}^{\theta_n+\epsilon} \frac{\sin\gamma}{\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)}\,d\theta = -\frac{\sin\gamma}{\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}}\int_{-\epsilon}^{\epsilon}\frac{d\phi}{\phi} = 0\ , \tag{F.20}$$

such that we can approximate the integral $\int_{\theta_n-\epsilon}^{\theta_n+\epsilon} \bar{f}(\omega,\theta)\,d\theta$ and hence $I_{\theta,2}(\omega)$ as

$$I_{\theta,2}(\omega) = \int_{\theta_n-\epsilon}^{\theta_n+\epsilon} \bar{f}(\omega,\theta)\,d\theta \approx 2\epsilon \cdot \lim_{\theta\to\theta_n} \bar{f}(\omega,\theta)\ . \tag{F.21}$$

We will now derive an expression for this (finite) limit value. The function $\bar{f}(\omega,\theta)$ can be written as

$$\begin{aligned}
\bar{f}(\omega,\theta) &= \frac{\sin\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right]}{\alpha+\beta\cos\theta} + \frac{\sin\gamma}{\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)} && \text{(F.22)}\\
&= \frac{\sin\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right]\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)+\sin\gamma\big(\alpha+\beta\cos\theta\big)}{\big(\alpha+\beta\cos\theta\big)\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)}\\
&= \frac{g_D(\theta)}{g_N(\theta)}\ . && \text{(F.23)}
\end{aligned}$$

Since

$$\lim_{\theta\to\theta_n} \bar{f}(\omega,\theta) = \frac{g_D(\theta_n)}{g_N(\theta_n)} = \frac{0}{0}\ , \tag{F.24}$$

we need L'Hôpital's rule to calculate the limit, i.e.

$$\lim_{\theta\to\theta_n}\frac{g_D(\theta)}{g_N(\theta)} = \frac{\lim\limits_{\theta\to\theta_n} g'_D(\theta)}{\lim\limits_{\theta\to\theta_n} g'_N(\theta)}\ . \tag{F.25}$$

The first order derivatives $g'_D(\theta)$ and $g'_N(\theta)$ are equal to

$$\begin{aligned} g'_D(\theta) &= \cos\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right](-\beta\omega\sin\theta)\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n) \\ &\quad +\sin\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right]\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}-\sin\gamma(\beta\sin\theta) \qquad \text{(F.26)} \\ g'_N(\theta) &= -\beta^2\sin\theta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)+\big(\alpha+\beta\cos\theta\big)\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}\,. \qquad \text{(F.27)} \end{aligned}$$

Since $g'_D(\theta_n) = g'_N(\theta_n) = 0$, we need L'Hôpital's rule once more and have to calculate the second-order derivatives $g''_D(\theta)$ and $g''_N(\theta)$,

$$\begin{aligned} g''_D(\theta) &= -\sin\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right](\beta\omega\sin\theta)^2\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n) \\ &\quad +\cos\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right](-\beta\omega\cos\theta)\beta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n) \\ &\quad +2\cos\left[\omega\big(\alpha+\beta\cos\theta\big)+\gamma\right](-\beta\omega\sin\theta)\beta\sqrt{1-\frac{\alpha^2}{\beta^2}} \\ &\quad -\sin\gamma(\beta\cos\theta) \qquad \text{(F.28)} \\ g''_N(\theta) &= -\beta^2\cos\theta\sqrt{1-\frac{\alpha^2}{\beta^2}}(\theta-\theta_n)-2\beta^2\sin\theta\sqrt{1-\frac{\alpha^2}{\beta^2}}\,. \qquad \text{(F.29)} \end{aligned}$$

The second-order derivatives $g''_D(\theta)$ and $g''_N(\theta)$ evaluated at θ_n give

$$\begin{aligned} g''_D(\theta_n) &= 2\omega\cos\gamma(\alpha^2-\beta^2)+\alpha\sin\gamma \qquad \text{(F.30)} \\ g''_N(\theta_n) &= 2(\alpha^2-\beta^2)\,, \qquad \text{(F.31)} \end{aligned}$$

such that

$$\lim_{\theta\to\theta_n}\bar{f}(\omega,\theta)=\omega\cos\gamma+\frac{\alpha\sin\gamma}{2(\alpha^2-\beta^2)}\,, \qquad \text{(F.32)}$$

and

$$\boxed{I_{\theta,2}(\omega)\approx\epsilon\left[2\omega\cos\gamma+\frac{\alpha\sin\gamma}{\alpha^2-\beta^2}\right]} \qquad \text{(F.33)}$$

If $\gamma = 0$, this integral reduces to $I_{\theta,2}(\omega) \approx 2\epsilon\omega$.

Remark F.1 Since for any γ, $\lim_{\theta\to\theta_n}\bar{f}(\omega,\theta)$ is finite, the function $\bar{f}(\omega,\theta)$ can be integrated numerically without any problem. In fact the total integral I can be written as

$$I = I_\theta(\omega_2)-I_\theta(\omega_1)=\int_{\theta_1}^{\theta_2}f(\omega_2,\theta)\,d\theta-\int_{\theta_1}^{\theta_2}f(\omega_1,\theta)\,d\theta \qquad \text{(F.34)}$$

$$= \int_{\theta_1}^{\theta_2} \bar{f}(\omega_2,\theta)\,d\theta - \int_{\theta_1}^{\theta_2} \bar{f}(\omega_1,\theta)\,d\theta\ , \tag{F.35}$$

which can be calculated numerically without any problem. △

Remark F.2 If $\theta_1 = \theta_n$ or $\theta_2 = \theta_n$, then the integral $I_\theta(\omega)$ cannot be decomposed as in (F.10), but is decomposed as (assuming $\theta_2 = \theta_n$)

$$I_\theta(\omega) = \underbrace{\int_{\theta_1}^{\theta_n-\epsilon} f(\omega,\theta)\,d\theta}_{I_{\theta,1}} + \underbrace{\int_{\theta_n-\epsilon}^{\theta_n} f(\omega,\theta)\,d\theta}_{I_{\theta,2}}\ . \tag{F.36}$$

According to (F.19), the integral $I_{\theta,2}(\omega)$ is equal to

$$I_{\theta,2}(\omega) = \int_{\theta_n-\epsilon}^{\theta_n} \bar{f}(\omega,\theta)\,d\theta + \int_{\theta_n-\epsilon}^{\theta_n} g(\theta)\,d\theta\ , \tag{F.37}$$

but since $\int_{\theta_n-\epsilon}^{\theta_n} g(\theta)\,d\theta$ is equal to $\pm\infty$ if $\gamma \neq 0$ and equal to 0 if $\gamma = 0$, the integral $I_{\theta,2}(\omega) = \pm\infty$ if $\gamma \neq 0$ and $I_{\theta,2}(\omega) = \epsilon\omega$ if $\gamma = 0$. However, the case $\theta_1 = \theta_n$ or $\theta_2 = \theta_n$ is very unlikely to occur (and slightly changing the value of θ_1 or θ_2 actually solves the problem). △

CASE 4: $\alpha = \beta \neq 0$

In this case, there is a singularity at $\theta_n = \pi$, such that $\sin\theta_n = 0$ and (F.14) becomes

$$\cos\theta = \cos\theta_n \cdot \cos(\theta - \theta_n) \tag{F.38}$$

$$= -\left[1 - \frac{(\theta-\theta_n)^2}{2} + \mathcal{O}\big((\theta-\theta_n)^4\big)\right]\ , \tag{F.39}$$

such that

$$\alpha + \beta\cos\theta = \alpha(1+\cos\theta) = \alpha\frac{(\theta-\theta_n)^2}{2} + \mathcal{O}\big((\theta-\theta_n)^4\big)\ . \tag{F.40}$$

In a first approximation the function $f(\omega,\theta)$ now can be approximated by

$$\frac{\sin\left[\omega\alpha\frac{(\theta-\theta_n)^2}{2} + \gamma\right]}{\alpha\frac{(\theta-\theta_n)^2}{2}}\ , \tag{F.41}$$

which can be further simplified to

$$g(\theta) = \frac{2\sin\gamma}{\alpha(\theta-\theta_n)^2}\ . \tag{F.42}$$

A problem now arises if $\theta_2 = \theta_n = \pi$ and $\gamma \neq 0$ (cf. remark F.2), since then the integral $\int_{\pi-\epsilon}^{\pi} g(\theta)\, d\theta = \pm\infty$. Therefore the integral $I_{\theta,2}(\omega) = \pm\infty$ if $\gamma \neq 0$ and $I_{\theta,2}(\omega) = \epsilon\omega$ if $\gamma = 0$. However, the case $\alpha = \beta \neq 0$ is very unlikely to occur, since α is always an integer number, whereas

$$\beta = \frac{\Delta d}{c} f_s \ , \tag{F.43}$$

with Δd a value related to inter-microphone distances.

G Calculation of expressions for near-field broadband beamforming

In this appendix, the calculation of the following expressions is discussed:

- Appendix G.1 (WLS criterion) : weighted LS, TLS eigenfilter

$$\boxed{\int_{\Theta_p}\int_{\Omega_p} \mathrm{Re}\big\{H(\omega,\theta,r)\big\}d\omega d\theta = \mathbf{w}^T \cdot \int_{\Theta_p}\int_{\Omega_p} \mathbf{g}_R(\omega,\theta,r)d\omega d\theta = \mathbf{w}^T\mathbf{a}}$$

- Appendix G.2 (Energy criterion): weighted LS, conventional eigenfilter, TLS eigenfilter, maximum energy array, non-linear criterion

$$\boxed{\int_{\Theta}\int_{\Omega} |H(\omega,\theta,r)|^2 d\omega d\theta = \mathbf{w}^T \cdot \int_{\Theta}\int_{\Omega} \mathbf{G}_R(\omega,\theta,r)d\omega d\theta \cdot \mathbf{w} = \mathbf{w}^T\mathbf{Q}_e\mathbf{w}}$$

- Appendix G.3 (Passband error) : conventional eigenfilter

$$\boxed{\begin{aligned}&\int_{\Theta_p}\int_{\Omega_p} |H(\omega_c,\theta_c,r) - H(\omega,\theta,r)|^2 d\omega d\theta = \mathbf{w}^T\mathbf{Q}_p\mathbf{w} = \\ &\mathbf{w}^T\int_{\Theta_p}\int_{\Omega_p}\mathrm{Re}\big\{[\mathbf{g}(\omega_c,\theta_c,r) - \mathbf{g}(\omega,\theta,r)][\mathbf{g}(\omega_c,\theta_c,r) - \mathbf{g}(\omega,\theta,r)]^H\big\}d\omega d\theta\,\mathbf{w}\end{aligned}}$$

- Appendix G.4 (Non-linear criterion) : non-linear criterion

$$\boxed{J_{sum}(\mathbf{w}) = \int_{\Theta}\int_{\Omega} |H(\omega,\theta,r)|^4 d\omega d\theta = \int_{\Theta}\int_{\Omega} \big(\mathbf{w}^T\mathbf{G}(\omega,\theta,r)\mathbf{w}\big)^2 d\omega d\theta}$$

G.1 WLS criterion

Using (9.4), (9.6) and (9.11), the ith element of $\mathbf{g}_R(\omega,\theta,r)$ is equal to

$$\mathbf{g}_R^i(\omega,\theta,r) = \frac{r\,\cos\Big[\omega\Big(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c}f_s\Big)\Big]}{\sqrt{p_n+q_n\cos\theta}}, \quad i = 1\ldots M\ , \tag{G.1}$$

with

$$k = \mathrm{mod}(i-1,L) \qquad n = \lfloor\frac{i-1}{L}\rfloor\ . \tag{G.2}$$

The ith element of $\mathbf{a}$ therefore is equal to

$$\mathbf{a}^i = \int_{\Theta_p}\int_{\Omega_p} \mathbf{g}_R^i(\omega,\theta,r)d\omega d\theta = \int_{\Theta_p}\int_{\Omega_p} \frac{r\,\cos\Big[\omega\Big(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c}f_s\Big)\Big]}{\sqrt{p_n+q_n\cos\theta}}d\omega d\theta\ . \tag{G.3}$$

M different integrals need to be calculated. The integral in (G.3) can be calculated as

$$I = \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \frac{r \, \cos\left[\omega\left(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c} f_s\right)\right]}{\sqrt{p_n + q_n \cos\theta}} d\omega d\theta = I_{WLS}(\omega_2) - I_{WLS}(\omega_1) \; ,$$

with

$$I_{WLS}(\omega) = \int_{\theta_1}^{\theta_2} \frac{r \, \sin\left[\omega\left(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c} f_s\right)\right]}{\sqrt{p_n + q_n\cos\theta}\left(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c} f_s\right)} d\theta \; , \qquad \text{(G.4)}$$

which can be integrated numerically without any problem. A special case occurs when $d_n = 0$, since then $\sqrt{p_n + q_n \cos\theta} = r$.

Special case 1: $d_n = 0$**,** $k \neq 0$**.**
In this case the integral I reduces to

$$I = \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \cos(\omega k) d\omega d\theta = \frac{\sin(\omega_2 k) - \sin(\omega_1 k)}{k} \cdot (\theta_2 - \theta_1) \; . \qquad \text{(G.5)}$$

Special case 2: $d_n = 0$**,** $k = 0$**.**
In this case the integral I reduces to

$$I = \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} 1 \, d\omega d\theta = (\omega_2 - \omega_1) \cdot (\theta_2 - \theta_1) \; . \qquad \text{(G.6)}$$

G.2 Energy criterion

Using (9.11) and (9.13), the (i, j)-th element of $\mathbf{G}(\omega, \theta, r)$ is equal to

$$\begin{aligned}
\mathbf{G}^{ij}(\omega,\theta,r) &= \mathbf{g}^i(\omega,\theta,r)\mathbf{g}^j(\omega,\theta,r)^* && \text{(G.7)} \\
&= \frac{r^2 \, e^{-j\omega\left((k-l)+\frac{r_n(\theta,r)-r_m(\theta,r)}{c} f_s\right)}}{r_n(\theta,r) r_m(\theta,r)} && \text{(G.8)} \\
&= \frac{r^2 \, e^{-j\omega\left((k-l)+\frac{\sqrt{p_n+q_n\cos\theta}-\sqrt{p_m+q_m\cos\theta}}{c} f_s\right)}}{\sqrt{p_n+q_n\cos\theta}\sqrt{p_m+q_m\cos\theta}} \; ,
\end{aligned}$$

with

$$k = \text{mod}(i-1, L) \qquad n = \lfloor \frac{i-1}{L} \rfloor \qquad \text{(G.9)}$$

$$l = \text{mod}(j-1, L) \qquad m = \lfloor \frac{j-1}{L} \rfloor \; . \qquad \text{(G.10)}$$

The (i, j)-th element of the real and the imaginary part of $\mathbf{G}(\omega, \theta, r)$ then are equal to

$$\mathbf{G}_R^{ij}(\omega,\theta,r) \; = \; \left[\mathbf{G}_R\right]_{nm}^{kl}(\omega,\theta,r) = \frac{r^2 \, \cos\left[\omega\left((k-l) + \frac{r_n(\theta,r)-r_m(\theta,r)}{c} f_s\right)\right]}{r_n(\theta,r) r_m(\theta,r)}$$

$$\mathbf{G}_I^{ij}(\omega,\theta,r) \;=\; [\mathbf{G}_I]_{nm}^{kl}(\omega,\theta,r) = -\frac{r^2\,\sin\left[\omega\Big((k-l)+\frac{r_n(\theta,r)-r_m(\theta,r)}{c}f_s\Big)\right]}{r_n(\theta,r)r_m(\theta,r)}\;.$$

The real part is symmetric and the imaginary part is anti-symmetric, since

$$\mathbf{G}_R^{ji}(\omega,\theta,r) = [\mathbf{G}_R]_{mn}^{lk}(\omega,\theta,r) \;=\; [\mathbf{G}_R]_{nm}^{kl}(\omega,\theta,r) = \mathbf{G}_R^{ij}(\omega,\theta,r) \qquad \text{(G.11)}$$

$$\mathbf{G}_I^{ji}(\omega,\theta,r) = [\mathbf{G}_I]_{mn}^{lk}(\omega,\theta,r) \;=\; -[\mathbf{G}_I]_{nm}^{kl}(\omega,\theta,r) = -\mathbf{G}_I^{ij}(\omega,\theta,r)\;. \qquad \text{(G.12)}$$

The spatial directivity spectrum $|H(\omega,\theta,r)|^2$ therefore can be written as

$$\boxed{|H(\omega,\theta,r)|^2 = \mathbf{w}^T\mathbf{G}_R(\omega,\theta,r)\mathbf{w}} \qquad \text{(G.13)}$$

which is symmetric in both ω and θ. The (i,j)-th element of $\mathbf{Q}_e$ is equal to

$$\mathbf{Q}_e^{ij} = \int_\Theta\int_\Omega \frac{r^2\,\cos\left[\omega\Big((k-l)+\frac{\sqrt{p_n+q_n\cos\theta}-\sqrt{p_m+q_m\cos\theta}}{c}f_s\Big)\right]}{\sqrt{p_n+q_n\cos\theta}\sqrt{p_m+q_m\cos\theta}}d\omega d\theta\;. \qquad \text{(G.14)}$$

Independent of the microphone configuration, $(2L-1)N^2$ different integrals need to be calculated. The integral in (G.14) can be calculated as

$$\begin{aligned} I &= \int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2} \frac{r^2\,\cos\left[\omega\Big((k-l)+\frac{\sqrt{p_n+q_n\cos\theta}-\sqrt{p_m+q_m\cos\theta}}{c}f_s\Big)\right]}{\sqrt{p_n+q_n\cos\theta}\sqrt{p_m+q_m\cos\theta}}d\omega d\theta \\ &= I_{en}(\omega_2)-I_{en}(\omega_1)\;, \qquad \text{(G.15)} \end{aligned}$$

with $I_{en}(\omega)$ equal to

$$\int_{\theta_1}^{\theta_2} \frac{r^2\,\sin\left[\omega\Big((k-l)+\frac{\sqrt{p_n+q_n\cos\theta}-\sqrt{p_m+q_m\cos\theta}}{c}f_s\Big)\right]}{\sqrt{p_n+q_n\cos\theta}\sqrt{p_m+q_m\cos\theta}\Big((k-l)+\frac{\sqrt{p_n+q_n\cos\theta}-\sqrt{p_m+q_m\cos\theta}}{c}f_s\Big)}d\theta\;,$$

which can be integrated numerically without any problem. A special case occurs when $d_n = d_m$, since then $\sqrt{p_n+q_n\cos\theta} = \sqrt{p_m+q_m\cos\theta}$.

Special case 1: $d_n = d_m$**,** $k \neq l$**.**
In this case the integral I reduces to

$$\begin{aligned} I &= \int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2} \frac{r^2\,\cos\left[\omega(k-l)\right]}{p_n+q_n\cos\theta}d\omega d\theta \qquad \text{(G.16)} \\ &= r^2\,\frac{\sin\left[\omega_2(k-l)\right]-\sin\left[\omega_1(k-l)\right]}{k-l}\cdot\int_{\theta_1}^{\theta_2}\frac{1}{p_n+q_n\cos\theta}d\theta\;. \qquad \text{(G.17)} \end{aligned}$$

Special case 2: $d_n = d_m$**,** $k = l$**.**
In this case the integral I reduces to

$$I = \int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2}\frac{r^2}{p_n+q_n\cos\theta}d\omega d\theta = r^2\,(\omega_2-\omega_1)\cdot\int_{\theta_1}^{\theta_2}\frac{1}{p_n+q_n\cos\theta}d\theta\;. \qquad \text{(G.18)}$$

G.3 Passband error

For calculating the passband error, the integrand $|H(\omega_c,\theta_c,r) - H(\omega,\theta,r)|^2$ can be written as

$$\begin{aligned}|H(\omega_c,\theta_c,r) - H(\omega,\theta,r)|^2 &= |\mathbf{w}^T\mathbf{g}(\omega_c,\theta_c,r) - \mathbf{w}^T\mathbf{g}(\omega,\theta,r)|^2 & \text{(G.19)}\\ &= \mathbf{w}^T\left[\mathbf{g}(\omega_c,\theta_c,r)\mathbf{g}^H(\omega_c,\theta_c,r) - \mathbf{g}(\omega,\theta,r)\mathbf{g}^H(\omega_c,\theta_c,r)\right.\\ &\quad \left. - \,\mathbf{g}(\omega_c,\theta_c,r)\mathbf{g}^H(\omega,\theta,r) + \mathbf{g}(\omega,\theta,r)\mathbf{g}^H(\omega,\theta,r)\right]\mathbf{w}\;. & \text{(G.20)}\end{aligned}$$

If we define $\tilde{\mathbf{G}}(\omega_1,\theta_1,\omega_2,\theta_2,r)$ as

$$\tilde{\mathbf{G}}(\omega_1,\theta_1,\omega_2,\theta_2,r) = \mathbf{g}(\omega_1,\theta_1,r)\mathbf{g}^H(\omega_2,\theta_2,r)\;, \tag{G.21}$$

then we can write (G.20) as

$$\mathbf{w}^T\underbrace{\Big[\mathbf{G}(\omega_c,\theta_c,r) - \tilde{\mathbf{G}}(\omega,\theta,\omega_c,\theta_c,r) - \tilde{\mathbf{G}}(\omega_c,\theta_c,\omega,\theta,r) + \mathbf{G}(\omega,\theta,r)\Big]}_{\hat{\mathbf{G}}(\omega_c,\theta_c,\omega,\theta,r) = \hat{\mathbf{G}}(\omega,\theta,\omega_c,\theta_c,r)}\mathbf{w}\;.$$

The (i,j)-th element of $\tilde{\mathbf{G}}(\omega_1,\theta_1,\omega_2,\theta_2,r)$ is equal to

$$\tilde{\mathbf{G}}^{ij}(\omega_1,\theta_1,\omega_2,\theta_2,r) = \frac{r^2\, e^{-j\omega_1\left(k + \frac{\sqrt{p_n+q_n\cos\theta_1}-r}{c}f_s\right)}\, e^{j\omega_2\left(l + \frac{\sqrt{p_m+q_m\cos\theta_2}-r}{c}f_s\right)}}{\sqrt{p_n+q_n\cos\theta_1}\,\sqrt{p_m+q_m\cos\theta_2}}\;,$$

with

$$\begin{aligned} k &= \mathrm{mod}(i-1,L) \qquad & n &= \lfloor\frac{i-1}{L}\rfloor & \text{(G.22)}\\ l &= \mathrm{mod}(j-1,L) \qquad & m &= \lfloor\frac{j-1}{L}\rfloor\;. & \text{(G.23)}\end{aligned}$$

Since $\hat{\mathbf{G}}(\omega_c,\theta_c,\omega,\theta,r)$ is complex Hermitian, the real part $\hat{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta,r)$ is symmetric and the imaginary part $\hat{\mathbf{G}}_I(\omega_c,\theta_c,\omega,\theta,r)$ is anti-symmetric, such that (G.20) can be written as

$$\boxed{\begin{aligned}|H(\omega_c,\theta_c,r) - H(\omega,\theta,r)|^2 &= \mathbf{w}^T\hat{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta,r)\mathbf{w}\\ &= \mathbf{w}^T\Big[\tilde{\mathbf{G}}_R(\omega_c,\theta_c,\omega_c,\theta_c,r) - \tilde{\mathbf{G}}_R(\omega,\theta,\omega_c,\theta_c,r)\\ &\quad - \tilde{\mathbf{G}}_R(\omega_c,\theta_c,\omega,\theta,r) + \tilde{\mathbf{G}}_R(\omega,\theta,\omega,\theta,r)\Big]\mathbf{w}\;.\end{aligned}} \tag{G.24}$$

The (i,j)-th element of the real part $\tilde{\mathbf{G}}_R(\omega_1,\theta_1,\omega_2,\theta_2,r)$ is equal to

$$\frac{r^2\,\cos\left[\omega_1\left(k + \frac{\sqrt{p_n+q_n\cos\theta_1}-r}{c}f_s\right) - \omega_2\left(l + \frac{\sqrt{p_m+q_m\cos\theta_2}-r}{c}f_s\right)\right]}{\sqrt{p_n+q_n\cos\theta_1}\,\sqrt{p_m+q_m\cos\theta_2}}\;, \tag{G.25}$$

such that the (i,j)-th element of $\mathbf{Q}_p$ is equal to

$$\int_{\Theta_p}\int_{\Omega_p}\hat{\mathbf{G}}_R^{ij}(\omega_c,\theta_c,\omega,\theta,r)d\omega d\theta =$$

$$(a) \int_{\Theta_p}\int_{\Omega_p} \frac{r^2 \cos\left[\omega_c\Big((k-l)+\frac{\sqrt{p_n+q_n\cos\theta_c}-\sqrt{p_m+q_m\cos\theta_c}}{c}f_s\Big)\right]}{\sqrt{p_n+q_n\cos\theta_c}\sqrt{p_m+q_m\cos\theta_c}}d\omega d\theta$$

$$(b) - \int_{\Theta_p}\int_{\Omega_p} \frac{r^2 \cos\left[\omega\Big(k+\frac{\sqrt{p_n+q_n\cos\theta}-r}{c}f_s\Big)-\omega_c\Big(l+\frac{\sqrt{p_m+q_m\cos\theta_c}-r}{c}f_s\Big)\right]}{\sqrt{p_n+q_n\cos\theta}\,\sqrt{p_m+q_m\cos\theta_c}}d\omega d\theta$$

$$(c) - \int_{\Theta_p}\int_{\Omega_p} \frac{r^2 \cos\left[\omega\Big(l+\frac{\sqrt{p_m+q_m\cos\theta}-r}{c}f_s\Big)-\omega_c\Big(k+\frac{\sqrt{p_n+q_n\cos\theta_c}-r}{c}f_s\Big)\right]}{\sqrt{p_m+q_m\cos\theta}\,\sqrt{p_n+q_n\cos\theta_c}}d\omega d\theta$$

$$(d) + \int_{\Theta_p}\int_{\Omega_p} \frac{r^2 \cos\left[\omega\Big((k-l)+\frac{\sqrt{p_n+q_n\cos\theta}-\sqrt{p_m+q_m\cos\theta}}{c}f_s\Big)\right]}{\sqrt{p_n+q_n\cos\theta}\sqrt{p_m+q_m\cos\theta}}d\omega d\theta\ .$$

The integrals (a) and (d) can be calculated numerically without any problem, whereas the integrals (b) and (c) are seen to be special cases of the integral

$$K_{ij}\int_{\theta_1}^{\theta_2}\int_{\omega_1}^{\omega_2} \frac{\cos\left[\omega\Big(\alpha_{ij}+\frac{\sqrt{\delta_{ij}+\epsilon_{ij}\cos\theta}-r}{c}f_s\Big)+\gamma_{ij}\right]}{\sqrt{\delta_{ij}+\epsilon_{ij}\cos\theta}}d\omega d\theta, \qquad \text{(G.26)}$$

with

$\alpha_{ij}=k$	$\alpha_{ij}=l$
$\delta_{ij}=p_n$	$\delta_{ij}=p_m$
$\epsilon_{ij}=q_n$	$\epsilon_{ij}=q_m$
$K_{ij}=\frac{r^2}{\sqrt{p_m+q_m\cos\theta_c}}$	$K_{ij}=\frac{r^2}{\sqrt{p_n+q_n\cos\theta_c}}$
$\gamma_{ij}=-\omega_c\Big(l+\frac{\sqrt{p_m+q_m\cos\theta_c}-r}{c}f_s\Big)$	$\gamma_{ij}=-\omega_c\Big(k+\frac{\sqrt{p_n+q_n\cos\theta_c}-r}{c}f_s\Big)$

Solving integrals of type (G.26) is discussed in Appendix H. For computing the passband error, $2(2L-1)N^2+2(LN)^2$ different integrals need to be calculated.

G.4 Non-linear criterion

For the non-linear criterion, the integrand $|H(\omega,\theta,r)|^4$ can be written as

$$\big(\mathbf{w}^T\mathbf{G}(\omega,\theta,r)\mathbf{w}\big)\big(\mathbf{w}^T\mathbf{G}(\omega,\theta,r)\mathbf{w}\big) \qquad \text{(G.27)}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w(i)w(j)w(k)w(l)\,\mathbf{g}^i(\omega,\theta,r)\mathbf{g}^j(\omega,\theta,r)^*\,\mathbf{g}^k(\omega,\theta,r)\mathbf{g}^l(\omega,\theta,r)^*$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w(i)w(j)w(k)w(l)\cdot$$

$$\frac{r^4\,e^{-j\omega\Big(\alpha_{ijkl}+\frac{\sqrt{\delta_i+\epsilon_i\cos\theta}-\sqrt{\delta_j+\epsilon_j\cos\theta}+\sqrt{\delta_k+\epsilon_k\cos\theta}-\sqrt{\delta_l+\epsilon_l\cos\theta}}{c}f_s\Big)}{\sqrt{\delta_i+\epsilon_i\cos\theta}\sqrt{\delta_j+\epsilon_j\cos\theta}\sqrt{\delta_k+\epsilon_k\cos\theta}\sqrt{\delta_l+\epsilon_l\cos\theta}}\ , \qquad \text{(G.28)}$$

with

$$\boxed{\begin{aligned}&\alpha_{ijkl} = \mathrm{mod}(i-1,L) - \mathrm{mod}(j-1,L) + \mathrm{mod}(k-1,L) - \mathrm{mod}(l-1,L)\\ &\delta_i = p_{\lfloor \frac{i-1}{L} \rfloor} = r^2 + d^2_{\lfloor \frac{i-1}{L} \rfloor}\\ &\epsilon_i = q_{\lfloor \frac{i-1}{L} \rfloor} = 2rd_{\lfloor \frac{i-1}{L} \rfloor}\end{aligned}}$$

Since $|H(\omega,\theta,r)|^4$ is real (and the filter coefficients are real), only the real part of the exponential function has to be considered, such that

$$\begin{aligned}|H(\omega,\theta)|^4 \; &= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w(i)w(j)w(k)w(l)\,\cdot\\ &\frac{r^4\,\cos\left[\omega\Big(\alpha_{ijkl} + \frac{\sqrt{\delta_i+\epsilon_i\cos\theta}-\sqrt{\delta_j+\epsilon_j\cos\theta}+\sqrt{\delta_k+\epsilon_k\cos\theta}-\sqrt{\delta_l+\epsilon_l\cos\theta}}{c}f_s\Big)\right]}{\sqrt{\delta_i+\epsilon_i\cos\theta}\sqrt{\delta_j+\epsilon_j\cos\theta}\sqrt{\delta_k+\epsilon_k\cos\theta}\sqrt{\delta_l+\epsilon_l\cos\theta}}\;,\end{aligned}$$

and $J_{sum}(\mathbf{w})$ can be written as

$$\boxed{J_{sum}(\mathbf{w}) = \int_\Theta\int_\Omega |H(\omega,\theta,r)|^4 d\omega d\theta = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} w_i w_j w_k w_l \rho_{ijkl}} \qquad \text{(G.29)}$$

with ρ_{ijkl} equal to

$$\int_\Theta\int_\Omega \frac{r^4\cos\left[\omega\Big(\alpha_{ijkl} + \frac{\sqrt{\delta_i+\epsilon_i\cos\theta}-\sqrt{\delta_j+\epsilon_j\cos\theta}+\sqrt{\delta_k+\epsilon_k\cos\theta}-\sqrt{\delta_l+\epsilon_l\cos\theta}}{c}f_s\Big)\right]}{\sqrt{\delta_i+\epsilon_i\cos\theta}\sqrt{\delta_j+\epsilon_j\cos\theta}\sqrt{\delta_k+\epsilon_k\cos\theta}\sqrt{\delta_l+\epsilon_l\cos\theta}} d\omega d\theta,$$

which can be calculated numerically without any problem and which only need to be computed once (since ρ_{ijkl} is independent of $\mathbf{w}$). As can be seen, α_{ijkl} can take on $4L-3$ distinct values $(-4L+2\ldots 4L-2)$ and because of the symmetry properties of ρ_{ijkl},

$$\rho_{ijkl} = \rho_{kjil} = \rho_{ilkj} = \rho_{klij} = \rho_{jilk} = \rho_{lijk} = \rho_{jkli} = \rho_{lkji}\;, \qquad \text{(G.30)}$$

only $(4L-3)\frac{N^4}{4}$ different integrals need to be calculated.

A special case occurs when

$$\sqrt{\delta_i+\epsilon_i\cos\theta} - \sqrt{\delta_j+\epsilon_j\cos\theta} + \sqrt{\delta_k+\epsilon_k\cos\theta} - \sqrt{\delta_l+\epsilon_l\cos\theta} = 0\;, \qquad \text{(G.31)}$$

which occurs when $\lfloor\frac{i-1}{L}\rfloor = \lfloor\frac{j-1}{L}\rfloor$ and $\lfloor\frac{k-1}{L}\rfloor = \lfloor\frac{l-1}{L}\rfloor$ or when $\lfloor\frac{i-1}{L}\rfloor = \lfloor\frac{l-1}{L}\rfloor$ and $\lfloor\frac{j-1}{L}\rfloor = \lfloor\frac{k-1}{L}\rfloor$. Suppose that $\lfloor\frac{i-1}{L}\rfloor = \lfloor\frac{j-1}{L}\rfloor$ and $\lfloor\frac{k-1}{L}\rfloor = \lfloor\frac{l-1}{L}\rfloor$, then ρ_{ijkl} reduces to

$$\rho_{ijkl} = \int_\Omega \cos(\omega\alpha_{ijkl})d\omega \int_\Theta \frac{r^4}{(\delta_i+\epsilon_i\cos\theta)\,(\delta_k+\epsilon_k\cos\theta)}d\theta\;. \qquad \text{(G.32)}$$

H Solving integrals for near-field assumption

This appendix discusses the calculation of the integral

$$I = K \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \frac{\cos\left[\omega\left(k + \frac{\sqrt{p_n + q_n \cos\theta} - r}{c} f_s\right) + \gamma\right]}{\sqrt{p_n + q_n \cos\theta}} d\omega d\theta \ , \tag{H.1}$$

which is required for computing the passband error in the near-field case (cf. Appendix G.3), with

$$\begin{array}{ll} p_n = r^2 + d_n^2 & q_n = 2rd_n \\ p_m = r^2 + d_m^2 & q_m = 2rd_m \\ K = \dfrac{r^2}{\sqrt{p_m + q_m \cos\theta_c}} & \gamma = -\omega_c\Big(l + \dfrac{\sqrt{p_m + q_m \cos\theta_c} - r}{c} f_s\Big) \end{array} \tag{H.2}$$

or with the indices n-m and k-l interchanged. The integral I can be calculated as

$$I = I_\theta(\omega_2) - I_\theta(\omega_1) \ , \tag{H.3}$$

with

$$\boxed{I_\theta(\omega) = K \int_{\theta_1}^{\theta_2} \frac{\sin\left[\omega\left(k + \frac{\sqrt{p_n + q_n \cos\theta} - r}{c} f_s\right) + \gamma\right]}{\sqrt{p_n + q_n \cos\theta}\left(k + \frac{\sqrt{p_n + q_n \cos\theta} - r}{c} f_s\right)} d\theta} \tag{H.4}$$

Normally this integral can be computed numerically without any problem (e.g. using the MATLAB commands quad or quad8), but some special cases occur.

CASE 1: $d_n = 0$, $k \neq 0$

If $d_n = 0$, then $\sqrt{p_n + q_n \cos\theta} = r$, such that the integral I now reduces to

$$\begin{aligned} I &= \frac{K}{r} \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \cos(\omega k + \gamma) d\omega d\theta & \text{(H.5)} \\ &= \frac{K}{r} \cdot \frac{\sin(\omega_2 k + \gamma) - \sin(\omega_1 k + \gamma)}{k} \cdot (\theta_2 - \theta_1) \ . & \text{(H.6)} \end{aligned}$$

CASE 2: $d_n = 0$, $k = 0$

The integral I now reduces to

$$I = \frac{K}{r} \int_{\theta_1}^{\theta_2} \int_{\omega_1}^{\omega_2} \cos\gamma \, d\omega d\theta = \frac{K}{r} \cdot \cos\gamma \cdot (\theta_2 - \theta_1) \cdot (\omega_2 - \omega_1) \ . \tag{H.7}$$

CASE 3: $\exists\, \theta_n \in]\theta 1, \theta 2[, k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c} f_s = 0$

The singularity θ_n in the denominator will occur when

$$\sqrt{p_n + q_n \cos\theta_n} = r - \frac{kc}{f_s} \triangleq \xi \;, \tag{H.8}$$

i.e.

$$\cos\theta_n = \frac{\xi^2 - p_n}{q_n} = \frac{\left(\frac{kc}{f_s}\right)^2 - 2\frac{kc}{f_s}r - d_n^2}{2rd_n} \triangleq \nu \tag{H.9}$$

Therefore a singularity will only occur if both conditions,

$$\xi \geq 0, \quad |\nu| \leq 1, \tag{H.10}$$

are satisfied. Because all considered functions are symmetric in θ, we do not need to consider negative angles θ and can assume $0 \leq \theta \leq \pi$. Therefore $\sin\theta_n$ will always be positive and can be written as

$$\sin\theta_n = \sqrt{1 - \cos^2\theta_n} = \sqrt{1 - \nu^2} \;. \tag{H.11}$$

Because of the singularity θ_n in the denominator, numerically calculating the integral $I_\theta(\omega)$ can give rise to some numerical problems (assuming that $\gamma \neq 0$). Similarly as for the far-field case, we can derive a function $g(\theta)$ which is an approximation for the function

$$\boxed{f(\omega,\theta) = K \frac{\sin\left[\omega\left(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c} f_s\right) + \gamma\right]}{\sqrt{p_n + q_n\cos\theta}\left(k + \frac{\sqrt{p_n+q_n\cos\theta}-r}{c} f_s\right)}} \tag{H.12}$$

around the singularity θ_n. Using (F.14), the Taylor expansion of $\cos\theta$ around θ_n is equal to

$$\begin{aligned}\cos\theta &= \cos\theta_n \cdot \cos(\theta - \theta_n) - \sin\theta_n \cdot \sin(\theta - \theta_n) && \text{(H.13)}\\ &= \cos\theta_n - \sin\theta_n \cdot (\theta - \theta_n) + \mathcal{O}\big((\theta - \theta_n)^2\big) \;, && \text{(H.14)}\end{aligned}$$

such that

$$\begin{aligned}\sqrt{p_n + q_n\cos\theta} &= \sqrt{p_n + q_n\cos\theta_n - q_n\sin\theta_n \cdot (\theta - \theta_n) + \mathcal{O}\big((\theta - \theta_n)^2\big)} && \text{(H.15)}\\ &= \xi\sqrt{1 - \frac{q_n\sin\theta_n}{\xi^2} \cdot (\theta - \theta_n) + \mathcal{O}\big((\theta - \theta_n)^2\big)} && \text{(H.16)}\\ &= \xi\left[1 - \frac{q_n\sin\theta_n}{2\xi^2} \cdot (\theta - \theta_n) + \mathcal{O}\big((\theta - \theta_n)^2\big)\right] \;, && \text{(H.17)}\end{aligned}$$

and

$$k + \frac{\sqrt{p_n + q_n\cos\theta} - r}{c} f_s = k + \frac{f_s}{c}\left[\xi - \frac{q_n\sin\theta_n}{2\xi} \cdot (\theta - \theta_n) - r\right] + \mathcal{O}\big((\theta - \theta_n)^2\big)$$

$$= k - \frac{f_s}{c}\left[\frac{kc}{f_s} + \frac{q_n \sin\theta_n}{2\xi}\cdot(\theta-\theta_n)\right] + \mathcal{O}\big((\theta-\theta_n)^2\big) \quad \text{(H.18)}$$

$$= -\frac{f_s}{c}\,\frac{q_n \sin\theta_n}{2\xi}\cdot(\theta-\theta_n) + \mathcal{O}\big((\theta-\theta_n)^2\big)\ . \quad \text{(H.19)}$$

In a first approximation the function $f(\omega,\theta)$ can be approximated by

$$K\frac{\sin\left[\omega\Big(-\frac{f_s}{c}\frac{q_n\sin\theta_n}{2\xi}\cdot(\theta-\theta_n)\Big)+\gamma\right]}{\xi\left(-\frac{f_s}{c}\frac{q_n\sin\theta_n}{2\xi}\cdot(\theta-\theta_n)\right)}\ , \quad \text{(H.20)}$$

which can be further simplified (since we assumed $\gamma \neq 0$) to

$$\boxed{g(\theta) = -K\frac{c}{f_s}\frac{2\sin\gamma}{q_n\sin\theta_n\cdot(\theta-\theta_n)}} \quad \text{(H.21)}$$

Note that $g(\theta)$ is independent of ω. We now define the function

$$\bar{f}(\omega,\theta) = f(\omega,\theta) - g(\theta)\ , \quad \text{(H.22)}$$

and will show that $\lim\limits_{\theta\to\theta_n} \bar{f}(\omega,\theta)$ is finite. The function $f(\omega,\theta)$ can be written as

$$f(\omega,\theta) \;=\; K\frac{\sin\left[\omega\Big(k+\frac{\sqrt{p_n+q_n\cos\theta}-r}{c}f_s\Big)+\gamma\right]}{\sqrt{p_n+q_n\cos\theta}\left(k+\frac{\sqrt{p_n+q_n\cos\theta}-r}{c}f_s\right)} \quad \text{(H.23)}$$

$$= K\frac{c}{f_s}\frac{\sin\left(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\right)}{p_n+q_n\cos\theta-\xi\sqrt{p_n+q_n\cos\theta}}\ , \quad \text{(H.24)}$$

such that the function $\bar{f}(\omega,\theta)$ can be written as

$$\bar{f}(\omega,\theta) \;=\; K\frac{c}{f_s}\left[\frac{\sin\left(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\right)}{p_n+q_n\cos\theta-\xi\sqrt{p_n+q_n\cos\theta}}+\frac{2\sin\gamma}{q_n\sin\theta_n(\theta-\theta_n)}\right]$$

$$= K\frac{c}{f_s}\left[\frac{\sin\left(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\right)q_n\sin\theta_n(\theta-\theta_n)}{\big(p_n+q_n\cos\theta-\xi\sqrt{p_n+q_n\cos\theta}\big)\,q_n\sin\theta_n(\theta-\theta_n)}\right.$$

$$\left.+\frac{2\sin\gamma\,\big(p_n+q_n\cos\theta-\xi\sqrt{p_n+q_n\cos\theta}\big)}{\big(p_n+q_n\cos\theta-\xi\sqrt{p_n+q_n\cos\theta}\big)\,q_n\sin\theta_n(\theta-\theta_n)}\right] \quad \text{(H.25)}$$

$$= K\frac{c}{f_s}\frac{g_D(\theta)}{g_N(\theta)}\ . \quad \text{(H.26)}$$

Since

$$\lim_{\theta\to\theta_n}\bar{f}(\omega,\theta) = K\frac{c}{f_s}\frac{g_D(\theta_n)}{g_N(\theta_n)} = \frac{0}{0}\ , \quad \text{(H.27)}$$

we need L'Hôpital's rule to calculate the limit, i.e.

$$\lim_{\theta\to\theta_n}\frac{g_D(\theta)}{g_N(\theta)}=\frac{\lim\limits_{\theta\to\theta_n}g'_D(\theta)}{\lim\limits_{\theta\to\theta_n}g'_N(\theta)}\ . \tag{H.28}$$

The first-order derivatives $g'_D(\theta)$ and $g'_N(\theta)$ are equal to

$$\begin{aligned} g'_D(\theta) \;=\; & -\frac{q_n^2\sin\theta_n}{2}\frac{\omega f_s}{c}(\theta-\theta_n)\cos\Big(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\Big)\cdot \\ & \frac{\sin\theta}{\sqrt{p_n+q_n\cos\theta}}+q_n\sin\theta_n\sin\Big(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\Big) \\ & -2q_n\sin\gamma\sin\theta\Big(1-\frac{\xi}{2\sqrt{p_n+q_n\cos\theta}}\Big) \end{aligned} \tag{H.29}$$

$$\begin{aligned} g'_N(\theta) \;=\; & -q_n^2\sin\theta_n(\theta-\theta_n)\sin\theta\Big(1-\frac{\xi}{2\sqrt{p_n+q_n\cos\theta}}\Big) \\ & +q_n\sin\theta_n\big(p_n+q_n\cos\theta-\xi\sqrt{p_n+q_n\cos\theta}\big) \end{aligned} \tag{H.30}$$

Since $g'_D(\theta_n)=g'_N(\theta_n)=0$, we need L'Hôpital's rule once more and have to calculate the second-order derivatives $g''_D(\theta)$ and $g''_N(\theta)$,

$$\begin{aligned} g''_D(\theta) \;=\; & -\frac{q_n^2\sin\theta_n}{2}\frac{\omega f_s}{c}\left[\frac{\omega f_s}{c}\sin\Big(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\Big)\cdot\right. \\ & \frac{q_n(\theta-\theta_n)\sin^2\theta}{p_n+q_n\cos\theta}+\frac{\cos\Big(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\Big)}{\sqrt{p_n+q_n\cos\theta}}\cdot \\ & \left.\left(\sin\theta+(\theta-\theta_n)\cos\theta+\frac{q_n(\theta-\theta_n)\sin^2\theta}{4(p_n+q_n\cos\theta)}\right)\right] \\ & -q_n^2\frac{\omega f_s}{c}\sin\theta_n\cos\Big(\omega\frac{f_s}{c}\big(\sqrt{p_n+q_n\cos\theta}-\xi\big)+\gamma\Big)\frac{\sin\theta}{2\sqrt{p_n+q_n\cos\theta}} \\ & -2q_n\sin\gamma\left[\cos\theta\Big(1-\frac{\xi}{2\sqrt{p_n+q_n\cos\theta}}\Big)-\frac{q_n\xi\sin^2\theta}{4(p_n+q_n\cos\theta)^{3/2}}\right] \end{aligned} \tag{H.31}$$

$$\begin{aligned} g''_N(\theta) \;=\; & -2q_n^2\sin\theta\sin\theta_n\Big(1-\frac{\xi}{2\sqrt{p_n+q_n\cos\theta}}\Big)-q_n^2\sin\theta_n(\theta-\theta_n)\cdot \\ & \left[\cos\theta\Big(1-\frac{\xi}{2\sqrt{p_n+q_n\cos\theta}}\Big)-\frac{q_n\xi\sin^2\theta}{4(p_n+q_n\cos\theta)^{3/2}}\right]\ . \end{aligned} \tag{H.32}$$

The second-order derivatives $g''_D(\theta)$ and $g''_N(\theta)$ evaluated at θ_n give

$$g''_D(\theta_n) \;=\; -\frac{q_n^2\sin^2\theta_n\cos\gamma}{\xi}\frac{\omega f_s}{c}-q_n\sin\gamma\left[\cos\theta_n-\frac{q_n\sin^2\theta_n}{2\xi^2}\right] \tag{H.33}$$

$$g''_N(\theta_n) \;=\; -q_n^2\sin^2\theta_n\ , \tag{H.34}$$

such that

$$\boxed{\lim_{\theta\to\theta_n}\bar{f}(\omega,\theta) = K\frac{c}{f_s}\frac{g_D''(\theta_n)}{g_N''(\theta_n)} = K\frac{\omega\cos\gamma}{\xi} + K\sin\gamma\frac{c}{f_s}\left[\frac{\cos\theta_n}{q_n\sin^2\theta_n} - \frac{1}{2\xi^2}\right]} \tag{H.35}$$

Since for any γ, $\lim\limits_{\theta\to\theta_n}\bar{f}(\omega,\theta)$ is finite, the function $\bar{f}(\omega,\theta)$ can be integrated numerically without any problem, and because $g(\theta)$ is independent of ω, the total integral I can be written as

$$I = I_\theta(\omega_2) - I_\theta(\omega_1) = \int_{\theta_1}^{\theta_2}\bar{f}(\omega_2,\theta)\,d\theta - \int_{\theta_1}^{\theta_2}\bar{f}(\omega_1,\theta)\,d\theta\ . \tag{H.36}$$

H.1 Far-field assumptions

When $r\to\infty$, i.e. for far-field, it can be shown that the near-field equations reduce to the far-field equations derived in Appendix F.

Singularity

For the singularity θ_n defined in (H.9),

$$\lim_{r\to\infty}\cos\theta_n = \lim_{r\to\infty}\frac{\left(\frac{kc}{f_s}\right)^2 - 2\frac{kc}{f_s}r - d_n^2}{2rd_n} = -\frac{kc}{f_sd_n}\ , \tag{H.37}$$

which corresponds to (F.7) in the far-field case.

Approximating function

For the approximating function $g(\theta)$ defined in (H.21),

$$\begin{aligned}\lim_{r\to\infty}g(\theta) \quad &= \quad \lim_{r\to\infty} -\frac{r^2}{\sqrt{r^2+d_m^2+2rd_m\cos\theta_c}}\frac{c}{f_s}\frac{2\sin\gamma}{2rd_n\sqrt{1-\cos^2\theta_n}\cdot(\theta-\theta_n)}\\ &= \quad -\frac{\sin\gamma}{\frac{f_s}{c}\,d_n\sqrt{1-\left(\frac{kc}{f_sd_n}\right)^2}\cdot(\theta-\theta_n)}\ ,\end{aligned} \tag{H.38}$$

which corresponds to (F.17) in the far-field case.

Limit value

For the limit value $L = \lim\limits_{\theta\to\theta_n}\bar{f}(\omega,\theta)$ defined in (H.35),

$$\begin{aligned}\lim_{r\to\infty}L &= \lim_{r\to\infty}\frac{r^2}{\sqrt{r^2+d_m^2+2rd_m\cos\theta_c}}\left(\frac{\omega\cos\gamma}{\xi}+\sin\gamma\frac{c}{f_s}\left[\frac{\cos\theta_n}{q_n\sin^2\theta_n}-\frac{1}{2\xi^2}\right]\right)\\ &= \lim_{r\to\infty}r\left(\frac{\omega\cos\gamma}{r-\frac{kc}{f_s}}+\sin\gamma\frac{c}{f_s}\frac{\cos\theta_n}{2rd_n(1-\cos^2\theta_n)}-\sin\gamma\frac{c}{f_s}\frac{1}{2\left(r-\frac{kc}{f_s}\right)^2}\right)\\ &= \omega\cos\gamma-\sin\gamma\frac{c}{f_s}\frac{\frac{kc}{f_sd_n}}{2d_n\left[1-\left(\frac{kc}{f_sd_n}\right)^2\right]} = \omega\cos\gamma+\frac{k\sin\gamma}{2\left[k^2-\left(\frac{f_sd_n}{c}\right)^2\right]}\ ,\end{aligned}$$

which corresponds to (F.32) in the far-field case.

List of Publications

Book Chapter

1. S. Doclo and M. Moonen, *GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement*, chapter 6 in “Microphone Arrays: Signal Processing Techniques and Applications” (Brandstein, M. S. and Ward, D. B., Eds.), pp. 111–132, Springer-Verlag, May 2001.

International Journal Papers

1. S. Doclo and M. Moonen, “GSVD-based optimal filtering for single and multimicrophone speech enhancement,” *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

2. S. Doclo and M. Moonen, “Multi-Microphone Noise Reduction Using Recursive GSVD-Based Optimal Filtering with ANC Postprocessing Stage,” *Accepted for publication (minor revision) in IEEE Trans. Speech and Audio Processing*, 2003.

3. S. Doclo and M. Moonen, “Design of far-field and near-field broadband beamformers using eigenfilters,” *Accepted for publication (minor revision) in Signal Processing*, 2003.

4. S. Doclo and M. Moonen, “Design of robust broadband beamformers for gain and phase errors in the microphone array characteristics,” *Accepted for publication in IEEE Trans. Signal Processing*, 2003.

5. S. Doclo and M. Moonen, “Robust adaptive time delay estimation for speaker localisation in noisy and reverberant acoustic environments,” *Submitted to EURASIP Journal on Applied Signal Processing, special issue on Signal Processing for Acoustic Communication Systems*, Sept. 2002.

National Journal Papers

1. S. Doclo and E. De Clippel, “Hoort u dat wel ?,” *Het Ingenieursblad*, vol. 67, no. 10, pp. 38–45, Oct. 1998.

International Conference Papers

1. S. Doclo, I. Dologlou, and M. Moonen, "A novel iterative signal enhancement algorithm for noise reduction in speech," in *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, Sydney, Australia, Dec. 1998, pp. 1435–1438.

2. S. Doclo and M. Moonen, "Robustness of SVD-based Optimal Filtering for Noise Reduction in Multi-Microphone Speech Signals," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Pocono Manor, Pennsylvania, USA, Sept. 1999, pp. 80–83.

3. S. Doclo and M. Moonen, "SVD-based optimal filtering with applications to noise reduction in speech signals," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, Oct. 1999, pp. 143–146.

4. S. Doclo and M. Moonen, "Noise Reduction in Multi-Microphone Speech Signals using Recursive and Approximate GSVD-based Optimal Filtering," in *Proc. of the IEEE Benelux Signal Proc. Symposium (SPS2000)*, Hilvarenbeek, The Netherlands, Mar. 2000.

5. S. Doclo, E. De Clippel, and M. Moonen, "Combined Acoustic Echo and Noise Reduction using GSVD-based Optimal Filtering," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, vol. 2, pp. 1061–1064.

6. S. Doclo, E. De Clippel, and M. Moonen, "Multi-microphone noise reduction using GSVD-based optimal filtering with ANC postprocessing stage," in *Proc. of DSP2000 Workshop*, Hunt TX, USA, Oct. 2000, pp. 383–388.

7. S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001, pp. 31–34.

8. S. Doclo and M. Moonen, "Robust time-delay estimation in highly adverse acoustic environments," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz NY, USA, Oct. 2001, pp. 59–62.

9. S. Doclo and M. Moonen, "Comparison of least-squares and eigenfilter techniques for broadband beamforming," in *Proc. of the IEEE Benelux Signal Processing Symposium (SPS2002)*, Leuven, Belgium, Mar. 2002, pp. 73–76.

10. S. Doclo and M. Moonen, "Design of far-field broadband beamformers using eigenfilters," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Toulouse, France, Sept. 2002, pp. III 237–240.

11. S. Doclo and M. Moonen, "Design of broadband speech beamformers robust against errors in the microphone array characteristics," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong SAR, China, Apr. 2003, pp. V 473–476.

12. S. Doclo and M. Moonen. "Design of broadband beamformers robust against microphone position errors," Submitted to *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003.

Abstracts

1. S. Doclo and M. Moonen, "SVD-based signal enhancement techniques for noise reduction in speech," in *NATO Advanced Study Institute : Signal Processing for Multimedia*, Castelvecchio Pascoli, Italy, July 1998.

2. S. Doclo and M. Moonen, "Multi-microphone noise reduction using GSVD-based optimal filtering," in *International Workshop on Microphone Array Systems*, Boston MA, USA, Oct. 2000.

3. S. Doclo and M. Moonen, "Design of robust broadband beamformers for speech applications," in *International Workshop on Microphone Array Systems*, Erlangen, Germany, May 2003.

Internal Reports

1. S. Doclo, I. Dologlou, and M. Moonen, "A novel iterative signal enhancement algorithm for noise reduction in speech," Tech. Rep. ESAT-SISTA/TR 1998-26, ESAT, Katholieke Universiteit Leuven, Belgium, Apr. 1998.

2. S. Doclo and M. Moonen, "SVD-based optimal filtering with applications to noise reduction in speech signals," Tech. Rep. ESAT-SISTA/TR 1999-33, ESAT, Katholieke Universiteit Leuven, Belgium, Apr. 1999.

3. K. Eneman, S. Doclo, and M. Moonen, "A comparison between iterative block-LMS and Partial Rank Algorithm," Tech. Rep. ESAT-SISTA/TR 1999-52, ESAT, Katholieke Universiteit Leuven, Belgium, Mar. 1999.

4. S. Doclo, K. Eneman, and M. Moonen, "Voice activity detection," Tech. Rep. ESAT-SISTA/TR 2001-62, ESAT, Katholieke Universiteit Leuven, Belgium, Jan. 2002.

5. S. Doclo and M. Moonen, "Far-field and near-field broadband beamformer design," Tech. Rep. ESAT-SISTA/TR 2002-109, ESAT, Katholieke Universiteit Leuven, Belgium, July 2002.

Curriculum Vitae

Simon Doclo was born in Wilrijk, Belgium, in 1974. In 1997 he received the electrical engineering degree (magna cum laude) from the Katholieke Universiteit Leuven, Belgium. From 1997 he has been employed as a research assistant at the Department of Electrical Engineering, KU Leuven, under supervision of Prof. dr. ir. Marc Moonen. His research interests are in the area of digital signal processing techniques for speech and audio applications.

From 1997 until 2002 he was funded by the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). He has been involved in the I.W.T. projects 'Multi-microphone Signal Enhancement Techniques for handsfree telephony and voice-controlled systems (MUSETTE)' in cooperation with Philips ITCL, and 'Performance improvement of cochlear implants by innovative speech processing algorithms' in cooperation with Cochlear Technology Centre Europe.

In 1998 he received the 1st prize 'KVIV-Studentenprijzen' (with Erik De Clippel) for his MSc thesis and in 2001 he received a Best Student Paper Award at the IEEE International Workshop on Acoustic Echo and Noise Control. He has been secretary of the IEEE Benelux Signal Processing Chapter (1997-2002).