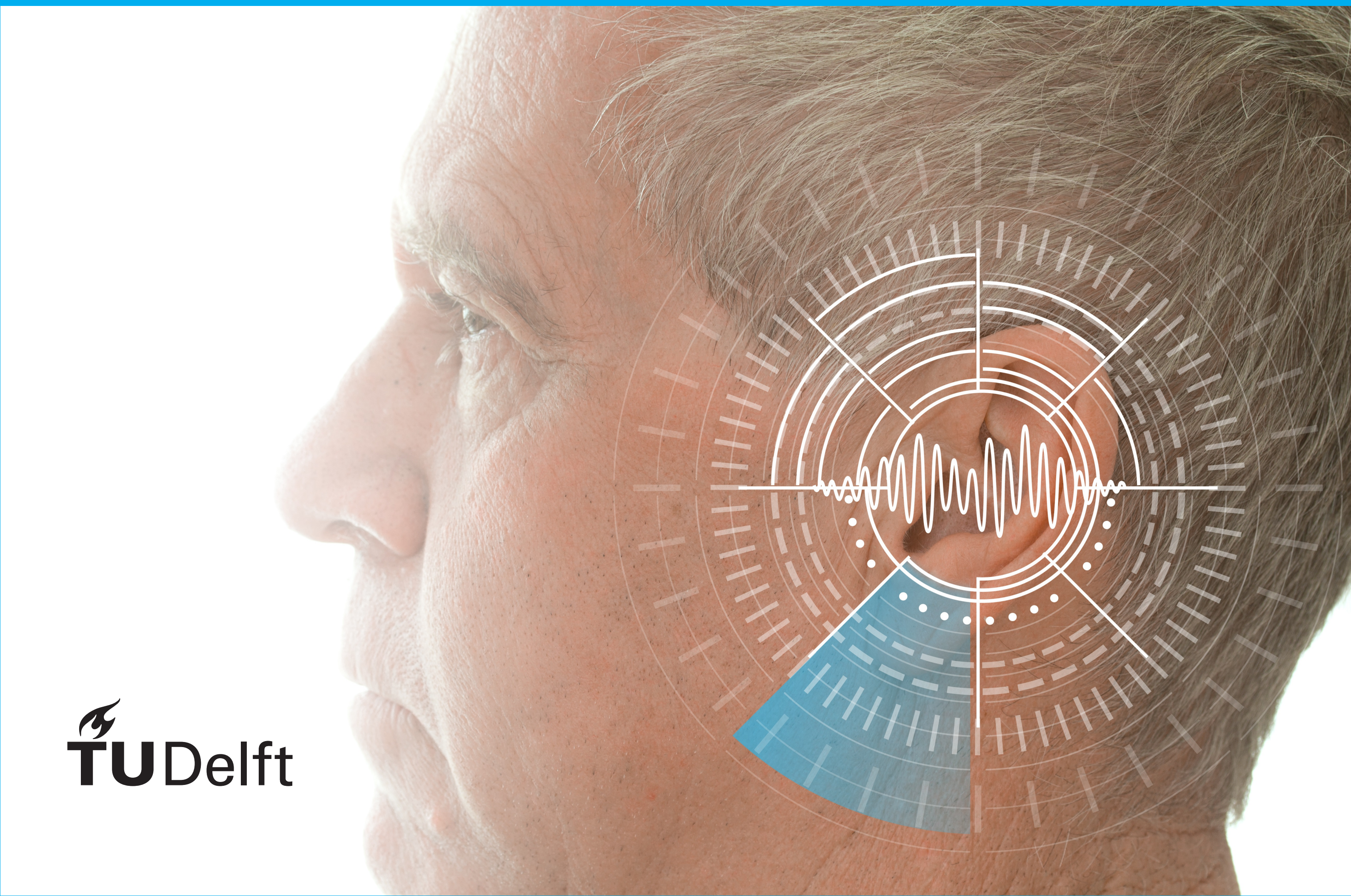
# Automatic Interferer Selection for Binaural Beamforming

## Master of Science Thesis

Costas A. Kokke



TU Delft

# Interferer selection for binaural cue preservation in joint binaural linearly constrained minimum variance beamforming

## Master of Science Thesis

by


Costas A. Kokke


in partial fulfilment of the requirements for the degree of

**Master of Science**
in Electrical Engineering

at the Delft University of Technology,
to be defended publicly on Thursday August 30, 2018 at 10:00 AM.

| | | |
|---|---|---|
| Student number: | 1505254 | |
| Specialisation: | Audio and Acoustic Signal Processing | |
| Faculty: | Electrical Engineering, Mathematics and Computer Science | |
| Project duration: | Januari 15, 2017 – August 30, 2018 | |
| Thesis committee: | Dr. ir. R. Heusdens, | TU Delft, Circuits and Systems group, chair |
| | Dr. ir. R. C. Hendriks, | TU Delft, Circuits and Systems group, supervisor |
| | Dr. A. Endo, | TU Delft, Terahertz Sensing group |
| | Ir. A. Koutrouvelis, | TU Delft, Circuits and Systems group |


An electronic version of this thesis is available at http://repository.tudelft.nl/.


TU Delft

# Abstract

Interferer selection for binaural cue preservation in joint binaural linearly constrained minimum variance beamforming

by

Costas A. Kokke
Master of Science in Electrical Engineering
Delft University of Technology, 2018
Dr.ir. Richard C. Hendriks, Supervisor

Spatial cues allow a listener to determine the direction sound is coming from. In addition, recognising spatially separated sound sources facilitate the listener to focus on specific sound sources. Because of this, preservation of spatial cues in multi-microphone hearing assistive devices is important to the listening experience and safety of the user. A number of linearly-constrained-minimum-variance-based methods exist for this purpose. Most of these are limited in the number of interfering sources for which they can preserve the spatial cues. In this thesis, a method of selecting the most important interfering point sources using convex optimisation is proposed. The method is presented based on two different convex relaxations, which are compared, using simulation experiments, to existing, exhaustive search and randomised methods in terms of noise suppression and localisation errors. Both methods are shown to improve the performance of the joint binaural linearly constrained minimum variance beamformer, an existing method for simultaneous noise reduction and spatial cue preservation, by giving it more degrees of freedom for noise reduction and allowing it to handle a larger number of (virtual) sources present in the scene.

# Contents

# List of Figures

# Preface

This thesis discusses two methods that solve the problem: Which interfering point sources should be selected for spatial cues preservation when using the joint binaural linearly constrained minimum variance (JBLCMV) beamformer? The JBLCMV has a limit on the number of sources for which it can do spatial cue preservation. Being able to select the optimal interfering sources for spatial cue preservation can greatly improve the listening experience for hearing assistive device users, as it allows the user to hear important sounds from their correct spatial origins. The proposed methods allow the JBLCMV to only constrain sources that are important to constrain. If the amount of sources that are important would exhaust the degrees of freedom of the filter, it allows the JBLCMV to constrain the sources that are most important, while still being able to do noise reduction. This allows the JBLCMV to be used well with pre-determined acoustic transfer functions, through the proposed framework. The performance of the proposed methods is evaluated by noise reducing capabilities and localisation errors, through simulation experiments.

The thesis report is written as part of the master of science graduation procedure within the Circuits and Systems group at the Delft University of Technology. The thesis committee, dr. ir. R. Heusdens (chair), dr. ir. R. C. Hendriks (supervisor), dr. A. Endo and ir. A. Koutrouvelis will judge the complete thesis project after it is defended on August 30, 2018.

## Acknowledgements

I would like to thank dr. ir. R. C. Hendriks and ir. A. Koutrouvelis for their supervision and advice over the course of the project, and dr. ir. R. Heusdens for his feedback and suggestions on the presentation and explanation of the proposed methods.

Part of the text on the BMVDR, GBLCMV, JBLCMV and pre-determined relative acoustic transfer functions in Sections 1.2.2 to 1.2.5, respectively, was written as part of a literature study written prior to starting the thesis project. The literature study was done as part of the free elective space of the Electrical Engineering master for 2 ECTS as the course *ET4399 Extra Project*.

*Costas A. Kokke*
*Delft, August 2018*



Circuits and Systems, TU Delft

# 1 Introduction

The use of hearing assistive devices (HADs) has significantly increased in our ageing society and the need for them will continue to increase, partly due to increased recreational exposure to loud sound [1]. The main goal of these devices is to compensate for the user's hearing loss and to improve speech intelligibility. Hearing loss compensation has some negative side effects though, as disturbing sounds that were inaudible before, suddenly become audible to the user. People with hearing loss also report difficulty with spatial source separation [2], the ability to separate their source of interest from environmental sounds and other interferers. This is due to the decreased temporal and spatial resolution of their hearing, and the compressive nature of the compensation done in existing HADs often amplifies this problem.

When a person suffers from hearing loss, it is not only the threshold of hearing that goes up, but the threshold of feeling also goes down. This means that the range of sound intensities a HAD can work with reduces as the hearing loss of the user progresses, causing the compression of sound, making it harder for the user to distinguish sound by intensity. This is one important aspect in spatial source separation. Therefore, modern HADs come with noise reduction algorithms, to reduce the loudness of environmental sounds.

Modern HADs typically come in pairs, called bilateral or binaural HADs, each equipped with multiple microphones. These microphones can be used to perform beamforming. Beamforming combines multiple microphone recordings, after properly changing their magnitudes and phases, resulting in an estimate of the target source. This is typically referred to as a binaural estimation of the target, meaning two signals are generated. One for the left and one for the right ear.

An important aspect of binaural HADs is their ability to preserve the spatial cues of the sound field. This helps make the output of the HADs appear more natural for the user. One reason to do this is the importance for the HAD user to localise sound in day-to-day situations, for example in traffic. Another reason is the ability of the human auditory system to distinguish spatially separated sources and to focus on sources by their location, perceptually attenuating other sources. For example, having a conversation with someone in front of you while being in a loud environment is possible because of that ability.

As mentioned before, beamforming is performed by changing the magnitudes and phases of the different microphone recordings before combining them. However, distorting the original magnitude and phase can potentially distort the spatial cues, as these are partly determined by the magnitude and phase relationships of the signals presented to the two ears. Practically, this could for example mean that the HAD user cannot localise traffic sounds and it may sound as if they heard a truck or car was coming towards them from a different direction. Such spatial cue distortions may have considerable consequences and should be prevented.

To this end, binaural beamformers are developed for simultaneous noise reduction and spatial cue preservation. In this thesis, only distortionless binaural beamformers are considered. Distortionless methods are those where the target is always undistorted. This is beneficial for the speech intelligibility. Most binaural beamforming techniques are limited in the amount of spatial cues they can retain, as the degrees of freedom for noise reduction and spatial cue preservation are limited and depend on the number of microphones. In general, noise reduction and spatial cue preservation come with a trade off of the available degrees of freedom. Spending all degrees of freedom on interferer spatial cue preservation implies no controlled noise reduction and vice versa. An important aspect of binaural noise reduction is therefore to be efficient in the use of constraints in order to preserve spatial cues.

Existing binaural noise reduction methods spend these degrees of freedom in different manners. The binaural minimum variance distortionless response (BMVDR) beamformer, for example, is a beamformer that spends all degrees of freedom on noise reduction, and thus none on spatial cue preservation. This causes all sound to appear to come from the direction of the target signal [3]. The BMVDR can be generalised to the general binaural linearly constrained minimum variance (GBLCMV) beamformer, which can have at most $2M-1$ constraints and do controlled noise reduction, where $M$ is the amount of microphones on both HADs. Typically, two constraints are used to binaurally constrain the target source, leaving $2M-3$ constraints to be used for other purposes. The joint binaural linearly constrained minimum variance (JBLCMV) beamformer is a distortionless binaural beamformer [3, 4] that fits in the GBLCMV framework. It uses one constraint to binaurally constrain an interfering point source, meaning is can at most binaurally constrain $2M-3$ interfering point sources, while still being able to do noise suppression.

These methods use constraints for each interferer in all time-frequency tiles, even in the tiles where an interferer might be inaudible, due to suppression by the beamformer, or due to masking by other sources. Since the amount of microphones on current hardware is low, typically $M = 4$, degrees of freedom are scarce. This not only means that degrees of freedom are quickly exhausted, but it also means that degrees of freedom are spent on inperceptual, and thus unnecessary, spatial cue constraints in some time-frequency tiles. To be able to take into account even more interfering sources, while still being able to do controlled noise reduction, a more efficient way to constrain the spatial cues of interfering sources is required. This thesis project aims to find such a method by automatic optimal interferer subset selection for JBLCMV, by choosing the optimal subset of known interferers for spatial cue preservation. The goals will be detailed fully in Section 1.3. Previous work on binaural beamforming that is relevant to this thesis will be discussed in Section 1.2. In Chapter 2 the problem will be formally introduced and the chapter will continue with methods of solving it. The validity of the methods will be experimentally shown in Chapter 3 with the proposed methods being compared to the methods introduced in Section 1.2. A discussion of the proposed methods and suggested future work will be discussed in Chapter 4. However, first Section 1.1 will lay out the signal model that will be used throughout this thesis.

## 1.1. Signal Model

Consider the binaural hearing aid setting, with two collaborating hearing aids that have a combined total of $M$ microphones installed. In this thesis, the signals are assumed to be processed on a frame-by-frame basis in the frequency domain. Since processing takes place independently per frame, time-frame indices are omitted for convenience.

Assuming an additive distortion model, a Fourier coefficient, $y_j[k]$, at the $j$th microphone is composed as follows,

$$y_j[k] = a_j[k]s[k] + \sum_{i=1}^{r} b_{ij}[k]u_i[k] + v_j[k]\,, \tag{1.1}$$

where $a_j[k]$ and $b_{ij}[k]$ are the acoustic transfer functions (ATFs) of the desired source and $i$th interferer to the $j$th microphone respectively, $s[k]$ and $u_i[k]$ are the desired source and $i$th interferer respectively, $r$ is the number of interferers, and $v_j[k]$ is additive uncorrelated noise. In the remainder of this work, the frequency variable $k$ will be omitted for simplicity, since all processing will be done per frequency bin, assuming frequency bins are mutually independent. All microphone signals can be combined using the following vector notation:

$$\begin{aligned}
\mathbf{y} &= \begin{bmatrix} y_1 & y_2 & \cdots & y_M \end{bmatrix}^T \in \mathbb{C}^M && (1.2a)\\
&= \mathbf{a}s + \sum_{i=1}^{r} \mathbf{b}_i u_i + \mathbf{v} && (1.2b)\\
&= \mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{v}\,, && (1.2c)
\end{aligned}$$

where $\mathbf{x} = \mathbf{a}s$, $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_r \end{bmatrix} \in \mathbb{C}^{M\times r}$ and $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \dots & u_r \end{bmatrix}^T$.

Assuming all interferers to be mutually uncorrelated, the cross power spectral density (CPSD) matrix $\mathbf{P}$ of all disturbances is defined as

$$\mathbf{P} = \sum_{i=1}^{r} \mathbf{P}_{\mathbf{n}_i} + \mathbf{P}_{\mathbf{v}} \in \mathbb{C}^{M\times M}\,, \tag{1.3}$$

where $\mathbf{P}_{\mathbf{n}_i}$ and $\mathbf{P}_{\mathbf{v}}$ are the CPSD matrices of $\mathbf{n}_i = \mathbf{b}_i u_i$ and $\mathbf{v}$ respectively. Similarly, the noisy recording CPSD matrix $\mathbf{P}_{\mathbf{y}}$ is composed as

$$\mathbf{P}_{\mathbf{y}} = \mathbf{P}_{\mathbf{x}} + \mathbf{P}\,, \tag{1.4}$$

where $\mathbf{P}_{\mathbf{x}}$ is the CPSD matrix of the target, $\mathbf{x}$.

Additionally, on each hearing aid we define one reference microphone. These microphones are used as a reference with respect to the preservation of the spatial cues of interfering point sources and the complete preservation of the target signal on both the left and right HAD.

## 1.2. Previous Work

Conventional (monaural) beamforming methods in hearing aids treat each ear separately, leading to the potential loss of spatial cues [5]. Binaural beamformers are a class of beamformers that aim to (partially) preserve the spatial cues of at least the target signal. Spatial cues are discussed in further detail in Section 1.2.1. Two existing binaural beamforming methods will be discussed in Sections 1.2.2 and 1.2.4. The general framework, the general binaural linearly constrained minimum variance (GBLCMV), in which they fit is discussed in Section 1.2.3.

### 1.2.1. Spatial Cues

To retain spatial cues of point sources, the interaural level difference (ILD) and interaural time difference (ITD) should be preserved. When working in the frequency domain, these correspond to the ILD and interaural phase difference (IPD). These can be obtained by defining the input and output interaural transfer functions (ITFs) as

$$\text{ITF}_{\mathbf{x}}^{\text{in}} = \frac{a_L}{a_R}, \qquad \text{ITF}_{\mathbf{x}}^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{a}}{\mathbf{w}_R^H \mathbf{a}}, \tag{1.5}$$

respectively, where $\mathbf{w}_L$ and $\mathbf{w}_R$ contain the beamformer coefficients for the left and right ear respectively, and $a_L$ and $a_R$ are the ATFs of the target source to the left and right reference microphone respectively. The ILD and IPD are then simply the magnitude squared and phase of the ITF respectively:

$$\text{ILD}_{\mathbf{x}}^{\text{in}} = \left|\text{ITF}_{\mathbf{x}}^{\text{in}}\right|^2, \qquad \text{ILD}_{\mathbf{x}}^{\text{out}} = \left|\text{ITF}_{\mathbf{x}}^{\text{out}}\right|^2, \tag{1.6}$$

$$\text{IPD}_{\mathbf{x}}^{\text{in}} = \angle\text{ITF}_{\mathbf{x}}^{\text{in}}, \qquad \text{IPD}_{\mathbf{x}}^{\text{out}} = \angle\text{ITF}_{\mathbf{x}}^{\text{out}}. \tag{1.7}$$

Note that if

$$\text{ITF}_{\mathbf{x}}^{\text{in}} = \text{ITF}_{\mathbf{x}}^{\text{out}}, \tag{1.8}$$

then

$$\text{ILD}_{\mathbf{x}}^{\text{in}} = \text{ILD}_{\mathbf{x}}^{\text{out}} \wedge \text{IPD}_{\mathbf{x}}^{\text{in}} = \text{IPD}_{\mathbf{x}}^{\text{out}}. \tag{1.9}$$

These relations can be used to retain spatial cues for point source interferers. Since this project is constrained to distortionless methods (the target may not be distorted), the transfer functions of the target should be preserved fully, instead of just the ITFs.

To quantify spatial cue errors, the input and output ILD difference and the input and output IPD difference can be used. These measures are defined as

$$\mathcal{E}_i^{\text{ILD}} = \left|\text{ILD}_i^{\text{out}} - \text{ILD}_i^{\text{in}}\right|, \tag{1.10}$$

$$\mathcal{E}_i^{\text{IPD}} = \frac{\left|\text{IPD}_i^{\text{out}} - \text{IPD}_i^{\text{in}}\right|}{\pi}, \tag{1.11}$$

where $\mathcal{E}_i^{\text{ILD}}$ and $\mathcal{E}_i^{\text{IPD}}$ are the ILD and IPD errors for the $i$th interferer, respectively. $\text{ILD}_i^{\text{in}}$ is the input ILD of the $i$the interferer. $\text{ILD}_i^{\text{out}}$, $\text{IPD}_i^{\text{in}}$ and $\text{IPD}_i^{\text{out}}$ are similarly defined. These are always determined using the true ATFs or true relative acoustic transfer functions (RTFs) of the interferer.

### 1.2.2. Minimum Variance Distortionless Response

A popular monaural beamformer is the minimum variance distortionless response (MVDR) beamformer, which also has a binaural variant [3]. It can also be seen as a specific case of the LCMV and binaural LCMV beamformers, as will be shown in Section 1.2.3. The monaural MVDR is determined by minimising the power in all directions, while perfectly retaining the signal in the target direction. That is,

$$\begin{aligned} &\text{minimise} \quad \mathbf{w}^H \mathbf{P} \mathbf{w} \\ &\text{subject to} \quad \mathbf{w}^H \mathbf{a} = a_{ref}, \end{aligned} \tag{1.12}$$

where $a_{ref}$ is the ATF of the target source to the reference microphone, of which there is only one in a monaural beamformer.

This can be extended to a binaural beamformer by constructing two beamformers using the same microphones,

$$\begin{aligned} \text{minimise} \quad & \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \\ \text{subject to} \quad & \mathbf{w}_L^H \mathbf{a} = a_L \\ & \mathbf{w}_R^H \mathbf{a} = a_R \,, \end{aligned} \tag{1.13}$$

where $\mathbf{w}_L$ and $\mathbf{w}_R$ are the beamformer for the left and right ear respectively, and $a_L$ and $a_R$ are the ATFs of the desired source to the left and right reference microphone respectively. The binaural MVDR method only retains the spatial cues of the target signal. As a result, all sound seems to come from the direction of the target, distorting the ITFs of the interferers [3].

If the two beamformers of Equation (1.13) are combined into one concatenated vector, by letting $\mathbf{w}^H = \left[\mathbf{w}_L^H, \mathbf{w}_R^H\right]$, the following equivalent expression is obtained:

$$\begin{aligned} \text{minimise} \quad & \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{C} = \begin{bmatrix} a_L & a_R \end{bmatrix} \\ \text{with} \quad & \tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{bmatrix} \\ & \mathbf{C} = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix} , \end{aligned} \tag{1.14}$$

which is a general (B)LCMV expression, as will be discussed in Section 1.2.3.

### 1.2.3. General Binaural Linearly Constrained Minimum Variance Framework

The LCMV formulation generalises the MVDR beamformer. With MVDR beamforming, only the gain of the target is constrained. The LCMV beamformer generalises this to any number of linear constraints, up to a number determined by the number of receivers, resulting in the following expression [6],

$$\begin{aligned} \text{minimise} \quad & \mathbf{w}^H \mathbf{P} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{C} = \mathbf{f}^H \,, \end{aligned} \tag{1.15}$$

where $\mathbf{C}$ is the constraint matrix and $\mathbf{f}$ is the response vector, which contains the desired responses for the beamformer applied to each of the constraint vectors in $\mathbf{C}$. This reduces to the MVDR if $\mathbf{C} = \mathbf{a}$ and $\mathbf{f}^H = a_{ref}$.

Depending on the number of constraints and the number of microphones available, a number of degrees of freedom are available for noise reduction. If the number of constraints is too large, no solution is possible. In general, given $M$ microphones and $L \leq M$ constraints, there is an analytical solution solution and $M - L$ degrees of freedom are left for noise reduction. When $M = L$, there are no degrees of freedom left for noise reduction, making the solution of the problem independent of the data. The LCMV solutions can be neatly summarised by

$$\mathbf{w} = \begin{cases} \mathbf{P}^{-1}\mathbf{C}\left(\mathbf{C}^H\mathbf{P}^{-1}\mathbf{C}\right)^{-1}\mathbf{f} \,, & \text{if} \quad L < M \\ \mathbf{C}^{-H}\mathbf{f} \,, & \text{if} \quad L = M \\ \text{no solution} \,, & \text{if} \quad L > M \,. \end{cases} \tag{1.16}$$

Similar to the binaural MVDR in Equation (1.13), concatenating the left and right beamformers as $\mathbf{w}^H = \left[\mathbf{w}_L^H, \mathbf{w}_R^H\right]$, the binaural LCMV beamformer can be formulated as

$$\begin{aligned} \text{minimise} \quad & \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{C} = \mathbf{f}^H \,. \end{aligned} \tag{1.17}$$

With the BMVDR $\mathbf{C} = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix}$ and $\mathbf{f}^H = \begin{bmatrix} a_L & a_R \end{bmatrix}$. A maximum of $2M - 1$ constraints can be taken into account if at least one degree of freedom for noise reduction is desired. The solution to the problem is still given by Equation (1.16).

### 1.2.4. The Joint BLCMV Beamformer

The binaural cues of a point source are given by its ILD and IPD, which are expressed by the ratio of magnitudes and phase differences of the received signals in the Fourier domain respectively. As discussed in Section 1.2.1, these can be preserved by preserving the ITFs. Preservation of both the ILD and IPD of $r$ interferers can be expressed as

$$\frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} = \frac{b_{iL}}{b_{iR}}\,, \quad \text{for} \quad i = 1, \ldots, r\,. \tag{1.18}$$

Equation (1.18) must hold for all $r$ interferers to retain the binaural cues. This equation can be equivalently expressed in a linear form, which will be useful when setting up linear constraints. That is,

$$\mathbf{w}_L^H \mathbf{b}_i b_{iR} - \mathbf{w}_R^H \mathbf{b}_i b_{iL} = 0\,, \quad \text{for} \quad i = 1, \ldots, r\,. \tag{1.19}$$

The joint BLCMV is a recently proposed method for binaural beamforming that aims to binaurally constrain as many interfering point sources as possible, while retaining at least some noise reducing capabilities [4]. The problem formulation of the JBLCMV, using Equation (1.19), is then given by:

$$\begin{aligned} &\text{minimise} && \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ &\text{subject to} && \mathbf{w}^H \mathbf{C} = \mathbf{f}^H \\ &\text{with} && \mathbf{C} = \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{b}_1 b_{1R} & \cdots & \mathbf{b}_r b_{rR} \\ \mathbf{0} & \mathbf{a} & -\mathbf{b}_1 b_{1L} & \cdots & -\mathbf{b}_r b_{rL} \end{bmatrix} \\ & && \mathbf{f}^H = \begin{bmatrix} a_L & a_R & 0 & \cdots & 0 \end{bmatrix}. \end{aligned} \tag{1.20}$$

The first two columns of the constraint matrix $\mathbf{C}$ shown in Equation (1.20) ensure that the target is not distorted. Each other column of $\mathbf{C}$ is used for binaural cue preservation of one interferer, as it satisfies Equation (1.19), which is equivalent to preserving the ITFs.

For convenience, the constraint matrix $\mathbf{C}$ and constraint vector $\mathbf{f}$ can be separated into components related to target and interferer preservation by letting

$$\mathbf{C} = \begin{bmatrix} \boldsymbol{\Lambda}_a & \boldsymbol{\Lambda}_b \end{bmatrix} \qquad \in \mathbb{C}^{2M \times (2+r)}\,, \tag{1.21}$$

$$\boldsymbol{\Lambda}_a = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix} \qquad \in \mathbb{C}^{2M \times 2}\,, \tag{1.22}$$

$$\boldsymbol{\Lambda}_b = \begin{bmatrix} \mathbf{b}_1 b_{1R} & \cdots & \mathbf{b}_r b_{rR} \\ -\mathbf{b}_1 b_{1L} & \cdots & -\mathbf{b}_r b_{rL} \end{bmatrix} \in \mathbb{C}^{2M \times r}\,, \tag{1.23}$$

$$\mathbf{f}^H = \begin{bmatrix} \mathbf{f}_a^H & \mathbf{f}_b^H \end{bmatrix} \qquad \in \mathbb{C}^{2+r}\,, \tag{1.24}$$

$$\mathbf{f}_a^H = \begin{bmatrix} a_L & a_R \end{bmatrix} \qquad \in \mathbb{C}^{2}\,, \tag{1.25}$$

$$\mathbf{f}_b = \mathbf{0}_r\,. \tag{1.26}$$

The JBLCMV uses one constraint per interferer. The maximum number of interferers that can be taken into account, while leaving at least one degree of freedom for noise reduction, is $r_{max} = 2M - 3$. The minus three comes from the two constraints to preserve the target on the left and right ears and the one degree of freedom that needs to be available at minimum. More details on the degrees of freedom is given in Section 1.2.3.

Since only the ratio of the transfer functions is considered, it is beneficial to consider relative transfer functions (RTFs). One reason it is beneficial is that the common delay (leading zeros that a set of impulse responses share) is removed when taking the ratio of ATFs. As a result, only the relative delay is considered. This means that the length of the DFT does not need to be as long as the entire impulse response to capture the desired amount of the tail end of the impulse responses. However, in practice RTF or ATF estimation of multiple interferers is rather challenging. Instead of estimating these transfer functions, one could develop algorithms based on pre-determined transfer functions, overcoming the need to estimate RTFs. Both of these subjects will be discussed in further detail in Section 1.2.5.

### 1.2.5. Pre-Determined Relative Acoustic Transfer Functions

In Section 1.2.4 it was specified that Equation (1.18) should hold for binaural cues to be retained. Equation (1.18) can be reformulated as

$$\mathbf{w}_L^H \mathbf{b}_{iL} = \mathbf{w}_R^H \mathbf{b}_{iR}\,, \quad \text{for} \quad i = 1, \ldots, r\,, \tag{1.27}$$

$$\mathbf{b}_{iL} = \frac{\mathbf{b}_i}{b_{iL}}\,, \tag{1.28}$$

where $\mathbf{b}_{iL}$ and $\mathbf{b}_{iR}$ are the left and right RTFs, respectively. Using RTFs instead of ATFs in JBLCMV beamforming turns $\mathbf{\Lambda}_b$ from Equation (1.23) into $\begin{bmatrix} \mathbf{b}_{1L} & \cdots & \mathbf{b}_{rL} \\ -\mathbf{b}_{1R} & \cdots & -\mathbf{b}_{rR} \end{bmatrix}$. The expression for the ITF preservation of the $i$the interfering point source from Equation (1.8) becomes

$$\frac{\mathbf{w}_L^H \mathbf{b}_{iL}}{\mathbf{w}_R \mathbf{b}_{iR}} = 1\,. \tag{1.29}$$

Using RTFs, the spatial cue error measures given by Equations (1.10) and (1.11) requires a redefinition of Equation (1.5) as

$$\text{ITF}_i^{\text{in}} = \frac{b_{iL}}{b_{iR}}\,, \qquad\qquad \text{ITF}_{\mathbf{x}}^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{b}_{iR}}{\mathbf{w}_R^H \mathbf{b}_{iR}}\,, \tag{1.30}$$

where $\frac{b_{iL}}{b_{iR}}$ is simply the first element from $\mathbf{b}_{iR}$. Since the delay from the source to the reference microphone is no longer present when using RTFs, the impulse response becomes shorter. This is beneficial when using the short-time Fourier transform, since windows are typically shorter than the ATFs.

In practice, these RTFs are unknown and have to be estimated. This is a challenging task, even under stationary and time-invariant conditions, but in particular when in dynamic scenarios. Therefore, instead of using the actual RTFs, pre-determined RTFs (PRTFs) could be used [7]. In [7] the effectiveness of PRTFs has been investigated. This was done by modelling $m$ virtual interferers equally spaced on a circle around the hearing aid user with sufficient radius to ensure the far field assumption can be made. The PRTFs were determined using a head-related impulse response database [8], which are independent of the room, but dependent of a head on which the hearing aids would be mounted. If the far field assumption is made, the PRTFs become approximately distance invariant. This means PRTFs only depend on the angle of the sources with respect to the microphones, when the far field assumption can be made.

When an actual interferer is in the direction of a PRTF that is used, the binaural cues should be perfectly preserved. When it is not in the direction of any of the PRTFs, an error is introduced. The error can be quantified as described in Section 1.2.1, when the filters $\mathbf{w}_L$ and $\mathbf{w}_R$ are based on the use of PRTFs. It has been shown in [7] that this error will grow, on average, when the number of PRTFs decreases. It was also shown, as expected, that the SNR lowers significantly when applying more constraints to the LCMV. This can be alleviated by using methods that allow for more interferers to be modelled [9].

## 1.3. Research Question

Binaural cue preservation for two collaborating hearing aids is challenging. Given the RTFs or PRTFs, binaural LCMV beamformers seem to be promising to preserve the spatial cues of a limited number of interferers, while preserving the target. The JBLCMV beamformer, in particular, can perform improved noise reduction compared to other binaural LCMV methods and can preserve the binaural cues of $2M - 3$ interferers. As a function of the number of interferers for which the binaural cues are preserved, the noise reduction capability goes down. Using the JBLCMV, an optimum in the trade-off between binaural cue preservation and noise reduction needs to be found.

Since JBLCMV does not constrain the attenuation of interferers, an interferer might actually become inaudible after beamforming. In this case, preserving its binaural cues might be useless and removing the constraint associated with that interferer would allow for more noise reduction. Especially when using PRTFs, a number of PRTFs might not even be associated with an interferer and as such should not be constrained binaurally. When an interferer is deemed inaudible, removing its associated constraints from the problem might improve the performance of the beamformer. It is however not guaranteed that the interferer stays inaudible after the constraints have changed, because of the unpredictable behaviour of the LCMV beampattern when constraints change. Here lies the challenge of choosing the best set of (P)RTFs to binaurally constrain, such that the maximum possible noise reduction and limited amount of possible audible spatial cue distortions are simultaneously achieved.

The above issues can be condensed in the following research question: “When performing joint binaural linearly constrained minimum variance beamforming, can performance in terms of noise reduction and spatial cue preservation be improved by lifting constraints on interferers, if they are deemed inaudible after processing?” This thesis will focus on methods to answer this question mainly through the application of convex optimisation. It should be noted that the audibility of interferers will be approximated with simple quantifiable measures that suffice for the proof-of-concept of the proposed methods.

# 2 Methods

In this chapter, the research question from Section 1.3 is used to formulate a suitable optimisation problem. This optimisation problem minimises the output noise power, based on the JBLCMV, with a subset of the interferer constraints as the minimisation variable. The problem, including the additional audibility constraint, will be formulated and discussed in Section 2.1.

Then, the problem will be approximated by two convex optimisation problems, since the problem formulated in Section 2.1 is non-convex. In Section 2.2 the relaxations to obtain these convex problems will be discussed. Since the problem introduced in Section 2.1 is also an integer program, a method to solve them as a convex optimisation is given in Section 2.2.4 as well. The parameters used in these methods are discussed in Sections 2.3.1 and 2.3.2 and the full proposed methods are summarised in Section 2.4.

## 2.1. Formulating the Optimisation Problem

Taking the JBLCMV as a basis, we can formulate an optimisation problem that finds the optimal set of interferers to select, according to the research question in Section 1.3. As before, $\mathbf{w} = \begin{bmatrix}\mathbf{w}_L^H & \mathbf{w}_R^H\end{bmatrix}^H$ and $\tilde{\mathbf{P}} = \begin{bmatrix}\mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}\end{bmatrix}$. In addition, a binary selection vector $\mathbf{p}$ is introduced that selects the set of interferers that should be binaurally constrained. An interferer should be selected when it is considered audible after filtering. As such, degrees of freedom are only used to preserve the spatial cues of audible sources.

In this thesis, audibility of interferers will be determined by setting a threshold on their summed power after filtering. For example, this threshold can be a fixed number, a different number for different frequency bins or a fraction of another power such as the target signal power. Section 2.3.1 will propose a choice for the threshold. Eventually, the audibility could be based on models of perception that reflect the audibility of the processed interferer within the total processed output. Until the proposed threshold is introduced in Section 2.3.1, the threshold will be considered a fixed number, independent of any minimisation variables.

Consider that $r$ interferers are present in the acoustic scene and that the selection vector $\mathbf{p} \in \{0,1\}^r$ is used to select the optimal set of interferers that are to be binaurally constrained. This needs to be done such that the output noise power is minimised and that, after processing, the summed power of the interferers that were not selected by $\mathbf{p}$ is below the power threshold $c$. Let $\mathbf{U} = \mathrm{diag}(\mathbf{u} \odot |\mathbf{b}_L|) \in \mathbb{R}^{r\times r}$ be the diagonal matrix with all individual interferer magnitudes, $\odot$ denotes the Hadamard product (also called the entrywise product), $\tilde{\mathbf{B}} = \begin{bmatrix}\mathbf{B}_L^H & \mathbf{B}_L^H\end{bmatrix}^H \in \mathbb{C}^{2M\times r}$, with $\mathbf{B}_L$ the matrix with the $r$ left RTFs of the $r$ interferers as its columns, and $\mathbf{w}_\mathbf{p} \in \mathbb{C}^{2M}$ the JBLCMV solution for the interferer set given by $\mathbf{p}$. The resulting interferer selection for the JBLCMV optimisation problem is given by

$$\begin{aligned}
&\underset{\mathbf{p}\in\{0,1\}^r}{\text{minimise}} && \mathbf{w}_\mathbf{p}^H \tilde{\mathbf{P}} \mathbf{w}_\mathbf{p} && \text{(2.1a)}\\
&\text{subject to} && \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H \mathbf{w}_\mathbf{p}\right\|_2^2 \le c && \text{(2.1b)}\\
& && \|\mathbf{p}\|_0 \le 2M - 3\,. && \text{(2.1c)}
\end{aligned}$$

$\tilde{\mathbf{B}}$ can contain either the true ATFs, the left RTFs or the right RTFs. When RTFs are used, the diagonal elements $u_i$ in $\mathbf{U}$ are appropriately scaled by either $|\mathbf{b}_L| = \begin{bmatrix}|b_{1L}| & |b_{2L}| & \dots & |b_{rL}|\end{bmatrix}$ or $|\mathbf{b}_R| = \begin{bmatrix}|b_{1R}| & |b_{2R}| & \dots & |b_{rR}|\end{bmatrix}$. In this report, the left RTFs are used when considering $\tilde{\mathbf{B}}$ and $\mathbf{U}$. The cost function, Equation (2.1a), is similar to the binaural LCMV methods discussed in Section 1.2, since minimising output noise power is still the objective. However, the filter now depends on the selection vector as will be shown in more detail in Section 2.1.1.

As specified by the research question, the interferers that are not binaurally constrained should not be audible. This is guaranteed by Equation (2.1b). The matrix $\mathbf{I} - \mathrm{diag}(\mathbf{p})$ selects the interferers that are not binaurally constrained. The filter $\mathbf{w}_\mathbf{p}$ is applied to these interferers and their power is summed. Their summed power should be below the power threshold $c$.

### 2.1.1. Adapting the LCMV Solution

The filter $\mathbf{w_p}$ is defined to be the JBLCMV beamformer for the interferer set given by $\mathbf{p}$. The JBLCMV solution, given in Equation (1.16), is adapted for the selection vector $\mathbf{p}$. To do so, the constraint matrix $\mathbf{C}$ and response vector $\mathbf{f}$ need to be redefined. We define a selection matrix $\mathbf{\Phi} \in \{0,1\}^{(r+2)\times(\|\mathbf{p}\|_0+2)}$ that removes the columns and elements from the matrix $\mathbf{\Lambda}$ (which contains all interferer constraints) and $\mathbf{f}$, respectively, that correspond to zero values in $\mathbf{p}$. Therefore, the following definitions can be made,

$$\mathbf{\Lambda} = \begin{bmatrix}\mathbf{\Lambda}_a & \mathbf{\Lambda}_b\end{bmatrix} \quad \in \mathbb{C}^{2M\times(r+2)}\,, \tag{2.2}$$

$$\mathbf{C} = \mathbf{\Lambda\Phi} \quad \in \mathbb{C}^{2M\times(\|\mathbf{p}\|_0+2)}\,, \tag{2.3}$$

$$\mathbf{f} = \mathbf{\Phi}^T\begin{bmatrix}a_L & a_R & \mathbf{0}_r^T\end{bmatrix}^H \in \mathbb{C}^{\|\mathbf{p}\|_0+2}\,. \tag{2.4}$$

The definition of $\mathbf{\Phi}$ is detailed in Section 2.1.2. The expression for $\mathbf{w_p}$ is obtained by combining the expressions for $\mathbf{C}$ and $\mathbf{f}$ from Equations (2.3) and (2.4) respectively and the LCMV solution from Equation (1.16):

$$\mathbf{w_p} = \tilde{\mathbf{P}}^{-1}\mathbf{\Lambda\Phi}\left(\mathbf{\Phi}^T\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f}\,. \tag{2.5}$$

Equation (2.5) is then substituted into the optimisation problem in Equation (2.1) to obtain

$$\underset{\mathbf{p}\in\{0,1\}^r}{\text{minimise}} \quad \mathbf{f}^H\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f} \tag{2.6a}$$

$$\text{subject to} \quad \left\|(\mathbf{I}-\mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda\Phi}\left(\mathbf{\Phi}^T\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f}\right\|_2^2 \leq c \tag{2.6b}$$

$$\|\mathbf{p}\|_0 \leq 2M-3 \tag{2.6c}$$

$$\begin{bmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi_p}\end{bmatrix} = \mathbf{\Phi} \tag{2.6d}$$

$$\mathbf{\Phi_p}\mathbf{\Phi_p}^T = \mathrm{diag}(\mathbf{p})\,, \quad \mathbf{\Phi_p}^T\mathbf{\Phi_p} = \mathbf{I}_{\|\mathbf{p}\|_0}\,. \tag{2.6e}$$

### 2.1.2. Selection Matrix

The selection matrix removes the columns from $\mathbf{\Lambda}$ and elements from $\mathbf{f}$, where the corresponding value in $\mathbf{p}$ equals zero. More specifically, only columns and elements from $\mathbf{\Lambda}_b$ and $\mathbf{f}_b$ are removed. Therefore, a selection matrix $\mathbf{\Phi_p}$ is defined that has the properties $\mathbf{\Phi_p}\mathbf{\Phi_p}^T = \mathrm{diag}(\mathbf{p})$ and $\mathbf{\Phi_p}^T\mathbf{\Phi_p} = \mathbf{I}_{|\mathbf{p}|}$. The matrix $\mathbf{\Phi_p}$ is then equal to $\mathrm{diag}(\mathbf{p})$ with its zero-columns removed [10].

Note that the matrix $\mathbf{\Lambda}$ in Equation (2.2) contains both $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_b$. Since the two columns of $\mathbf{\Lambda}_a$ should always remain, the selection matrix $\mathbf{\Phi}$ is defined as $\mathbf{\Phi} = \mathrm{diag}(\mathbf{I}_2, \mathbf{\Phi_p})$. As illustration, consider the following example, where $r = 3$:

$$\begin{aligned}
\mathbf{p} = \begin{bmatrix}1\\0\\1\end{bmatrix} \Longrightarrow \mathbf{\Lambda\Phi} &= \begin{bmatrix}\mathbf{\Lambda}_a & \mathbf{\Lambda}_b\end{bmatrix}\begin{bmatrix}\mathbf{I}_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi_p}\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{\Lambda}_a & \mathbf{\Lambda}_b\mathbf{\Phi_p}\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{\Lambda}_a & \begin{bmatrix}\mathbf{b}_1 b_{1R} & \mathbf{b}_2 b_{2R} & \mathbf{b}_3 b_{3R}\\ -\mathbf{b}_1 b_{1L} & -\mathbf{b}_2 b_{2L} & -\mathbf{b}_3 b_{3L}\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\\0 & 1\end{bmatrix}\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{\Lambda}_a & \begin{matrix}\mathbf{b}_1 b_{1R} & \mathbf{b}_3 b_{3R}\\ -\mathbf{b}_1 b_{1L} & -\mathbf{b}_3 b_{3L}\end{matrix}\end{bmatrix},
\end{aligned} \tag{2.7}$$

and similarly for $\mathbf{f}^H = \begin{bmatrix}\mathbf{f}_a^H & \mathbf{f}_b^H\end{bmatrix}$.

## 2.2. Problem Relaxation

To solve Equation (2.6), a convex relaxation and subsequent convex solving method would be ideal, as it would allow the global solution to the relaxed problem to be found efficiently. However, the problem in Equation (2.6) is non-convex for a number of reasons.

First, the cost function is problematic as it depends on $\mathbf{\Phi}$, which is defined with a non-linear operation (removing zero-columns from $\mathrm{diag}(\mathbf{p})$) and, as a function of $\mathbf{p}$, its associated equality constraints in Equation (2.6e) are non-affine. Second, the power constraint, Equation (2.6b), is non-convex in the minimisation variable $\mathbf{p}$ and also suffers from $\mathbf{\Phi}$, similar to in the cost function. The occurrence of $\mathbf{\Phi}$ in this constraint is very similar to how it appears in the cost function, meaning the issue can be resolved in the cost function and power constraint simultaneously. Third, the vector variable $\mathbf{p}$ is binary, which makes the feasible region disjoint and thus non-convex. Finally, the cardinality function is by definition non-convex.

The $l_1$-norm has been shown to be a good convex substitute of the cardinality function, promoting sparse solutions [11]. In the case of a binary selection vector, these two functions produce the same results. Since the selection vector is also non-negative, the cardinality constraint

$$\|\mathbf{p}\|_0 \leq 2M - 3\,, \tag{2.8}$$

can immediately be relaxed to

$$\mathbf{1}^T\mathbf{p} \leq 2M - 3\,. \tag{2.9}$$

### 2.2.1. Cost Function Relaxation

In this section, the cost function in Equation (2.6a) is convexified. The cost function is currently non-convex due to the occurrence of $\mathbf{\Phi}$, which is non-linearly dependent of the minimisation variable $\mathbf{p}$. First, consider the following decomposition of the data dependent constant[1] matrix $\mathbf{R}$:

$$\begin{aligned}\mathbf{R} &= \mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\\ &= \lambda\mathbf{I} + \mathbf{G}\,,\end{aligned} \tag{2.10}$$

with $\lambda > \lambda_{max}(\mathbf{R})$ and $\lambda_{max}(\mathbf{R})$ the largest eigenvalue of $\mathbf{R}$. Because $\mathbf{R}$ is positive definite due to self-noise, $\lambda$ is positive and $\mathbf{G}$ is negative definite. These properties are important later, in Equation (2.15). Equation (2.10) is substituted into the cost function in Equation (2.6a) to obtain

$$\begin{aligned}f_0(\mathbf{p}) &= \mathbf{f}^H\mathbf{\Phi}(\mathbf{\Phi}^T\overbrace{\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}}^{\mathbf{R}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{f}\\ &= \mathbf{f}^H\mathbf{\Phi}(\mathbf{\Phi}^T(\lambda\mathbf{I}_{r+2} + \mathbf{G})\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{f}\\ &= \mathbf{f}^H\underbrace{\mathbf{\Phi}(\lambda\mathbf{I}_{\|\mathbf{p}\|_0+2} + \mathbf{\Phi}^T\mathbf{G}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T}_{\mathbf{Q}}\mathbf{f}\,.\end{aligned} \tag{2.11}$$

Then, the matrix $\mathbf{Q}$ in Equation (2.11) is rewritten using the matrix inversion lemma [12, p.18]:

$$\begin{aligned}\mathbf{Q} &= \mathbf{\Phi}(\lambda\mathbf{I} + \mathbf{\Phi}^T\mathbf{G}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\\ &= \mathbf{G}^{-1} - \mathbf{G}^{-1}\big(\mathbf{G}^{-1} + \lambda^{-1}\,\mathrm{diag}\big(\begin{bmatrix}\mathbf{1}_2^T & \mathbf{p}^T\end{bmatrix}\big)\big)^{-1}\mathbf{G}^{-1}\,.\end{aligned} \tag{2.12}$$

**Semi-definite relaxation**

The cost function in Equation (2.11) is convex in $\mathbf{Q}$, so it can be written as

$$f_0(\mathbf{Q}) = \mathbf{f}^H\mathbf{Q}\mathbf{f}\,, \tag{2.13}$$

with Equation (2.12) as an additional constraint. Equation (2.12) is not affine in $\mathbf{p}$, and is thus not a convex constraint. To relax Equation (2.12), the equality is changed to an inequality. Since the matrix $\mathbf{Q}$ appears in the cost function in Equation (2.13), which is minimised, the relaxation of Equation (2.12) should be choosen such that the cost function containing $\mathbf{Q}$ bounds the original cost function in Equation (2.11) from above. That is,

$$\mathbf{Q} \succeq \mathbf{G}^{-1} - \mathbf{G}^{-1}\big(\mathbf{G}^{-1} + \lambda^{-1}\,\mathrm{diag}\big(\begin{bmatrix}\mathbf{1}_2^T & \mathbf{p}^T\end{bmatrix}\big)\big)^{-1}\mathbf{G}^{-1}\,, \tag{2.14}$$

which is reformulated as a linear matrix inequality (LMI) using the Schur Complement [13, p.650]:

$$\begin{bmatrix}-\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\begin{bmatrix}\mathbf{1}_2^T & \mathbf{p}^T\end{bmatrix}\big) & \mathbf{G}^{-1}\\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1}\end{bmatrix} \succeq 0\,. \tag{2.15}$$

[1] $\mathbf{R}$ is constant in the sense that it does not depend on any minimisation variables. Of course, it does change over time and across frequency bins.

Since each matrix block on the main diagonal of a positive semi-definite matrix is positive semi-definite and $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ needs to be invertable, $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ must be positive definite. It can be proven that it is, given the choice of $\lambda > \lambda_{max}(\mathbf{R})$, as shown in the following section.

This results in the following optimisation problem:

$$\underset{\mathbf{p}\in\{0,1\}^r,\ \mathbf{Q}\in\mathbf{S}_{++}^{r+2}}{\text{minimise}} \quad \mathbf{f}^H\mathbf{Q}\mathbf{f} \tag{2.16a}$$

$$\text{subject to} \quad \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right) & \mathbf{G}^{-1} \\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1} \end{bmatrix} \succeq 0 \tag{2.16b}$$

$$\left\|(\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f}\right\|_2^2 \leq c \tag{2.16c}$$

$$\mathbf{1}^T\mathbf{p} \leq 2M - 3\,. \tag{2.16d}$$

The cost function and its associated positive semi-definiteness constraint are convex and the matrix $\mathbf{\Phi}$ has been removed from the problem. The power constraint remains non-convex, as does the binary $\mathbf{p}$ parameter.

**Proof of Definiteness for Schur Complement in Equation (2.15)**

To be able to successfully transform the constraint in Equation (2.14) in the LMI in Equation (2.15), the matrix block $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ must always be positive definite. To prove that it is, we show that the smallest eigenvalue of $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ is always positive, when $\lambda > \lambda_{max}(\mathbf{R})$.

Matrix $\mathbf{G}$ is guaranteed to be negative definite when $\mathbf{R}$ is positive definite and $\lambda\mathbf{I}$ is subtracted from $\mathbf{R}$, with $\lambda > \lambda_{max}(\mathbf{R})$. The negative definiteness is due to the following relationship:

$$\begin{aligned} \mathbf{R} &\in \mathbb{H}^n\,, \\ \boldsymbol{\lambda}_{\mathbf{R}} &= \left[\lambda_1, \lambda_2, \ldots, \lambda_n\right]\,, \\ \boldsymbol{\lambda}_{\mathbf{R}-a\mathbf{I}} &= \left[\lambda_1 - a, \lambda_2 - a, \ldots, \lambda_n - a\right]\,, \end{aligned} \tag{2.17}$$

where $\boldsymbol{\lambda}_{\mathbf{R}}$ is the set eigenvalues of $\mathbf{R}$ in non-increasing order, meaning $\lambda_{max}(\mathbf{R}) = \lambda_1$. The set of eigenvalues of $\mathbf{G}^{-1}$ is, in non-increasing order,

$$\boldsymbol{\lambda}_{\mathbf{G}^{-1}} = \left[(\lambda_n - \lambda)^{-1}, (\lambda_{n-1} - \lambda)^{-1}, \ldots, (\lambda_1 - \lambda)^{-1}\right]\,, \tag{2.18}$$

with $\lambda_i > 0 \quad \text{for} \quad i = 1, 2, \ldots, n$, because of the positive definiteness of $\mathbf{R}$. This results in $\mathbf{G}^{-1}$ being negative definite and $-\mathbf{G}^{-1}$ being positive definite, because multiplying any negative definite matrix with any negative scalar produces a positive definite matrix.

Applying one of Weyl's inequalities for matrix theory [14, p.157], a lower bound on the smallest eigenvalue of $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ can be obtained. When this lower bound is larger than zero, $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ is always positive definite. The relevant inequality by Weyl is

$$\begin{aligned} &\text{given} \quad \mathbf{C} = \mathbf{A} + \mathbf{B}\,, \\ &\lambda_{min}(\mathbf{C}) \geq \lambda_{min}(\mathbf{A}) + \lambda_{min}(\mathbf{B})\,, \end{aligned} \tag{2.19}$$

where matrices $\mathbf{A}$ and $\mathbf{B}$ are both self-adjoint. The eigenvalues of $\left[-\lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ are either 0 or $-\lambda^{-1}$, because of the binary elements in $\mathbf{p}$. This results in the following relations:

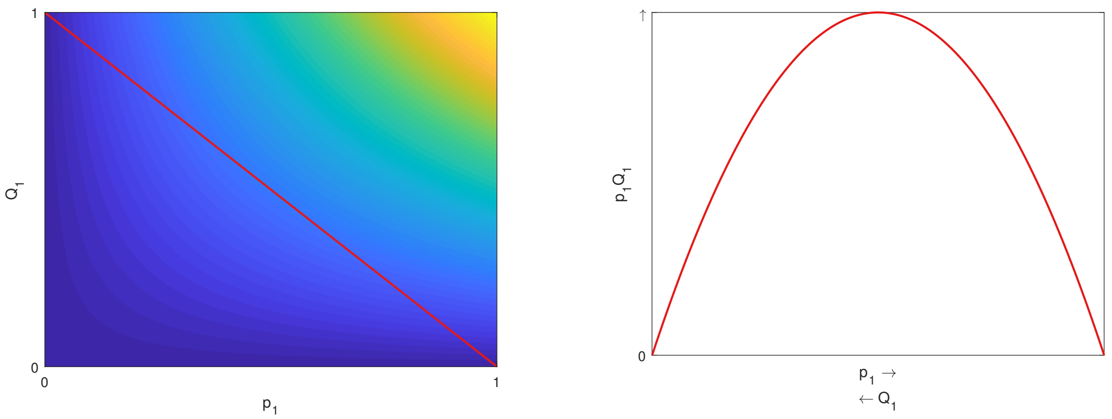
$$\lambda_{min}\left(-\lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right) = -\lambda^{-1}\,, \tag{2.20}$$

$$\lambda_{min}\left(-\mathbf{G}^{-1}\right) = (\lambda - \lambda_n)^{-1}\,, \tag{2.21}$$

$$\begin{aligned} \lambda_{min}\left(-\mathbf{G}^{-1}\right) + \lambda_{min}\left(-\lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right) &= (\lambda - \lambda_n)^{-1} - \lambda^{-1} \\ &> 0\,, \end{aligned} \tag{2.22}$$

$$\begin{aligned} \lambda_{min}\left(-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right) &\geq \lambda_{min}\left(-\mathbf{G}^{-1}\right) + \lambda_{min}\left(-\lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right) \\ &> 0\,, \end{aligned} \tag{2.23}$$

which proves that $\left[-\mathbf{G}^{-1} - \lambda^{-1}\operatorname{diag}\left(\left[\mathbf{1}_2^T \quad \mathbf{p}^T\right]\right)\right]$ is positive definite for any $\lambda > \lambda_{max}(\mathbf{R})$.




(a) Top down view of the surface describing $p_1 Q_1$. Yellow and blue correspond to a high and low function value, respectively.

(b) Cross section of the surface describing $p_1 Q_1$, at the red line in Figure 2.1a.

Figure 2.1: Visuals to illustrate that Equation (2.24), containing the multiplication of the minimisation variables $\mathbf{p}$ and $\mathbf{Q}$, is non-convex. In these figures $p_1 = \mathbf{p} \in [0,1]^1$ and $Q_1 = \mathbf{Q} \in [0,1]^1$.

### 2.2.2. Relaxing the Power Constraint

In this section, the inequality constraint in Equation (2.24) is convexified. The constraint is written out using the results from Section 2.2.1 to obtain

$$\begin{aligned} f_1(\mathbf{p},\mathbf{Q}) &= \left\| (\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f} \right\|_2^2 \\ &= \mathbf{f}^H\mathbf{Q}\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U}\underbrace{(\mathbf{I} - \operatorname{diag}(\mathbf{p}))(\mathbf{I} - \operatorname{diag}(\mathbf{p}))}_{(\mathbf{I} - \operatorname{diag}(\mathbf{p}))}\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f}\,. \end{aligned} \tag{2.24}$$

This is not a convex inequality constraint, due to the multiplication of the two different minimisation variables, $\mathbf{p}$ and $\mathbf{Q}$. Consider the case where $\mathbf{p} \in [0,1]^1$ and $\mathbf{Q} \in [0,1]^1$. Now Figure 2.1 clearly shows that Equation (2.24) is non-convex in this case, and thus definitely also for $\mathbf{p} \in [0,1]^r$ and $\mathbf{Q} \in \mathbf{S}_{++}^{r+2}$.

Two different methods of relaxing this expression are presented in this section. The first is based on the Schur Complement and setting up an LMI by first applying a regularisation to the selection matrix $(\mathbf{I} - \operatorname{diag}(\mathbf{p}))$, while the second is based on overestimating the total binaurally unconstrained output interferer power.

**Relaxation using selection matrix regularisation**

Applying a relaxation specifically for the occurrence of $\mathbf{p}$ in the middle of Equation (2.24), such that the Schur Complement can be used here as well, would open up the possibility for an LMI positive semi-definiteness constraint using a linear matrix inequality:

$$\begin{gathered} \mathbf{f}^H\mathbf{Q}\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U}\mathbf{X}^{-1}\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f} \leq c \\ \text{if and only if} \\ \begin{bmatrix} \mathbf{X} & \mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f} \\ \mathbf{f}^H\mathbf{Q}\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U} & c \end{bmatrix} \succeq 0 \\ \text{and} \\ \mathbf{X} \succ 0\,, \end{gathered} \tag{2.25}$$

where $\mathbf{X}$ is an unknown positive definite matrix that should approximate and/or bound $(\mathbf{I} - \operatorname{diag}(\mathbf{p}))$. The matrix $(\mathbf{I} - \operatorname{diag}(\mathbf{p}))$ is not invertable, and thus, also not positive definite, when $\mathbf{p}$ is anything but the zero vector. This is why the matrix $(\mathbf{I} - \operatorname{diag}(\mathbf{p}))$ cannot be used directly.

For the Schur complement to hold, the selection matrix in Equation (2.24) needs to be positive definite. To achieve this, a small error can be introduced on the diagonal:

$$\mathbf{f}^H\mathbf{Q}\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U}(\mathbf{I} - (1-\epsilon)\operatorname{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f} \leq c\,, \tag{2.26}$$

with $\epsilon \ll 1$. Then, the Schur Complement can be used, as in Equation (2.25), to obtain

$$\begin{bmatrix} \mathbf{I} + (\frac{1}{\epsilon} - 1)\,\mathrm{diag}(\mathbf{p}) & \mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{f} \\ \mathbf{f}^H\mathbf{Q}\boldsymbol{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U} & c \end{bmatrix} \succeq 0\,, \tag{2.27}$$

as a relaxation of the power constraint. This relaxation should not deteriorate the optimal point much when $\epsilon$ is chosen sufficiently small, as binaurally constrained interferers would only be counted towards the power constraint by a very small fraction.

The constraint in Equation (2.16c) of is now expressed as an LMI constraint. The optimisation problem can now be expressed as

$$\begin{aligned} \underset{\mathbf{p}\in\{0,1\}^r,\ \mathbf{Q}\in\mathbf{S}_{++}^{r+2}}{\text{minimise}} \quad & \mathbf{f}^H\mathbf{Q}\mathbf{f} \\ \text{subject to} \quad & \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\begin{bmatrix}\mathbf{1}_2^T & \mathbf{p}^T\end{bmatrix}\big) & \mathbf{G}^{-1} \\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1} \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} \mathbf{I} + (\frac{1}{\epsilon} - 1)\,\mathrm{diag}(\mathbf{p}) & \mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{f} \\ \mathbf{f}^H\mathbf{Q}\boldsymbol{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U} & c \end{bmatrix} \succeq 0 \\ & \mathbf{1}^T\mathbf{p} \leq 2M - 3\,. \end{aligned} \tag{2.28}$$

**Relaxing by over-estimating the interferer power after filtering**

As stated before, Equation (2.24) is not a convex constraint due to the multiplication of the minimisation variables, $\mathbf{p}$ and $\mathbf{Q}$. Instead of using the Schur complement to set up an LMI as done above in Equations (2.26) to (2.28), the constraint can also be relaxed such that the two variables are no longer multiplied with each other, or, more specifically, by over-estimating Equation (2.24).

Consider the scenario where $r \leq 2M - 3$, and thus, where the JBLCMV can constrain all RTFs. Then, it is known that in this scenario, the total interferer power when binaurally constraining all interferers is equal or higher than the total interferer power when constraining a subset of the interferers binaurally:

$$\left\|\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{f}\right\|_2^2 \leq \left\|\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{JB}\right\|_2^2, \tag{2.29}$$

with $\mathbf{w}_{JB}$ being the JBLCMV filter that preserves the spatial cues of all $r$ interferers. By doing so, it uses degrees of freedom for spatial cue preservation instead of for noise reduction. As a result, the output of the filter will have more noise than a filter that spatially constrains less interferers. This is expressed by Equation (2.29). Based on Equation (2.29), it can be argued that

$$\left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{f}\right\|_2^2 \lessapprox \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{JB}\right\|_2^2. \tag{2.30}$$

So then Equation (2.24) can be relaxed to

$$\left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{f}\right\|_2^2 \lessapprox \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{JB}\right\|_2^2 \leq c\,, \tag{2.31}$$

which makes it a convex approximation of an over-estimator of $\left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\boldsymbol{\Lambda}\mathbf{Q}\mathbf{f}\right\|_2^2$. Notice that due to the relaxation, it is not an exact over-estimator anymore, since Equation (2.29) only holds for the complete set of interferers. On a subset only Equation (2.30) holds. The constraint in Equation (2.16c) is now expressed as the norm of an affine expression, which is convex. The optimisation problem can now be expressed as

$$\begin{aligned} \underset{\mathbf{p}\in\{0,1\}^r,\ \mathbf{Q}\in\mathbf{S}_{++}^{r+2}}{\text{minimise}} \quad & \mathbf{f}^H\mathbf{Q}\mathbf{f} \\ \text{subject to} \quad & \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\begin{bmatrix}\mathbf{1}_2^T & \mathbf{p}^T\end{bmatrix}\big) & \mathbf{G}^{-1} \\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1} \end{bmatrix} \succeq 0 \\ & \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{JB}\right\|_2^2 \leq c \\ & \mathbf{1}^T\mathbf{p} \leq 2M - 3\,. \end{aligned} \tag{2.32}$$

Since the relaxation variable $\mathbf{Q}$ is no longer needed in the relaxation of the power constraint in Equation (2.32), the cost function can be relaxed more efficiently, by taking the epigraph of the original cost function. Combining Equations (2.11) and (2.12), the cost function can be written as

$$f_0(\mathbf{p}) = \mathbf{f}^H\mathbf{G}^{-1}\mathbf{f} - \mathbf{f}^H\mathbf{G}^{-1}\big(\mathbf{G}^{-1} + \lambda^{-1}\,\mathrm{diag}\big(\big[\mathbf{1}_2^T \quad \mathbf{p}^T\big]\big)\big)^{-1}\mathbf{G}^{-1}\mathbf{f}\,. \tag{2.33}$$

The constant first term can be neglected, since it can only change the value of the function in the optimal point, not the argument of the optimal point, and the value of the function is not of interest here. This leads to the following epigraph expression and LMI:

$$\mathbf{f}^H\mathbf{G}^{-1}\big(-\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\big[\mathbf{1}_2^T \quad \mathbf{p}^T\big]\big)\big)^{-1}\mathbf{G}^{-1}\mathbf{f} \leq t_0\,, \tag{2.34}$$

$$\begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\big[\mathbf{1}_2^T \quad \mathbf{p}^T\big]\big) & \mathbf{G}^{-1}\mathbf{f} \\ \mathbf{f}^H\mathbf{G}^{-1} & t_0 \end{bmatrix} \succeq 0\,. \tag{2.35}$$

This relaxation has the advantage of a lower dimensional relaxation variable and LMI. The optimisation problem can now be more efficiently expressed as

$$\begin{aligned} \underset{\hat{\mathbf{p}}\in[0,1]^r,\ t_0}{\text{minimise}} \quad & t_0 \\ \text{subject to} \quad & \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\big[\mathbf{1}_2^T \quad \mathbf{p}^T\big]\big) & \mathbf{G}^{-1}\mathbf{f} \\ \mathbf{f}^H\mathbf{G}^{-1} & t_0 \end{bmatrix} \succeq 0 \\ & \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{JB}\right\|_2^2 \leq c \\ & \mathbf{1}^T\mathbf{p} \leq 2M-3\,. \end{aligned} \tag{2.36}$$

Both the relaxation of Equations (2.28) and (2.36) will be compared in Chapter 3. Both suffer from an imperfect relaxation, namely the relaxation parameter $\epsilon$ in Equation (2.28) and the inexact over-estimator in Equation (2.36). The first relaxation can handle a large amount of (P)RTFs, unlike the second relaxation, since it depends on the JBLCMV beamformer, which is constrained to $2M-3$ (P)RTFs. To be able to compare them anyway, when using the second relaxation, the JBLCMV beamformer is used with the $2M-3$ strongest (P)RTFs, meaning the (P)RTFs with the highest associated value $u_i$ or $\hat{u}_i$. This choice is further discussed in Section 3.1.

### 2.2.3. Power Threshold Relaxation for $r > 2M-3$

This relaxation is not done to convexify an expression, but to prevent infeasibility of the optimisation problem. When the number of interferers is larger than the number of interferers that a JBLCMV beamformer can constrain ($2M-3$), there is no longer a guarantee that the power constraint, Equation (2.1b), repeated below, can be met.

$$\left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{\mathbf{p}}\right\|_2^2 \leq c\,.$$

When the number of interferers is less than or equal to $2M-3$, Equation (2.1b) can always be met, because when $\mathbf{p}$ equals the all ones vector, the norm equals zero. The cardinality of $\mathbf{p}$ is constrained by Equation (2.1c), however. This means that when the number of interferers exceeds $2M-3$, the norm in Equation (2.1b) cannot be made zero anymore. Furthermore, the minimal value of the norm in Equation (2.1b) might be above the set threshold. If this is the case, the problem is infeasible.

Because an infeasible problem is of no use in determining the beamformer coefficients, the threshold should be relaxed when the number of known interferers (or pre-determined relative acoustic transfer function, as discussed in Section 2.3.2) might exceed $2M-3$. This can be done by introducing a penalty to the cost function when the threshold is exceeded. The original, unrelaxed, optimisation problem from Equation (2.1) is used to demonstrate the threshold relaxation. That is,

$$\underset{\mathbf{p}\in\{0,1\}^r,\ t_1}{\text{minimise}} \quad \mathbf{w}_{\mathbf{p}}^H\tilde{\mathbf{P}}\mathbf{w}_{\mathbf{p}} + \alpha\max\left(0, t_1 - c\right)^2 \tag{2.37a}$$

$$\text{subject to} \quad \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_{\mathbf{p}}\right\|_2^2 \leq t_1 \tag{2.37b}$$

$$\|\mathbf{p}\|_0 \leq 2M-3\,, \tag{2.37c}$$

where $t_1 \in \mathbb{R}^1$ is the new relaxation variable.

When $t_1$ is lower than $c$, the penalty term is zero and as such, the cost function is unchanged. When it is larger than the set threshold, the penalty equals the squared error multiplied by the weight parameter $\alpha$. The parameter $\alpha$ should likely be chosen such that the set threshold is not exceeded, unless necessary.

The penalty expression that is added to the cost function is convex, as is the appearance of the new variable $t_1$ in all proposed relaxations. As such, it can be added to any of the proposed methods when necessary. The optimal $t_1$ should also be provided to the random rounding algorithm, Algorithm 1, since the randomised rounding algorithm will try to find a feasible solution according to the constraint given. As discussed before, a feasible solution might not exist when $r > 2M - 3$. The relaxed threshold $t_1$ in Equation (2.37) is a new value for the power threshold, such that a feasible solution can exist, and so the randomised rounding algorithm should check feasibility according to this updated threshold, instead of the original one given by $c$ in Equation (2.37).

### 2.2.4. Binary Parameter

The variable $\mathbf{p}$ contains binary elements. However, a binary variable leads to a non-convex problem, because the feasible region is disjoint. The parameter $\mathbf{p}$ can be relaxed in a number of ways. One method of solving would be an exhaustive or greedy search, which is feasible for a small number of interferers. Of course, convex relaxations are not necessary when applying search methods. One exhaustive search method is described in Section 3.2.1, to be used when evaluating simulation results in Chapter 3. Another method is based on solving the optimisation problems using continuous values for $\mathbf{p}$, i.e. $\hat{\mathbf{p}} \in [0, 1]^r$. Then, specific rounding methods can be used to find binary solution vectors.

**Continuous selector**

Making $\mathbf{p}$ continuous instead of binary introduces a problem in the power constraint as the power is then scaled by the value that is assigned to the respective $\hat{p}_i$, leading to a fraction of an interferer's power to be summed in the total in Equation (2.16c). This causes low values of $\hat{\mathbf{p}}$ to become prevalent as optimal solutions. Continuous values of $\hat{\mathbf{p}}$ do not have the same significant effect on the cost function, since the variable is applied to both the constraint matrix and response vector in the original LCMV problem. Furthermore, when selector values are continuous, the Schur Compliment on the relaxed power constraint from Equation (2.27) technically does not hold, since

$$\left[\mathbf{I} + (\tfrac{1}{\epsilon} - 1)\,\mathrm{diag}(\hat{\mathbf{p}})\right]^{-1} \neq \mathbf{I} - (1 - \epsilon)\,\mathrm{diag}(\hat{\mathbf{p}}) \;, \tag{2.38}$$

and the properties that were used to eliminate $\mathbf{\Phi}$ also only held for binary $\mathbf{p}$. Previous work suggests that a randomised rounding method can find solutions regardless of these issues [10, 15, 16].

A randomised rounding algorithm generates a random uniformly distributed vector with values between zero and one. It will compare this random vector to the continuous solution vector $\hat{\mathbf{p}}^\star$. For elements where $\hat{p}_i^\star$ is larger then the corresponding random value, $p_i^\star$ will be rounded to one. Otherwise, it is rounded to zero. The obtained rounded solution vector $\mathbf{p}^\star$ is then checked for feasibility using Equation (2.6b). If it is infeasible, the rounding step is repeated. Otherwise, the solution is accepted as the optimal solution. The randomised rounding algorithm is given in Algorithm 1. It is called for each frequency bin separately. Instead of checking feasibility per random rounding realisation and accepting the first feasible set, a number $L_{rr}$ random realisations can be determined and the best one in terms of cost can be picked from the feasible realisations. The difference between both methods is further discussed in Section 3.4.4.

## 2.3. Practical Considerations

The proposed methods in Equations (2.28) and (2.36) require interferer RTFs, interferer powers and a binaurally unconstrained interferer output power setting to function. First, a power threshold setting will be discussed, based on a simple audibility measure. Then, the problem of the unknown interferer RTFs and powers will be solved using PRTFs in Section 2.3.2.

### 2.3.1. Defining a Power Threshold

The two proposed methods from Equations (2.28) and (2.36) contain a parameter $c$ to constrain the total binaurally unconstrained interferer power after filtering (see Equations (2.28) and (2.36)). The threshold $c$ will be set such that it describes a minimal SNR for the binaurally unconstrained sources. This specific SNR measure is determined by taking the ratio of the target signal power and the summed power of the binaurally

**Algorithm 1** Randomised Rounding

**function** RANDOMROUNDING($\hat{\mathbf{p}}^\star, k_r, c, \tilde{\mathbf{B}}, \mathbf{U}, \tilde{\mathbf{P}}^{-1}, \mathbf{C}, \mathbf{f}$)
  $constraint \leftarrow \inf$
  $i \leftarrow 0$
  **while** $constraint > c \wedge i \leq k_r$ **do** ▷ Stop when constraint is met or after $k$ iterations
    **if** $i = k_r$ **then** ▷ Maximum number of iterations
      $\mathbf{p}^\star \leftarrow \mathbf{1}$ ▷ Round up to guarantee feasibility
    **else**
      $i \leftarrow i + 1$
      **for all** $p_j$ **do**
        $p_j^\star \leftarrow 1$ with probability $\hat{p}_j^\star$ ($0$ otherwise) ▷ Random rounding
      **end for**
      $\mathbf{\Phi}_\mathbf{p} \leftarrow \operatorname{diag}(\mathbf{p}^\star)$ with its zero columns removed.
      $\mathbf{\Phi} \leftarrow \begin{bmatrix} \mathbf{I}_2 & 0 \\ 0 & \mathbf{\Phi}_\mathbf{p} \end{bmatrix}$
      $\mathbf{w}_\mathbf{p} \leftarrow \tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{C}^H\tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f}$ ▷ Actual filter
      $constraint \leftarrow \left\|(\mathbf{I} - \operatorname{diag}(\mathbf{p}^\star))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_\mathbf{p}\right\|_2^2$ ▷ Binaurally unconstrained interferer power sum
    **end if**
  **end while**
  **return** $\mathbf{p}^\star$
**end function**

unconstrained interfering sources after filtering. The total power of all unconstrained interferers is given by Equation (2.1b), repeated below.

$$\left\|(\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_\mathbf{p}\right\|_2^2 \leq c\,.$$

The minimal SNR is then described by

$$\frac{P_s}{\left\|(\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_\mathbf{p}\right\|_2^2} \geq \beta\,, \tag{2.39}$$

where $\beta$ is the desired minimal SNR and $P_s$ is the target signal power. This expression can be re-written to resemble the power constraint expressions as they have been presented before, similar to Equation (2.1b):

$$\left\|(\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w}_\mathbf{p}\right\|_2^2 \leq \beta^{-1}P_s\,, \tag{2.40}$$

and as such, the threshold parameter can be set as

$$c = \beta^{-1}P_s\,. \tag{2.41}$$

### Target Power Estimation

Equation (2.41) depends on the target signal power, which needs to be estimated. This can be done through power subtraction. First, the total output power, consisting of the target, interferers and additive noise can be determined by

$$P_y = \mathbf{w}^H\tilde{\mathbf{P}}_\mathbf{y}\mathbf{w}\,. \tag{2.42}$$

Second, since the total CPSD matrix is assumed to be given by Equation (1.4), the total output noise power is subtracted to obtain an estimate of the target signal power:

$$P_s = \mathbf{w}^H\tilde{\mathbf{P}}_\mathbf{y}\mathbf{w} - \mathbf{w}^H\tilde{\mathbf{P}}\mathbf{w}\,. \tag{2.43}$$

The choice of the filter $\mathbf{w}$ is preferably such that it does not depend on the minimisation variables and it should be distortionless. For that reason, the BMVDR is chosen, which will be called $\mathbf{w}_{BM}$ from here on.

In frequency bins where there is no or hardly any target signal present, the threshold in Equation (2.41) will be a very low value. This will cause the power threshold Equation (2.40) to not allow (hardly) any unconstrained interferer power to be present at the output in that frequency bin. However, if the interferers are heavily suppressed in those frequency bins as well, some interferers might be inaudible, regardless of low SNR, simply by how low their absolute power is. To take this into account, a simple heuristic is used. A minimal value is assigned to $\beta^{-1}P_s$. In this report, $-50$ dB is used as a minimum value for $\beta^{-1}P_s$, approximately corresponding to the level difference between an ordinary conversation and barely audible sounds. Since effects like interband masking (masking that occurs across critical bands) and non-simultaneous masking are otherwise not taken into consideration, setting a minimal threshold value as proposed allows exceptionally quiet interferers, that are highly likely masked, to be binaurally unconstrained.

### 2.3.2. Pre-determined Relative Acoustic Transfer Functions

The algorithms derived before all depend on the noise CPSD matrix, the interferer powers and their ATFs or RTFs. In practice, these are unknown. The estimation of these parameters in dynamic scenarios is highly challenging [7].

Pre-determined relative acoustic transfer functions (PRTFs), as proposed in [7], are used to eliminate the need for estimation of the ATFs of the sources. The PRTFs are chosen to be equally spaced on a circle in the horizontal plane around the microphones. This means that no direction is considered to be more important and no elevation is considered. If some directions are deemed more important than others, the PRTFs can be more dense in that area, making the mismatches that occur when a source is not exactly in the direction of a PRTF smaller.

In this section, the changes to the algorithms that are needed to use PRTFs instead of ATFs will be presented. A database of head related impulse responses [8] is used to obtain the required PRTFs. Taking $\mathbf{q}_i$ to be the $i$th PRTF, the PRTFs can be obtained by

$$\begin{aligned} \mathbf{q}_{iL} &= \frac{\mathbf{q}_i}{q_{iL}}\,, \\ \mathbf{q}_{iR} &= \frac{\mathbf{q}_i}{q_{iR}}\,, \end{aligned} \tag{2.44}$$

where $q_{iL}$ and $q_{iR}$ are the pre-determined ATFs at the left and right reference microphone respectively. Using these PRTFs, new constraint matrices can be set-up:

$$\mathbf{\Lambda_q} = \begin{bmatrix} \mathbf{q}_{1L} & \mathbf{q}_{2L} & \cdots & \mathbf{q}_{mL} \\ -\mathbf{q}_{1R} & -\mathbf{q}_{2R} & \cdots & -\mathbf{q}_{mR} \end{bmatrix}, \tag{2.45}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{\Lambda_a} & \mathbf{\Lambda_q} \end{bmatrix}, \tag{2.46}$$

where $m$ is the total number of PRTFs. The target constraints are still using the true ATFs, such that all methods are distortion-less. This means BMVDR remains unchanged (as it uses zero PRTFs) and JBLCMV can use the constraint matrix from Equation (2.46) with $m \leq 2M - 3$. The proposed methods can use the constraint matrix from Equation (2.46) with any number of PRTFs, since the maximum allowed number of binaurally constrained sources is already constrained by the optimisation problems.

If the far-field assumption can be made, PRTFs on a single circle suffice, as they become approximately distant invariant. Equation (2.47) gives the Fraunhofer distance, where $D$ is the maximum distance between microphones and $\lambda_{min}$ is the smallest wavelength, given by Equation (2.48).

$$d_f = \frac{2D^2}{\lambda_{min}}\,, \tag{2.47}$$

$$\lambda_{min} = \frac{2c}{F_s}\,, \tag{2.48}$$

where $F_s$ is the sampling frequency. The Fraunhofer distance is related to the far field assumption as follows [17]: The far field assumption holds when

$$d \gg d_f \approx 2.16\,\text{m}\,, \tag{2.49}$$

$$d \gg D \approx 215\,\text{mm}\,, \tag{2.50}$$

where $d$ is the distance of the source to the receivers. The approximation of $d_f$ was determined with $c = 343\,\text{m}\,\text{s}^{-1}$ and $F_s = 16.0\,\text{kHz}$, which is typical in hearing aids. Since the database that is used to obtain the

PRTFs offers impulse responses for distances of 0.80 m and 3.00 m, all PRTFs are determined using the head related impulse responses at distance 3.00 m. Only considering the horizontal plane that the microphones are in is a choice for simplicity. In reality, PRTFs for different elevations would improve performance in terms of localisation errors.

The different powers in the directions of the PRTFs, used in the diagonal matrix $\mathbf{U}$ in Equations (2.28) and (2.36), need to be known as well. To estimate these powers, MVDR beamscanning is used [18]. The beamscanner is described by

$$\hat{u}_i = \sqrt{\frac{1}{\mathbf{q}_{iL}^H \mathbf{P}^{-1} \mathbf{q}_{iL}}}\,, \quad \text{for} \quad i = 1, 2, \ldots, m\,. \tag{2.51}$$

The interferer power matrix for use in the proposed optimisation problems, Equations (2.28) and (2.36), is then set-up as

$$\hat{\mathbf{U}} = \begin{bmatrix} \hat{u}_1 & & & \\ & \hat{u}_2 & & \\ & & \ddots & \\ & & & \hat{u}_m \end{bmatrix}, \tag{2.52}$$

and the PRTF matrix for the power constraint as

$$\hat{\mathbf{B}} = \begin{bmatrix} \mathbf{q}_{1L} & \mathbf{q}_{2L} & \cdots & \mathbf{q}_{mL} \\ \mathbf{q}_{1L} & \mathbf{q}_{2L} & \cdots & \mathbf{q}_{mL} \end{bmatrix}. \tag{2.53}$$

Note that $\hat{\mathbf{U}}$ and $\hat{\mathbf{B}}$ are both determined using the left PRTFs. They can be determined using the right PRTFs, which will yield equivalent results.

## 2.4. Proposed Methods

The proposed methods and associated algorithms to solve them are briefly summarised here. Both can be implemented with estimated ATFs, RTFs or PRTFs by choosing the appropriate constraint matrices as discussed in Sections 1.2.5 and 2.3.2. This thesis only considers the RTFs and PRTFs. The proposed methods below are written as though the true RTFs are used, but can be changed to use PRTFs by changing all occurences of $\tilde{\mathbf{B}}$ and $\mathbf{U}$ to $\hat{\mathbf{B}}$ and $\hat{\mathbf{U}}$ from Equations (2.52) and (2.53), respectively.

### 2.4.1. First Proposed Method

This proposed method is based on the relaxations that resulted in Equation (2.28) and completing it using the methods discussed in Sections 2.2.4 and 2.3.1. This method is based on introducing an error to the matrix $\left[\mathbf{I} - \mathrm{diag}(\hat{\mathbf{p}})\right]$, such that is becomes positive definite. The complete convex optimisation problem is given by

$$\underset{\hat{\mathbf{p}}\in[0,1]^r,\ \mathbf{Q}\in\mathbf{S}_{++}^{r+2},\ t_1}{\text{minimise}} \quad \mathbf{f}^H\mathbf{Q}\mathbf{f} + \alpha \max(0, t_1 - c) \tag{2.54a}$$

$$\text{subject to} \quad \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1}\,\mathrm{diag}\big(\begin{bmatrix}\mathbf{1}_2^T & \hat{\mathbf{p}}^T\end{bmatrix}\big) & \mathbf{G}^{-1} \\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1} \end{bmatrix} \succeq 0 \tag{2.54b}$$

$$\begin{bmatrix} \mathbf{I} + \left(\epsilon^{-1} - 1\right)\mathrm{diag}(\hat{\mathbf{p}}) & \mathbf{U}\tilde{\mathbf{B}}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{Q}\mathbf{f} \\ \mathbf{f}^H\mathbf{Q}\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\tilde{\mathbf{B}}\mathbf{U} & t_1 \end{bmatrix} \succeq 0 \tag{2.54c}$$

$$\mathbf{1}^T\hat{\mathbf{p}} \leq 2M - 3 \tag{2.54d}$$

$$\text{with} \quad \beta^{-1}\mathbf{w}_{BM}^H(\mathbf{P}_\mathbf{y} - \mathbf{P})\mathbf{w}_{BM} = c\,, \tag{2.54e}$$

where $\mathbf{w}_{BM}$ is the BMVDR beamformer. The parameter $\epsilon$ is chosen to be 0.01. The weight $\alpha$ can be set very high if the threshold relaxation should not be used, or it can be be left out completely. Section 3.1 will detail when this is the case.

Algorithm 2 describes the method that is used to obtain the binary solution vector $\mathbf{p}^\star$ from the continuous $\hat{\mathbf{p}}^\star$ obtained from Equation (2.54). When $L = 1$ it is identical to Algorithm 1. The parameter $L$ will be discussed further in Section 3.4.4. Furthermore, the input argument $c$ of Algorithm 2 should be $\max(t_1^\star, c)$ from Equation (2.54).

**Algorithm 2** Randomised Rounding for $L_{rr}$ realisations

**function** RANDOMROUNDING($\hat{\mathbf{p}}^{\star}, k_r, L_{rr}, c, \tilde{\mathbf{B}}, \mathbf{U}, \tilde{\mathbf{P}}^{-1}, \mathbf{C}, \mathbf{f}$)
  $constraint \leftarrow \inf$
  $i \leftarrow 0$
  **while** $i \leq k_r$ **do** ▷ Stop after $k$ iterations
    **if** $i = k_r$ **then** ▷ Do a safe guess after maximum number of iterations
      **return** $\mathbf{p}^{\star} \leftarrow$ ones for the $2M-3$ strongest sources after BMVDR filtering
    **else**
      $i \leftarrow i + 1$
      **for all** $p_{jl}$ **do**
        $p^{\star}_{jl} \leftarrow 1$ with probability $\hat{p}^{\star}_{j}$ $(0$ otherwise$)$ ▷ Random rounding for $l = 1, \ldots, L_{rr}$
      **end for**
      $\Omega \triangleq \{l \mid f_1(\mathbf{p}_l) \leq c, \|\mathbf{p}_l\|_1 \leq 2M-3, l = 1, \ldots, L\}$ ▷ Set of feasible realisations [16]
      **if** $\Omega$ is not empty **then**
        **return** $\mathbf{p}^{\star} \leftarrow \arg\min f_0(\mathbf{p}_l)$ ▷ Return the best feasible realisation
      **end if**
    **end if**
  **end while**
**end function**

### 2.4.2. Second Proposed Method

The second proposed method is obtained from Equation (2.36) and adding the discussed methods from Sections 2.2.4 and 2.3.1. This method used a relaxation of the power constraint Equation (2.24) that over-estimated the summed binaurally unconstrained interferer power after filtering by applying the JBLCMV on the unconstrained interferers, instead of $\mathbf{w}_p$. This yields

$$\underset{\hat{\mathbf{p}} \in [0,1]^r,\, t_0,\, t_1}{\text{minimise}} \quad t_0 + \alpha \max(0, t_1 - c) \tag{2.55a}$$

$$\text{subject to} \quad \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1} \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_2^T & \hat{\mathbf{p}}^T \end{bmatrix}\right) & \mathbf{G}^{-1}\mathbf{f} \\ \mathbf{f}^H \mathbf{G}^{-1} & t_0 \end{bmatrix} \succeq 0 \tag{2.55b}$$

$$\left\| (\mathbf{I} - \operatorname{diag}(\hat{\mathbf{p}})) \mathbf{U} \tilde{\mathbf{B}}^H \mathbf{w}_{JB} \right\|_2^2 \leq t_1 \tag{2.55c}$$

$$\mathbf{1}^T \hat{\mathbf{p}} \leq 2M - 3 \tag{2.55d}$$

$$\text{with} \quad \beta^{-1} \mathbf{w}_{BM}^H (\mathbf{P}_{\mathbf{y}} - \mathbf{P}) \mathbf{w}_{BM} = c\,, \tag{2.55e}$$

where $\mathbf{w}_{BM}$ is the BMVDR beamformer. To obtain the binary solution vector, Algorithm 2 is used exactly as discussed in Section 2.4.1. Again, the weight $\alpha$ can be set very high if the power threshold relaxation should only be used if no other feasible solution exists. This is likely the desired behaviour, though it can be tuned if desired by the user.

# 3 Results

A number of experiments have been done to verify the proposed methods. These experiments have been run for the comparison methods as well, namely BMVDR, JBLCMV, Exhaustive Search (ES), Random Constrained Interferer Selection (R1), and Random Interferer Set Permutation (R2). ES, R1 and R2 are detailed in Section 3.2. The proposed methods are named as follows:

**Proposed 1** Proposed relaxation using selection matrix regularisation, as in Section 2.4.1.

**Proposed 2** Proposed relaxation using over-estimation of output interferer power, as in Section 2.4.2.

The experiments test two different scenarios. Using PRTFs or true RTFs.

## 3.1. Experiments

To verify the proposed methods a number of experiments were set up. The simulation experiments are built to illustrate the behaviour of the proposed methods compared to existing methods and comparison methods when certain variables change. Performance measures are the SNR gains and weighted ILD and IPD errors. How these are measured is detailed in Sections 3.3.1 and 3.3.2.

The experiments are run using four microphones, two on each ear. This means the JBLCMV constraint bound is $2M - 3 = 5$ (P)RTFs. The experiments that have been done can be categorised as:

1. Using true RTFs, to test the theoretical validity of the proposed methods.
2. Using PRTFs, to evaluate the practical application of the method.

The experiments are otherwise kept similar, to be able to make a fair comparison of true RTF versus PRTF performance. Details on the specific environment that was simulated are given in Section 3.1.2. The number of interferers will be varied between 1 and 9. In the experiment where PRTFs are used, 8, 12 and 24 PRTFs are tested.

The choice of the threshold $c$ in Equations (2.54) and (2.55) contains an SNR parameter $\beta$. As this parameter increases, SNR gains are expected to drop and spatial error should drop as well, with both eventually equalling JBLCMV performance. Whether it influences the optimality gap between the proposed methods and the exhaustive search will be tested as well.

In the experiments where the bound $5$ is crossed, different values of the penalty weight $\alpha$ in Equations (2.54) and (2.55) are tested. A low value of $\alpha$ means the originally set power threshold $c$ is allowed to be crossed more than when $\alpha$ is high. This might allow for higher SNR gains, but poorer ILD and IPD errors.

Additionally, the differences in performance due to the value of $L_{rr}$ in randomised rounding is evaluated, by doing the first experiment for six different values of $L_{rr}$. The values tested are $1$, $10$, $20$, $30$, $40$ and $50$. These results are discussed separately in Section 3.4.4.

### 3.1.1. Analysis and Synthesis

To simulate recordings in the acoustic scene, four microphones from [8] are used. The front and back microphones on each ear to be specific. The corresponding impulse responses that are used for the desired sources, described in Section 3.1.2, are resampled to 16 kHz, which is a typical sampling frequency in HADs. The front microphones are labelled as the reference microphones.

Figure 3.1 shows how and when the signals are obtained and processed. The target signal is a speech recording from the TIMIT database [19], a database of various speakers and English text. The interferers are either WGN or speech-shaped Gaussian noise. These sources are convolved with their respective impulse responses. All obtained signals corresponding to one microphone are added together and AWGN is added as microphone self noise. This is described equivalently in the discrete time domain by

$$y_j[n] = s[n] \star h_{sj}[n] + \sum_{i=1}^{r} u_i[n] \star h_{ij}[n] + v_j[n] \,, \quad \text{for} \quad j = 1, \ldots, 4 \,, \tag{3.1}$$

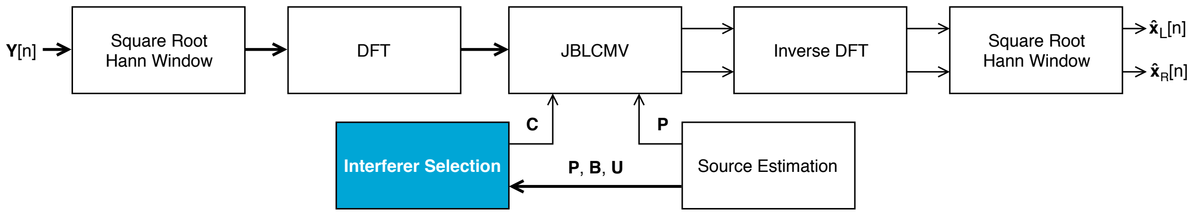



Figure 3.1: The beamformer, interferer selection and source estimation process, including the Analysis and Synthesis process. The matrix $\mathbf{Y}[n]$ contains the 4 microphone recording segments. Each is 16 ms long and is updated every 8 ms, for 50 % overlap. Source estimation is done using either the true RTFs and powers estimated using Welch's method, or using PRTFs and beamscanning. The vectors $\hat{\mathbf{x}}_L[n]$ and $\hat{\mathbf{x}}_R[n]$ are combined with those of the previous time segment, using 50 % overlap-and-add.

| Source | Distance [cm] | Angle [°] | Type |
|---|---|---|---|
| $\mathbf{x}$ | 80 | 95 | Speech |
| $\mathbf{n}_1$ | 300 | 15 | SSGN |
| $\mathbf{n}_2$ | 300 | 45 | WGN |
| $\mathbf{n}_3$ | 300 | 75 | WGN |
| $\mathbf{n}_4$ | 300 | 105 | SSGN |
| $\mathbf{n}_5$ | 300 | 165 | WGN |
| $\mathbf{n}_6$ | 300 | 225 | SSGN |
| $\mathbf{n}_7$ | 300 | 240 | WGN |
| $\mathbf{n}_8$ | 300 | 270 | SSGN |
| $\mathbf{n}_9$ | 300 | 300 | SSGN |

Table 3.1: All possible sources in the acoustic scene. Each interfering source is either white Gaussian noise (WGN) or speech-shaped Gaussian noise (SSGN). All sources have equal power.

where $y_j[n]$ is the $j$th microphone received signal, $s[n]$ is the target signal, $h_{sj}[n]$ is the impulse response from the target to the $j$th microphone, $u_i[n]$ is the $i$th interferer and $v_j[n]$ is the AWGN on the $j$th microphone.

An STFT is performed on each of the received signals, using 16 ms frame length, a square-root Hann window and 50 % overlap. This is where the beamformer processing is performed, per frequency bin, per time frame. After processing, the inverse Fourier transform is performed on the results. Then, another square-root Hann window is applied to each time frame. After addition of the appropriate time frames, with 50 % overlap, the left and right recovered signals are obtained.

The CPSD matrices are estimated across the entire recording. In applications, the noise CPSD matrix could for example be estimated during noise-only windows (using voice activity detection, for example). The interferer noise CPSD $\mathbf{P}_\mathbf{n}$ is determined by

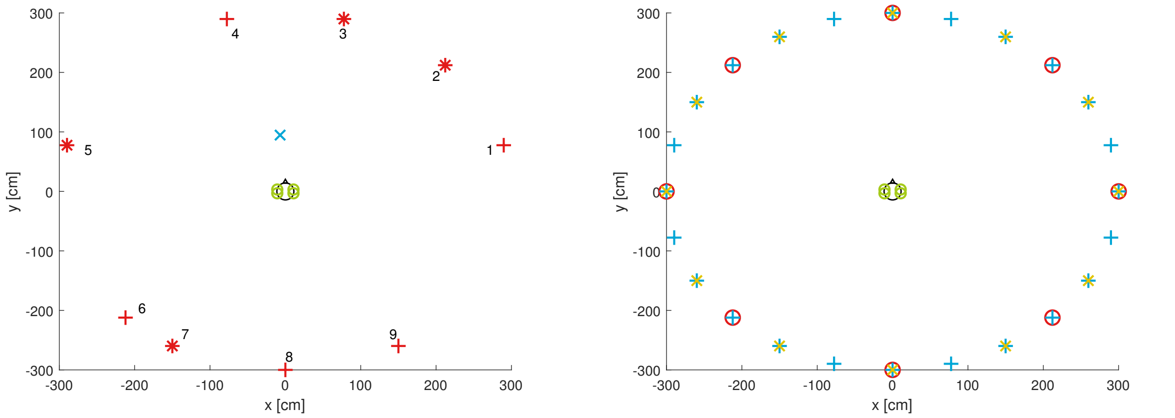
$$\mathbf{P}_\mathbf{n} = \mathbf{B}_L \mathbf{U}^2 \mathbf{B}_L^H \,, \tag{3.2}$$

where $\mathbf{U}$ is appropriately scaled to the left RTFs. The diagonal power matrix $\mathbf{U}$ is determined using Welch's method, when true RTFs are used. When PRTFs are used, beamscanning is used to determine the powers, as described in Section 2.3.2, and no scaling is required.

### 3.1.2. Acoustic Scene

The experiments are done using a number of different acoustic scenes. The difference between the acoustic scenes is how many interfering point sources are present. The sources that can be present in the scene are shown in Table 3.1 and Figure 3.2a. The target signal $\mathbf{x}$ is present in all experiments. Interferers are present or not, based on the experiment. If an experiment states that 3 interferers are present, sources $\mathbf{n}_1$ through $\mathbf{n}_3$ are present, for example. The number of interferers present is varied accordingly in all experiments, as described in Section 3.1.

For experiments using PRTFs, the same sources are used as described in Table 3.1 and Figure 3.2a. The PRTFs are placed uniformly on a circle around the microphones, using the head related impulse responses at 300 cm distance, as described in Section 2.3.2 and illustrated in Figure 3.2b. Three sets of PRTFs are




(a) Locations of the point sources. The target is indicated by ×, SSGN interferers by + and WGN interferers by ∗.

(b) The locations of the PDRTFs. ○ indicates a PRTF in $\mathcal{P}_8$, × one in $\mathcal{P}_{12}$ and + a PRTF in $\mathcal{P}_{24}$. The definitions of $\mathcal{P}_8$, $\mathcal{P}_{12}$ and $\mathcal{P}_{24}$ are given by Equations (3.3) to (3.5), respectively.

Figure 3.2: Top-down view of the acoustic scene. Microphones are indicated by ○ and are centered around the origin.

used in the experiments. Let

$$\mathcal{P}_8 \triangleq \{(\mathbf{q}_L(\gamma), \mathbf{q}_R(\gamma)) \,|\, \gamma = 45°k\,,\ k = 0, 1, \ldots, 7\}\,, \tag{3.3}$$

$$\mathcal{P}_{12} \triangleq \{(\mathbf{q}_L(\gamma), \mathbf{q}_R(\gamma)) \,|\, \gamma = 30°k\,,\ k = 0, 1, \ldots, 11\}\,, \tag{3.4}$$

$$\mathcal{P}_{24} \triangleq \{(\mathbf{q}_L(\gamma), \mathbf{q}_R(\gamma)) \,|\, \gamma = 15°k\,,\ k = 0, 1, \ldots, 23\}\,, \tag{3.5}$$

be the PRTF sets for $8$, $12$ and $24$ PRTFs respectively, where $(\mathbf{q}_L(\gamma), \mathbf{q}_R(\gamma))$ is the PRTF couple corresponding to an azimuth angle of $\gamma$.

## 3.2. Comparison Methods

Since the proposed methods include randomness, due to the randomised rounding algorithm, two additional random selection methods are given here, which will be used to compare the results to. By comparing the proposed methods to these random methods, their validity can be shown more evidently. An exhaustive search is also described.

These methods are intentionally kept very simple, as they are not proposed methods, they are only introduced to show an optimal case and two different random selection methods, to check whether the proposed methods approach optimal and outperform simple random selection methods.

### 3.2.1. Exhaustive Search

The exhaustive search determines feasibility in Equation (2.6) for every possible realisation of $\mathbf{p}$. Of all feasible solution vectors, the solution vector that gives the lowest value of Equation (2.6a) is selected as the optimal solution. The implementation is shown in Algorithm 3.

### 3.2.2. Random Constrained Interferer Selection

Instead of being based on convex relaxations or using a cost function, this method randomly selects interferers or PRTFs to binaurally constrain. For each interferer or PRATF, there is a 50% chance that it is selected or not. As such, the number of interferers or PRATFs selected is also random. The obtained set is checked for feasibility though, to ensure a fair comparison to the proposed methods, since non-feasible $\mathbf{p}$ vectors are likely to produce better SNR gains. Algorithm 4 shows the implementation of the method.

### 3.2.3. Random Interferer Set Permutation

The third comparison method is based on a random permutation of a solution set from one of the proposed methods. This method is used to check whether the right interferers were selected to binaurally constrain, and not just the right number.

**Algorithm 3** Exhaustive Search

**function** EXHAUSTIVESEARCH($c, \tilde{\mathbf{B}}, \mathbf{U}, \tilde{\mathbf{P}}^{-1}, \mathbf{C}, \mathbf{f}$)
  $best \leftarrow \inf$
  **for** $i \leftarrow 0, 2^r - 1$ **do** ▷ All possible combinations
    $\mathbf{p} \leftarrow i_{(10 \mapsto 2)}$ ▷ $\mathbf{p}$ becomes the binary representation of $i$
    $\mathbf{\Phi_p} \leftarrow \text{diag}(\mathbf{p})$ with its zero columns removed.
    $\mathbf{\Phi} \leftarrow \begin{bmatrix} \mathbf{I}_2 & 0 \\ 0 & \mathbf{\Phi_p} \end{bmatrix}$
    $\mathbf{w_p} \leftarrow \tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{C}^H\tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f}$ ▷ Actual filter
    $constraint \leftarrow \left\|(\mathbf{I} - \text{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w_p}\right\|_2^2$ ▷ Binaurally unconstrained interferer power sum
    $power \leftarrow \mathbf{w_p}^H\tilde{\mathbf{P}}\mathbf{w_p}$ ▷ Determine output noise power
    **if** $power < best \wedge constraint < c$ **then** ▷ Check if feasible and better
      $best \leftarrow power$ ▷ Save new best
      $\mathbf{p}^\star \leftarrow \mathbf{p}$
    **end if**
  **end for**
  **return** $\mathbf{p}^\star$
**end function**

**Algorithm 4** Random Constrained Interferer Selection

**function** RANDOMSELECTION($k_r, c, \tilde{\mathbf{B}}, \mathbf{U}, \tilde{\mathbf{P}}^{-1}, \mathbf{C}, \mathbf{f}$)
  $constraint \leftarrow \inf$
  $i \leftarrow 0$
  **while** $constraint > c \wedge i \leq k_r$ **do** ▷ Stop when constraint is met or after $k$ iterations
    **if** $i = k_r$ **then** ▷ Maximum number of iterations
      $\mathbf{p} \leftarrow \mathbf{1}$ ▷ Guaranteed feasibility
    **else**
      $i \leftarrow i + 1$
      **for all** $p_i$ **do**
        $p_i \leftarrow 1$ with probability $0.5$ ($0$ otherwise)
      **end for**
      $\mathbf{\Phi_p} \leftarrow \text{diag}(\mathbf{p})$ with its zero columns removed.
      $\mathbf{\Phi} \leftarrow \begin{bmatrix} \mathbf{I}_2 & 0 \\ 0 & \mathbf{\Phi_p} \end{bmatrix}$
      $\mathbf{w_p} \leftarrow \tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{C}^H\tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f}$ ▷ Actual filter
      $constraint \leftarrow \left\|(\mathbf{I} - \text{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w_p}\right\|_2^2$ ▷ Binaurally unconstrained interferer power sum
    **end if**
  **end while**
  **return** $\mathbf{p}$
**end function**

**Algorithm 5** Random Interferer Set Permutation

**function** RANDOMPERMUTATION($\mathbf{p}^\star, k_r, c, \tilde{\mathbf{B}}, \mathbf{U}, \tilde{\mathbf{P}}^{-1}, \mathbf{C}, \mathbf{f}$)
  $constraint \leftarrow \inf$
  $i \leftarrow 0$
  **while** $constraint > c \wedge i \leq k_r$ **do** ▷ Stop when constraint is met or after $k$ iterations
    **if** $i = k_r$ **then** ▷ Maximum number of iterations
      $\mathbf{p} \leftarrow \mathbf{1}$ ▷ Guaranteed feasibility
    **else**
      $i \leftarrow i + 1$
      $\mathbf{p} \leftarrow$ Random permutation of $\mathbf{p}^\star$
      $\mathbf{\Phi_p} \leftarrow \mathrm{diag}(\mathbf{p})$ with its zero columns removed.
      $\mathbf{\Phi} \leftarrow \begin{bmatrix} \mathbf{I}_2 & 0 \\ 0 & \mathbf{\Phi_p} \end{bmatrix}$
      $\mathbf{w_p} \leftarrow \tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{C}^H\tilde{\mathbf{P}}^{-1}\mathbf{C}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{f}$ ▷ Actual filter
      $constraint \leftarrow \left\|(\mathbf{I} - \mathrm{diag}(\mathbf{p}))\mathbf{U}\tilde{\mathbf{B}}^H\mathbf{w_p}\right\|_2^2$ ▷ Binaurally unconstrained interferer power sum
    **end if**
  **end while**
  **return** $\mathbf{p}$
**end function**

This method receives a solution vector $\mathbf{p}^\star$ from one of the proposed methods, performs a random permutation on the elements of that vector and checks the new vector for feasibility. The procedure is repeated until a feasible vector is found (this may be the original vector). The method is detailed fully by Algorithm 5.

## 3.3. Performance Measures

The SNR gain and spatial cue errors are the primary performance measures. In Section 3.4.4 the number of iterations of randomised rounding is evaluated as well, which requires no further introduction. The method of determining SNR gain is defined in Section 3.3.1. The spatial cue errors are weighted ILD and IPD errors. The weights are based on perception models and are detailed in Section 3.3.2.

### 3.3.1. SNR gain comparisons

The SNR gain describes the improvement of SNR of the target compared to all noise at the input and output of the beamformer [4]. Input and output SNR are given by

$$\mathrm{SNR}_{\mathrm{in}}[k] = \frac{\mathbf{e}^H\tilde{\mathbf{P}}_{\mathbf{x}}[k]\mathbf{e}}{\mathbf{e}^H\tilde{\mathbf{P}}[k]\mathbf{e}}, \tag{3.6}$$

$$\mathrm{SNR}_{\mathrm{out}}[k] = \frac{\mathbf{w}^H[k]\tilde{\mathbf{P}}_{\mathbf{x}}[k]\mathbf{w}[k]}{\mathbf{w}^H[k]\tilde{\mathbf{P}}[k]\mathbf{w}[k]}, \tag{3.7}$$

respectively, where $\tilde{\mathbf{P}}_{\mathbf{x}}$ is the true target CPSD matrix and $\mathbf{e} = \begin{bmatrix} 1 & \mathbf{0}_{2M-2}^T & 1 \end{bmatrix}^T$ is a selection vector to select the left and right reference microphones, which are indexed as 1 and $M$ respectively. Let the SNR gain in the $k$th frequency bin be

$$G_{\mathrm{SNR}}[k] = \frac{\mathrm{SNR}_{\mathrm{out}}[k]}{\mathrm{SNR}_{\mathrm{in}}[k]}. \tag{3.8}$$

This gain is averaged over all frequency bins to obtain the average SNR gain, resulting in

$$\bar{G}_{\mathrm{SNR}} = \frac{1}{\frac{1}{2}K + 1}\sum_{k=1}^{\frac{1}{2}K+1} \frac{\mathbf{w}^H[k]\tilde{\mathbf{P}}_{\mathbf{x}}[k]\mathbf{w}[k]}{\mathbf{w}^H[k]\tilde{\mathbf{P}}[k]\mathbf{w}[k]}\left(\frac{\mathbf{e}^H\tilde{\mathbf{P}}_{\mathbf{x}}[k]\mathbf{e}}{\mathbf{e}^H\tilde{\mathbf{P}}[k]\mathbf{e}}\right)^{-1}, \tag{3.9}$$

where $K$ is the total number of frequency bins.

| Band No. | Bandwidth [Hz] | Band No. | Bandwidth [Hz] | Band No. | Bandwidth [Hz] |
|---|---|---|---|---|---|
| 1 | $0-100$ | 9 | $920-1080$ | 16 | $2700-3150$ |
| 2 | $100-200$ | 10 | $1080-1270$ | 17 | $3150-3700$ |
| 3 | $200-300$ | 11 | $1270-1480$ | 18 | $3700-4400$ |
| 4 | $300-400$ | 12 | $1480-1720$ | 19 | $4400-5300$ |
| 5 | $400-510$ | 13 | $1720-2000$ | 20 | $5300-6400$ |
| 6 | $510-630$ | 14 | $2000-2320$ | 21 | $6400-7700$ |
| 7 | $630-770$ | 15 | $2320-2700$ | 22 | $7700-9500$ |
| 8 | $770-920$ | | | | |

Table 3.2: Idealised Critical Band Filter Bank [21], for a system using $F_s \leq 19$ kHz.

### 3.3.2. Spatial Cue Errors

The spatial cue errors are weighted according to their respective critical band SNRs (CBSNRs) compared to the target [20]. The CBSNR can be used as a measure to determine how audible noise is in the presence of the target, and vice-versa. When the SNR in a critical band is better than 24 dB, the noise is considered inaudible, and when it is poorer than $-4$ dB, the target is considered masked [21].

The CBSNR is determined by

$$\text{CBSNR}_{i,j} = \frac{\sum_{k\in\text{CB}_i} \mathbf{w}^H[k]\tilde{\mathbf{P}}_{\mathbf{x}}[k]\mathbf{w}[k]}{\sum_{k\in\text{CB}_i} \mathbf{w}^H[k]\tilde{\mathbf{P}}_{\mathbf{n}_j}[k]\mathbf{w}[k]}\,, \tag{3.10}$$

where $\text{CB}_i$ is the set of frequency bins corresponding to the $i$th critical band, given by Table 3.2, and $\text{CBSNR}_i$ is the CBSNR in the $i$th critical band. The target is binaurally constrained in all methods, so only spatial cue errors on the interferers need to be considered. When the target is masking the noise, it does not matter whether the spatial cues of the interferer are preserved. To reflect this, the spatial cue errors are weighted by how audible the interferer is estimated to be. Let

$$\phi_{i,j} = \begin{cases} 1\,, & \text{if} \quad \text{CBSNR}_{i,j} \leq \lambda \\ 1 - \frac{\text{CBSNR}_{i,j} - \lambda}{\rho - \lambda}\,, & \text{if} \quad \lambda < \text{CBSNR}_{i,j} < \rho \\ 0\,, & \text{if} \quad \text{CBSNR}_{i,j} \geq \rho\,, \end{cases} \tag{3.11}$$

where $\lambda = -4$ dB is the noise-masks-tone threshold, $\rho = 24$ dB is the tone-masks-noise threshold and $\phi_{i,j}$ is the error weight for the $i$th critical band and $j$th interferer. The weight is a linear expression between $\lambda$ and $\rho$ dependent on the CBSNR, with zero and one as minimum and maximum weight respectively.

Using the error weight $\phi_{i,j}$ the weighted average ILD and IPD error are defined. Let

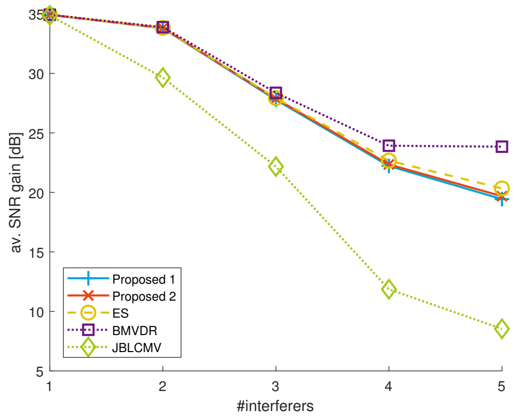
$$\bar{\mathcal{E}}^{\text{ILD}} = \frac{1}{r}\sum_{j=1}^{r} \frac{\sum_{i=1}^{22} \phi_{i,j} \sum_{k\in\text{CB}_i} \mathcal{E}_j^{\text{ILD}}[k]}{\sum_{i=1}^{22} \sum_{k\in\text{CB}_i} \phi_{i,j}}\,, \tag{3.12}$$

$$\bar{\mathcal{E}}^{\text{IPD}} = \frac{1}{r}\sum_{j=1}^{r} \frac{\sum_{i=1}^{22} \phi_{i,j} \sum_{k\in\text{CB}_i} \mathcal{E}_j^{\text{IPD}}[k]}{\sum_{i=1}^{22} \sum_{k\in\text{CB}_i} \phi_{i,j}}\,, \tag{3.13}$$

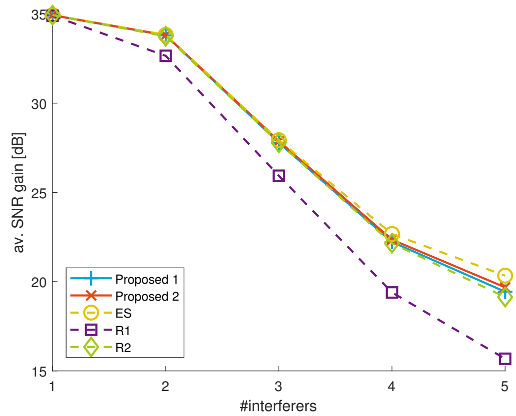
where $\bar{\mathcal{E}}^{\text{ILD}}$ and $\bar{\mathcal{E}}^{\text{IPD}}$ are the weighted average ILD and IPD respectively. These are the two spatial cue error measures that will be used to evaluate the methods.

## 3.4. Simulation Results

Simulation results are categorised by what type of transfer functions were used, true or pre-determined. In the figures, the solid curves correspond to the proposed methods, the dashed curves are comparison method results and dotted lines represent the existing methods. Unless stated otherwise, $L_{rr} = 10$ was used in randomised rounding Algorithm 2, $k_r = 1000$ was used in Algorithms 2, 4 and 5, $\epsilon = 100^{-1}$ is used in Equation (2.54) and $\lambda = 1.001 \cdot \lambda_{max}(\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda})$ is used in Equations (2.54) and (2.55).




(a) Average SNR gains (Equation (3.9)) compared to existing methods.




(b) Average SNR gains (Equation (3.9)) compared to random methods.

Figure 3.3: The proposed methods compared to the existing and random methods in terms of average SNR gains for $\beta = 10$ dB and varying number of interfering point sources present in the acoustic scene.

### 3.4.1. Using True Transfer Functions

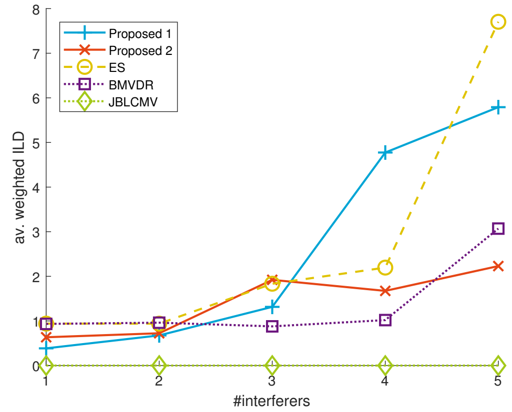
In the first experiment, true transfer functions are used in the constraint matrix $\mathbf{\Lambda}_b$. Using the results from experiment 1, the SNR gain and spatial cues errors will be compared between the different methods. Additionally, the optimality gap is examined, which is the difference in SNR gain between the proposed methods and the exhaustive search.

**With up to five interferers present in the acoustic scene**

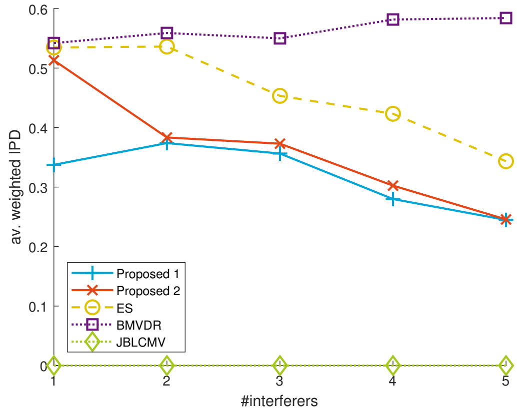
Figure 3.3 shows the average SNR gains of the different methods. When few interfering sources are present, the input SNR is relatively high. As such, all methods are close to each other, and approach the BMVDR performance. As the amount of sources, and thus noise power, in the scene increases, the average SNR gain of the methods drifts apart, with the proposed methods in between BMVDR and JBLCMV performance, as shown in Figure 3.3a. This is to be expected, since the proposed methods chose to spend degrees of freedom on noise reduction or spatial cue preservation, per frequency bin, per interfering source. Additionally, the proposed methods outperform the random methods here, in Figure 3.3b, and come very close to the optimal solution provided by the exhaustive search. The fact that the random interferer set permutation method is so close in performance to the proposed methods implies that it does not matter much which interfering sources are selected for spatial cues preservation, with regards to average SNR gain. Since all sources in the acoustic scene are equal in power, this is partially expected, though in frequency bins where speech is dominant, it should matter more which interferer is selected. It is likely that in those frequency bins, many interferer set permutations are infeasible. This would cause the random interferer set permutation method to select the same set as the proposed methods with relatively high likelihood.

Looking at the spatial cue errors in Figure 3.4, using the same setup, we notice that the ILD errors of the proposed methods and the exhaustive search can be greater than those of the BMVDR. The ILD errors can be very unpredictable, since anything can happen to the magnitude of an unconstrained source after beamforming. Additionally, the noise reducing ability of the proposed methods is lower or equal to the BMVDR, which can cause unconstrained sources to be more audible with the proposed method than with BMVDR.

This behaviour is not an issue however, as it simply indicates that the parameter $\beta$, which is the SNR threshold in the proposed methods, is not strict enough. In Figure 3.5 the results are shown when $\beta = 24$ dB. The spatial cues error are now all well below the BVMDR and much closer to JBLCMV. It should be noted though, that the spatial cues are not directly constrained in the proposed methods, which might cause some potentially unexpected behaviour when analysing the spatial cues errors. Of course, setting the value of $\beta$ to a higher, stricter, value has a negative effect as well, since it is essentially a trade-off parameter between noise reduction and spatial cue preservation. Figure 3.6 shows the SNR gain performance of the proposed methods and, comparing it to Figure 3.3, lower SNR gains of the proposed methods are immediately noticeable.

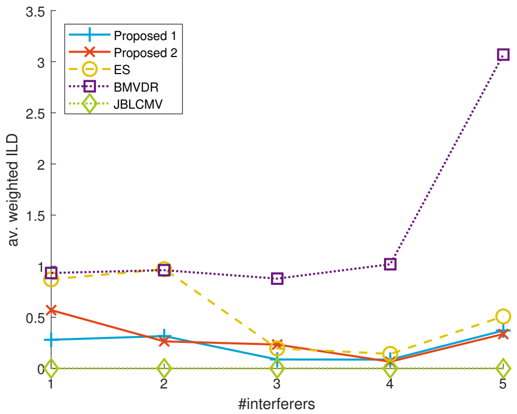



(a) Average weighted ILD errors (Equation (3.12)).

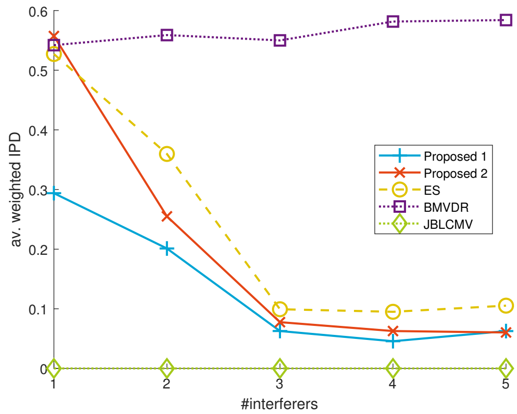



(b) Average weighted IPD errors (Equation (3.13)).

Figure 3.4: The proposed methods compared to the existing methods in terms of average weighted ILD and IPD errors for $\beta = 10$ dB and varying number of interfering point sources present in the acoustic scene.

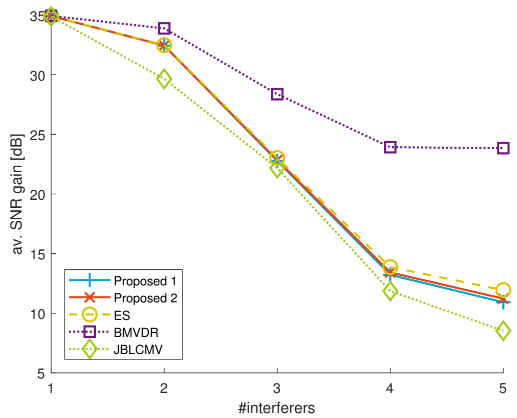



(a) Average weighted ILD errors (Equation (3.12)).

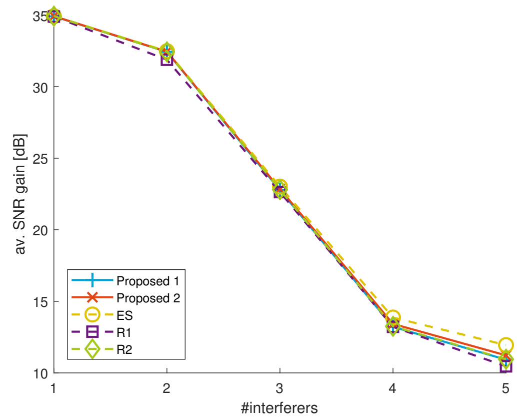



(b) Average weighted IPD errors (Equation (3.13)).

Figure 3.5: The proposed methods compared to the existing methods in terms of average weighted ILD and IPD errors for $\beta = 24$ dB and varying number of interfering point sources present in the acoustic scene.

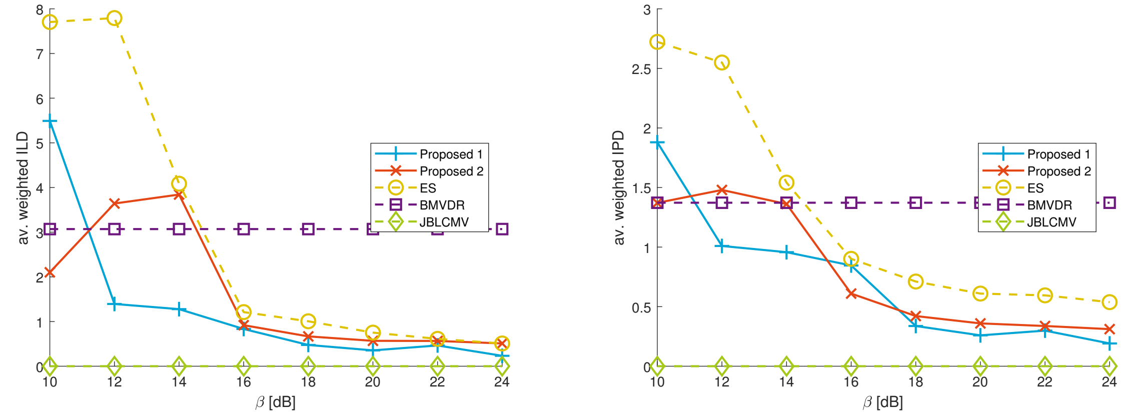



(a) Average SNR gains (Equation (3.9)) compared to existing methods.




(b) Average SNR gains (Equation (3.9)) compared to random methods.

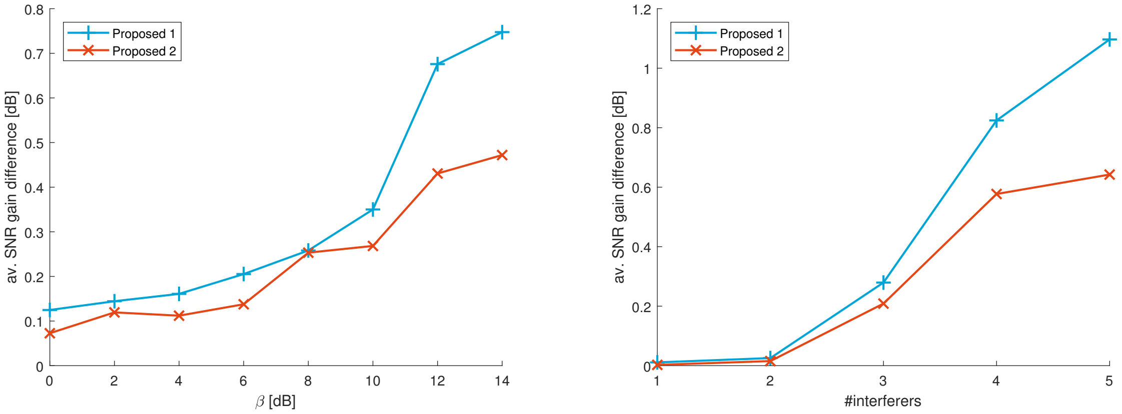
Figure 3.6: The proposed methods compared to the existing and random methods in terms of average SNR gains for $\beta = 24$ dB and varying number of interfering point sources present in the acoustic scene.




(a) ILD error with 5 interferers present, with varying $\beta$.

(b) ILD error, averaged over all amounts of interferers present (1 through 5), with varying $\beta$.

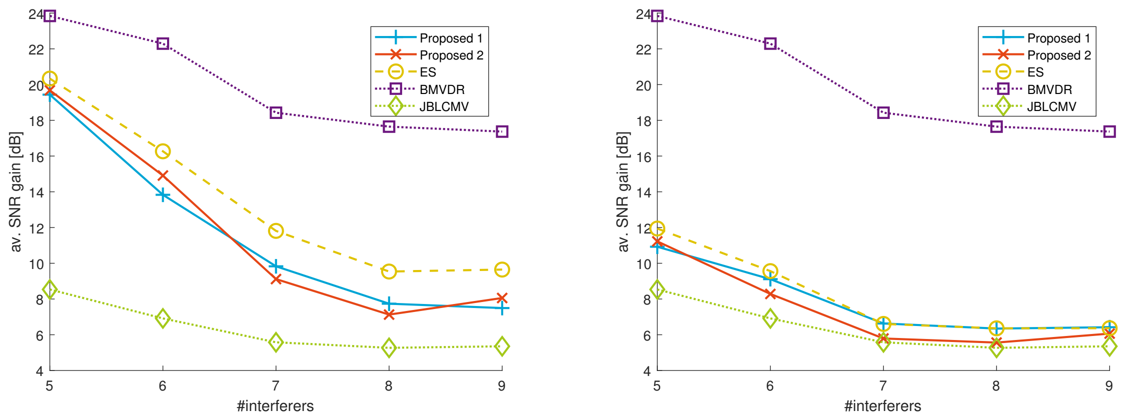
Figure 3.7: The average ILD error of the proposed methods for $L_{rr} = 50$ and varying $\beta$.




(a) Optimality gap due to $\beta$, averaged over all amounts of interferers present (1 through 5).

(b) Optimality gap due to the amount of interferers present, averaged over all values of $\beta$ (0 dB to 20 dB in steps of 2 dB).

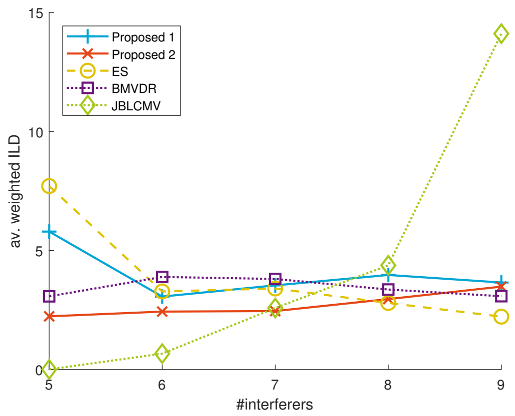
Figure 3.8: The optimality gap of the proposed methods for $L_{rr} = 50$.



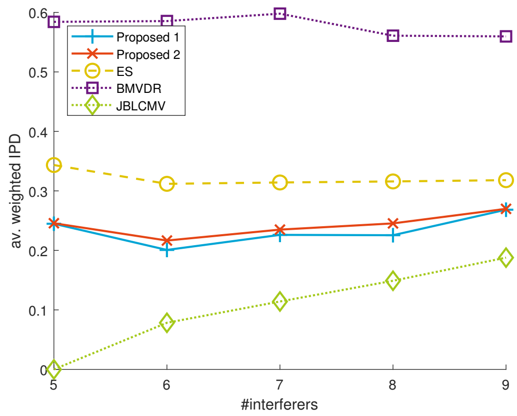

(a) Average SNR gains (Equation (3.9)) for $\beta = 10$ dB.

(b) Average SNR gains (Equation (3.9)) for $\beta = 24$ dB.

Figure 3.9: The proposed methods compared to the existing methods in terms of average SNR gains for varying number of interfering point sources present in the acoustic scene and $\alpha = 10^6$.




(a) Average weighted ILD errors (Equation (3.12)).




(b) Average weighted IPD errors (Equation (3.13)).

Figure 3.10: The proposed methods compared to the existing methods in terms of average weighted spatial cue errors for varying number of interfering point sources present in the acoustic scene, $\beta = 10\,\text{dB}$ and $\alpha = 10^6$.

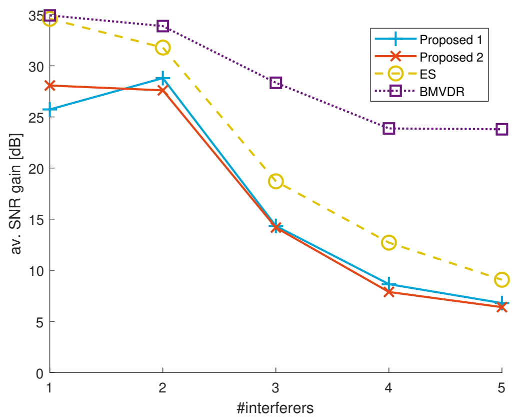
To further investigate the influence of the parameter $\beta$, Figure 3.7 shows how the average ILD error changes for different values of $\beta$. The figures clearly show that for reliable spatial cue preservation, the threshold $\beta$ should be strict enough, or the proposed method will trade-off too much spatial cue preservation for better noise reduction. With only few sources binaurally constrained, the potential spatial cue errors grow, because of effects described in [7]. In summary, the binaurally unconstrained sources get projected to a linear combination of binaurally constrained RTFs. The closer a binaurally constrained source is to a binaurally constrained RTF, the lower the expected error will be. Hence a low amount of binaurally constrained RTFs can increase the expected spatial cue errors. Additionally, Figure 3.8a shows the optimality gap increases for increasing $\beta$ as well, though it stays comparatively small. This can be attributed to the fact that, for a strict value of $\beta$, larger continuous values of $\hat{\mathbf{p}}$ become more prevalent to satisfy the power constraint in Equations (2.54c) and (2.55c). Due to this, feasible rounding results may contain more 1s than necessary. Figure 3.8 also shows that with increasing total noise present or increasing dimension of $\mathbf{p}$, and increasing values of $\beta$, the optimality gap of proposed method 1, using matrix regularisation, grows faster than that of proposed method 2, using output interferer power over-estimation. This is theorised to be due to the relaxation parameter $\epsilon$ in Equation (2.54b), which is further discussed in Section 4.1.1.

#### With up to nine interferers present in the acoustic scene

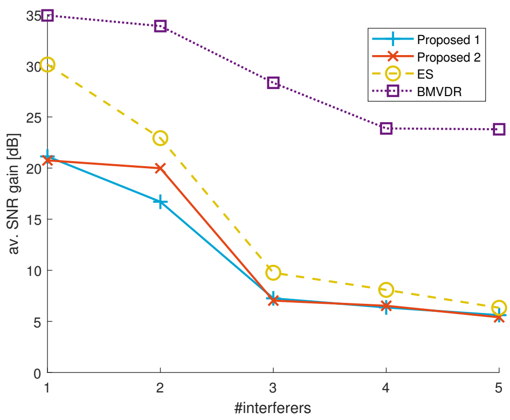
To further show the theoretical validity of the proposed methods, the number of interfering point sources in the acoustic scene is increased to a maximum of 9, before evaluating the performance for PRTFs. These experiments can show if the proposed methods still work when the JBLCMV is no longer able to binaurally constrain all interferers present in the acoustic scene. The maximum number of interferers that the JBLCMV can constrain is 5 when 4 microphones are used. Even though the JBLCMV cannot constrain all the sources present in the scene, it is included in the experiments to compare the proposed methods to. The JBLCMV here simply constrains the 5 strongest, in terms of power, interfering point sources in each frequency bin. This is a simplified choice to find the most important interferers to binaurally constrain and we will see that the proposed methods outperform this simplified choice.

Figure 3.9 shows how the SNR gain performance of the proposed methods is better than JBLCMV still, where JBLCMV uses simplified interferer selection as described above. Even when the value of $\beta$ is stricter, this is still the case. The optimality gap jumps up going from 5 to 6 interferers present, but does not appear to increase further as the amount of interferers present continues to increase. When $\beta = 24\,\text{dB}$ in Figure 3.9b, the SNR gain of the first proposed method, using matrix regularisation, even closes almost completely.

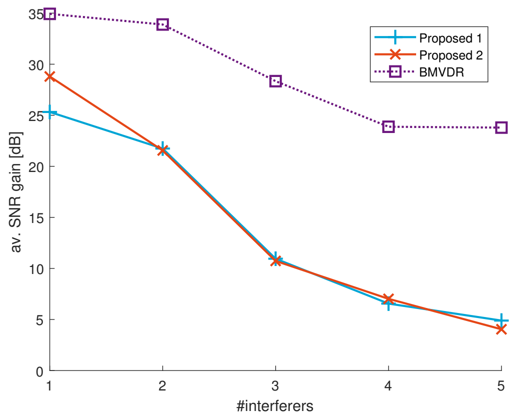
Interestingly, when observing the spatial cue errors in Figure 3.10, the JBLCMV with the simplified interferer selection performs very poorly, while the proposed methods do not increase much as the amount of interferers present increases. This further shows that the power constraint in Equation (2.24), through indirectly constraining spatial cue errors, improves spatial cue performance compared to just selecting the 5 strongest interferers. This results in the proposed methods improving both the SNR gain performance and spatial cue error performance.



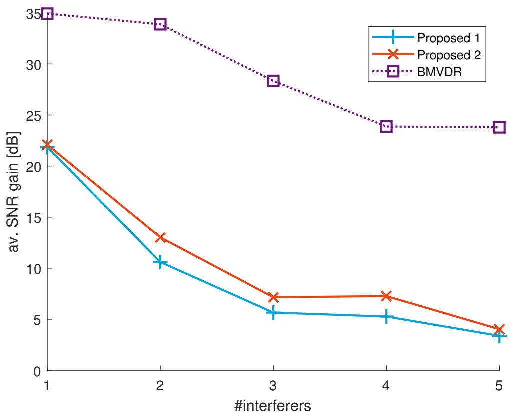

(a) Using $\mathcal{P}_8$ and $\beta = 10\,\mathrm{dB}$.



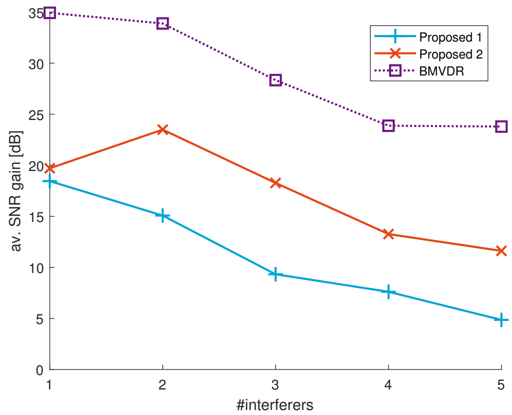

(b) Using $\mathcal{P}_8$ and $\beta = 18\,\mathrm{dB}$.



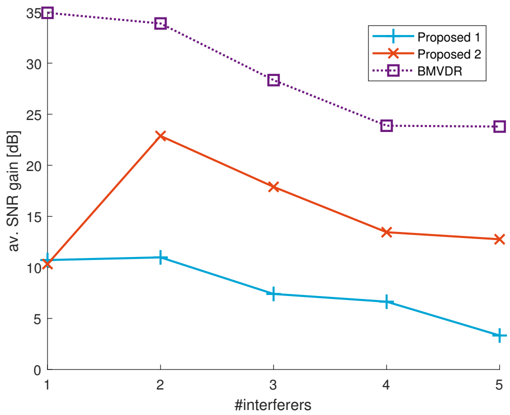

(c) Using $\mathcal{P}_{12}$ and $\beta = 10\,\mathrm{dB}$.




(d) Using $\mathcal{P}_{12}$ and $\beta = 18\,\mathrm{dB}$.

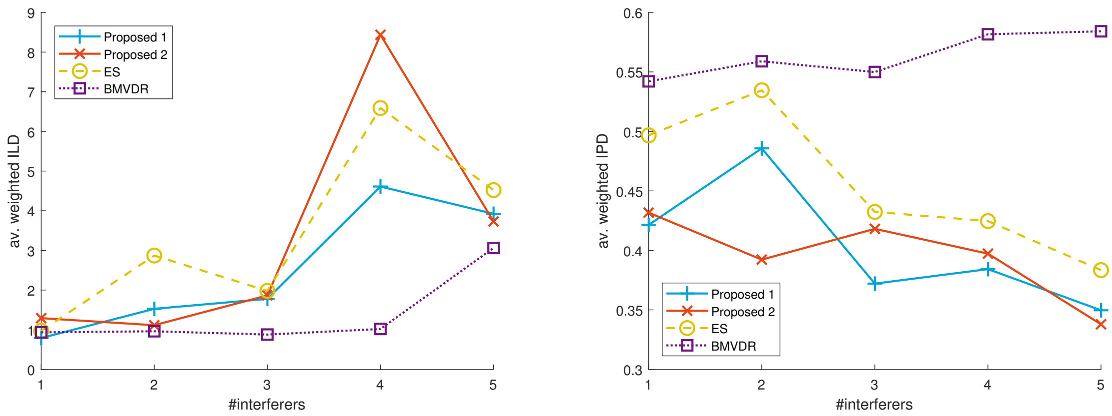



(e) Using $\mathcal{P}_{24}$ and $\beta = 10\,\mathrm{dB}$.




(f) Using $\mathcal{P}_{24}$ and $\beta = 18\,\mathrm{dB}$.

Figure 3.11: The SNR gain of the proposed methods for $\alpha = 10^6$, using PRTFs and a varying number of interfering point sources present in the acoustic scene.




(a) Weighted ILD errors.

(b) Weighted IPD errors.

Figure 3.12: The proposed methods compared to the existing methods in terms of weighted ILD and IPD errors for $\mathcal{P}_8$, $\beta = 10$ dB and varying number of interfering point sources present in the acoustic scene.

### 3.4.2. Using Pre-determined Relative Acoustic Transfer Functions

Next, PRTFs are used in the constraint matrix $\mathbf{\Lambda}_b$. As an additional variable, the effect of the number of PRTFs that are set up is also examined. The random selection methods are no longer considered here. They are very inefficient in their discussed implementation to handle cases where $m > 2M - 3$. Also, their behaviour, and how they are outperformed by the proposed methods has already been evaluated in Section 3.4.1 and Figures 3.3b and 3.6b. The exhaustive search is only used for $\mathcal{P}_8$, where the number of options to consider is still comparatively low.

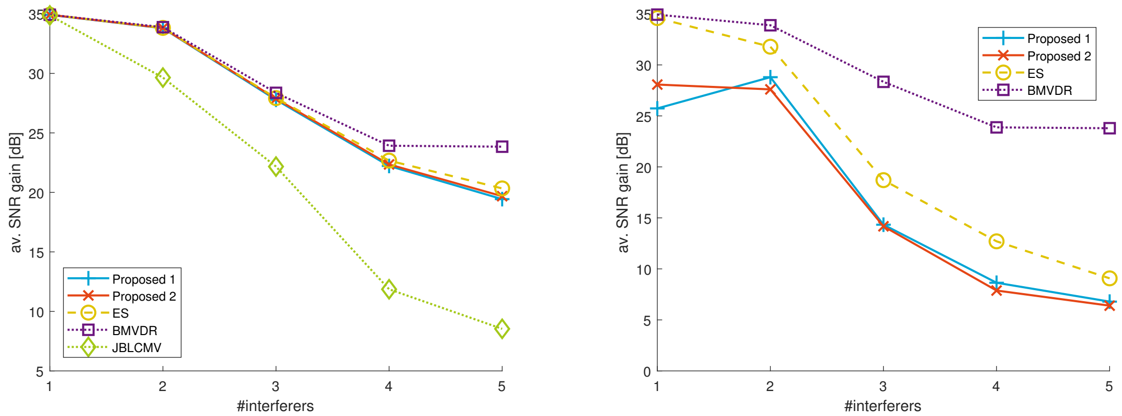
Figure 3.11 gives the SNR gain results for all three PRTFs sets that were considered. In Figures 3.11a and 3.11b it is clear that the proposed methods are still close to the optimal solution, though the optimality gap has slightly increased compared to the true RTF case, in Figure 3.8. Increasing the number of used PRTFs to 12 still yields expected results, with slightly lower, but very similar, SNR gains as with $\mathcal{P}_8$. When 24 PRTFs are used however, it is clear that proposed method 2 no longer functions properly. This is especially clear when comparing Figures 3.11b and 3.11f, where proposed method 2 achieves higher SNR gains than the exhaustive search in Figure 3.11b. This means the power threshold relaxation from Section 2.2.3 was likely used heavily. This is due to the approximate over-estimation of the output interferer power, in Equation (2.55c), being a poor over-estimation when $m \gg 2M - 3$.

Proposed method 1 however has very consistent SNR gains for the different PRTF sets. Only when the amount of interfering point sources present in the scene is low, the performance goes down for the larger PRTF sets. This is due to the power constraint relaxation in Equation (2.54c), which will add the relaxation parameter $\epsilon$ to the power sum more often when the total amount of virtual interferers present is higher. When there are few actual interferers present, this causes more PRTFs to be selected for spatial cue preservation than might be necessary for the unrelaxed power constraint, Equation (2.16c). There is a discussion on the relaxation parameter $\epsilon$ specifically in Section 4.1.1.

The spatial cues error when using PRTFs is expected to be higher, as an interfering point source will only be preserved spatially if its azimuth angle corresponds with the azimuth angle of a binaurally constrained PRTF. The locations of interfering point sources is chosen such that some correspond to PRTF azimuth angles, and some do not. Regardless of that, which PRTFs are binaurally constrained is variable, so spatial cues errors are expected. Figure 3.12 confirms these expectations. Since the maximum allowed number of binaurally constrained sources is $2M - 3$, increasing the number of PRTFs does not improve the spatial cue errors as much as in [7], though they do improve slightly. It should be noted that proposed method one also performs better here, following the trend of the exhaustive search, while staying below it, as expected and desired.

### 3.4.3. True RTFs compared to PRTFs

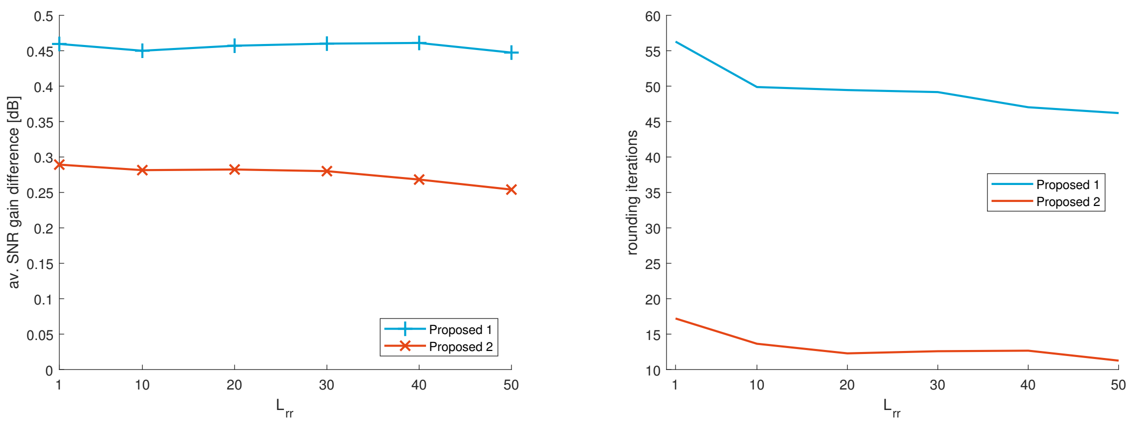
Figure 3.13 provides a comparison of two experiments, one using true RTFs and the other using PRTFs, that are otherwise equivalent. When using PRTFs, the optimal SNR gain drops, as do those of the proposed methods. The optimality gap also increases, though in general the trend of the exhaustive search is followed.




(a) Using true RTFs. Figure 3.3a repeated.

(b) Using 8 PRTFs. Figure 3.11a repeated.

Figure 3.13: The proposed methods compared to the existing methods in terms of average SNR gains for $\beta = 10$ dB and varying number of interfering point sources present in the acoustic scene.




(a) The optimality gap, dependent on $L_{rr}$. Averaged across $\beta = 0$ through 24 dB, $r = 1$ through 5.

(b) The number of random rounding iterations needed, depending on the number of realisations per iteration $L_{rr}$. $\beta = 10$ dB, $r = 5$.

Figure 3.14: The behaviour of the proposed methods for different values of $L_{rr}$.

It does confirm the suspicion that the optimality gap increases with the length of $\mathbf{p}$. Both proposed relaxations suffer from this, though, as discussed in Section 3.4.2, proposed method 2 starts to show very poor performance once the number of PRTFs increases more. The relaxation of proposed method 1 could be improved to lessen the increasing optimality gap by tuning the parameter $\epsilon$, which is discussed in Section 4.1.1.

### 3.4.4. Randomised Rounding

To evaluate the proposed methods further and to be able to comment on the performance of randomised rounding in this application, the number of randomised rounding iterations needed to come to a feasible solution were evaluated for the two proposed methods, for different parameter settings.

Figure 3.14a shows how the optimality gap changes when different values of $L_{rr}$ are used in Algorithm 2. The expectation was that the optimality gap would shrink as $L_{rr}$ got larger. What is observed is however that there is only negligible gains when $L_{rr}$ increases and purely due to the randomness, the optimality gap might even grow slightly for larger $L_{rr}$.

Looking at the number of random rounding iterations to come to a feasible solution, seen in Figure 3.14b, it is immediately obvious that the first proposed method performs worse in terms of distribution of the values of $\hat{\mathbf{p}}^\star$. It should also be noted that the amount of iterations needed to come to a feasible solution, as $L_{rr}$ increases, barely decreases. It should be noted that attempts that did not result in a feasible solution after $k_{rr}$ iterations have not been counted in this average. This is something that happened rarely, but because of the large value of $k_{rr}$ influenced the average significantly. Considering that each iteration of rounding is computationally more complex when $L_{rr}$ is larger, it may be beneficial to keep $L_{rr}$ small. For this reason, most results presented in the chapter are generated using $L_{rr} = 10$.

More sophisticated random rounding methods may improve the performance in terms of number of iterations needed, as discussed in Section 4.3.1.

# 4 Discussion

This thesis set out to find a method to answer, and solve, the question: “When performing joint binaural linearly constrained minimum variance beamforming, can performance in terms of noise reduction and spatial cue preservation be improved by lifting constraints on interferers, if they are deemed inaudible after processing?”

A method of interferer selection for preservation of spatial cues, based on a simple audibility measure has been proposed, with two convex optimisation problem statements to find the approximately optimal constraint subset efficiently. Additionally, these methods allow an interferer subset to be found that can be constrained by the JBLCMV, even when the full set of interferer transfer functions is larger than the JBLCMV can normally constrain. This allows the JBLCMV to be used with PRTFs, of which there should be a relatively larger amount for good results, more than can typically be handled by the JBLCMV.

Finding the best interferer subset for the JBLCMV to constrain when the full set of interferers has less than or equal to $2M-3$ transfer functions is feasible with search methods, since the amount of options to consider is limited. When using PRTFs however, this is no longer feasible. For example, when considering only the horizontal plane, with uniformly spaced PRTFs $15°$ apart and 4 microphones, the number of options to consider is already 55455. Considering elevation and/or more tightly packed PRTFs will quickly make this infeasible, if it was not already. The amount of options that exists is given by

$$N = \frac{m!}{(2M-3)!(m-2M+3)!} + \frac{m!}{(2M-4)!(m-2M+4)!} + \cdots + \frac{m!}{m!}. \tag{4.1}$$

In contrast, the proposed methods use semi-definite programming methods to be solved, which can be solved in polynomial time [22].

## 4.1. First Proposed Method

The first proposed relaxation of the initial optimisation problem in Equation (2.1) uses two semi-definite relaxations and a matrix regularisation to make the problem convex. It adds an error on the diagonal of the selection matrix $(\mathbf{I} - \mathrm{diag}(\mathbf{p}))$ to allow the Schur complement to exist, such that the semi-definite relaxation can be made, as described by Equations (2.26) and (2.27).

The method has been shown to approximate the optimal solution, given by the exhaustive search method, well when using both true and pre-determined RTFs. While the optimality gap was on average slightly higher for this method than for the second proposed method when $r \leq 2M-3$, the method showed consistent performance when a larger number of PRTFs was introduced. It is expected that the performance of this method can be further increased by tuning the relaxation parameter $\epsilon$, as discussed in Section 4.1.1.

### 4.1.1. The Relaxation Parameter, $\epsilon$

The relaxation parameter adds a small error to the power constraint in Equation (2.16c). Through small-scale experiments, it has been observed that the chosen value of $\epsilon$ can have a significant effect on the solution vector found. Choosing it too small causes all selector values to be extremely similar in value, while choosing it too close to 1 causes the error to be added to the power constraint to be too large. For the purpose of the simulation experiments in Chapter 3, a value for $\epsilon$ which performed nicely has been found empirically. What the optimal value of $\epsilon$ would be however, is still unknown.

When the threshold of the power constraint is strict and/or the total amount of noise is large, values that are found for the selection vector $\hat{\mathbf{p}}$ are again very close together. In those specific cases, adjusting the value of $\epsilon$ can again allow values for $\hat{\mathbf{p}}$ that are more similar to the solutions provided by the exhaustive search. This implies that the value of $\epsilon$ is somehow related to the total amount of noise and the threshold itself, or perhaps to the condition number of the matrix in Equation (2.54c). Researching this relation would likely make this proposed relaxation more reliable.

### 4.1.2. Using Pre-determined Relative Acoustic Transfer Functions

When this proposed method is implemented using PRTFs, the performance changes little, besides the normal performance differences that are expected when going from true RTFs to PRTFs, in particular due to steering vector mismatches. The most notable performance difference comes from the power threshold relaxation from Section 2.2.3. The beamscanning method that is used will assign a power to each PRTF that is set up, even if it is not associated with or near a true interfering point source. When there are 24 PRTFs and each of them has some power, the amount of binaurally unconstrained interferer power can seemingly be very large. Another method of directional power estimation might provide more accurate power for each of the PRTFs, such that the PRTFs that are further removed from true interferers get an even lower power estimation assigned to them. Otherwise, a high penalty weight $\alpha$ will at least ensure the threshold is crossed as little as possible.

Another consideration might be to implement a linear penalty term instead of a quadratic one as proposed in Section 2.2.3. Linear penalty terms are typically implemented when a constraint may not be violated at all, while a quadratic penalty will typically allow small violations. In this context the quadratic term was chosen specifically because violating the threshold is expected to be inevitable, yet large violations should be prevented in the interest of spatial cues preservation.

Furthermore, the aforementioned steering vector mismatches do cause a new issue that is particular to both proposed methods. Since the power is only estimated in the directions of the PRTFs, feasibility is only checked with those powers and directions. The actual amount of binaurally unconstrained interferer power might be larger. The true feasibility cannot be checked however and as such this is a necessary concession that needs to be made when using PRTFs.

## 4.2. Second Proposed Method

This proposed method simplifies the optimisation problem in Equation (2.6) by using a fixed filter expression to determine interferer power at the output of the beamformer. This leads to a smaller LMI in the final optimisation problem, Equation (2.55), and better distributions of the values in $\hat{\mathbf{p}}^{\star}$. The use of a fixed filter has been experimentally shown to be a good over-estimator when $r$ or $m$ is small. When the number of (P)RTFs in $\mathbf{\Lambda}_b$ grows however, the estimation becomes worse, to the point where the performance is unreliable, as shown by Figure 3.11e for example. This means that this method is less suited to be used with PRTFs, of which you typically want a large number. Because of the approximate over-estimation of the binaurally unconstrained interferer power, the spatial cues error, in particular the ILD error, that can occur is less well behaved as well.

In conclusion, this method works very well, better than the first proposed method, when the amount of (P)RTFs in $\mathbf{\Lambda}_b$ is less than $2M-3$. When this bound is crossed, performance eventually drops and the relaxation is not accurate anymore. If there is another way to estimate the worst case filter, instead of the JBLCMV for the $2M-3$ strongest sources, this method might perform better.

## 4.3. Recommendations

As stated in Section 4.1.1, an expression for $\epsilon$ or reformulation of the LMI in Equation (2.54c) should be found that performs better than a fixed value that was found heuristically. Also, it was mentioned that a better method of beamscanning, or other type of directional power estimation, will be beneficial when using a large amount of PRTFs. Following are two other subjects that could warrant further research.

### 4.3.1. Randomised Rounding

The randomised rounding algorithm, in Algorithm 2, that was used does not guarantee that the optimal solution is covered by the probabilities given in $\hat{\mathbf{p}}^{\star}$. Methods exist that ensure that the optimal solution is a possible outcome of the randomised rounding procedure. Such methods include multiplying the solution vector by a small constant larger than $1$ after a failed iteration, or adding a small constant to the solution vector after each iteration [23]. Since, particularly with the proposed methods, when the optimal solution is not covered, it is typically due to values of $\hat{\mathbf{p}}^{\star}$ that are too small or even zero, these method might prove very beneficial.

Consider an extremely simplified example: two interfering point sources that, after processing, each have a power of $1$. Suppose the binaurally unconstrained power threshold is $0.5$, then a theoretical solution could be $\hat{\mathbf{p}}^{\star} = \begin{bmatrix} 0 & 0.5 \end{bmatrix}^T$. However, when performing the randomised rounding, there exists no rounded solution that is feasible, since $\mathbf{p}^{\star} = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ and $\mathbf{p}^{\star} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ lead to a binaurally unconstrained power of $1$ and $2$ respectively. These scenarios occur rarely, but they do occur and a more sophisticated method of randomised rounding would ensure that also in these cases a feasible, or even optimal, solution can be found.

### 4.3.2. Audibility Measures

Currently, audibility of sources in the proposed methods was estimated by the SNR on a per frequency bin basis. Literature suggests that audibility can be more accurately estimated using masking principles [21]. Simultaneous masking using idealised critical bands was used to estimate the audibility of sources when determining ILD and IPD weights. Using this principle in the optimisation problem as well might improve spatial cues error performance of the proposed methods. This would require redefining the optimisation problem for critical bands instead of per frequency bin, though this should be a trivial step.

In Section 2.3.1, a minimum threshold value was proposed to compensate for not taking these additional masking effects into account. Setting up better constraints for these masking effect using psycho-acoustic theories would likely result in more accurate binaurally unconstrained noise thresholds.

More importantly, the simultaneous masking principle that was used to evaluate the methods only considers tone versus noise. In the problem discussed in this thesis, there are multiple sources: the target, interferers and noise. Additionally, there is non-simultaneous masking, which is to say, masking of tones in the current time frame by tones or noise in previous time frames, and interband masking, masking that happens across different critical bands.

It is also difficult to relate the objective measure that is the weighted average ILD and IPD from Section 3.3.2 to perception. Intuitively, the BMVDR should be expected to have the worst spatial performance, since it collapses all sources to the target ATFs. It is however commonly observed that the weighted average ILD for the proposed methods is higher than that of the BMVDR. What exactly the means for the perception of the acoustic scene is unclear without extensive listening tests and more realistic simulations. It should be noted though, that the observed weighted average IPD error of the proposed methods were below those of the BMVDR, as expected. Additionally, literature suggests that the interaural time differences are more important to localising sound overall, but especially in lower frequencies [24, 25]. Using this information in objective measures for sound localisation could help show the objective performance of binaural beamformers more clearly, which in turn can help in developing better beamformer methods.

# Glossary

The next lists describe several symbols and abbreviations that are used within the body of the document. Typically, boldface uppercase symbols indicate matrices, boldface lowercase symbols indicate vectors and other symbols are scalars. The values and equations associated with symbols in this list are those that are used for the majority of this work.

## Abbreviations

ATF Acoustic Transfer Function.

AWGN Additive White Gaussian Noise.

BMVDR Binaural Minimum Variance Distortionless Response [3]

CPSD Cross Power Spectral Density.

DFT Discrete Fourier Transform.

GBLCMV General Binaural Linearly Constrained Minimum Variance.

HAD Hearing Assistive Device.

ILD Interaural Level Difference.

IPD Interaural Phase Difference.

ITD Interaural Time Difference.

ITF Interaural Transfer Function.

JBLCMV Joint Binaural Linearly Constrained Minimum Variance [4].

LCMV Linearly Constrained Minimum Variance.

PRTF Pre-determined Relative acoustic Transfer Function.

RTF Relative acoustic Transfer Function.

SNR Signal to Noise Ratio.

SSGN Speech-Shaped Gaussian Noise.

WGN White Gaussian Noise.

## Constants and Parameters

$\alpha$ Power threshold penalty weight.

$\beta$ Minimal desired binaurally unconstrained SNR. The SNR of the target point source over the binaurally unconstrained interfering point sources.

$\epsilon$ Power constraint relaxation parameter for proposed method 1.

$c$ Generic power threshold parameter.

$k_r$ Maximum number of rounding iterations. $1000$ iterations

$L$ Number of LCMV constraints.

$L_{rr}$ Amount of random rounding realisations to try each iteration of Algorithm 2

$M$ Number of microphones on the left and right HAD combined. 4 microphones

$m$ Number of PRTFs used.

$r$ Total number of interfering point sources.

## Indices

$i$ Interfering point source index.

$j$ Microphone index.

$k$ Frequency bin index.

## Number Sets

$\mathbb{C}$ Set of all complex numbers.

$\mathbb{H}$ Hermitian matrices.

$\mathbb{R}$ Real numbers.

$\mathbf{S}_+$ Positive semi-definite matrices.

$\mathbf{S}_{++}$ Positive definite matrices.

## Operators

$(\cdot)^H$ Conjugate transpose.

$\mathrm{diag}(\cdot)$ Provides a diagonal matrix with the elements of $(\cdot)$ on the main diagonal.

$\lambda_{max}(\cdot)$ The largest eigenvalue of $(\cdot)$

$\lambda_{min}(\cdot)$ The smallest eigenvalue of $(\cdot)$

$\|\cdot\|_0$ $l_0$-“norm”, more accurately called the cardinality. The number of non-zero elements.

$\|\cdot\|_1$ $l_1$-norm.

$\|\cdot\|_2$ $l_2$-norm, also called the Euclidean norm.

$\odot$ The Hadamard product, also known as the entrywise product.

$(\cdot)^\star$ Optimal value.

$\star$ Convolution.

$\mathsf{E}[\cdot]$ Expected value.

## Other Symbols

$\hat{\mathbf{p}}$ The continuous selection vector. A relaxation of $\mathbf{p}$. $\hat{\mathbf{p}} \in [0,1]^r$

$\mathcal{E}_i^{\mathsf{ILD}}$ ILD error of the $i$th interferer. $\mathcal{E}_i^{\mathsf{ILD}} = \left|\mathsf{ILD}_i^{\mathsf{out}} - \mathsf{ILD}_i^{\mathsf{in}}\right|$

$\mathcal{E}_i^{\mathsf{IPD}}$ IPD error of the $i$th interferer. $\mathcal{E}_i^{\mathsf{IPD}} = \frac{\left|\mathsf{IPD}_i^{\mathsf{out}} - \mathsf{IPD}_i^{\mathsf{in}}\right|}{\pi}$

$\mathsf{ILD}_i^{\mathsf{in}}$ Input ILD of the $i$th interferer. $\mathsf{ILD}_i^{\mathsf{in}} = \left|\mathsf{ITF}_i^{\mathsf{in}}\right|^2$

| | | |
|---|---|---|
| $\mathsf{ILD}_i^{\mathsf{out}}$ | Output ILD of the $i$th interferer. | $\mathsf{ILD}_i^{\mathsf{out}} = \left\lvert \mathsf{ITF}_i^{\mathsf{out}} \right\rvert^2$ |
| $\mathsf{IPD}_i^{\mathsf{in}}$ | Input IPD of the $i$th interferer. | $\mathsf{IPD}_i^{\mathsf{in}} = \angle \mathsf{ITF}_i^{\mathsf{in}}$ |
| $\mathsf{IPD}_i^{\mathsf{out}}$ | Output IPD of the $i$th interferer. | $\mathsf{IPD}_i^{\mathsf{out}} = \angle \mathsf{ITF}_i^{\mathsf{out}}$ |
| $\mathsf{ITF}_i^{\mathsf{in}}$ | Input ITF of the $i$th interferer. | $\mathsf{ITF}_i^{\mathsf{in}} = \frac{b_{iL}}{b_{iR}}$ |
| $\mathsf{ITF}_i^{\mathsf{out}}$ | Output ILD of the $i$th interferer. | $\mathsf{ITF}_{\mathbf{x}}^{\mathsf{out}} = \frac{\mathbf{w}_L^H \mathbf{b}_{iR}}{\mathbf{w}_R^H \mathbf{b}_{iR}}$ |
| $\mathsf{ITF}_{\mathbf{x}}^{\mathsf{in}}$ | Input ITF of the target point source. | $\mathsf{ITF}_{\mathbf{x}}^{\mathsf{in}} = \frac{a_L}{a_R}$ |
| $\mathsf{ITF}_{\mathbf{x}}^{\mathsf{out}}$ | Output ITF of the target point source. | $\mathsf{ITF}_{\mathbf{x}}^{\mathsf{out}} = \frac{\mathbf{w}_L^H \mathbf{a}}{\mathbf{w}_R^H \mathbf{a}}$ |
| $\tilde{\mathbf{B}}$ | Stacked interfering point source ATFs/RTFs matrix. | $\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_L^H & \mathbf{B}_L^H \end{bmatrix}^H \in \mathbb{C}^{2M \times r}$ |
| $\tilde{\mathbf{P}}$ | Binaural diagonal block CPSD matrix for all disturbances. | $\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{bmatrix} \in \mathbb{C}^{2M \times 2M}$ |
| $\tilde{\mathbf{P}}_{\mathbf{y}}$ | Binaural diagonal block CPSD matrix for the recorded signals. | $\tilde{\mathbf{P}}_{\mathbf{y}} = \begin{bmatrix} \mathbf{P}_{\mathbf{y}} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\mathbf{y}} \end{bmatrix} \in \mathbb{C}^{2M \times 2M}$ |
| $\mathbf{\Lambda}$ | Constraint matrix containing the constraint to preserve the target point source and all potential constraints to preserve the spatial cues of the (virtual) interfering point sources. | $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_a & \mathbf{\Lambda}_b \end{bmatrix} \in \mathbb{C}^{2M \times (r+2)} \left( \vee\, \mathbb{C}^{2M \times (m+2)} \right)$ |
| $\mathbf{\Lambda}_a$ | JBLCMV constraint matrix for preservation of the target point source. | $\mathbf{\Lambda}_a = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix} \in \mathbb{C}^{2M \times 2}$ |
| $\mathbf{\Lambda}_b$ | JBLCMV constraint matrix for preservation of the spatial cues of the (virtual) interfering point sources. | $\mathbf{\Lambda}_b \in \mathbb{C}^{2M \times r} \left( \vee\, \mathbb{C}^{2M \times m} \right)$ |
| $\mathbf{\Phi}$ | Selection matrix that selects columns and elements from $\mathbf{\Lambda}$ and $\mathbf{f}$, respectively, depending on $\mathbf{p}$. | $\mathbf{\Phi} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\mathbf{p}} \end{bmatrix}$ |
| $\mathbf{\Phi}_{\mathbf{p}}$ | Selection matrix that selects columns and elements from $\mathbf{\Lambda}_b$ and $\mathbf{f}_b$, respectively, depending on $\mathbf{p}$. | $\mathbf{\Phi}_{\mathbf{p}} \mathbf{\Phi}_{\mathbf{p}}^T = \mathrm{diag}(\mathbf{p})\ , \quad \mathbf{\Phi}_{\mathbf{p}}^T \mathbf{\Phi}_{\mathbf{p}} = \mathbf{I}_{\lVert \mathbf{p} \rVert_0}$ |
| $\mathbf{0}_x$ | All zeros vector of length $x$. | |
| $\mathbf{1}_x$ | All ones vector of length $x$. | |
| $\mathbf{a}$ | Stacked ATF vector of the target point source. | $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_M \end{bmatrix}^T \in \mathbb{C}^M$ |
| $\mathbf{B}$ | Matrix containing all interfering point source ATFs. | $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_r \end{bmatrix} \in \mathbb{C}^{M \times r}$ |
| $\mathbf{b}_i$ | Stacked ATF vector of the $i$th interfering point source. | $\mathbf{b}_i = \begin{bmatrix} b_{i1} & b_{i2} & \cdots & b_{iM} \end{bmatrix}^T \in \mathbb{C}^M$ |
| $\mathbf{b}_L$ | Vector containing the transfer functions of all interfering point sources to the left reference microphone. | $\mathbf{b}_L = \begin{bmatrix} b_{1L} & b_{2L} & \dots & b_{rL} \end{bmatrix} \in \mathbb{C}^r$ |
| $\mathbf{b}_{iL}$ | Left relative acoustic transfer functions of the $i$th interferer. | $\mathbf{b}_{iL} = \frac{\mathbf{b}_i}{b_{iL}} \in \mathbb{C}^M$ |
| $\mathbf{b}_{iR}$ | Left relative acoustic transfer functions of the $i$th interferer. | $\mathbf{b}_{iR} = \frac{\mathbf{b}_i}{b_{iR}} \in \mathbb{C}^M$ |
| $\mathbf{C}$ | GBLCMV constraint matrix. | $\mathbf{C} \in \mathbb{C}^{2M \times L}$ |
| $\mathbf{f}$ | GBLCMV response vector. | $\mathbf{f} \in \mathbb{C}^L$ |
| $\mathbf{f}_a$ | JBLCMV response vector for the preservation of the target. | $\mathbf{f}_a^H = \begin{bmatrix} a_L & a_R \end{bmatrix} \in \mathbb{C}^2$ |

$\mathbf{f}_b$ JBLCMV response vector for the preservation of the spatial cues of the (virtual) interfering point sources. $\mathbf{f}_b = \mathbf{0}_r \in \mathbb{C}^r (\vee\ \mathbf{f}_b = \mathbf{0}_m \in \mathbb{C}^m)$

$\mathbf{I}_x$ Identity matrix of size $x \times x$.

$\mathbf{n}_i$ The $i$th interfering point source at the microphone array. $\mathbf{n}_i = \mathbf{b}_i u_i \in \mathbb{C}^M$

$\mathbf{P}_\mathbf{v}$ CPSD matrix of the additive uncorrelated noise. $\mathbf{P}_\mathbf{v} = \mathsf{E}\big[\mathbf{v}\mathbf{v}^H\big] \in \mathbb{C}^{M\times M}$

$\mathbf{P}$ CPSD matrix of all disturbances. $\mathbf{P} = \sum_{i=1}^r \mathbf{P}_{\mathbf{n}_i} + \mathbf{P}_\mathbf{v} \in \mathbb{C}^{M\times M}$

$\mathbf{p}$ The binary selection vector that selects (virtual) interfering point sources to keep and remove in the JBLCMV solution. A 1 keeps the source and a 0 removes it. $\mathbf{p} \in \{0,1\}^r (\vee\ \{0,1\}^m)$

$\mathbf{P}_{\mathbf{n}_i}$ CPSD matrix of the $i$th interfering point source. $\mathbf{P}_{\mathbf{n}_i} = p_{u_i} \mathbf{b}_i \mathbf{b}_i^H \in \mathbb{C}^{M\times M}$

$\mathbf{P}_\mathbf{x}$ CPSD matrix of the target point source. $\mathbf{P}_\mathbf{x} = p_s \mathbf{a}\mathbf{a}^H \in \mathbb{C}^{M\times M}$

$\mathbf{P}_\mathbf{y}$ Received signal CPSD matrix. $\mathbf{P}_\mathbf{y} = \mathbf{P}_\mathbf{x} + \mathbf{P} \in \mathbb{C}^{M\times M}$

$\mathbf{U}$ Diagonal matrix containing the interfering point source magnitudes on the main diagonal, scaled appropriately to the RTFs used. $\mathbf{U} = \operatorname{diag}(\mathbf{u}) \operatorname{diag}(|\mathbf{b}_L|) \in \mathbb{R}^{r\times r}$

$\mathbf{u}$ Stacked interfering point source vector. $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix}^T \in \mathbb{C}^r$

$\mathbf{v}$ Stacked additive uncorrelated noise vector. $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_M \end{bmatrix}^T \in \mathbb{C}^M$

$\mathbf{w}_L$ Beamformer coefficients for the left ear. $\mathbf{w}_L \in \mathbb{C}^M$

$\mathbf{w}_R$ Beamformer coefficients for the right ear. $\mathbf{w}_R \in \mathbb{C}^M$

$\mathbf{w}_\mathbf{p}$ The JBLCMV solution for the interfering point source set given by $\mathbf{p}$.

$$\mathbf{w}_\mathbf{p} = \tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{\Phi}\Big(\mathbf{\Phi}^T\mathbf{\Lambda}^H\tilde{\mathbf{P}}^{-1}\mathbf{\Lambda}\mathbf{\Phi}\Big)^{-1}\mathbf{\Phi}^T\mathbf{f} \in \mathbb{C}^{2M}$$

$\mathbf{w}_{BM}$ The BMVDR filter.

$\mathbf{w}_{JB}$ The JBLCMV filter that preserves the spatial cues of all $r$ (virtual) interferers, or the $2M-3$ strongest (virtual) interferers, if there are more than $2M-3$.

$\mathbf{x}$ Target signal at the microphone array. $\mathbf{x} = \mathbf{a}s \in \mathbb{C}^M$

$\mathbf{y}$ Stacked received signal vector in one frequency bin. $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_M \end{bmatrix}^T \in \mathbb{C}^M$

$a_j$ The ATF of the target point source to the $j$th microphone.

$a_L$ ATF of the target source to the left reference microphone.

$a_R$ ATF of the target source to the right reference microphone.

$b_{ij}$ The ATF of the $i$th interfering point source to the $j$th microphone.

$p_s$ Power spectral density of the target point source.

$p_{u_i}$ Power spectral density of the $i$ interfering point source.

$s$ The target point source.

$u_i$ The $i$th interfering point source.

$v_j$ Additive uncorrelated noise on the $j$th microphone.

$y_j[k]$ The Fourier coefficient in the $k$th frequency bin of the received signal at the $j$th microphone.

# Bibliography

[1] World Health Organization *et al.*, “Hearing loss due to recreational exposure to loud sounds: a review,” 2015.

[2] H. Glyde, L. Hickson, S. Cameron, and H. Dillon, “Problems hearing in noise in older adults: a review of spatial processing disorder,” *Trends in amplification*, vol. 15, no. 3, pp. 116–126, 2011.

[3] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, “Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 12, pp. 2449–2464, 2015.

[4] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, “Improved multi-microphone noise reduction preserving binaural cues,” in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 460–464.

[5] T. Van den Bogaert, T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters, “Horizontal localization with bilateral hearing aids: Without is better than with,” *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 515–526, 2006.

[6] B. D. Van Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” *IEEE assp magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[7] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, J. Jensen, and M. Guo, “Binaural beamforming using pre-determined relative acoustic transfer functions,” in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 1–5.

[8] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses,” *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 6, 2009.

[9] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, “Relaxed binaural lcmv beamforming,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 137–152, 2017.

[10] J. Zhang, S. P. Chepuri, R. C. Hendriks, and R. Heusdens, “Microphone subset selection for mvdr beamformer based noise reduction,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 550–563, 2018.

[11] B. Polyak, M. Khlebnikov, and P. Shcherbakov, “An lmi approach to structured sparse feedback design in linear control systems,” in *Control Conference (ECC), 2013 European*. IEEE, 2013, pp. 833–838.

[12] K. B. Petersen, M. S. Pedersen *et al.*, “The matrix cookbook,” *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2008.

[13] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[14] J. Franklin, *Matrix theory*. Mineola, N.Y: Dover Publications, 2000.

[15] J. Zhang, R. Heusdens, and R. C. Hendriks, “Rate-distributed spatial filtering based noise reduction in wireless acoustic sensor networks,” *arXiv preprint arXiv:1712.07941*, 2017.

[16] S. P. Chepuri and G. Leus, “Sparse sensing for distributed detection,” *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1446–1460, 2016.

[17] T. Rappaport, *Wireless communications : principles and practice*. Upper Saddle River, N.J: Prentice Hall PTR, 2002.

[18] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[19] J. Garofolo *et al.* (1993) TIMIT Acoustic-Phonetic Continuous Speech Corpus. [Online]. Available: https://catalog.ldc.upenn.edu/LDC93S1

[20] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, "Binaural speech enhancement with spatial cue preservation utilising simultaneous masking," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 598–602.

[21] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.

[22] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[23] P. Raghavan and C. D. Tompson, "Randomized rounding: a technique for provably good algorithms and algorithmic proofs," *Combinatorica*, vol. 7, no. 4, pp. 365–374, 1987.

[24] W. M. Hartmann, "How we localize sound," *Physics today*, vol. 52, pp. 24–29, 1999.

[25] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002.