Politecnico di Milano
Dipartimento di Elettronica e Informazione
Dottorato di Ricerca in Ingegneria dell'Informazione

Un approccio geometrico per l'analisi e la sintesi di campi acustici per applicazioni multimediali

Tesi di dottorato di:
Antonio Canclini
ID: 738568

Relatore:
Prof. Augusto Sarti
Tutore:
Prof. Antonio Capone
Coordinatore del programma di dottorato:
Prof. Carlo Fiorini

XXIV ciclo

Politecnico di Milano
Dipartimento di Elettronica e Informazione
Piazza Leonardo da Vinci 32 I 20133 — Milano

Politecnico di Milano
Dipartimento di Elettronica e Informazione
Dottorato di Ricerca in Ingegneria dell'Informazione

A geometric approach to analysis and synthesis of acoustic wave fields for multimedia applications

Doctoral Dissertation of:
Antonio Canclini
ID: 738568

Advisor:
Prof. Augusto Sarti
Tutor:
Prof. Antonio Capone
Supervisor of the Doctoral Program:
Prof. Carlo Fiorini

XXIV edition

To Arrigo, Giovanna, Amanda and Margherita

Acknowledgments

I would like to thank my supervisor, Prof. Augusto Sarti, who supported and encouraged me during the entire Ph.D. attending, offering me many interesting (and challenging) research activities within the E.C. founded SCENIC project. I am also grateful to Prof. Stefano Tubaro for his valuable advices.

A very special thank goes to Dr. Fabio Antonacci, who supported me as well, with his precious suggestions and great helpfulness, even during the most difficult times.

I would also like to thank Prof. Rudolf Rabenstein of the Friedrich-Alexander University in Erlangen, Germany, for his kind hospitality and supervision during my staying at the LNT-LMS laboratory. I'm also grateful to all the LMS staff, for their patience and assistance in setting up all the experiments.

I greatly appreciated the helpful suggestions and improvements given by the thesis reviewers, Profs. Raffaele Parisi (La Sapienza University, Rome, Italy) and Emanuël Habets (Friedrich-Alexander University, Erlangen, Germany): many thanks for the careful work and interest.

I am very grateful to Dr. Marco Compagnoni for his mathematical advices and for many fruitful discussions.

I am also grateful to all the (past and present) ISPG/SMCLab people: it has been a pleasure to work with them, as well as sharing many funny (and less serious) moments.

A very special thank goes to my parents Giovanna and Arrigo, and to my sister Amanda: their support during my studies has been invaluable.

I would like to thank all my friends for their sincere support and for many memorable moments shared with me: thanks to everybody, I can't list you simply because you are really so many!

The deepest and most grateful thank goes to Margherita, for always standing by my side and helping in keeping me serene, even when problems and worries seemed to be insurmountable. Many thanks for the beautiful days we spent together in the last three years.

Sommario

In questa tesi si propone un approccio innovativo per l'analisi e la sintesi di campi acustici, basato sulla decomposizione geometrica di un campo sonoro. La propagazione di onde sonore in un mezzo omogeneo può essere descritta tramite raggi acustici che hanno origine in un punto sorgente e che si propagano in tutte le direzioni. L'interazione con l'ambiente è quindi modellizzata secondo le leggi dell'acustica geometrica, ovvero come riflessione dei raggi acustici sugli ostacoli presenti lungo il cammino dei raggi stessi. Gruppi di raggi acustici contigui, originati dallo stesso punto sorgente, possono essere rappresentati come fasci acustici. Ogni qualvolta un fascio acustico incontra un ostacolo, esso viene riflesso generandone di nuovi che saranno a loro volta verranno riflessi, producendo una struttura ramificata. Per questo motivo, un campo acustico può essere interpretato come sovrapposizione di fasci acustici, ognuno dei quali caratterizzato da un'origine (il punto sorgente), dalla direzione, e dall'apertura angolare. Un modo naturale per descrivere i raggi è la geometria proiettiva, che ben si adatta alla natura proiettiva dei raggi stessi. Lo stesso formalismo risulta inoltre essere adatto per descrivere oggetti acustici più complessi quali sorgenti, ricevitori, riflettori e fasci acustici.

Per quanto riguarda l'analisi dei campi acustici, la geometria proiettiva permette di convertire misure acustiche standard (come i tempi di arrivo, le differenze dei tempi di arrivo, e le direzioni di arrivo) in vincoli quadratici espressi in coordinate omogenee. Essi sono tutti caratterizzati dalla medesima formulatione matematica, che permette di combinare più vincoli in una funzione di costo la cui forma risulta essere equivalente per i diversi problemi di stima introdotti, in particolare per problemi standard come la calibrazione delle schiere di microfoni e la localizzazione di sorgente, e per problemi innovativi come l'inferenza della geometria dell'ambiente.

Per quanto riguarda la sintesi di campi acustici, la decomposizione geometrica permette di sviluppare una metodologia per la riproduzione di campi acustici arbitrari per mezzo di una o più schiere di altoparlanti. La conoscenza della geometria

dell'ambiente (che può essere stimata nella fase di analisi) può essere sfruttata per compensare parte della riverberazione (le prime riflessioni) prodotta dalle pareti, al fine di rendere il sistema "consapevole" dell'ambiente. Dal punto di vista geometrico, le prime riflessioni sono determinate da un'insieme di altoparlanti immagine, i quali, una volta inseriti nel modello, contribuiscono alla riproduzione sonora.

La tesi include inoltre una serie di nuove metodologie per la validazione, teorica e sperimentale, delle tecniche proposte. Per quanto riguarda l'analisi dei campi acustici, si mostrerà che l'accuratezza degli algoritmi di stima può essere predetta tramite una funzione lineare che lega l'errore di misura a quello di stima. Per quanto concerne la sintesi di campi acustici, viene proposta una metodologia per valutare sperimentalmente la bontà dei campi riprodotti da sistemi di altoparlanti. Il campo acustico viene campionato su una circonferenza per mezzo di una coppia di microfoni rotanti, ed estrapolato mediante decomposizione in armoniche circolari. L'eventuale degradazione del campo misurato (estrapolato) rispetto a quello desiderato viene quindi quantificata tramite metriche standard (basate sull'errore quadratico medio) e metriche innovative (analisi modale). Nella tesi sono riportati i risultati di alcune simulazioni ed esperimenti di sintesi di campi acustici in ambienti riverberanti.

Le soluzioni proposte in questo lavoro trovano potenziale applicazione in diversi campi, quali le telecomunicazioni avanzate (telepresenza), i video-giochi immersivi, la produzione musicale distribuita, la spazializzazione sonora per sistemi di home entertainment.

Abstract

In this thesis we propose a novel approach to wave field analysis and synthesis based on the geometric description of a sound field. Under the hypothesis of homogeneous medium, wave propagation can be approximated by means of acoustic rays that originate from a point source and spread in all the directions. The interaction with the environment causes the rays to reflect over the obstacles following the geometrical acoustic laws. Contiguous bundle of rays can be represented as acoustic beams that split and branch during the propagation as they encounter reflectors. An acoustic wave field can be seen as the superposition of acoustic beams, each characterized by the origin (source), direction and angular aperture. Due to the projective nature of rays, the most suitable tool for their representation is projective geometry, which gives also a compact and efficient description of more complex acoustic entities such as sources, receivers, reflectors and beams.

As far as the wave field analysis is concerned, projective geometry makes it possible to convert standard acoustic measurements (TOAs, TDOAs, DOAs) into homogeneous quadratic constraints, which all share the same mathematical formulation. This way, through the combination of multiple constraints it is possible to formulate a cost function whose form is equivalent for several estimation problems, ranging from standard ones (e.g. acoustic source localization), to novel ones such as the inference of the geometry of the environment.

As far as the sound field synthesis is concerned, the geometric description of a wave field leads to developing a methodology that aims at reproducing a complex wave field by superimposing elementary beams rendered by means of a loudspeaker array. The knowledge of the map of the environment, which can be estimated in the analysis stage, can be exploited to compensate for the environment hosting the loudspeakers, thus making the rendering system environment-aware. In particular, the reverberations (in terms of early reflections) are seen as determined by a set of image loudspeakers, which are inserted into the model and contribute to the

reproduction of the wave field.

This thesis also includes various novel validation methodologies for the proposed techniques. As far as wave field analysis is concerned, we show that it is possible to predict the accuracy of estimation algorithms as a linear mapping between the error on the measurements and the estimation error. As for the sound field synthesis, we propose a methodology for assessing the quality of wave fields reproduced by real loudspeakers. The wave field is sampled over a circle with a pair of rotating microphones and extrapolated by means of the circular harmonic decomposition. The deviation between the extrapolated and the expected wave fields are then evaluated by means of standard (MSE-based) and novel (modal analysis) metrics. In this thesis we also show the results of some simulations and experiments of room compensation and virtual environment rendering.

The solutions proposed in this thesis find potential application in a wide range of fields, including advanced telecommunications (telepresence), immersive gaming, distributed music production, spatial audio at home.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivations

The last decade has been characterized by an exponential growing of multimedia applications employing multichannel input/output audio systems. The availability of low-cost and small-size sensors and emitters contributed to make these systems widespread, also in everyday home applications. Such systems typically employ set of microphones, possibly organized in clusters or arrays, and distributions of loudspeakers. Microphone arrays are extensively used for wave field analysis applications, see for example [1, 2]. The signals acquired by the sensors are first converted into acoustic measurements, which are then employed to obtain information about acoustic elements present in the environment. For instance, modern laptops typically include microphone arrays that are used to improve the quality of the recorded sound. Furthermore, source localization and tracking is becoming more and more frequent in more complex tele-conference scenarios [3, 4, 5]. On the other hand, arrays or distributions of loudspeakers are adopted to generate synthetic wave fields (acoustic rendering) within a certain region of space, with the aim of increasing the immersivity of one or more listeners. Stereophony, which constituted the standard for audio reproduction for many years, is being replaced by 5.1 and 7.1 surround systems. If this trend will be confirmed, we expect massive reproduction systems such as Wave Field Synthesis (WFS) [6] and Higher Order Ambisonics (HOA) [7] to start becoming attractive for everyday applications.

In the literature, wave field analysis and synthesis methodologies are often handled as two separated topics. We believe, however, that synthesis algorithms could highly benefit from the information gathered during the wave field analysis stage. Consider, for example, a scenario in which one or more microphone arrays acquire the signals generated by controlled or uncontrolled acoustic stimuli in a reverberant environment. A suitable analysis of the acquired signals could be used to determine

the positions of the loudspeakers (array calibration) constituting the reproduction system. The same sensors could then be employed to extract information about the environment (geometry, reflective properties, ...), which can be exploited to design a sound field reproduction system taking into account the acoustic properties of the environment in which it operates.

In this thesis we propose a novel approach to wave field analysis and synthesis, with the aim of making analysis algorithms functional to the synthesis phase. Moreover, we introduce novel methodologies and metrics for assessing the effectiveness of the proposed algorithms. These methodologies are adopted for evaluating the proposed techniques, but they are general enough to be employed for the evaluation of different algorithms. In the following Section we introduce the main contributions given in this work, and we describe the application scenario considered for testing the proposed methodologies.

1.2 Original contributions and application scenario

A geometric approach for modelling the wave field

The methodologies that we propose are based on a geometric modelling of the wave field. We start from the fact that, following the laws of the geometrical acoustics [8], a propagating wave front can be described by means of acoustic rays. The direction of a ray coincides with the local normal to the wave front. Since air is commonly considered as an homogeneous propagation medium, rays can be represented as lines, which originate at the acoustic source position. Although ray acoustics is not novel for modelling applications (e.g. for ray and beam tracing applications [9, 10]), in this thesis we propose an alternate representation of acoustic propagation by means of projective geometry, whose validity goes beyond modelling purposes, as it is used also for analysis, inference and rendering applications. Projective geometry is a common tool in computer vision applications [11], since an image can be seen as formed by the projection on a plane of all the rays that pass through the center of projection. In the same way, in audio applications acoustic rays share the same projective nature. In fact, an acoustic source can be seen as the center of projection of all the rays originated from it. Similarly, an acoustic receiver can be interpreted as the center of projection of all the gathered rays. Exploiting these properties, projective geometry allows us to describe more complex objects, such as acoustic reflectors

(rays passing through the reflector surface) and acoustic beams (compact bundles of rays, originated from the same source but with a limited visibility in space). Moreover, an extension of the classical projective geometry, called oriented projective geometry [12], makes it possible to distinguish between different orientations of lines, as well as the two opposite sides of a reflector. The geometric representation through (oriented) projective geometry reveals to be suitable for both wave field analysis and synthesis purposes. We discuss about this fact in the next two paragraphs.

A geometric approach for wave field analysis

As far as wave field analysis is concerned, projective geometry allows us to define a general theoretical framework which enables to express acoustic measurements into a compact mathematical representation. More specifically, standard acoustic measurements such as Times Of Arrival (TOAs), Time Differences Of Arrival (TDOAs), and Directions Of Arrival (DOAs) can be converted into projective constraints, which embed informations about acoustic geometric primitives. In particular, acoustic measurements relative to the direct path between sources and microphones brings to projective constraints acting on projective points (representing either sources or sensors). On the other hand, measurements relative to reflective paths (i.e. paths determined by acoustic reflections) lead to constraints acting on projective lines (representing the line on which a planar reflector lies). All the projective constraints share the same mathematical form. More precisely, they all consist of quadratic forms, which correspond to different conic sections. For example, we will see that TOAs relative to direct paths are associated to circumferences; DOAs relative to reflective paths generate parabolas. However, in projective geometry there is no distinction among different conic sections, since they can be seen as different projections of a circumference [13]. For this reason, the combination of multiple projective quadratic constraints leads to a common representation valid for different estimation problems. Indeed, the proposed theory is suitable for addressing both standard problems, such as array calibration (i.e., the estimation of the pose of sensors within an array) and acoustic source localization; and novel problems such as the inference of the geometry of the environment from acoustic measurements. Notice that the problem of estimating the geometry of the environment is not novel at all, since it has been addressed with other stimuli (e.g. computer vision [14], or radar tomography [15]). However, acoustic measurements give more insight into the acoustic properties of the

environment. For instance, the same wall could be detected using an image based technique, while being completely transparent from the acoustic standpoint.

A geometric approach for wave field rendering

Sound field reconstruction can be approached in rather different fashions, depending on the underlying mathematical representation of the acoustic wave field. WFS, for example, relies on the Kirchhoff-Helmholtz integral to reconstruct the acoustic wave field from a spatial sampling on a closed surface surrounding the spatial region of interest [6]. HOA, on the other hand, relies on cylindrical (in the 2D case) or spherical (in the 3D case) harmonic decompositions of the sound field to perform sound reproduction with a spatial distribution of loudspeakers [7].

In this thesis we focus on the rendering of two-dimensional (horizontal) acoustic wave fields by means of a distribution of loudspeakers located on the hearing plane. Although three-dimensional reproduction is becoming attractive for the movie industry and 3D sound reproduction systems exist (see for example *IOSONO*® and *Auro Technologies*® installations[1]), an accurate 3D reproduction over a wide region still requires an enormous number of loudspeakers [16] (e.g., a uniform sampling of a sphere), making this solution extremely costly and not trivial to implement. For this reason, 2D sound field reproduction turns out to be more tractable, especially in everyday scenarios (video-conferencing, spatial sound at home, ...).

The rendering technique proposed in this thesis presents several points of novelty. First of all, the wave field representation is based neither on spatial sampling nor on harmonic decomposition of the sound field. Instead, we consider the wave field as a geometric superposition of acoustic beams, each accounting for an image (wall-reflected) source. Acoustic beams are assumed as being the result of the angular windowing of the radial pattern of the actual source, on the part of the visibility function that describes the chain of wall reflections that originate the image source. The computation of this angular windowing can be performed using beam tracing techniques [17], which was originally conceived for tracing acoustic paths in real time in the presence of a moving receiver. This method, in fact, is suitable for swiftly computing the branching of acoustic beams originated from an acoustic source, as they encounter reflectors during their propagation. This technique was later generalized to a method for computing the visibility between reflectors, which can be exploited

[1] www.iosono-sound.com, www.auro-3d.com

for tracing beams in real time even in the presence of moving sources in the environment [18]. The use of this beam tracing process allows us to correctly emulate the impact of an arbitrary 2D virtual environment onto the rendered wave field. This is accomplished through the superposition of the acoustic beams generated by virtual wall reflections, which can be traced in real time even if the acoustic source is moving within the virtual environment. The generation of acoustic beams using a spatial distribution of loudspeakers can be obtained with any beamforming method. One rather effective method that enables an arbitrary placement of the image source and a controllable beam shape is described in [19].

The geometric approach to sound field synthesis makes intensively use of the geometry of the environment in order to compensate for the effect of the reverberations (*room-compensation*). We assume this knowledge to be estimated during the analysis stage by means of microphone arrays. In the last years, researchers started to investigate room-compensation techniques for rendering applications. For example, in [20], room compensation is achieved through a multichannel adaptive filtering in the wave-domain. Although this system is capable of capturing almost instantaneous variations in the acoustic properties of the environment, it does not embed any information about the environment itself (which can be considered mostly static). The work in [21] shows that room compensation in environments characterized by a convex geometry can be performed by modelling the acoustic reflections using the image source method [22]. The proposed rendering methodology goes a step beyond the one in [21], being suitable for arbitrarily complex environments. In fact, once the geometry of the real environment hosting the loudspeakers is known, the very same beam tracing engine described before can be used for predicting the effect of the reverberations in the listening area. In particular, we propose to exploit the existing acoustic reflectors to enrich the set of actual arrays with numerous image (wall-reflected) speaker arrays. Using again the beam tracing technique, we can estimate the position of each image loudspeaker, as well as its visibility inside the listening region. This information is included into the data model, and therefore image loudspeakers contribute, together with the real ones, to generate the synthetic wave field. As a result, room compensation is achieved in a fully geometric fashion.

Validation techniques for analysis and synthesis algorithms

In this thesis we introduce innovative methodologies for assessing the accuracy of analysis and synthesis algorithms. As discussed before, wave field analysis is mainly accomplished by estimation algorithms (e.g. source localization, localization of acoustic reflectors). Such algorithms rely on acoustic measurements, which are naturally subject to a certain amount of measurement error. This is due to a series of factors, ranging from the environmental noise to errors in positioning the arrays, as well as the non-idealities of the sensors. In the context of estimation theory, researchers typically resort to Monte-Carlo simulations and to the Cramer-Rao Lower Bound (CRLB) [23] for assessing the accuracy of an estimation methodology. The first alternative is often time-demanding, as for each potential value of the unknown variables a huge number of simulations is required. CRLB is then used as a theoretical limit of the achievable accuracy under the hypothesis that no bias is introduced in the estimation process. In this thesis we propose a novel method that combines the advantages of Monte-Carlo simulations (algorithm-dependent prediction of the accuracy) and CRLB (low computational cost) to predict the covariance matrix of the estimated variables. We will see that, under the same hypothesis of CRLB (i.e., small bias), there exists a linear relationship between the covariance matrix of the measurement and that of the estimation. Moreover, the proposed error propagation analysis goes beyond the information provided the CRLB. On one hand, the proposed method provides a lower bound for a specific estimation technique, instead that for the general problem. On the other hand, the CRLB turns out to be a particular case of the error propagation analysis when it is applied to a Maximum-Likelihood cost function.

As far as wave field rendering is concerned, in this thesis we propose novel solutions for both theoretical and experimental evaluations. In the literature, the accuracy of rendering systems is typically evaluated by computing the Mean Square Error (MSE) between the theoretical prediction of the reproduced wave field reproduced and the desired wave field (see [16, 21], for example). MSE based metrics, however, provide only a global information on the error. For a better characterization of the error, we introduce a modal analysis of the rendered wave field in the wave number domain. This solution allows us to analyze a wave field in terms of its propagating components, thus allowing us to determine which propagation directions are actually corrupted in the rendering process. In this thesis we also extend such theoretical metrics to

experimental data. In order to do so, we propose a technique for measuring the wave field reproduced by a real distribution of loudspeakers, based on the Circular Harmonic Decomposition [24]. Through the combination of standard (MSE) and novel (modal) analysis of the measured wave field, it is possible to quantify the degradation introduced by the non-idealities of a real system.

Application scenario

The main methodologies proposed in this thesis are evaluated through a number of simulations and experiments, within the context of an application scenario. In particular, we consider a wave field rendering system operating into an arbitrarily shaped reverberant environment. In order to achieve a good reproduction of the desired wave field, we need to compensate for the environment, i.e. reverberations have to be dampened. We do so in a geometrical way. In a first stage, wave field analysis techniques are used to infer the geometry of the environment from acoustic measurements. This knowledge, together with the position of the loudspeakers, is then provided to the beam tracing engine, which is used to determine the set of image loudspeakers, along with their visibility in the listening region. As a result, this information enables the implementation of a fully environment-aware rendering system. This scenario may be realistic in an everyday living-room, where a distribution of loudspeakers is installed for home-entertainment purposes. Whenever the environment is modified (e.g. because of a change in the disposition of the furniture, or after covering windows with curtains), the loudspeakers could be exploited for generating a set acoustic stimuli, which can be processed by one or more microphones in order to extract information about the acoustic reflectors.

As far as the evaluation is concerned, we consider the following simulations and experiments. The accuracy of reflector localization algorithms (room inference) is first evaluated by means of the theoretical error propagation analysis presented in this thesis. Experiments are then performed to asses the localization accuracy for different room geometries. We then evaluate the accuracy of the wave field reproduced by the geometric rendering system. In particular, we perform a theoretical and experimental evaluation of a loudspeaker system reproducing the effect of an acoustic source, possibly collocated inside a virtual environment whose acoustics is rendered through superposition of acoustic beams. We will see that the results of the analysis stage reveal to be highly accurate, providing an effective estimate of the room geometry.

For this reason, we omit to test the robustness of the rendering technique against errors in estimating the map of the environment.

1.3 Application fields

The solutions proposed in this work find potential applications in several fields. As an example, let us consider a tele-conference scenario. With the proposed rendering methodology it could be possible to realistically emulate the presence of a virtual acoustic source (telepresence), along with the virtual environment surrounding it. A talker in a first room, therefore, could be localized and virtually reproduced in a second room at the correct position by means of a loudspeaker array, giving the impression of his presence. In order to compensate for the reverberation of the conference room, a set of sensors could be employed for estimating the position of the main reflectors present in the environment.

Music and artistic applications may also take advantage from the proposed solutions. As an example, a geometric rendering system installed into a recording studio could be employed for simulating the acoustics of an arbitrary environment. The performance of a musician, therefore, can be virtually and instantaneously moved from a small concert hall to a precise cathedral, and so on. The same scenario could also be exploited for distributed music production, where several musicians could perform in the same virtual environment while being in different recording studios.

A further potential application could be in the field of home-entertainment scenarios (3D audio at home). Multiple small-size speaker arrays installed in the wall are first calibrated through a set of small-size microphone arrays, which could be then used for inferring the geometry of the environment.

The geometric rendering system could also find potential applications in the field of immersive gaming. Indeed, the realistic reproduction of acoustic sources and virtual environments is suitable for the rendering of arbitrary synthetic acoustic scenes.

1.4 Outline of the thesis

This thesis is organized as follows.

Chapter 2, after an overview of some basic acoustic laws, describes the geometric representation of wave fields. By means of the (oriented) projective geometry framework, a mathematical description of acoustic elements such as rays, beams, sources,

receivers, and reflectors is then provided.

Chapter 3 is devoted to geometric wave field analysis solutions. Starting from standard acoustic measurements, it presents the procedure for formulating and solving a wide set of estimation problems, ranging from array calibration and acoustic source localization to room geometry estimation.

Chapter 4 focuses on the geometric approach to wave field rendering. After an introduction on the beam tracing technique, geometric room compensation and the rendering of virtual environments are described.

In Chapter 5 the theoretical and experimental validation techniques are discussed, focusing on the theoretical error propagation analysis for estimation algorithms; on the wave field measuring methodology; and on the modal analysis for assessing the accuracy of rendering techniques. Moreover, the evaluation metrics are presented.

Chapter 6 reports the results of simulations and experiments conducted considering the application scenario introduced in Section 1.2, focusing on reflector localization techniques used for estimating the room geometry; and on the geometric rendering methodology.

Finally, Chapter 7 draws the conclusions and outlines possible future research directions.

1.5 List of publications

The publications relative to the work presented in the thesis are listed in the following:

- Antonacci F., Calatroni A., Canclini A., Galbiati A., Sarti. A., Tubaro S., *"Rendering of an acoustic beam through an array of loudspeakers"*, in Proceedings of the 12th International Conference on Digital Audio Effects, DAFx-09, Como, Italy, September 2009;

- Antonacci, F., Calatroni, A., Canclini, A., Galbiati, A., Sarti, A., Tubaro, S., *"Soundfield rendering with loudspeaker arrays through multiple beam shaping"*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. , vol., no., pp.313-316, 18-21 Oct. 2009;

- Markovic D., Canclini A., Antonacci F., Sarti A.; Tubaro S., *"Visibility-based beam tracing for soundfield rendering"*, Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on , vol., no., pp.40-45, 4-6 Oct. 2010;

- Annibale P., Antonacci F., Canclini A., Rabenstein R., Sarti A., *"Wave-based and Geometric Representations of Sound Fields"*, 9. ITG Fachtagung Sprachkommunikation, Bochum, Germany, Oct. 2010;

- Canclini A., Annibale P., Antonacci F., Sarti, A., Rabenstein R., Tubaro S., *"From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry"*, Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on , vol., no., pp.2620-2623, 22-27 May 2011;

- Canclini, A., Annibale, P., Antonacci, F., Sarti, A., Rabenstein, R., Tubaro, S., *"A methodology for evaluating the accuracy of wave field rendering techniques"* Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on , vol., no., pp.69-72, 22-27 May 2011;

- Canclini A., Annibale P., Antonacci F., Sarti A., Rabenstein R., Tubaro S., *"An Angular Frequency Domain Metric for the Evaluation of Wave Field Rendering Techniques"*, in proceedings of European Signal Processing Conf. (EUSIPCO), Aug. 2011;

- Filos J., Canclini A., Thomas M. R. P., Antonacci F., Sarti A., Naylor P. A., *"Robust Inference of Room Geometry from Acoustic Impulse Responses"*, in proceedings of European Signal Processing Conf. (EUSIPCO), Aug. 2011;

- Canclini A., Antonacci F., Thomas R.P., Filos J., Sarti A., Naylor P. A., Tubaro S., *"Exact localization of acoustic reflectors from Quadratic constraints"*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011. WASPAA '11.

2 Geometric representation of wave fields

This chapter introduces the terminology and the tools used throughout this thesis for representing complex wave fields. Starting from the wave equation and the Helmholtz equation, we first derive a simple model for acoustic point sources. Moreover, we will see how, under certain conditions, a wave field can be represented in terms of acoustic rays. Due to the projective nature of rays, the most natural representation for describing the propagation is given by *projective geometry*. Although projective geometry is employed especially in the computer vision community [11], its use in audio is becoming important. For instance, in [25] the author formulates the problem of acoustic scene reconstruction (localization of both M sensors and N sound sources active in the environment) following the idea of structure from motion, typical in computer vision problems. In [26] the authors extend the approach followed in [25] to keep into account the uncertainties inherent in acoustic measurements. In our context, the use of *homogeneous coordinates* instead of Cartesian ones gives us a number of advantages such as the uniform handling of points to infinity (avoiding their handling as special cases), the perfect duality between points and lines, as well as a very efficient representation of conic sections.

We notice that traditional projective geometry does not allow to discriminate the orientation of lines (i.e., the direction of a ray). Analysis and synthesis applications often require to distinguish between rays traveling towards opposite directions. As an example, such a distinction would make possible to manage separately the two sides of a reflective surface, as described later on in this Chapter. For these reasons, we introduce a set of acoustic geometric primitives defined on the base of the *oriented projective geometry* [12], which extends classical projective geometry by explicitly assigning an orientation to geometric entities. Such a description reveals to be particularly useful for developing the beam tracing engine described in Chapter

4.

This Chapter is organized as follows: Section 2.1 provides a background of some basic physical acoustic concepts such as the wave equation and the model of point sources, and defines acoustic rays by means of a high frequency approximation of the wave equation; Section 2.2, after an overview on classical projective geometry, introduces the framework of oriented projective geometry; finally, in Section 2.3 we define the set of acoustic geometric primitives employed throughout this work.

2.1 Modelling the acoustic propagation

In this Section we first derive the equation that governs the propagation of acoustic waves. Discussing its formulation in the frequency domain, i.e. the Helmholtz equation, we present the model of a point source and the spherical propagation described by the Green's function. We then introduce a particular high frequency approximation of the wave equation called Eikonal equation. We will see how this approximation is suitable for representing the propagating wave fronts as acoustic rays.

2.1.1 The acoustic wave equation

The derivation of the acoustic wave equation is based on three basic physical principles:

- the equation of motion,
- the continuity equation,
- the state equation (gas law).

Consider an infinitesimal volume element V of a fluid, whose center of gravity is located at $\boldsymbol{x} = [x, y, z]^T$. Let m be the mass of the volume. Applying a force $\boldsymbol{F}$ to the fluid, we can write the Newton's second law

$$\boldsymbol{F} = m\frac{\partial \boldsymbol{v}(\boldsymbol{x}, t)}{\partial t} , \tag{2.1}$$

where $\boldsymbol{v}(t, \boldsymbol{x})$ denotes the particle velocity and t is the time. Since a force can be expressed in terms of the pressure on a surface, the Newton's law can be rewritten

as

$$\boldsymbol{F} = -\nabla p(t, \boldsymbol{x}) V \ , \tag{2.2}$$

where $p(t, \boldsymbol{x})$ denotes the acoustic pressure at $\boldsymbol{x}$; and ∇ is the gradient operator. Inserting (2.2) into (2.1), we finally obtain the *Euler's equation of motion* [27]

$$\nabla p(t, \boldsymbol{x}) = -\varrho_0 \frac{\partial \boldsymbol{v}(t, \boldsymbol{x})}{\partial t} \ , \tag{2.3}$$

where $\varrho_0 = \frac{m}{V}$ is the static density of the fluid[1]. Next, we consider the conservation of mass principle: since the total mass of fluid in a deformable volume must remain constant, letting $\tau(t, \boldsymbol{x})$ be the volume variation due to the pressure changing, we can derive the *continuity equation* [27]

$$\frac{\partial \tau(t, \boldsymbol{x})}{\partial t} = V \left[\nabla \cdot \boldsymbol{q}(t, \boldsymbol{x})\right] \ . \tag{2.4}$$

Finally, we focus on the properties of the propagation medium. Assuming that the fluid is an ideal gas, the Charles-Boyle gas law applies to the volume, i.e.

$$PV = RT \ ,$$

where P is the total pressure in the volume; T is the absolute temperature in degrees Kelvin; and R is the specific gas constant[2]. Since the heat exchange in the wave in the audible frequency range is negligible [27], the thermodynamic process can be considered adiabatic. Therefore, the relation between the total pressure and the volume is given by

$$PV^{\gamma} = \text{constant} \ , \tag{2.5}$$

where γ is the adiabatic index[3]. The total pressure can be expressed as $P = P_0 + p$, where P_0 is the undisturbed pressure. Analogously, the volume can be written as $V = V_0 + \tau$, where V_0 is the undisturbed volume[4]. Differentiating (2.5) and observing

[1] For air $\varrho_{0_{\text{air}}} = 1.18 \frac{\text{kg}}{\text{m}^3}$.
[2] For air $R_{\text{air}} = 286.9 \frac{\text{J}}{\text{kg T}}$.
[3] For air $\gamma_{\text{air}} = 1.4$.
[4] The dependency of P, p, V and τ upon space and time $(t, \boldsymbol{x})$ is here omitted for the sake of compactness.

that $p << P_0$ and $\tau << V_0$ the *state equation* becomes [27]

$$\frac{\partial p(t,\boldsymbol{x})}{\partial t} = -\gamma \frac{P_0}{V_0} \frac{\partial \tau(t,\boldsymbol{x})}{\partial t} \ . \tag{2.6}$$

We now derive the acoustic wave equation through the combination of the equation of motion (2.3), the continuity equation (2.4) and the state equation (2.6). In particular, the combination of (2.4) and (2.6) leads to

$$\frac{\partial p(t,\boldsymbol{x})}{\partial t} = -\gamma P_0 \left[\nabla \cdot \boldsymbol{q}(t,\boldsymbol{x})\right] \ ,$$

and differentiating with respect to t we obtain

$$\frac{\partial^2 p(t,\boldsymbol{x})}{\partial t} = -\gamma P_0 \left[\nabla \cdot \frac{\partial \boldsymbol{q}(t,\boldsymbol{x})}{\partial t}\right] \ . \tag{2.7}$$

Taking the divergence of each side of (2.3) we get $\nabla \cdot \nabla p(t,\boldsymbol{x}) = -\varrho_0 \left[\nabla \cdot \boldsymbol{q}(t,\boldsymbol{x})\right]$, which can be rewritten as

$$\nabla^2 p(t,\boldsymbol{x}) = -\varrho_0 \left[\nabla \cdot \boldsymbol{q}(t,\boldsymbol{x})\right] \ , \tag{2.8}$$

where ∇^2 is the Laplacian operator. Finally, the combination of (2.7) and (2.8) leads to the acoustic wave equation

$$\nabla^2 p(t,\boldsymbol{x}) = \frac{1}{c^2} \frac{\partial^2 p(t,\boldsymbol{x})}{\partial t^2} \ , \tag{2.9}$$

where the speed of sound c is related to the propagation medium properties through

$$c^2 = \frac{\gamma P_0}{\varrho_0}.$$

Further details on the derivation of the acoustic wave equation are available for example in [27, 28].

2.1.2 Helmholtz equation and dispersion relation

The wave equation (2.9) is expressed in the space-time domain $(t,\boldsymbol{x})$. We now investigate the properties of the spatio-temporal Fourier transform of this equation. Let $P(\omega,\boldsymbol{x})$ be the temporal Fourier trasform of $p(t,\boldsymbol{x})$, ω being the temporal fre-

quency. Applying the temporal Fourier transform to (2.9) we obtain the *Helmholtz equation* [27]

$$\nabla^2 P(\omega, \boldsymbol{x}) + k_0^2 P(\omega, \boldsymbol{x}) = 0 \,, \tag{2.10}$$

where $k_0 = \omega/c$. The Helmholtz equation can be expressed also in the spatio-temporal frequency domain. Taking the multi-dimensional Fourier transform of (2.10) with respect to $\boldsymbol{x}$ we get

$$-k^2 \bar{P}(\omega, \boldsymbol{k}) + k_0^2 \bar{P}(\omega, \boldsymbol{k}) = 0 \,,$$

where $\bar{P}(\omega, \boldsymbol{k})$ is the multi-dimensional spatio-temporal Fourier transform of $p(t, \boldsymbol{x})$; and $k = \|\boldsymbol{k}\|$ is the wave number and denotes the length of the spatial frequency vector $\boldsymbol{k} = [k_x, k_y, k_z]^T$. We notice that this representation is satisfied only when the *dispersion relation* [27] holds, i.e.

$$k^2 = k_0^2 = \left(\frac{\omega}{c}\right)^2 \,. \tag{2.11}$$

There exists a single case for which the dispersion relation is not fulfilled, corresponding to the trivial solution $\bar{P}(\omega, \boldsymbol{k}) = 0$. It is important to observe how (2.11) links the length k of the spatial frequency vector $\boldsymbol{k}$ to the temporal frequency ω. This means that the solutions of the wave equation are not independent in time and space.

2.1.3 Point sources

The wave equation (2.9) and the Helmholtz equation (2.10) assume a source-free medium. They are therefore homogeneous, as their right-hand terms are zero. However, complex wave fields are, in general, generated by the effect of multiple acoustic sources. As a consequence, some non-zero term has to be added to the right-hand terms in the wave equation and Helmholtz equation. A very simple and widely used model is the point source. It can be seen as an infinitesimally small pulsating sphere radiating acoustic energy. Since the medium is assumed to be homogeneous, the energy spreads spherically. A point source located at $\boldsymbol{x}_0 = [x_0, y_0, z_0]^T$ is defined as the solution of the inhomogeneous Helmholtz equation [29]

$$\nabla^2 G_\omega(\boldsymbol{x}|\boldsymbol{x}_0) + k^2 G_\omega(\boldsymbol{x}|\boldsymbol{x}_0) = -\delta(\boldsymbol{x} - \boldsymbol{x}_0) \,, \tag{2.12}$$

where $\boldsymbol{x} \neq \boldsymbol{x}_0$ represents a receiver point. The term $G_\omega(\boldsymbol{x}|\boldsymbol{x}_0)$ is called a *Green's function* and it is also a solution of the Helmholtz equation (2.10). The source is described by $\delta(\boldsymbol{x} - \boldsymbol{x}_0)$, which is a three-dimensional Dirac delta function. It can be shown [29] that the solution of (2.12) representing outward travelling waves is given by

$$G_\omega(\boldsymbol{x}|\boldsymbol{x}_0) = \frac{e^{-j\frac{\omega}{c}\|\boldsymbol{x}-\boldsymbol{x}_0\|}}{4\pi\|\boldsymbol{x}-\boldsymbol{x}_0\|} \,. \tag{2.13}$$

Eq.(2.13) can be interpreted as the free-field transfer function between the source at $\boldsymbol{x}_0$ and the receiver at $\boldsymbol{x}$.

2.1.4 Eikonal equation and acoustic rays

The Eikonal equation is a particular high frequency approximation of the wave equation, and it represents the basis of the ray theory [28, 30]. It is widely used in seismic and geophysics applications [31, 32] since it constitutes a powerful tool for dealing with non-homogeneous propagation media. Although, in general, the acoustic propagation in air is considered homogeneous, the Eikonal equation is at the base of the geometrical acoustics [8], for which the propagation is described through acoustic rays. Ray acoustics is typically employed in geometric methods used for modelling the acoustics of complex environments, see for example [9, 33, 10]. In our context, acoustic rays represent the foundation of all the geometric acoustic primitives introduced in the next Section.

For the sake of completeness, here we report the derivation for non-homogeneous media. In this case, the Helmholtz equation is given by [8]

$$\nabla^2 P(\omega, \boldsymbol{x}) + \frac{\omega^2}{c^2(\boldsymbol{x})} P(\omega, \boldsymbol{x}) = 0 \,, \tag{2.14}$$

where the space-dependent sound speed $c(\boldsymbol{x})$ accounts for the non-homogeneousness of the medium. The most general solution of (2.14) representing a propagating wave front is given by

$$P(\boldsymbol{x}, \omega) = S(\omega) A(\boldsymbol{x}, \omega) e^{j\omega T(\boldsymbol{x})} \,, \tag{2.15}$$

where $S(\omega)$ is the Fourier transform of the propagating signal $s(t)$; $A(\boldsymbol{x}, \omega)$ is an amplitude term which depends on both the position and the frequency; $T(\boldsymbol{x})$ is a function of the position representing the phase, and it is called Eikonal. Notice that $T(\boldsymbol{x})$ does not depend on the temporal frequency, and this condition is verified

for low spectral dispersion. In other words, this fact guarantees the existence of the wave front. We now analize the properties of the amplitude term $A(\boldsymbol{x}, \omega)$, whose dependency on position and frequency can be separated by expressing it as an inverse power series, also called Debye expansion [8]:

$$A(\boldsymbol{x}, \omega) = \sum_{n=0}^{\infty} \frac{A_n(\boldsymbol{x})}{(-j\omega)^n} \ ,$$

where $A_n(\boldsymbol{x})$ represents the nth series coefficient. We observe that, for high frequencies, $A(\boldsymbol{x}, \omega) \approx A_0(\boldsymbol{x})$ and therefore (2.15) becomes

$$P(\boldsymbol{x}, \omega) = S(\omega) A_0(\boldsymbol{x}) e^{j\omega T(\boldsymbol{x})} \ . \tag{2.16}$$

Using (2.16), we can apply the Laplacian operator to $P(\omega, \boldsymbol{x})$ obtaining

$$\begin{aligned} \nabla^2 P(\omega,\boldsymbol{x}) \quad = \quad & S(\omega)\nabla^2 A_0(\boldsymbol{x}) e^{j\omega T(\boldsymbol{x})} + 2S(\omega)\nabla A_0(\boldsymbol{x}) j\omega \nabla T(\boldsymbol{x}) e^{j\omega T(\boldsymbol{x})} + \\ & + j\omega S(\omega) A_0(\boldsymbol{x}) \nabla^2 T(\boldsymbol{x}) e^{j\omega T(\boldsymbol{x})} - \omega^2 S(\omega) A_0(\boldsymbol{x}) [\nabla T(\boldsymbol{x})]^2 e^{j\omega T(\boldsymbol{x})} \ . \end{aligned}$$

Inserting this result into the Helmholtz equation (2.14), after some passages we get

$$\omega^2 S(\omega) \Big\{ [\nabla T(\boldsymbol{x})]^2 - \tfrac{1}{c^2(\boldsymbol{x})} \Big\} + \omega S(\omega) \big[2A_0(\boldsymbol{x}) \nabla A_0(\boldsymbol{x}) \nabla T(\boldsymbol{x}) + A_0(\boldsymbol{x}) \nabla^2 T(\boldsymbol{x}) \big] + S(\omega) \nabla^2 A_0(\boldsymbol{x}) = 0 \ ,$$

which is a second order equation in ω. At very high frequencies the linear and constant terms are negligible and therefore the propagation is governed solely by

$$[\nabla T(\boldsymbol{x})]^2 - \frac{1}{c^2(\boldsymbol{x})} = 0 \ . \tag{2.17}$$

Eq.(2.17) is called *Eikonal equation* and represents a high-frequency approximation of the Helmholtz equation. While (2.14) clearly operates on all the directions of the term $P(\omega, \boldsymbol{x})$, (2.17) acts only on the gradient of the phase term $T(\boldsymbol{x})$. This means that the solution of the Eikonal equation is insensitive to possible variations of the medium which are orthogonal to the direction of propagation. As a consequence, the solution of (2.17) can be interpreted as a ray which propagates in the direction of maximum variation of $T(\boldsymbol{x})$, i.e. orthogonally to the wave front. For instance, a point source can be seen as the origin of an infinite number of rays, which uniformly spread towards all directions in space.

2.2 Introduction to the 2D projective geometry

In this Section we give an overview on the projective geometry, following the approach proposed in [11]. We will focus on the duality between points and lines, as well as on the representation of conic sections. After discussing the limitations of the classical formulation of projective geometry, we will present the oriented projective geometry as introduced by Stolfi in [12].

2.2.1 Points, lines and duality

The 2D Euclidian space is a plane denoted by $\mathbb{R}^2$. A point in this plane is represented univocally by a pair of Cartesian coordinates (x_1, x_2) that can be seen as the vector $\boldsymbol{x} = [x_1, x_2]^T$. A line is represented by $l_1x_1 + l_2x_2 + l_3 = 0$, which is referred as the line implicit equation. A line is therefore univocally identified by the parameter vector $\mathbf{l} = [l_1, l_2, l_3]^T$. Notice that the correspondence between lines and vectors is not one to one: the vector $k\mathbf{l} = [kl_1, kl_2, kl_3]^T$, $k \neq 0$, represents the same line, as its associated line equation $kl_1x_1 + kl_2x_2 + kl_3 = 0$ can be divided by k to obtain the original one. This is due to the fact that the implicit line equation is homogeneous. We are therefore defining an equivalence class of vectors, called *homogeneous vectors*. The set of all equivalence classes of vectors forms a *projective space* $\mathbb{P}^2$ (essentially $\mathbb{R}^3$ except $[0, 0, 0]^T$ that does not correspond to any line).

A point $\boldsymbol{x} = [x_1, x_2]^T$ lies on the line $\mathbf{l} = [l_1, l_2, l_3]^T$ if and only if $l_1x_1+l_2x_2+l_3 = 0$. If we add a third coordinate and represent the point as the vector $\mathbf{x} = [x_1, x_2, 1]^T$ we can write this equation in the matrix form

$$\mathbf{l}^T\mathbf{x} = 0 \, .$$

The above equation still holds if we replace $\mathbf{x}$ with $k\mathbf{x}$, $k \neq 0$. As a consequence, the homogeneous vectors $[kx_1, kx_2, k]^T \in \mathbb{P}^2$ represent the same point $\boldsymbol{x} = [x_1, x_2]^T \in \mathbb{R}^2$. Using the homogeneous representation of lines, the intersection between two lines $\mathbf{l}$ and $\mathbf{l}'$ is readily found as the cross-product of $\mathbf{l}$ and $\mathbf{l}'$:

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' = \det \begin{bmatrix} \underline{\mathtt{i}} & \underline{\mathtt{j}} & \underline{\mathtt{k}} \\ l_1 & l_2 & l_3 \\ l'_1 & l'_2 & l'_3 \end{bmatrix} = [l_2l'_3 - l_3l'_2,\ l_3l'_1 - l_1l'_3,\ l_1l'_2 - l_2l'_1]^T \, ,$$

where $\underline{\mathtt{i}} = [1, 0, 0]^T$, $\underline{\mathtt{j}} = [0, 1, 0]^T$ and $\underline{\mathtt{k}} = [0, 0, 1]^T$. Analogously, the line $\mathbf{l}$ joining

two points $\mathbf{x}$ and $\mathbf{x}'$ is given by the cross product of $\mathbf{x}$ and $\mathbf{x}'$:

$$\mathbf{l} = \mathbf{x} \times \mathbf{x}' = \det \begin{bmatrix} \underline{\mathtt{i}} & \underline{\mathtt{j}} & \underline{\mathtt{k}} \\ x_1 & x_2 & 1 \\ x'_1 & x'_2 & 1 \end{bmatrix} = [x_2 - x'_2,\ x'_1 - x_1,\ x_1 x'_2 - x_2 x'_1]^T \ .$$

Though it may seem useless to use an extra coordinate to represent a point in the 2D space, this constitutes a significant conceptual step with some important implications. In this representation an arbitrary homogeneous vector $\mathbf{x} = [x_1, x_2, x_3]^T$ represents the point $\boldsymbol{x} = [x_1/x_3, x_2/x_3]^T$ in the euclidian space $\mathbb{R}^2$. When $x_3 = 0$ the two Cartesian coordinates go to infinity. This way we introduce the concept of *points at infinity* (or *ideal points*). We observe the importance of points at infinity considering the problem met in euclidian geometry in finding the intersection of two parallel lines, $\mathbf{l} = [l_1, l_2, l_3]^T$ and $\mathbf{l}' = [l_1, l_2, l'_3]^T$. If we use projective coordinates, the third coordinate of the intersection point obtained with the cross product of two lines is zero:

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' = \det \begin{bmatrix} \underline{\mathtt{i}} & \underline{\mathtt{j}} & \underline{\mathtt{k}} \\ l_1 & l_2 & l_3 \\ l_1 & l_2 & l'_3 \end{bmatrix} = (l'_3 - l_3)[l_2, -l_1, 0]^T \ .$$

This result agrees with the general idea that parallel lines meet at infinity and allows us to treat parallel lines in the same way as non parallel ones, without the need to distinguish different cases. Furthermore, this allows us to exchange the role of points and lines in statements concerning their properties, without the risk of incurring into special cases. For example, the equation $\mathbf{x}^T\mathbf{l} = 0$ is symmetrical and consequently the role of points and lines can be switched to obtain $\mathbf{l}^T\mathbf{x} = 0$. The duality is also present in the definitions of the intersection of two lines ($\mathbf{x} = \mathbf{l} \times \mathbf{l}'$) and of the line passing through two given points ($\mathbf{l} = \mathbf{x} \times \mathbf{x}'$). Similarly, we can change the statement "two distinct points lie on a single line" to "two distinct lines intersect in a single point". This duality principle is not true in standard Euclidian geometry, where we need to distinguish special cases.

All points at infinity $\mathbf{x}_\infty = [x_1, x_2, 0]^T$ lie on a single line known as *line at infinity* $\mathbf{l}_\infty = [0, 0, 1]^T$, in fact

$${\mathbf{x}_\infty}^T \mathbf{l}_\infty = [x_1, x_2, 0][0, 0, 1]^T = 0, \quad \forall x_1, \forall x_2 \ .$$

Analogously, every line $\mathbf{l} = [l_1, l_2, l_3]^T$ intersects this line at a point at infinity:

$$\mathbf{x}_\infty = \mathbf{l} \times \mathbf{l}_\infty = \det \begin{bmatrix} \underline{\mathtt{i}} & \underline{\mathtt{j}} & \underline{\mathtt{k}} \\ l_1 & l_2 & l_3 \\ 0 & 0 & 1 \end{bmatrix} = [l_2, -l_1, 0]^T \ .$$

Notice that the intersection does not depend on l_3 and since the vector $[l_2, -l_1]^T$ is orthogonal to the line's normal $[l_1, l_2]^T$, it represents the line's direction and therefore lines at infinity can be seen as the set of all directions of lines in the plane.

2.2.2 Projective representation of conics

In this Section we review the general form of a conic and we provide its matrix-form description using projective coordinates. Such a description is useful for dealing with the quadratic constraints arising from the acoustic measurements introduced in Chapter 3. A conic is represented by a quadratic form

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f = 0 \ , \tag{2.18}$$

where $\mathbf{c} = [a, b, c, d, e, f]^T$ is the parameter vector. As (2.18) is homogeneous, the vectors $\mathbf{c} = [a, b, c, d, e, f]$ and $k\mathbf{c} = [ka, kb, kc, kd, ke, kf]$, $k \neq 0$, represent the same conic. Therefore the parameter vector $\mathbf{c}$ has five degrees of freedom.

If $\mathbf{x} = [x_1, x_2, 1]^T$ are the homogeneous coordinates corresponding to the point $\boldsymbol{x} = [x_1, x_2]^T$, (2.18) can be readily written as

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0 \ , \tag{2.19}$$

where

$$\mathbf{C} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \ . \tag{2.20}$$

A conic is said to be degenerate if $\text{rank}(\mathbf{C}) < 3$. In particular, if $\text{rank}(\mathbf{C}) = 2$, the conic-matrix can be written as

$$\mathbf{C} = \mathbf{l}\mathbf{m}^T + \mathbf{m}\mathbf{l}^T \ .$$

In this case the conic reduces to the pair of lines $\mathbf{l}, \mathbf{m}$. If $\text{rank}(\mathbf{C}) = 1$, the conic

matrix can be expressed in the form

$$\mathbf{C} = \mathbf{l}\mathbf{l}^T \ ,$$

that means that the conic corresponds to the line $\mathbf{l}$.

2.2.3 Dual conics

The conic matrix $\mathbf{C}$ introduced in the previous Section is also called *point conic*, since it is defines an equation on points. We saw in Section 2.2.1 that there exists a strong duality between points and lines in the projective space. For this reason, conics in the projective space can be defined starting from lines. More specifically, we call a *line conic* the conic identified by the set of lines tangent to it.

A line $\mathbf{l} = [l_1, l_2, l_3]^T$ tangent to the conic $\mathbf{C}$ satisfies the equation

$$\mathbf{l}^T \mathbf{C}^\diamond \mathbf{l} = 0 \ , \tag{2.21}$$

where, for a non-singular matrix (i.e. a non-degenerate conic), $\mathbf{C}^\diamond = \det(\mathbf{C})\mathbf{C}^{-1}$ is the adjoint of the matrix $\mathbf{C}$. The formulation in (2.21) is the dual of the formulation in (2.19) as it expresses the conic in the line-space. Equation (2.21) will be extensively used in Chapter 3 to derive the constraints used to infer the geometry of the environment in which we operate.

2.2.4 Projective transformations

Projective geometry is useful also for dealing with geometric transformations, whose general form is given by

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \ ,$$

where $\mathbf{H}$ is a non-singular 3×3 matrix. Such a transformation is called *homography* or *projectivity* [11], and it represents an invertible mapping from $\mathbf{x} \in \mathbb{P}^2$ to $\mathbf{x}' \in \mathbb{P}^2$. Notice that the above equation still holds for any non-zero scaling factor. This means that $\mathbf{H}$ is homogeneous, i.e. only the ratio of the matrix elements is significant. The total number of independent ratios in $\mathbf{H}$ is eight, corresponding to a transformation with eight degrees of freedom (dof). An exhaustive discussion on this topic can be found in [11]. In this Section we will introduce some basic transformations which are useful to describe the rigid motion and the specular reflection, which are both

described by an isometric transformation.

2.2.4.1 Rigid motion

The rigid motion of an object is a transformation of the plane $\mathbb{R}^2$ that preserves Euclidean distances. Such a transformation is called isometry, and it can be seen as the composition of a rotation and a translation. In Euclidean geometry, the rigid motion from the point $\boldsymbol{x}$ to $\boldsymbol{x}'$ is represented by

$$\boldsymbol{x}' = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{t} \ , \quad \boldsymbol{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \ , \quad \boldsymbol{t} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \ ,$$

where $\boldsymbol{R}$ describes a rotation through an angle θ about the origin of the Cartesian plane; and $\boldsymbol{t}$ denotes a planar translation. Projective geometry allows us to represent an isometric transformation using a single matrix that embeds the rotation and the translation. Denoting with $\mathbf{x} = [\boldsymbol{x}, 1]^T$ and $\mathbf{x}' = [\boldsymbol{x}', 1]^T$ the homogeneous representation of $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively, the isometry is described by

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \ , \quad \mathbf{H} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \ . \tag{2.22}$$

2.2.4.2 Specular reflection

An interesting transformation which can be described by an isometry is the specular reflection with respect to a line. This reveals to be very useful for the description of acoustic reflections, as common in geometrical acoustics [22]. With reference to figure 2.1, consider a line with parameters $\mathbf{l} = [l_1, l_2, l_3]^T$ and a point $\mathbf{x} = [x_1, x_2, 1]^T$. The reflection of $\mathbf{x}$ over $\mathbf{l}$ is denoted as $\mathbf{x}' = [x'_1, x'_2, 1]$. The points $\mathbf{x}$ and $\mathbf{x}'$ are equidistant from $\mathbf{l}$ and lie in the two opposite half planes delimited by the line. Moreover, $\mathbf{x}'$ is bounded to lie on the line $\mathbf{l}'$ perpendicular to $\mathbf{l}$ and passing through $\mathbf{x}$. In other words, the line $\mathbf{l}$ coincides with the axis of the segment $\mathbf{x}\mathbf{x}'$. These considerations lead to write

$$\begin{cases} x'_1 = \frac{-l_1^2 x_1 + l_2^2 x_1 - 2 l_1 l_2 x_2 - 2 l_1 l_3}{l_1^2 + l_2^2} \\ x'_2 = \frac{l_1^2 x_2 - l_2^2 x_2 - 2 l_1 l_2 x_1 - 2 l_2 l_3}{l_1^2 + l_2^2} \end{cases} \ , \tag{2.23}$$

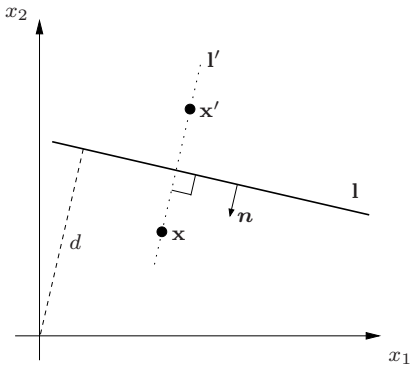


Figure 2.1: Specular reflection of the point $\mathbf{x}$ over the line $\mathbf{l}$: the mirror point $\mathbf{x}'$ lies on the line $\mathbf{l}'$ perpendicular to $\mathbf{l}$ and passing through $\mathbf{x}$. The line $\mathbf{l}$ turns out to be the axis of the segment $\mathbf{x}\mathbf{x}'$.

which can be expressed in matrix form as

$$\mathbf{x}' = \mathbf{H}\mathbf{x} = \begin{bmatrix} -\frac{l_1^2-l_2^2}{l_1^2+l_2^2} & -\frac{2l_1l_2}{l_1^2+l_2^2} & -\frac{2l_1l_3}{l_1^2+l_2^2} \\ -\frac{2l_1l_2}{l_1^2+l_2^2} & \frac{l_1^2-l_2^2}{l_1^2+l_2^2} & -\frac{2l_2l_3}{l_1^2+l_2^2} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} \,. \tag{2.24}$$

If the scaling of the vector $\mathbf{l}$ is such that $l_1^2 + l_2^2 = 1$ and the third coordinate equals the distance d of the line from the origin (i.e. $d = l_3$), the transformation describing the specular reflection can be rewritten as

$$\mathbf{H} = \begin{bmatrix} 1-2l_1^2 & -2l_1l_2 & -2dl_1 \\ -2l_1l_2 & 1-2l_2^2 & -2dl_2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} - 2\boldsymbol{n}\boldsymbol{n}^T & -2d\boldsymbol{n} \\ \mathbf{0}^T & 1 \end{bmatrix} ,$$

where $\boldsymbol{n} = [l_1, l_2]^T$ is a unit vector denoting the normal to $\mathbf{l}$. Notice that the transformation is fully determined from the knowledge of the line parameters, as it does not depends on the point coordinates.

2.2.5 Oriented Projective Geometry

One of the major drawbacks of the classical projective geometry is that it is not possible to discriminate between a ray travelling from $\mathbf{x} = k[x_1, x_2, 1]^T$ to $\mathbf{x}' = k[x_1', x_2', 1]^T$ from the ray with opposite direction. Moreover, projective geometry does not provide any mathematical tool to discriminate in which of the two half-spaces determined by a line a given point belongs to. To overcome these problems

we will use a slightly modified version of the classical projective geometry called *Oriented Projective Geometry* (OPG) [12]. This leads to the definition of the *Oriented Projective Space*, denoted by $\mathbb{T}^2$ for the two-dimensional case. OPG preserves the properties of the classical projective geometry but at the same time allows us to talk about oriented lines and planes, signed angles and convex figures (the latter have no real relevance for our work and thus will not be discussed here). The change consists in taking into account the sign of the scaling coefficient: points $k\mathbf{x}$ and $k'\mathbf{x}$ are not coincident if $\text{sign}(k) \neq \text{sign}(k')$. More specifically, the line going from $\mathbf{x}$ to $\mathbf{x}'$ is represented in OPG by

$$\mathbf{l}^+ = k[l_1, l_2, l_3]^T = k[x_2 - x_2', x_1' - x_1, x_1x_2' - x_2x_1']^T, \quad k > 0 \,,$$

while the line going from $\mathbf{x}'$ to $\mathbf{x}$ is represented by

$$\mathbf{l}^- = k[l_1, l_2, l_3]^T = k[x_2 - x_2',\ x_1' - x_1,\ x_1x_2' - x_2x_1']^T, \quad k < 0 \,.$$

Notice also that

$$\mathbf{l}^- = -\mathbf{l}^+ = -k[l_1, l_2, l_3]^T = k[x_2' - x_2,\ x_1 - x_1',\ x_2x_1' - x_1x_2']^T, \quad k > 0 \,.$$

Therefore, given two points $\mathbf{x}$ and $\mathbf{x}'$ we can define two lines joining them. The non-oriented line $\mathbf{l}$ is given by the union of two oriented lines $\mathbf{l}^+$ and $\mathbf{l}^-$. This also means that when we are performing operations such as finding the line passing through two points, we need to pay attention to the order of operands. Given two points, $\mathbf{x}$ and $\mathbf{x}'$, the oriented lines $\mathbf{l}^+$ and $\mathbf{l}^-$ are obtained as

$$\begin{aligned} \mathbf{l}^+ &= \mathbf{x}\times\mathbf{x}' \,, \\ \mathbf{l}^- &= \mathbf{x}'\times\mathbf{x} = -(\mathbf{x}\times\mathbf{x}') = -\mathbf{l}^+ \,. \end{aligned}$$

By distinguishing line's direction we can also distinguish its left and right sides. Given the oriented line $\mathbf{l} = k[l_1, l_2, l_3]^T$, $k > 0$, and the point $\mathbf{x} = [x_1, x_2, 1]$, this point is on the line's left side if $\mathbf{x}^T\mathbf{l} > 0$. Conversely, $\mathbf{x}$ is on the line's right side if $\mathbf{x}^T\mathbf{l} < 0$. In order to make this fact evident, consider a point $\mathbf{x}' = [x_1', x_2', 1]^T$ on the line, such that $\mathbf{x}'^T\mathbf{l} = 0$. With reference to Figure 2.2, we represent $\mathbf{x}$ and $\mathbf{x}'$ in the three-dimensional Euclidean space $\mathbb{R}^3$ as $\boldsymbol{X} = [x_1, x_2, 0]^T$ and $\boldsymbol{X}' = [x_1', x_2', 0]^T$, respectively. Let $\boldsymbol{V} = [l_2, -l_1, 0]^T$ be the vector representing the line direction, and

$\boldsymbol{U} = [x_1 - x'_1, x_2 - x'_2, 0]^T$ the vector going from $\boldsymbol{X}'$ to $\boldsymbol{X}$. In order to determine

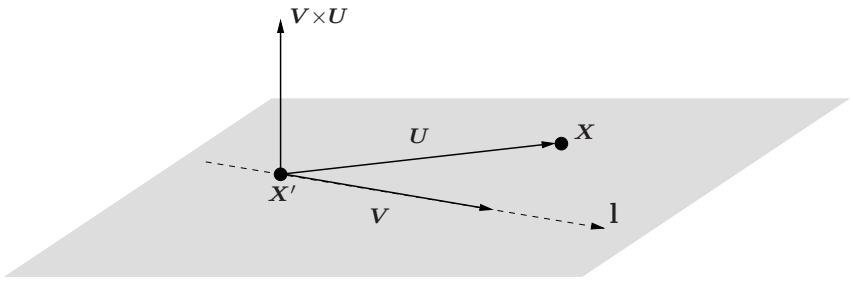


Figure 2.2: The point $\boldsymbol{X}$ lies on the left side of the oriented line $\mathbf{l}$. This means that the third coordinate of the cross-product $\boldsymbol{V} \times \boldsymbol{U}$ is positive.

whether $\boldsymbol{X}$ lies on the left or right half-plane of $\mathbf{l}$, it is sufficient to evaluate the orientation (i.e., the sign of the third coordinate) of the vector

$$\boldsymbol{V}\times\boldsymbol{U} = \det \begin{bmatrix} \underline{\mathtt{i}} & \underline{\mathtt{j}} & \underline{\mathtt{k}} \\ l_2 & -l_1 & 0 \\ x_1 - x'_1 & x_2 - x'_2 & 0 \end{bmatrix} = [0,\ 0,\ l_2x_2 - l_2x'_2 + l_1x_1 - l_1x'_1]^T \ .$$

More specifically, $\boldsymbol{V}\times\boldsymbol{U}$ points upwards if $\boldsymbol{X}$ is on the left side of $\mathbf{l}$ and, vice-versa, it points downwards if $\boldsymbol{X}$ is on the line's right side. Switching back to the projective space, since $\mathbf{x}'^T\mathbf{l} = x'_1l_1 + x'_2l_2 + l_3 = 0$ it is readily verified that

$$l_3 = -x'_1l_1 - x'_2l_2 \ . \tag{2.25}$$

Moreover, computing the product

$$\mathbf{x}^T\mathbf{l} = [x_1, x_2, 1]^T[l_1, l_2, l_3] = x_1l_1 + x_2l_2 + l_3$$

and using (2.25), it is easy to see that

$$\boldsymbol{V}\times\boldsymbol{U} = [0,\ 0,\ \mathbf{x}^T\mathbf{l}]^T \ .$$

Therefore, the sign of the product $\mathbf{x}^T\mathbf{l}$ indicates if the point $\mathbf{x}$ lies on the left or right side of the oriented line $\mathbf{l}$.

2.3 Acoustic geometric primitives in the ray parameter space

In this Section we introduce the concept of ray parameter space, which constitutes the most natural choice for representing acoustic rays. Acoustic rays can be considered as elementary entities whose composition leads to the definition of more complex acoustic entities such as sources, receivers, reflectors and beams, which are efficiently defined in the ray parameter space as well.

2.3.1 The ray parameter space

The acoustic ray can be seen as an oriented line in the geometric space. A line in $\mathbb{R}^2$ is represented by the equation

$$l_1 x_1 + l_2 x_2 + l_3 = 0.$$

We parameterize a ray with the coordinates $[l_1, l_2, l_3]^T$ of the line on which the ray lies. Since the vectors $[l_1, l_2, l_3]^T$ and $k[l_1, l_2, l_3]^T, k \neq 0$ represent the same ray, this parametrization defines a class of equivalence, as it uses scalable (homogeneous) coordinates. With reference to Figure 2.3, a generic point in the (l_1, l_2, l_3) space corresponds to a ray in the geometric space and thus this parametrization is here referred as the *ray space*. The equivalence class inherent in the ray space implies that the ray space is a projective space $\mathbb{P}^2$. As already discussed in Section 2.2, using the classical projective geometry it is not possible to distinguish different travel directions. Adopting the concepts of the oriented projective geometry introduced in Section 2.2.5 it is possible to assign different orientations to lines. Rays on the same line but with opposite orientations can therefore be interpreted as two distinct lines in the oriented projective space $\mathbb{T}^2$

$$\mathbf{l}^+ = k[l_1, l_2, l_3]^T, k > 0$$
$$\mathbf{l}^- = k[l_1, l_2, l_3]^T, k < 0\ .$$

2.3.1.1 Reduced ray space

For clarity of visualization, rather than visualizing the whole three dimensional ray space, we depict the primitives in a reduced 2D ray space, obtained by intersecting

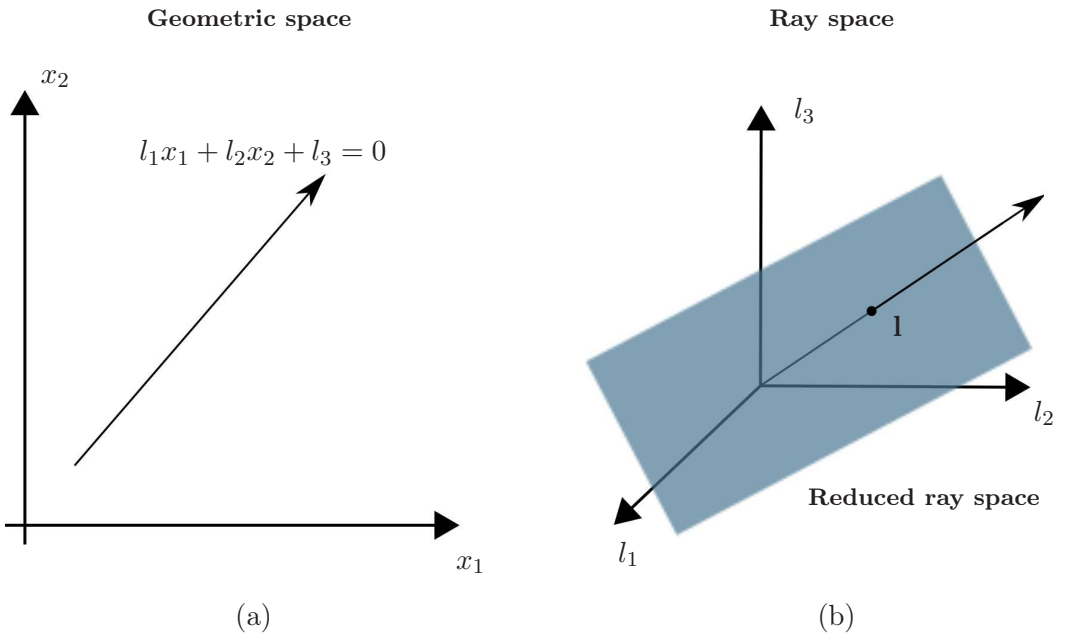


Figure 2.3: A ray in geometric (a) and ray spaces (b).

the ray space with a prescribed plane, as shown in Figure 2.3 (b). We notice, however, that in the reduced ray space we cannot distinguish rays with the same direction but with opposite orientations.

2.3.2 Geometric primitives

In order to model the environment we are working in (i.e. the acostic scene), we need to define a minimal set of elementary objects. The geometric primitives we are interested in are acoustic *rays*, *sources*, *receivers*, *reflectors* and *beams*. We will see how these objects are conveniently described in the ray space. Moreover, we will show that the ray space turns out to play the role of dual of the geometric space. Exploiting this duality, it is also possible to model complex acoustic scenes by taking into account the mutual visibility of the reflectors present in the environment.

2.3.2.1 Acoustic ray

As described before, the acoustic ray can be seen as an oriented line in the geometric space and thus as a half line passing through the origin in the ray space. An example

of the representation of a point in the geometric and reduced ray spaces is shown in Figure 2.5-(a).

2.3.2.2 Acoustic source and receiver

Sources and receivers can be seen as points in the geometric space. In the ray space, the Euclidean point $\boldsymbol{x}_P = [x_{1_P}, x_{2_P}]^T$ can be represented as the set of all rays that pass through it. The homogeneous coordinates of $\boldsymbol{x}_P$ are $\mathbf{x}_P = \lambda[\boldsymbol{x}_P, 1]^T, \quad \lambda > 0$. The set of rays passing through $\boldsymbol{x}_P$ is

$$\mathcal{P} = \{(l_1, l_2, l_3) \in \mathbb{R}^3 \mid l_1 x_{1_P} + l_2 x_{2_P} + l_3 = 0\} = \{\mathbf{l} \in \mathbb{P}^2 \mid \mathbf{x}_P^T \mathbf{l} = 0\}. \tag{2.26}$$

Note that (2.26) defines a plane in the ray space passing through the origin, as $\mathbf{x}_P$ is known. An example of the representation of a point in the geometric and reduced ray spaces is shown in Figure 2.5-(b).

2.3.2.3 Reflector

In the geometric domain the reflector R is a line segment and it is completely defined by the two endpoints $\mathbf{x}_A = [\boldsymbol{x}_A, 1]^T$ and $\mathbf{x}_B = [\boldsymbol{x}_B, 1]^T$, denoted as $\mathcal{A}$ and $\mathcal{B}$ in the ray space. As for points, we represent the reflector with the set of all the rays that intersect it, i.e. the rays that pass through an intermediate point between $\mathbf{x}_A$ and $\mathbf{x}_B$. Examples of two such rays, $\mathbf{l}^+$ and $\mathbf{l}^-$, are shown in Figure 2.4. As each point

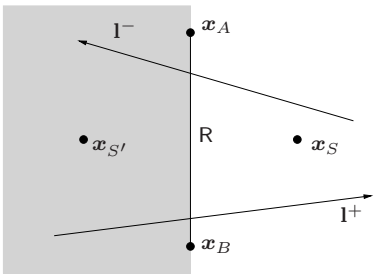


Figure 2.4: The source $\boldsymbol{x}_S$ is mirrored over the reflector in order to obtain the source $\boldsymbol{x}_{S'}$. According to the image source principle, the obstacles in the half-space in which $\boldsymbol{x}_{S'}$ lies (depicted with the shaded wiring) are not considered in the visibility evaluation from $\boldsymbol{x}_{S'}$. Two rays $\mathbf{l}^+$ and $\mathbf{l}^-$ have with different orientations since they cross the obstacle at opposite sides.

of the line segment maps to a plane in the ray space, the reflector corresponds to the set of all the planes representing the infinite intermediate points between $\mathcal{A}$ and $\mathcal{B}$,

i.e.

$$\mathcal{R} = \mathcal{A} \cup ... \cup \mathcal{P}_i \cup ... \cup \mathcal{B}.$$

Orienting the reflector According to the image source principle [22], an acoustic source located at $\boldsymbol{x}_S$ in the geometric space is mirrored over the reflector R in the so-called mirror position $\boldsymbol{x}_{S'}$. We notice that, when we evaluate the visibility of the environment from a mirrored source, we do not consider the reflectors in the half-space where the mirrored source lies, i.e. the grey shaded area in Figure 2.4. When we evaluate the visibility from the mirrored source $\boldsymbol{x}_{S'}$ we consider only rays that pass through the line segment $\boldsymbol{x}_A\boldsymbol{x}_B$ and go from the grey to the white half-space. This motivates us in distinguishing rays that have different directions with respect to two half-spaces and thus in defining two reflectors, one for each face of the line segment. With reference to Figure 2.5-(c), the two rays $\mathbf{l}^+$ and $\mathbf{l}^-$ fall onto the opposite faces of the line segment $\mathcal{AB}$, i.e. onto two different reflectors defined by the same line segment but characterized by different directions of incident rays. We observe that $\mathbf{l}^+$ and $\mathbf{l}^-$ have the endpoints $\mathcal{A}$ and $\mathcal{B}$ on opposite sides: $\mathcal{A}$ is on the left for $\mathbf{l}^+$ and on the right for $\mathbf{l}^-$ and therefore $\mathbf{x}_A^T\mathbf{l}^+ > 0$ and $\mathbf{x}_A^T\mathbf{l}^- < 0$. We exploit this inequalities when we represent the two *oriented reflectors* $\mathcal{R}_+$ and $\mathcal{R}_-$ corresponding to the *non-oriented reflector* $\mathcal{R}$. We define them in the ray space as the set of all rays that fall to the corresponding face of the line segments:

$$\mathcal{R}_+ = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} > 0\} \cap \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_B^T\mathbf{l} < 0\} = \mathcal{A}_+ \cap \mathcal{B}_-,$$

$$\mathcal{R}_- = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} < 0\} \cap \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_B^T\mathbf{l} > 0\} = \mathcal{A}_- \cap \mathcal{B}_+.$$

The non oriented reflector can be expressed in a closed form as the union of two oriented reflectors that compose it

$$\mathcal{R} = \mathcal{R}_+ \cup \mathcal{R}_- = \{\mathcal{A}_+ \cap \mathcal{B}_-\} \cup \{\mathcal{A}_- \cap \mathcal{B}_+\}.$$

We notice that the rays in $\mathcal{R}_-$ correspond to the rays in $\mathcal{R}_+$ but with opposite travel directions. We denote this relationship as $\overline{\mathcal{R}}_- \triangleq \{-\mathbf{l} \mid \mathbf{l} \in \mathcal{R}_-\}$. Therefore we have that $\overline{\mathcal{R}}_- = \mathcal{R}_+$ and $\overline{\mathcal{R}}_+ = \mathcal{R}_-$.

Visibility region All the rays originated from an oriented reflector $\mathcal{R}_i$ form the region of visibility from that reflector, $\Gamma(\mathcal{R}_i)$. These rays correspond to the rays

that define the oriented reflector, i.e. the ray that fall onto it, but with opposite travel directions and thus, using the notation introduced above, $\Gamma(\mathcal{R}_i) = \overline{\mathcal{R}}_i$. By intersecting this region with the rays that fall onto another oriented reflector $\mathcal{R}_j$ we obtain the visibility region of $\mathcal{R}_j$ from $\mathcal{R}_i$

$$\Upsilon(\mathcal{R}_i, \mathcal{R}_j) = \Gamma(\mathcal{R}_i) \cap \mathcal{R}_j = \overline{\mathcal{R}}_i \cap \mathcal{R}_j \ .$$

With reference to Figure 2.5-(d), in the reduced ray space the visibility region $\Upsilon(\mathcal{R}_i, \mathcal{R}_j)$ is given by the intersection of four half spaces that form a pyramid with the apex at the origin of the ray space.

Visibility diagram If the environment is composed of more than two reflectors, mutual occlusions may arise. This corresponds to an overlapping of visibility regions in the ray space. Sorting out which reflector occludes which in the geometric space means determining which visibility region overrides which in their overlap as described in Figure 2.5-(e). We perform this operation casting a test ray in the overlap region and finding which reflector it hits first, as illustrated in Figure 2.5-(e). The resulting collection of visibility regions constitutes the visibility diagram of the reflector $\mathcal{R}_i$

$$\Omega(\mathcal{R}_i) = \{\Upsilon^{\circledast}(\mathcal{R}_i, \mathcal{R}_j) \neq \emptyset,\ j \neq i\}_{j=1}^{N_i} \ ,$$

where $\mathcal{R}_j$ are the N_i reflectors visible from $\mathcal{R}_i$ and $^{\circledast}$ indicates that visibility regions have been overridden according to the front-to-back order. The collection of visibility diagrams of all the reflectors gives us the information on the mutual visibility between all reflectors in the environment. Notice that this operation requires only the knowledge of the geometry of the environment, as it is independent from the presence of potential sources or receivers.

2.3.2.4 Beam

With the term *beam* we refer to a compact bundle of acoustic rays that originate from the same point and fall onto the same reflector. A beam W, originated from reflector R_i that falls onto reflector R_j, is completely specified by an origin $\mathbf{x}_S$ (image source) and by the (connected) illuminated region of the reflector R_j. In the ray space, the beam corresponds to the intersection of the representations $\mathcal{S}$ of the image source

and of the illuminated portion of the visibility region $\Upsilon(\mathcal{R}_i, \mathcal{R}_j)$:

$$\mathcal{W} = \mathcal{S} \cap \Upsilon(\mathcal{R}_i, \mathcal{R}_j)|_{\mathbf{l}_1}^{\mathbf{l}_2},$$

where $|_{\mathbf{l}_1}^{\mathbf{l}_2}$ indicates that the visibility region is limited by the rays $\mathbf{l}_1$ and $\mathbf{l}_2$. Figure 2.5-(f) shows an example of a beam in the geometric and reduced ray space.

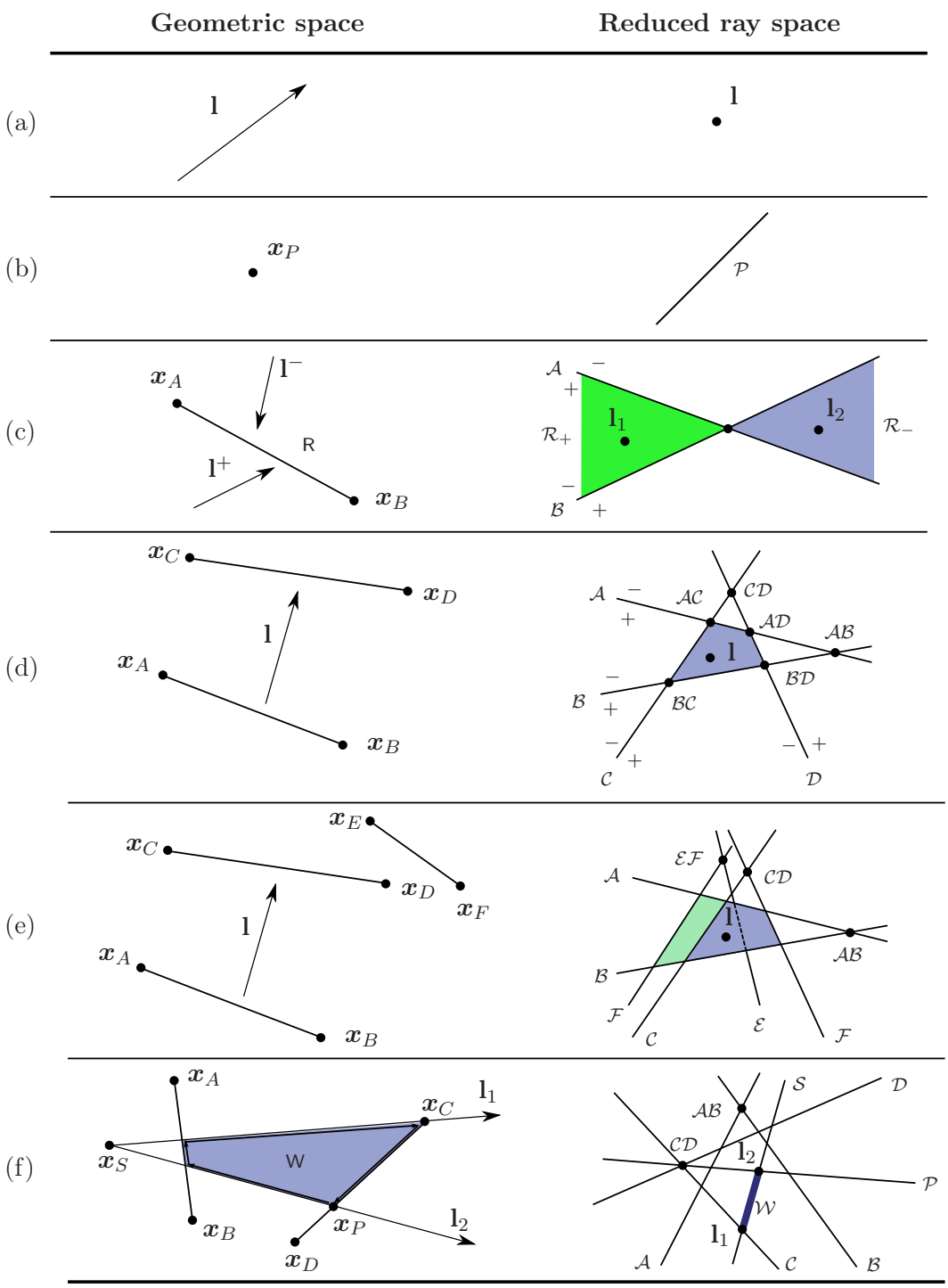


Figure 2.5: Transformations from geometric to ray space: (a) ray; (b) point; (c) reflector; (d) visibility region; (e) visibility diagram; (f) beam.

3 Geometric wave field analysis

This Chapter is devoted to geometric solutions for wave field analysis. Starting from classical acoustic measurements (Times of Arrival, Time Differences of Arrival, and Directions of Arrival), we will devise algorithms for estimating the position of acoustic primitives such as microphones, sources, and reflectors. We will see how the use of projective geometry is suitable for turning acoustic measurements into homogeneous quadratic constraints that directly act on the primitives to be estimated. Multiple constraints are then collected into an homogeneous cost function, whose minimization leads to an estimate of the primitive of interest. Two are the main advantages of using a projective representation: on one hand, since points and lines have the identical homogeneous representation, they can be managed in the same way; on the other hand, all the homogeneous constraints arising from the measurements share the same mathematical form, independently from the specific estimation problem. This also means that different estimation problems can be solved through the minimization of a cost function collecting multiple constraints and whose form is problem independent as well. As a consequence, projective geometry leads to an unified theory valid for a huge set of problems. It is therefore possible to identify an unified optimization strategy that can be applied to all the estimation problems.

The Chapter is organized as follows. In Section 3.1 we present an overview of acoustic measurements such as Time of Arrival, Time Difference of Arrival, and Direction of Arrival. In particular, we give some hints for their measurement for both direct and indirect acoustic paths. In Section 3.2 we introduce the projective constraints arising from acoustic measurements. Section 3.3 is devoted to the development of the theory which is at the base of all the estimation problems. After the definition of a suitable cost function, a set of possible common optimization strategies is discussed. Finally, Section 3.4 describes the estimation algorithms used for array calibration, source localization, and reflector localization tasks. Moreover, such estimation problems will be discussed within the context of acoustic scene analysis.

3.1 Acoustic measurements on direct and indirect paths

In this Section we describe the acoustic measurements on which geometric constraints rely. Times Of Arrival (TOAs), Time Differences Of Arrival (TDOAs), and Direction Of Arrival (DOAs) are considered for both direct and indirect paths generated by the presence of an acoustic source in a reverberant environment. The measurement of these quantities is performed by means of one or more microphones possibly arranged in an array structure.

3.1.1 Time Of Arrival

In this paragraph we assume that the reproduction and the acquisition systems are synchronized, i.e. the analog-to-digital and the digital-to-analog converters share the same clock. This means that from the acquisitions we can measure the time of flight of the signal that propagates from a controlled acoustic source located at $\boldsymbol{x}_s$ (e.g., a loudspeaker) to a microphone located at $\boldsymbol{x}_m$. An example is shown in Figure 3.1, which depicts the direct source-microphone path and two reflective paths in a rectangular room. We aim at measuring the Time Of Arrival of direct and reflective

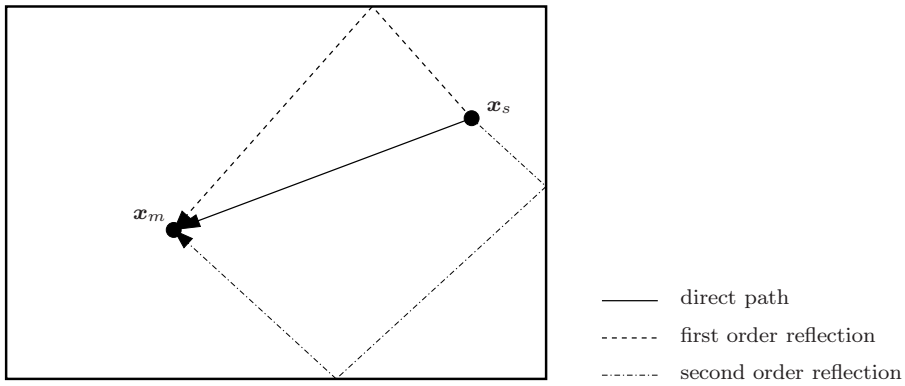


Figure 3.1: Examples of direct and reflective paths between a source located at $\boldsymbol{x}_s$ and a microphone located at $\boldsymbol{x}_m$ in a rectangular room.

acoustic paths that link $\boldsymbol{x}_s$ and $\boldsymbol{x}_m$ from the estimation of the corresponding acoustic impulse response $\bar{h}_{sm}(t)$. In order to do so, the loudspeaker produces a known time-continuous sequence $\bar{s}(t)$. The time-continuous signal $\bar{u}_m(t)$ acquired by the sensor is commonly modelled as the summation of delayed and attenuated replicas of the

signal $\bar{s}(t)$ [2]:

$$\bar{u}_m(t) = \bar{\alpha}_m \bar{s}(t - \tau_m) + \sum_{l=1}^{L} \bar{\beta}_{lm} \bar{s}(t - \tau_{lm}) + \bar{\nu}_m(t) \ . \tag{3.1}$$

The coefficient $\bar{\alpha}_m$ is the attenuation of the direct path from source to receiver; τ_m is the corresponding delay; the summation over l models the presence in the impulse response of reflections from obstacles, which are delayed by τ_{lm} and attenuated by $\bar{\beta}_{lm}$; finally $\bar{\nu}_m(t)$ is an additive noise that alters the measurement. The number of reflective paths L is generally greater than the number of obstacles in the environment, since paths that bounce on more than one obstacle are present in $\bar{u}_m(t)$.

After sampling by the analog-to-digital converter we write the time-discrete signal as

$$u_m(n) \triangleq \bar{u}_m(nT_s) = \alpha_m s(n - i_m) + \sum_{l=1}^{L} \beta_{lm} s(n - i_{lm}) + \nu_m(n) \ , \tag{3.2}$$

where T_s is the sampling period; n is the time-discrete index; $s(n) \triangleq \bar{s}(nT_s)$ is the discrete version of the source signal; $\nu_m(n) \triangleq \bar{\nu}_m(nT_s)$ is the discrete noise signal; i_m and i_{lm} are the discrete versions of τ_m and τ_{lm}, respectively; α_m and β_{lm} are the attenuation coefficients associated to $\bar{\alpha}_m$ and $\bar{\beta}_{lm}$, respectively (in general, $\alpha_m \neq \bar{\alpha}_m$ and $\beta_{lm} \neq \bar{\beta}_{lm}$). We observe that, since the measurement relies on the discrete delays i_m and i_{lm} rather than on τ_m and τ_{lm}, we introduce an error. This is due to the fact that, after time-sampling, the time-axis becomes discrete.

In the literature different methodologies have been proposed for the estimation of the acoustic impulse response. They range from algorithms devoted to acoustic channel identification [34, 35, 36], to techniques dedicated to the estimation of the Time of Arrival [37, 38]. One of the simplest methods rely on the cross-correlation between the emitted signal $s_m(n)$ and the acquired signal $u_m(n)$ [39]. Although the cross-correlation gives only a rough estimate of the impulse response, it reveals to be suitable for assessing the TOA of the direct and the first reflective paths [39], corresponding to i_m and $i_{lm}, l = 1, \ldots, L$, respectively. In mathematical terms, this corresponds to

$$\hat{h}_m(n) = u_m(n) \otimes s(n) \triangleq \sum_{k=-\infty}^{+\infty} u_m(k) s(n - k) \ , \tag{3.3}$$

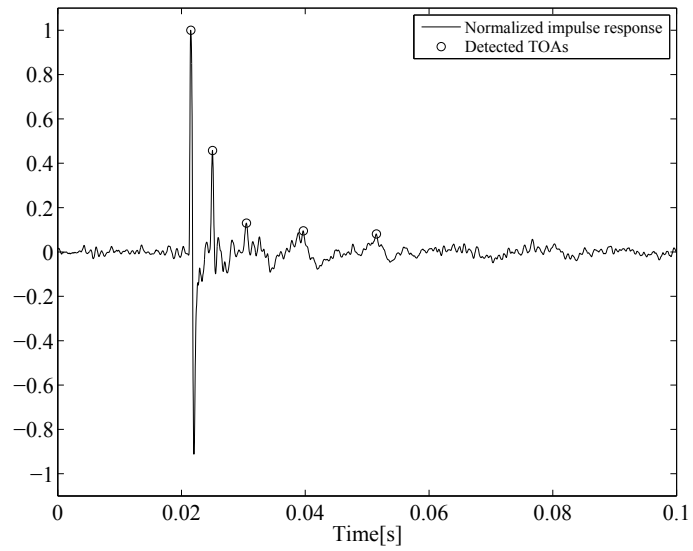


Figure 3.2: Impulse response acquired in a dry room in which two parallel reflective panels have been placed. The peaks corresponding to the direct path and the first two orders of reflections are marked with circles.

where the symbol $\otimes$ denotes the cross-correlation operator. In order to obtain the maximum resolution along with an acceptable robustness against additive noise, the most suitable sequence $s(n)$ for our purposes is a band-limited white noise signal of arbitrary duration.

Figure 3.2 shows an example of impulse response acquired when source and sensor are placed in an acoustically dry room and some reflectors are placed in arbitrary positions; the peaks corresponding to the direct path and the first two orders of reflections are marked with circles. Amplitudes have been normalized to the highest peak of the impulse response. We observe a different duration of the impulses corresponding to direct and reflective paths: while the direct-path impulse is defined on a short time interval, the peaks corresponding to the reflective paths are spread over a longer time range. This low-pass filtering effect is due to the frequency-dependent reflection coefficient of obstacles in the environment.

It is important to notice that the response of the whole sound processing chain (digital-to-analog converter, loudspeaker, microphone and analog-to-digital converter) is not explicitly modeled, therefore in $\hat{h}_m(n)$ we observe not only the effect of the acoustic channel between $\boldsymbol{x}_s$ and $\boldsymbol{x}_m$, but also the convolutive effect of emission and acquisition devices. For our purposes, however, we consider the processing chain to be completely transparent.

Under the assumptions introduced above, TOAs can be measured by estimating the time positions of peaks in the impulse response. In particular the algorithm must select the discrete time lags $i_m, i_{1m}, i_{2m}, \dots i_{Lm}$ corresponding to the most relevant local maxima of $\hat{h}_m(n)$. An estimate $\hat{\tau}_m, \hat{\tau}_{1m}, \hat{\tau}_{2m}, \dots \hat{\tau}_{Lm}$ of the continuous time positions associated to the discrete lags $i_m, i_{1m}, i_{2m}, \dots i_{Lm}$ can be obtained through parabolic interpolation [40]. The TOAs, expressed in seconds, are finally given by $t_m = \hat{\tau}_m T_s$ (direct path); and $t_{lm} = \hat{\tau}_{lm} T_s \ , l = 1, \dots L$ (reflective paths).

In the following we will consider that source and sensor are in line of sight. This means that the direct path is always present in the impulse response. Since the direct path always has the shortest propagation time, the lag i_m corresponds to the location of the first relevant peak in the impulse response. Consequently, as a final result of the measurement process, it is possible to organize TOAs in two sets: the first one, t_m, contains only the TOA associated to the direct path; the second one, $t_{1m}, \dots t_{Lm}$, arranges the TOAs of reflective paths.

3.1.2 Time Difference Of Arrival

We now remove the hypothesis that the source is synchronized with the acquisition device, which is now composed of a pair of microphones located at $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$. This scenario accounts for situations in which an arbitrary source (e.g. a person uttering a sentence, a ring of a mobile phone, etc.) is active in the environment. Even if we cannot measure the Time Of Arrival, many cues about the source position can still be extracted from the joint knowledge of the synchronized signals available at sensors. After time sampling, the signals acquired by the sensors are

$$\begin{aligned} u_p(n) &= \alpha_p s(n - i_p) + \sum_{l=1}^{L} \beta_{lp} s(n - i_{lp}) + \nu_p(n) \\ u_q(n) &= \alpha_q s(n - i_q) + \sum_{l=1}^{L} \beta_{lq} s(n - i_{lq}) + \nu_q(n) \ . \end{aligned} \tag{3.4}$$

The model in the above equation is equivalent to the model in (3.2); notice, however, that the source is unknown, let alone its delay. In the following we discuss about the possibility of estimating the Time Difference Of Arrival on the two microphones, which corresponds, for the direct path, to the time delay $t_{pq} = (i_p - i_q)T_s$, where T_s denotes the sampling period. A simple way for estimating TDOAs is again the cross-

correlation [39], which is computed, in this case, between the signals received on a pair of synchronized microphones. The position of the peaks in the cross-correlation give an estimate of the direct and indirect TDOAs. If the source and the microphones are in line of sight, we expect the main peak to be related to the direct path.

We start our discussion by neglecting, for the moment, the presence of reflections from obstacles, therefore the signals acquired by sensors become

$$\begin{aligned} u_p(n) &\approx \alpha_p s(n - i_p) + \nu_p(n) \\ u_q(n) &\approx \alpha_q s(n - i_q) + \nu_q(n) \ . \end{aligned} \tag{3.5}$$

If the source is sufficiently far from the microphones we can assume that the attenuation is the same on both the sensors, i.e. $\alpha = \alpha_p = \alpha_q$. Assuming also that additive noises at different microphones are uncorrelated, the cross-correlation of $u_p(n)$ and $u_q(n)$ gives

$$R_{pq}(n) = \alpha^2 s(n - i_p) \otimes s(n - i_q) \ .$$

Different source signals can be used. We consider here two cases: sinusoidal (or narrowband) signal and white noise.

As far as sinusoidal signals are concerned the cross-correlation becomes

$$R_{pq}(n) = \alpha^2 \cos\{2\pi T_s[n - (i_p - i_q)]\} \ .$$

The signal $R_{pq}(n)$ has maxima at $n = i_p - i_q + \frac{k}{T_s}$, $k \in \mathbb{Z}$. We notice that, according to the geometric arrangement of the array, only a limited range of delays is allowed. In order to make this fact more evident, we consider the case shown in Figure 3.3. A distant source generates a planar wavefront propagating towards the microphones in direction θ. The propagation delay on the pair of sensors is

$$t_{pq} = (i_p - i_q)T_s = \frac{D}{c} \cos\theta \ , \tag{3.6}$$

where D is the distance between the microphones; T_s is the sampling period; and c is the speed of sound. Notice from (3.6) that the delay is in the range $\left[-\frac{D}{c}, \frac{D}{c}\right]$. Therefore, we search for maxima of the cross-correlation only in this interval.

In order to accommodate the case of moving sources, in real scenarios it is common to work with frames of finite length. When using rectangular windows, each estimation of the cross-correlation turns out to be multiplied by a Bartlett (triangular)

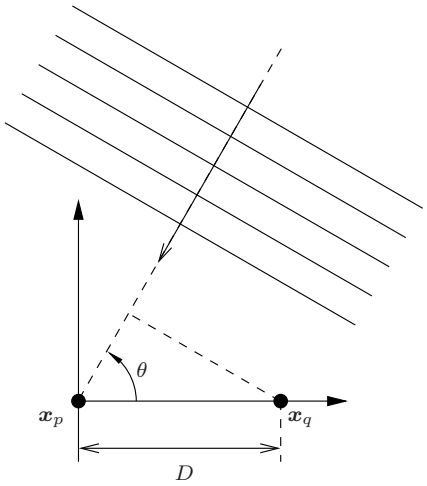


Figure 3.3: The delay between two microphones depends on the distance between microphones in the array and the Direction of Arrival of the signal.

window:

$$R_{pq}(n) = \alpha^2 \cos\{2\pi T_s[n - (i_p - i_q)]\}w(n) \ , \tag{3.7}$$

where $w(n)$ is the window. Notice that windowing does not affect the estimation of the propagation delay to sensors $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$, as local maxima are still present at lags $k = i_p - i_q + \frac{k}{T_s}$, $k \in \mathbb{Z}$.

When the source signal is a white noise, we obtain

$$R_{pq}(n) = \alpha^2 \delta[n - (i_p - i_q)] \ , \tag{3.8}$$

which exhibits a global maximum for $n = i_p - i_q$.

The cases of sinusoidal signals and white noise represent two extreme situations: in the first case the cross-correlation is periodic and multiple local maxima are present, while with white noise a single maximum is present. Arbitrary signals (e.g a person uttering a sentence, a musical instrument, etc.) exhibit an intermediate behaviour between these two cases.

When reflectors are present in the environment, along with the direct signal we observe reflections from walls. The cross-correlation $R_{pq}(n)$ between $u_p(n)$ and $u_q(n)$

gives

$$R_{pq}(n) = \alpha^2 s(n - i_p) \otimes s(n - i_q) + \alpha \sum_{l_1=1}^{L} \beta_{l_1 p} s(n - i_p) \otimes s(n - i_{l_1 q}) + \\ + \alpha \sum_{l_2=1}^{L} \beta_{l_2 p} s(n - i_q) \otimes s(n - i_{l_2 p}) + \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} \beta_{l_1 p} \beta_{l_2 p} s(n - i_{l_1 q}) \otimes s(n - i_{l_2 p}) \ . \tag{3.9}$$

The first term in the right-hand member of $R_{pq}(n)$ is the cross-correlation of the direct paths. The second (third) term accounts for the cross-correlation of the direct signal in $u_p(n)$ ($u_q(n)$) with the reflected ones in $u_q(n)$ ($u_p(n)$). Finally, the last term is the cross-correlation between reflected replicas in $u_p(n)$ and $u_q(n)$. The TDOAs associated to the L reflective paths are given by

$$t_{l,pq} = (i_{l_1 p} - i_{l_2 q}) T_s \ , \quad l = l_1 = l_2 \ , \quad l = 1, \dots L \ .$$

We observe that $R_{pq}(n)$ presents $(L+1)^2$ local maxima. As a consequence, from the observation of a single impulse response, it is impossible to assign a label to local maxima, which identifies the walls that generate reflections.

3.1.3 Direction Of Arrival

The Direction Of Arrival is an acoustic measurement alternative to the TDOA. As for TDOAs, the measurement of DOAs requires no synchronism of the source with respect to the acquisition system. As shown in Figure 3.3, when the source is located far from the microphones, its Direction Of Arrival is related to the time delay on the sensors. The straightforward approach for estimating the DOA of a signal is therefore based on (3.6). However, this solution is not robust against measurement noise, since a small error in the estimation of the time delay may cause a severe error on the estimated DOA, especially when the source is almost aligned with the sensors (i.e. when $\theta \approx 0 + k\pi \ , k \in \mathbb{Z}$).

The literature presents many solutions for the robust estimation of the DOAs relative to one or more acoustic sources by means of a microphone array. These are typically subdivided into two categories [41]: non-parametric methods based on spatial filtering; and parametric methods. Non-parametric techniques rely on a sim-

ple array data model which does not specify any particular assumption on the data covariance. The delay and sum beamformer, as well as the Capon (MVDR) beamformer [42, 43, 41] belong to this category. More sophisticated techniques such as MUSIC [44, 45], ESPRIT [46], EB-ESPRIT [47] belong to the category of parametric algorithms, and they are formulated under restrictive hypothesis on the statistics of the array data. In particular, to work properly they require source signals to be uncorrelated each other. Moreover, they assume that the number of sources to be localized is known in advance and it must be smaller than the number of microphones. When all these assumptions are verified, parametric algorithms outperform non-parametric techniques. Conversely, in situations where it is not possible to assume the uncorrelatedness of the sources, non-parametric algorithms reveal to be more robust than parametric ones.

In this paragraph we give some hints for the estimation of the DOAs, in particular for the detection of reflected sources. DOAs related to the reflected sources constitute the basis for the geometric constraint described in Section 3.2.3. Reflected sources can be seen as mirrored versions of the principal source. As a consequence, their signals are characterized by strong mutual correlation with the direct path one. For this reason, in order to estimate the DOAs related to the reflected paths the natural choice falls on non-parametric algorithms. A convenient approach for such estimation is the wideband Capon algorithm proposed in [48]. Since, in general, no a-priori information about the position of the mirror source is available, the algorithm is adapted for a circular uniform array, which guarantees a constant resolution for all the directions. The algorithm is summarized in the following.

With reference to Figure 3.4, the sensors are uniformly arranged on a circle with radius ρ at positions $\boldsymbol{x}_i = [\rho\cos\phi_i, \rho\sin\phi_i]^T,\ i = 1, \dots M$. The ith microphone acquires the signal $u_i(n)$, where n is the discrete time-index. We refer to the filterbank analysis of $u_i(n)$ with the symbol $y_i(\omega_k, n)$, where $\omega_k,\ k = 1, \dots, K$ denotes the kth frequency sub-band. The signals are then organized in the vector

$$\mathbf{y}(\omega_k, n) = \begin{bmatrix} y_1(\omega_k, n) \\ y_2(\omega_k, n) \\ \vdots \\ y_M(\omega_k, n) \end{bmatrix},$$

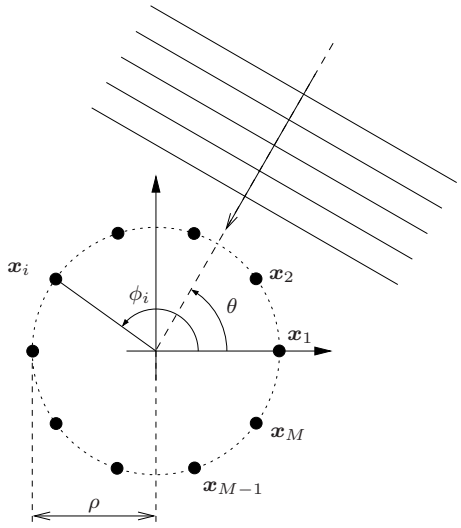


Figure 3.4: A planar wave front coming from the direction θ impinges on a uniform circular array.

and the frequency-dependent autocorrelation matrix is estimated as [48]

$$\mathbf{R}_k = \frac{1}{K}\sum_{t=1}^{T}\mathbf{y}(\omega_k, n)\mathbf{y}^H(\omega_k, n)\ , k = 1\ldots, K. \tag{3.10}$$

For each sub-band, we compute the function

$$f(\vartheta, k) = \frac{1}{\mathbf{a}(\vartheta, k)^H \mathbf{R}_k^{-1}\mathbf{a}(\vartheta, k)}\ , \qquad \mathbf{a}(\vartheta, k) = \begin{bmatrix} e^{j\frac{\omega_k}{c}\rho\cos(\vartheta-\phi_1)} \\ e^{j\frac{\omega_k}{c}\rho\cos(\vartheta-\phi_2)} \\ \vdots \\ e^{j\frac{\omega_k}{c}\rho\cos(\vartheta-\phi_M)} \end{bmatrix}\ , \tag{3.11}$$

where c is the speed of sound. The function $f(\vartheta, k)$ gives an estimate of the energy in the kth sub-band coming from the direction ϑ, and it is known as *spatial-spectrum* [41]. The term $\mathbf{a}(\vartheta, k)$ is called *transfer vector* [41], and models the propagation of a monochromatic plane wave with frequency ω_k, traveling with direction ϑ towards the sensors. The overall spatial-spectrum is computed as the geometric mean of the spatial-spectra of the different sub-bands [48]:

$$F(\vartheta) = \left[\prod_{k=1}^{K} f(\vartheta, k)\right]^{\frac{1}{K}}\ . \tag{3.12}$$

Finally, the DOAs are selected as the most relevant local maxima of the function $F(\vartheta)$ in the range $[0, 2\pi]$. For example, when a single source (e.g. a source in a low-reverberation chamber) impinges on the array from the direction θ as in Figure 3.4, $F(\vartheta)$ is expected to exhibit a single peak located in proximity of $\vartheta = \theta$. In presence of reverberations, letting θ be the DOA relative to the direct path and $\theta_1, \theta_2, \ldots$ the DOAs relative to the reflected paths, the function $F(\vartheta)$ will present peaks in proximity of $\vartheta = \theta, \theta_1, \theta_2, \ldots$.

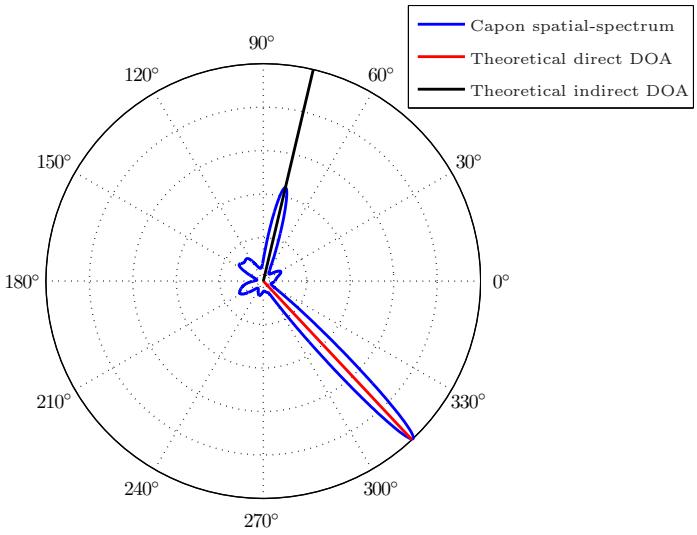


Figure 3.5: Example of Capon spatial-spectrum for measuring direct and indirect DOAs.

An example of measurement of direct and indirect DOAs is shown in Figure 3.5. The acquisitions relative to the example were made in a low-reverberation room whose acoustics has been altered by inserting a rigid wooden panel. The array was composed of 10 microphones disposed on a circle with radius 4 cm, perpendicular to the plane of the panel. From the plot of the spatial-spectrum one can identify the DOAs relative to the direct and reflective paths.

3.2 Constraints deriving from acoustic measurements

In this Section we describe how to turn measurements (TOAs, TDOAs, DOAs) into constraints acting on geometric primitives such as sources, microphones, and reflectors. In particular, we show that each measurement on acoustic paths turns out to generate a projective constraint in $\mathbb{P}^2$ that takes the shape of a quadratic form.

3.2.1 Constraints related to TOAs

In this paragraph we will derive the geometric constraints arising from the measurement of the Time of Arrival. We will see that the TOA related to the direct path leads to a the quadratic constraint on the position of sources or microphones. On the other hand, a reflective TOA generates a constraint acting on the line on which the reflector generating the acoustic path lies.

3.2.1.1 Direct signal

Let t_m be the Time Of Arrival measured by a microphone located at $\boldsymbol{x}_m = [x_{1_m}, x_{2_m}]^T$. If the source location $\boldsymbol{x}_s = [x_{1_s}, x_{2_s}]^T$ is given, the TOA constrains the microphone to be placed on a circumference centered at $\boldsymbol{x}_s$ and with radius $\rho_m = t_m c$, where c is the sound speed. The equation of the circumference is

$$(x_1 - x_{1_s})^2 + (x_2 - x_{2_s})^2 = \rho_m^2.$$

Expanding the terms, we obtain

$$x_1^2 + x_{1_s}^2 - 2x_1x_{1_s} + x_2^2 + x_{2_s}^2 - 2x_2x_{2_s} - \rho_m^2 = 0\ ,$$

and through comparison with the expression of a general conic in (2.18) we get the parameter vector $\mathbf{c}_m = [1,\ 0,\ 1,\ -2x_{1_s},\ -2x_{2_s},\ -\rho_m^2 + x_{1_s}^2 + y_{1_s}^2]^T$ and the corresponding conic matrix

$$\mathbf{C}_m = \begin{bmatrix} 1 & 0 & -x_{1_s} \\ 0 & 1 & -x_{2_s} \\ -x_{1_s} & -x_{2_s} & -\rho_m^2 + x_{1_s}^2 + y_{1_s}^2 \end{bmatrix}\ . \tag{3.13}$$

As a consequence, the TOA leads to a quadratic constraints that, adopting homogeneous coordinates, can be written as

$$\mathbf{x}^T \mathbf{C}_m \mathbf{x} = 0 \, ,$$

where $\mathbf{x} = [\boldsymbol{x}, 1]^T$ is the homogeneous representation of the point $\boldsymbol{x} = [x_1, x_2]^T$.

Notice that the roles of microphone and source can be interchanged: if we know the microphone position, the TOA constrains the source to be located on a circumference centered at $\boldsymbol{x}_m$, and the conic matrix (3.13) is modified inserting (x_{1_m}, x_{2_m}) in place of (x_{1_s}, x_{2_s}).

3.2.1.2 Indirect signal

We now consider the measurement of the Time Of Arrival related to a reflected signal. For the sake of simplicity, we assume that only one reflector is present in the environment, so that there exists a single reflected path. With reference to Figure 3.6, we consider the acoustic path that links the source located at $\boldsymbol{x}_s = [x_{1_s}, x_{2_s}]^T$ and the receiver located at $\boldsymbol{x}_m = [x_{1_m}, x_{2_m}]^T$ through a reflection. The reflection point $\boldsymbol{x}_P$ on the reflector honors the Snell's law. We observe that the corresponding

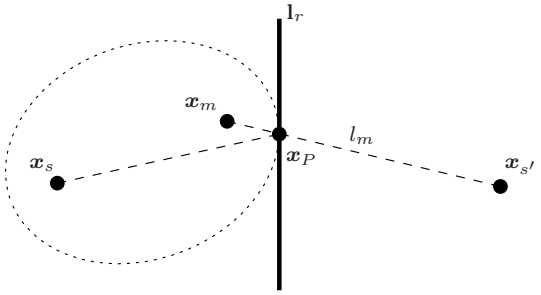


Figure 3.6: An acoustic path links $\boldsymbol{x}_s$ and $\boldsymbol{x}_m$ through the reflector. The reflection point $\boldsymbol{x}_P$ honors the Snell's law, and it is bounded to lie on an ellipse whose foci are $\boldsymbol{x}_s$ and $\boldsymbol{x}_m$ and its major axis is l_m. Moreover, the reflector line $\mathbf{l}_r$ corresponds to the tangent to the ellipse at $\boldsymbol{x}_P$.

TOA t'_m is the sum of the time-of-flight from $\boldsymbol{x}_s$ to $\boldsymbol{x}_P$ and the time-of-flight from $\boldsymbol{x}_P$ to $\boldsymbol{x}_m$, and it is equivalent to the TOA relative to the mirror source located at $\boldsymbol{x}_{s'}$. Consequently, we are constraining the reflection point $\boldsymbol{x}_P$ to lie on an ellipse whose foci are $\boldsymbol{x}_s$ and $\boldsymbol{x}_m$ and whose major axis is $l_m = t'_m c$, where c is the sound speed. In order to find the equation of the ellipse, we start from the fact that it is defined as the locus of all points of the plane whose distances from the foci add to

the same constant. In our case, this is equivalent to

$$\sqrt{(x_1-x_{1_s})^2+(x_2-x_{2_s})^2}+\sqrt{(x_1-x_{1_m})^2+(x_2-x_{2_m})^2}=l_m \ ,$$

which can be written also as

$$\sqrt{(x_1-x_{1_s})^2+(x_2-x_{2_s})^2}=l_m-\sqrt{(x_1-x_{1_m})^2+(x_2-x_{2_m})^2} \ . \tag{3.14}$$

Taking the square power of both sides of (3.14), after some arithmetics we get

$$2l_m\sqrt{(x_1-x_{1_m})^2+(x_2-x_{2_m})^2}=l_m^2+2(xx_{1_s}+xx_{2_s}-xx_{1_m}-xx_{2_m})+x_{1_m}^2+x_{2_m}^2-x_{1_s}^2-x_{2_s}^2 \ . \tag{3.15}$$

Finally, squaring both the sides of (3.15) and comparing the result with the general conic equation (2.18), we derive the homogeneous parameter vector of the ellipse $\mathbf{c}_m = [a_m, b_m, c_m, d_m, e_m, f_m]^T$, where

$$\begin{cases} a_m &= -4[l_m^2-(x_{1_m}-x_{1_s})^2] \\ b_m &= 8(x_{1_m}-x_{1_s})(x_{2_m}-x_{2_s}) \\ c_m &= -4[l_m^2-(x_{2_m}-x_{2_s})^2] \\ d_m &= 4[l_m^2(x_{1_m}+x_{1_s})-(x_{1_m}-x_{1_s})(x_{1_m}^2+x_{2_m}^2-x_{1_s}^2-x_{2_s}^2)] \\ e_m &= 4[l_m^2(x_{2_m}+x_{2_s})-(x_{2_m}-x_{2_s})(x_{1_m}^2+x_{2_m}^2-x_{1_s}^2-x_{2_s}^2)] \\ f_m &= l_m^4-2l_m^2(x_{1_m}^2+x_{2_m}^2+x_{1_s}^2+x_{2_s}^2)+(x_{1_m}^2+x_{2_m}^2-x_{1_s}^2-x_{2_s}^2)^2 \end{cases} \ . \tag{3.16}$$

As for the direct path, we can encode the constraint relative to the reflective path in the quadratic form

$$\mathbf{x}^T\mathbf{C}_m\mathbf{x}=0 \ , \ \mathbf{C}_m = \begin{bmatrix} a_m & b_m/2 & d_m/2 \\ b_m/2 & c_m & e_m/2 \\ d_m/2 & e_m/2 & f_m \end{bmatrix} \tag{3.17}$$

where, as usual, $\mathbf{x}$ is the homogeneous representation of the point $\boldsymbol{x} = [x_1, x_2]^T$.

The constraint given by (3.17) imposes the presence of the reflection point on an ellipse, but it does not involve any geometric primitive. If we are interested in posing a constraints on the line on which the reflector lies, it is convenient to consider the dual of the ellipse $\mathbf{C}_m$, which is given by the set of lines tangent it:

$$\mathbf{l}^T\mathbf{C}_m^{\diamond}\mathbf{l}=0 \ , \tag{3.18}$$

where $\mathbf{C}_m^{\diamond} = \det(\mathbf{C}_m)\mathbf{C}_m^{-1}$ is the line-conic matrix; and $\mathbf{l} = [l_1, l_2, l_3]^T$ is the homogeneous representation of a line tangent to the ellipse. This constraint now acts directly on the reflector line $\mathbf{l}_r = [l_{1_R}, l_{2_R}, l_{3_R}]^T$, which has to be a solution of (3.18).

3.2.2 Constraints related to TDOAs

In this Section we transform the measurement of the Time Difference of Arrival into a constraint that acts directly on the position of the source (localization problems) or on the position of the reflector (geometry-inference problem).

3.2.2.1 Direct signal

We consider the Time Difference of Arrival t_{pq} estimated from the lag of the maximum of the cross-correlation $R_{pq}(n)$, measured from the signals on two microphones located at $\boldsymbol{x}_p = [x_{1_p}, x_{2_p}]^T$ and $\boldsymbol{x}_q = [x_{1_q}, x_{2_q}]^T$. The knowledge of t_{pq} constrains the source to lie on an hyperbola whose foci are $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ and whose major axis is $l_{pq} = |t_{pq}|c$, as depicted in Figure 3.7. In fact, the hyperbola is defined as the locus

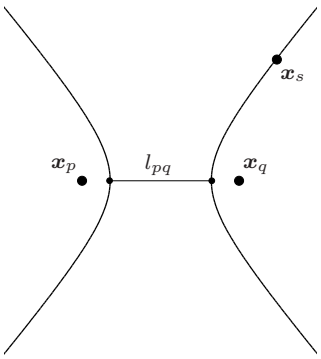


Figure 3.7: The TDOA t_{pq} measured on the microphone pair located at $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ constrains the acoustic source $\boldsymbol{x}_s$ to lie on a hyperbola with foci at the sensor positions and major axis $l_{pq} = |t_{pq}|c$.

of points where the difference of the distances from the two foci is constant. This means that

$$\sqrt{(x_1-x_{1_p})^2+(x_2-x_{2_p})^2}-\sqrt{(x_1-x_{1_q})^2+(x_2-x_{2_q})^2}=l_{pq}\ ,$$

or equivalently that

$$\sqrt{(x_1-x_{1_p})^2+(x_2-x_{2_p})^2}=l_{pq}+\sqrt{(x_1-x_{1_q})^2+(x_2-x_{2_q})^2}\ . \tag{3.19}$$

As for the ellipse, we take the square power of both sides of (3.19) and, after some calculation we obtain

$$2l_{pq}\sqrt{(x_1-x_{1_q})^2+(x_2-x_{2_q})^2}=-\left[l_{pq}^2+2(xx_{1_p}+xx_{2_p}-xx_{1_q}-xx_{2_q})+x_{1_q}^2+x_{2_q}^2-x_{1_p}^2-x_{2_p}^2\right]. \tag{3.20}$$

After taking again the square of both sides and expanding the terms, through comparison with the general conic equation (2.18) we obtain the parameter vector of the hyperbola $\mathbf{c}_{pq}=[a_{pq},b_{pq},c_{pq},d_{pq},e_{pq},f_{pq}]^T$, where

$$\begin{cases} a_{pq} &= -4[l_{pq}^2-(x_{1_q}-x_{1_p})^2] \\ b_{pq} &= 8(x_{1_q}-x_{1_p})(x_{2_q}-x_{2_p}) \\ c_{pq} &= -4[l_{pq}^2-(x_{2_q}-x_{2_p})^2] \\ d_{pq} &= 4[l_{pq}^2(x_{1_q}+x_{1_p})-(x_{1_q}-x_{1_p})(x_{1_q}^2+x_{2_q}^2-x_{1_p}^2-x_{2_p}^2)] \\ e_{pq} &= 4[l_{pq}^2(x_{2_q}+x_{2_p})-(x_{2_q}-x_{2_p})(x_{1_q}^2+x_{2_q}^2-x_{1_p}^2-x_{2_p}^2)] \\ f_{pq} &= l_{pq}^4-2l_{pq}^2(x_{1_q}^2+x_{2_p}^2+x_{1_q}^2+x_{2_p}^2)+(x_{1_q}^2+x_{2_q}^2-x_{1_p}^2-x_{2_p}^2)^2 \end{cases}. \tag{3.21}$$

Once again, the quadratic constraint can be expressed using homogeneous coordinates as

$$\mathbf{x}^T\mathbf{C}_{pq}\mathbf{x}=0\ ,\ \mathbf{C}_{pq}=\begin{bmatrix} a_{pq} & b_{pq}/2 & d_{pq}/2 \\ b_{pq}/2 & c_{pq} & e_{pq}/2 \\ d_{pq}/2 & e_{pq}/2 & f_{pq} \end{bmatrix} \tag{3.22}$$

where $\mathbf{x}=[x_1,x_2,1]^T$.

From comparison between (3.21) and (3.16), we observe that the structures of the hyperbola and ellipse parameter vectors are equivalent. The quantity that makes it possible to distinguish between the two different conics is the length of the major axis. In order to prove this fact, we consider a general non-degenerate conic with parameters $\mathbf{c}=[a,b,c,d,e,f]^T$. According to [13], the shape of the conic is determined by the discriminant $\Delta=b^2-4ac$. In particular, if $\Delta>0$ the conic is an ellipse; if $\Delta=0$ the conic is a parabola; if $\Delta<0$ the conic is an hyperbola. Computing the discriminant of the ellipse (3.16) and of the hyperbola (3.21) we obtain

$$\Delta_m=-16l_m^2[(x_{1_m}-x_{1_s})^2+(x_{2_m}-x_{2_s})^2-l_m^2] \tag{3.23}$$

and

$$\Delta_{pq} = -16l_{pq}^2[(x_{1_p} - x_{1_q})^2 + (x_{2_p} - x_{2_q})^2 - l_{pq}^2] \ , \tag{3.24}$$

respectively. It is clear that $\Delta_m > 0$ if and only if $l_m^2 > (x_{1_m} - x_{1_s})^2 + (x_{2_m} - x_{2_s})^2$, i.e. if the major axis l_m is greater than the distance between the foci $\boldsymbol{x}_m = [x_{1_m}, x_{2_m}]^T$ and $\boldsymbol{x}_s = [x_{1_s}, x_{2_s}]^T$. The constraint arising from the indirect TOA presented in Section 3.2.1 reflects this situation: the indirect path is always longer than the distance between the source and the microphone. On the other hand, $\Delta_{pq} < 0$ if and only if $l_{pq}^2 < (x_{1_p} - x_{1_q})^2 + (x_{2_p} - x_{2_q})^2$, i.e. when the major axis l_{pq} is smaller than the distance between the foci $\boldsymbol{x}_p = [x_{1_q}, x_{2_p}]^T$ and $\boldsymbol{x}_q = [x_{1_q}, x_{2_q}]^T$. Indeed, this is the case of the constraint on the TDOA: the time delay t_{pq} is always in the range $\left[-\frac{D_{pq}}{c}, \frac{D_{pq}}{c}\right]$, where D_{pq} is the distance between the sensors.

3.2.2.2 Indirect signal

Like for the case of TOAs, in this Section we consider a scenario in which a single reflector is present in the environment. This causes the presence of a reflective acoustic path that links source and microphones through a specular reflection. As a consequence, only two replicas of the source signal are present in the microphone signals $u_p(n)$ and $u_q(n)$. With reference to (3.9) $L = 1$ and therefore $R_{pq}(n)$ exhibits $(L+1)^2 = 4$ local maxima. We consider here the Time Difference of Arrival t'_{pq} that is directly related to the reflection, neglecting the presence of other local maxima in the cross-correlation. Figure 3.8 depicts the geometry of the problem. The source is located at $\boldsymbol{x}_s$, and the microphone locations are $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$. The point $\boldsymbol{x}_{s'}$ represents the image source, which is obtained by mirroring $\boldsymbol{x}_s$ over the reflector. Denoting with d_r the distance of the reflector line $\mathbf{l}_r$ from the origin, and letting $\boldsymbol{n}_r$ be the unit vector normal to the reflector and pointing towards the half-space containing $\boldsymbol{x}_s$, the image source position is obtained as

$$\mathbf{x}_{s'} = \mathbf{H}_r \mathbf{x}_s \ , \quad \mathbf{H}_r = \begin{bmatrix} \mathbf{I} - 2\boldsymbol{n}_r\boldsymbol{n}_r^T & -2d_r\boldsymbol{n}_r \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{3.25}$$

where $\mathbf{x}_s$ and $\mathbf{x}_{s'}$ are the homogeneous representations of $\boldsymbol{x}_s$ and $\boldsymbol{x}_{s'}$, respectively. As seen in Section 2.2.4.2, the homography $\mathbf{H}_r$ contains all the information about the reflector, since $\boldsymbol{n}_r$ and d_r univocally define the line parameter vector $\mathbf{l}_r = [l_{1_r}, l_{2_r}, l_{3_r}]^T$ associated to the reflector.

We now exploit the knowledge of the TDOA t'_{pq} relative to the indirect path, which

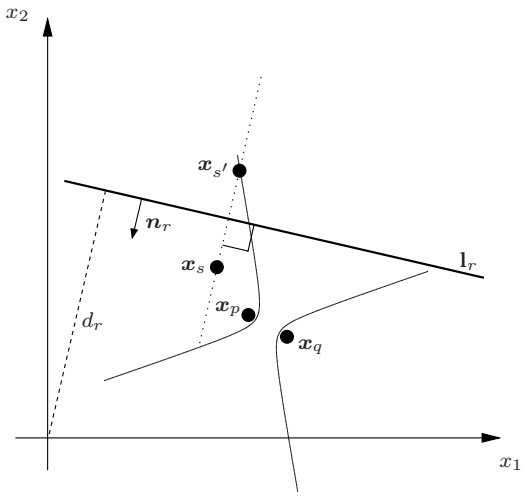


Figure 3.8: An acoustic path links $\boldsymbol{x}_s$ to $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ through the reflector $\mathbf{l}_r$. The reflective path can be thought as generated from the image source $\boldsymbol{x}_{s'}$, obtained by mirroring $\boldsymbol{x}_s$ over the reflector. The image source is bounded to lie on a hyperbola with foci in $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$.

constrains the image source $\boldsymbol{x}_{s'}$ to lie on a hyperbola with foci in $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ and with major axis $|t'_{pq}|c$. This is equivalent to writing

$$\mathbf{x}_{s'}^T \mathbf{C}'_{pq} \mathbf{x}_{s'} = 0 \ , \tag{3.26}$$

where the conic matrix $\mathbf{C}'_{pq}$ is obtained as in (3.22). Replacing (3.25) into (3.26) gives

$$\mathbf{x}_s^T \mathbf{H}_r^T \mathbf{C}'_{pq} \mathbf{H}_r \mathbf{x}_s = 0 \ . \tag{3.27}$$

We observe that if the source position $\boldsymbol{x}_s$ is known, the only unknowns in (3.27) are the distance d_r and the versor $\boldsymbol{n}_r$, embedded into the matrix $\mathbf{H}_r$.

3.2.3 Constraints from DOAs

In this Section we investigate the constraints related to Direction Of Arrival measurements on direct and reflected paths. In the following we will assume that the DOAs are measured using the algorithm presented in Section 3.1 by means of a circular array, whose geometry is given. We will refer to DOAs as lines linking the corresponding sources (real or image ones) and the center of the array $\boldsymbol{x}_0$.

3.2.3.1 Direct signal

A Direction Of Arrival can be seen as the line linking the corresponding source with the array reference point. For the wide-band Capon algorithm introduced in Section 3.1 the reference point coincides with the center of the circular array $\boldsymbol{x}_c = [x_{1_c}, x_{2_c}]^T$. Letting θ_s be the DOA of an acoustic source located at $\boldsymbol{x}_s$, the source is constrained to lie on the line

$$\mathbf{l}_s = [-\sin\theta_s,\ \cos\theta_s,\ x_{1_c}\sin\theta_s - x_{2_c}\cos\theta_s]^T\ , \tag{3.28}$$

which is the line passing through $\boldsymbol{x}_c$ and directed as θ_s. The source position is therefore a solution of

$$\mathbf{x}^T\mathbf{l}_s = 0\ ,$$

where $\mathbf{x}$ is a point on the line $\mathbf{l}_s$.

3.2.3.2 Indirect signal

We now consider the Direction Of Arrival relative to the reflective path originated by the presence of an acoustic source at $\boldsymbol{x}_s$ and of a single reflector lying on the line $\mathbf{l}_r$. As usual, we can think of the reflective path as generated by the image source $\boldsymbol{x}_{s'}$, obtained by mirroring the source $\boldsymbol{x}_s = [x_{1_s}, x_{2_s}]^T$ against the reflector. Therefore, the measured DOA $\theta_{s'}$ unequivocally defines the line $\mathbf{l}_{s'} = [-\sin\theta_{s'}, \cos\theta_{s'}, x_{1_c}\sin\theta_{s'} - x_{2_c}\cos\theta_{s'}]^T$ passing through the centre of the microphone array and with direction $\theta_{s'}$, as shown in Figure 3.9. We notice that if we make an hypothesis on the position of the image source $\boldsymbol{x}_{s'}$, the reflector line $\mathbf{l}_r$ is constrained to be the axis of the segment $\boldsymbol{x}_s\boldsymbol{x}_{s'}$ in order to honor the Snell's law. Let $\boldsymbol{x}_{ps}$ be the intersection between the line perpendicular to $\mathbf{l}_{s'}$ through $\boldsymbol{x}_{s'}$ and the reflector line $\mathbf{l}_r$. By construction the triangle $\boldsymbol{x}_s\boldsymbol{x}_{s'}\boldsymbol{x}_{ps}$ is isosceles and $\mathbf{l}_r$ is the bisector of the angle $\widehat{\boldsymbol{x}_s\boldsymbol{x}_{ps}\boldsymbol{x}_{s'}}$. We recall that the focal property of a parabola with focus $\boldsymbol{x}_s$ and directrix $\mathbf{l}_{s'}$ states that the line tangent to a parabola at $\boldsymbol{x}_{ps}$ is also the bisector of the angle formed by the line joining $\boldsymbol{x}_s$ and $\boldsymbol{x}_{ps}$ and the perpendicular to $\mathbf{l}_{s'}$ through $\boldsymbol{x}_{ps}$, as shown in Figure 3.9.

The parabola is also described as the locus of points equidistant from $\mathbf{l}_{s'}$ and from $\boldsymbol{x}_s$, which is equivalent to writing

$$|-x_1\sin\theta_{s'} + x_2\cos\theta_{s'} + x_{1_c}\sin\theta_{s'} - x_{2_c}\cos\theta_{s'}| = \sqrt{(x_{1_s}-x_1)^2 + (x_{2_s}-x_2)^2}, \tag{3.29}$$

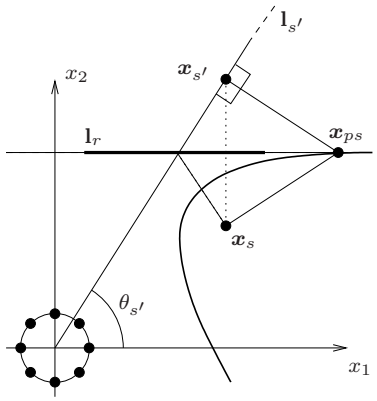


Figure 3.9: An acoustic source at $\boldsymbol{x}_s$ is mirrored over the reflector to its image position $\boldsymbol{x}_{s'}$. A sensor array centered in the origin measures the DOA $\theta_{s'}$ related to the reflected path. The reflector line $\mathbf{l}_r$ is the tangent at $\boldsymbol{x}_{ps}$ of a parabola with directrix $\mathbf{l}_{s'}$ and focus at $\boldsymbol{x}_s$.

where the left side is the Euclidean distance between the generic point $\boldsymbol{x} = [x_1, x_2]^T$ and the directrix; the right side corresponds to the Euclidean distance between $\boldsymbol{x}$ and the focus. Squaring both the sides and expanding the terms, through comparison with the general conic equation (2.18) we obtain the parameter vector $\mathbf{c}_{s'} = [a_{s'}, b_{s'}, c_{s'}, d_{s'}, e_{s'}, f_{s'}]$, where

$$\begin{cases} a_{s'} &= \cos^2\theta s' \\ b_{s'} &= 2\cos\theta_{s'}\sin\theta s' \\ c_{s'} &= \sin^2\theta s' \\ d_{s'} &= -2x_{1_s} + 2\sin\theta_{s'}(x_{1_c}\sin\theta_{s'} - x_{2_c}\cos\theta_{s'}) \\ e_{s'} &= -2x_{1_s} - 2\cos\theta_{s'}(x_{1_c}\sin\theta_{s'} - x_{2_c}\cos\theta_{s'}) \\ f_{s'} &= x_{1_s}^2 + x_{2_s}^2 - (x_{1_c}\sin\theta_{s'} - x_{2_c}\cos\theta_{s'})^2 \end{cases} . \tag{3.30}$$

The homogeneous representation of the parabolic constraint is therefore given by

$$\mathbf{x}^T\mathbf{C}_{s'}\mathbf{x} = 0\ ,\ \mathbf{C}_{s'} = \begin{bmatrix} a_{s'} & b_{s'}/2 & d_{s'}/2 \\ b_{s'}/2 & c_{s'} & e_{s'}/2 \\ d_{s'}/2 & e_{s'}/2 & f_{s'} \end{bmatrix} \tag{3.31}$$

where $\mathbf{x} = [\boldsymbol{x}, 1]^T$.

As for the case of the indirect TOA discussed in Section 3.2.1, it is convenient to formulate the constraint given by (3.31) in terms of the reflector line $\mathbf{l}_r$. To do so,

we consider the dual representation of the conic given by

$$\mathbf{l}^T \mathbf{C}^{\diamond}_{s'} \mathbf{l} = 0 \ , \ \mathbf{C}^{\diamond}_{s'} = \det(\mathbf{C}_{s'}) \mathbf{C}^{-1}_{s'} \ , \tag{3.32}$$

whose solutions are all the lines $\mathbf{l}$ tangent to the parabola with focus in $\boldsymbol{x}_s$ and directrix $\mathbf{l}_{s'}$. The reflector line $\mathbf{l}_r$ is one of the infinite solutions of 3.32.

3.3 From constraints to estimation

The projective relationships derived in the previous Section provide a complete theoretical framework for turning acoustic measurements into constraints that act directly on geometric acoustic primitives such as sources, microphones and reflectors. In this Section we exploit the common formulation of the derived constraints. In particular, we consider constraints that are described by a conic section, which are characterized by the same quadratic form. This way, it is possible to abstract from a specific problem (estimation of source or microphone positions, as well as the position of a reflector) for defining a general estimation problem relying on the minimization of a cost function that combines multiple quadratic constraints. The global minimum of the cost function gives an estimate of the geometric primitive of interest, being either a point (source or microphone position) or a line (reflector position). Moreover, we will develop a set of minimization strategies which are not strictly problem-dependent, but whose performances and feasibility depend on the structure of the conic matrices.

3.3.1 Definition of the cost function

Let the geometric primitive under estimation (position of a source, a sensor or line parameters of a reflector) be denoted by $\mathbf{v} = [v_1, v_2, v_3]^T$. The matrix associated to the constraint (it can be $\mathbf{C}$ or $\mathbf{C}^\diamond$) is denoted by the symbol $\mathbf{G}$. The generic constraint, therefore, has the form

$$\mathbf{v}^T\mathbf{G}\mathbf{v} = 0\ . \tag{3.33}$$

When multiple measurements concerning the same variable of interest are considered, we can organize the constraints in the following system equation

$$\begin{cases} \mathbf{v}^T\mathbf{G}_1\mathbf{v} & = 0 \\ \mathbf{v}^T\mathbf{G}_2\mathbf{v} & = 0 \\ \quad\vdots & \\ \mathbf{v}^T\mathbf{G}_N\mathbf{v} & = 0 \end{cases}, \tag{3.34}$$

where $\mathbf{G}_1 \dots \mathbf{G}_N$ are the conic matrices associated to N measurements. However, acoustic measurements are affected by different kinds of error, due as an example to outliers or the finite sampling frequency. Therefore, it is convenient to define a cost

function that sums the squared residuals of the individual constraints:

$$J(\mathbf{v}) = \sum_{n=1}^{N} \left(\mathbf{v}^T \mathbf{G}_n \mathbf{v}\right)^2 . \tag{3.35}$$

The minimization of this function is not, in general, an easy task, since $J(\mathbf{v})$ is a multivariate polynomial of degree 4 which presents a huge number of local minima. Furthermore, since the function is built from projective constraints, an infinite number of solutions correspond to its global minimum. Finally, a constraint has to be posed on the norm of the searched solution, in order to prevent us from finding the trivial null solution $\mathbf{v}_0 = [0, 0, 0]^T$. In the next Section we will present a set of methodologies for addressing these problems.

3.3.2 Minimization strategies

The cost function (3.35) introduced above is obtained as the squared sum of a number of homogeneous quadratic forms. As a consequence, $J(\mathbf{v})$ turns out to be a multivariate polynomial of order 4 in the variables v_1, v_2 and v_3. The minimization of multivariate polynomials is a very debated topic in the literature, see for example [49, 50, 51]. Exact methods exist [52, 53] for the unconstrained case, however they are not applicable to our kind of problem, which is constrained. Other minimization techniques [54, 55, 56, 57] rely on a semidefinite relaxation [58] of the original problem, which is suitable for the constrained optimization. Unfortunately, such methods provide only approximated solutions, which can by highly inaccurate [59]. In this Section we propose a set of alternate approaches for minimizing an homogeneous cost function in the form of (3.35). In particular, we will introduce two exact methods and two iterative ones.

3.3.2.1 Exact minimization through cutting of the parameter space

As stated before, the cost function in (3.35) is a multivariate polynomial of order 4 in v_1, v_2, v_3. Denoting the coefficients of the conic associated to the nth measurement

with $\mathbf{c}_n = [a_n\ b_n\ c_n\ d_n\ e_n\ f_n]^T$, the cost function can be expanded as

$$\begin{aligned} J(\mathbf{v}) = \sum_{n=1}^{N} & [a_n^2 v_1^4 + c_n^2 v_2^4 + f_n^4 v_3^4 + 2a_n b_n v_1^3 v_2 + 2a_n d_n v_1^3 v_3 + 2b_n c_n v_1 v_2^3 + \\ & + 2c_n e_n v_2^3 v_3 + 2d_n f_n v_1 v_3^3 + 2e_n f_n v_2 v_3^3 + (2a_n c_n + b_n^2) v_1^2 v_2^2 + \\ & + (2a_n f_n + d_n^2) v_1^2 v_3^2 + (2c_n f_n + e_n^2) v_2^2 v_3^2 + 2(a_n e_n + b_n d_n) v_1^2 v_2 v_3 + \\ & + 2(b_n e_n + c_n d_n) v_1 v_2^2 v_3 + 2(b_n f_n + d_n e_n) v_1 v_2 v_3^2] \ . \end{aligned} \tag{3.36}$$

The vector $\mathbf{v}$ is homogeneous (i.e. it is defined up to a scaling factor), and therefore the direct minimization of (3.36) leads to an infinite number of solutions representing the same point. The scaling ambiguity can be solved by noticing that $\mathbf{v}$ can be defined by the ratio of its elements. This corresponds to cutting the parameter space with a plane orthogonal to one of the parameter axes, e.g. $v_i = 1$, where $i \in \{1, 2, 3\}$. The cost function can therefore be redefined as

$$J_i(\mathbf{v}) \triangleq J(\mathbf{v})|_{v_i=1} \ , \qquad i \in \{1, 2, 3\} \ . \tag{3.37}$$

We observe that the plane $v_i = 1$ excludes potential solutions having $v_i = 0$. This situation can be circumvented by cutting the parameter space with a further plane $v_j = 1,\ j \neq i$, which leads to a second cost function $J_j(\mathbf{v})$. The reduced cost functions $J_i(\mathbf{v})$ and $J_j(\mathbf{v})$ have the form of non-homogeneous fourth-order polynomial in two variables. As an example, the unknowns in $J_i(\mathbf{v})$ are v_l and v_k, where $l, k \in \{1, 2, 3\} \setminus \{i\},\ l \neq k$. In order to find the global minimum of $J(\mathbf{v})$, we define two unconstrained minimization problems

$$\begin{aligned} \hat{\mathbf{v}}_i &= \arg\min_{\mathbf{v}} J_i(\mathbf{v}) \\ \hat{\mathbf{v}}_j &= \arg\min_{\mathbf{v}} J_j(\mathbf{v}) \end{aligned} \ , \tag{3.38}$$

whose solutions $\hat{\mathbf{v}}_i$ and $\hat{\mathbf{v}}_j$ make null the gradients of $J_i(\mathbf{v})$ and $J_j(\mathbf{v})$, respectively. In mathematical terms, this means that $\hat{\mathbf{v}}_i$ is contained in the set

$$V_i = \left\{ \mathbf{v} : \frac{\partial J_i(\mathbf{v})}{\partial v_l} = 0 \ \wedge \ \frac{\partial J_i(\mathbf{v})}{\partial v_k} = 0 \right\} \ .$$

Similarly, $\hat{\mathbf{v}}_j$ is contained in

$$V_j = \left\{ \mathbf{v} : \frac{\partial J_j(\mathbf{v})}{\partial v_l} = 0 \ \ \wedge \ \ \frac{\partial J_j(\mathbf{v})}{\partial v_k} = 0 \right\} \ .$$

Since the gradient of $J_i(\mathbf{v})$ ($J_j(\mathbf{v})$) is a vector of two bivariate polynomials of order 3, V_i (V_j) contains 9 elements. Some of them may be imaginary, thus we denote with $\bar{V}_i$ ($\bar{V}_j$) the subset of purely real solutions of V_i (V_j). In the light of the above considerations, the global minimum of $J(\mathbf{v})$ is finally given by

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} J(\mathbf{v}) \ , \ \mathbf{v} \in \bar{V}_i \cup \bar{V}_j \ . \tag{3.39}$$

The choice of the planes used for cutting the parameter space depends on the meaning of the vector $\mathbf{v}$, i.e. on the type of primitive it represents. As an example, when $\mathbf{v}$ is the homogeneous representation of an Euclidean point (suitable for sources and microphones), it is sufficient to minimize a single cost function in the form of (3.37) by posing $v_3 = 1$. In this case we are not interested to points having $v_3 = 0$, as they represent points at infinity. When $\mathbf{v}$ corresponds, instead, to a line parameter vector (a reflector line), without any a-priori information on the position of the line to be estimated, it is necessary to select two planes (e.g. $v_1 = 1$ and $v_2 = 1$), and solve a problem in the form of (3.39).

The proposed methodology guarantees to find the exact solution of the minimization problem. Its main drawback, however, is the fact that its implementation requires the use of a symbolic mathematical solver. Since the computation time may be considerable, this solution is suitable only when the estimation time is not crucial (e.g. for array calibration or reflector estimation). For applications where real-time is a requirement, such as source localization, more efficient solutions must be applied, such as the iterative one described in the next paragraph.

3.3.2.2 Iterative minimization on the unit sphere

Here we tackle the problem of the minimization of the cost function using a different reduction of the parameter space. Instead of cutting the parameter space with one or more planes, we formulate a single optimization problem which looks for solutions lying on the unit sphere, i.e.

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} J(\mathbf{v}) \ \text{ subject to } \|\mathbf{v}\| = 1 \ . \tag{3.40}$$

As depicted in Figure 3.10, a point $\mathbf{v} = [v_1, v_2, v_3]$ on the unit sphere is described by two parameters, namely the azimuth angle ϕ and the elevation θ, while the radius is kept fixed to $\rho = 1$.

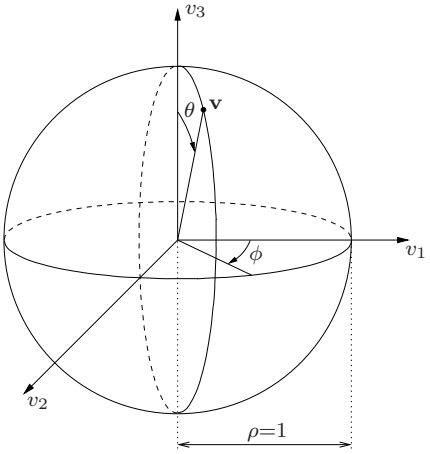


Figure 3.10: Representation of points on the unit sphere.

The conversion to spherical coordinates is given by

$$\begin{cases} v_1 &= \cos\phi \sin\theta \\ v_2 &= \sin\phi \sin\theta \\ v_3 &= \cos\theta \end{cases} . \tag{3.41}$$

This means that the parameter space reduces to $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi]$. It represents a bounded search space, which reveals to be particularly suitable for iterative minimization algorithms, such as the non-linear least squares algorithm described in [60]. The constrained minimization problem in (3.40) can be reformulated as an unconstrained one:

$$(\hat{\phi}, \hat{\theta}) = \underset{\phi\in[0,2\pi],\ \theta\in[0,\pi]}{\arg\min} \; J\left(\mathbf{v}(\phi, \theta)\right) \ ,$$

whose solution, in the projective domain, corresponds to the vector

$$\hat{\mathbf{v}} = [\cos\hat{\phi}\sin\hat{\theta},\ \sin\hat{\phi}\sin\hat{\theta},\ \cos\hat{\theta}]^T \ .$$

This method could be sub-optimal in some cases and it could get trapped into local minima. However, we will see that, if properly initialized, it leads to accurate solutions and it can be adopted for source localization purposes.

3.3.2.3 Iterative minimization on the unit cylinder

As an alternative to the minimization on the unit sphere we can consider the optimization on the unit cylinder. Eq.(3.40) is modified as follows

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} J(\mathbf{v}) \;\text{ subject to } \|\mathbf{v}_{1,2}\| = 1 \;, \tag{3.42}$$

where $\mathbf{v}_{1,2} \triangleq [v_1, v_2]^T$ denotes the first two coordinates of $\mathbf{v} = [v_1, v_2, v_3]^T$. The constraint on $\mathbf{v}_{1,2} = [v_1, v_2]^T$ forces $\mathbf{v}$ to lie on the unit cylinder, in fact

$$\begin{cases} v_1 &= \cos\phi \\ v_2 &= \sin\phi \\ v_3 &= z \end{cases} ,$$

where $\phi \in [0, 2\pi]$ denotes the azimuth angle; and $z \in \mathbb{R}$ is the signed distance from the plane $v_3 = 0$, as shown in Figure 3.11.

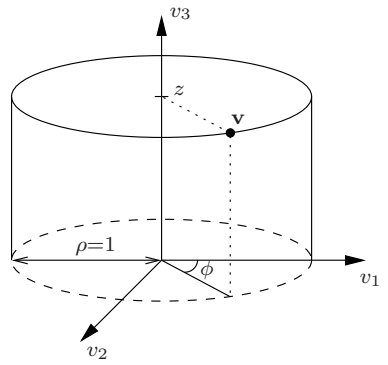


Figure 3.11: Representation of points on the unit cylinder.

The unconstrained version of the problem in (3.42) is given by

$$(\hat{\phi}, \hat{z}) = \arg\min_{\phi\in[0,2\pi],\; z\in\mathbb{R}} J\left(\mathbf{v}(\phi, \mathbf{z})\right) \;,$$

which leads to the final solution

$$\hat{\mathbf{v}} = [\cos\hat{\phi},\; \sin\hat{\phi},\; \hat{z}]^T.$$

Notice that this methods is suitable especially when the primitive to be estimated

is a line: in this case, the parameters ϕ and z have a direct geometrical interpretation. More specifically, the angle ϕ corresponds to the orientation of the normal of the line; $|z|$ coincides with the distance of the line from the origin.

3.3.2.4 Exact minimization using the GTRS approach

In this paragraph we discuss about the possibility of turning the minimization of the generic cost function (3.35) into a linear Least Square problem with a single quadratic constraints. As noted in [59], problems of these type are referred to as *Generalized Trust Region Subproblems* (GTRS), whose exact solution can be derived quite efficiently. We will see that, whenever the cost function combines constraints having the form of a circumference, the estimation problem can be brought to a GTRS one. It is important to notice that this approach is not limited to the case of direct TOA measurements described in Section 3.2.1, which directly leads to circumferences. Indeed, we will see that, under suitable conditions, the dual representations of ellipses, parabolas, and hyperbolas can be brought to circumferences.

We start our discussion by observing that a generic conic parameter vector $\mathbf{c}_{\mathrm{C}} = [a_{\mathrm{C}}, b_{\mathrm{C}}, c_{\mathrm{C}}, d_{\mathrm{C}}, e_{\mathrm{C}}, f_{\mathrm{C}}]^T$ represents a circumference if and only if $a_{\mathrm{C}} = c_{\mathrm{C}}$ and $b_{\mathrm{C}} = 0$ [13], i.e. when the conic matrix has the form

$$\mathbf{C}_{\mathrm{C}} = \begin{bmatrix} a_{\mathrm{C}} & 0 & d_{\mathrm{C}}/2 \\ 0 & a_{\mathrm{C}} & e_{\mathrm{C}}/2 \\ d_{\mathrm{C}}/2 & e_{\mathrm{C}}/2 & f_{\mathrm{C}} \end{bmatrix} . \tag{3.43}$$

As seen in Section 3.2, this applies for TOAs related to the direct path, which bound a source (microphone) to lie on circumference centered at the microphone (source) position.

We now investigate how other type of conics can be brought to a circumference in the dual domain. We first consider the conic matrix $\mathbf{C}_{\mathrm{E}}$ of an ellipse, whose parameters are given in (3.16). Computing the corresponding dual matrix we obtain

$$\mathbf{C}_{\mathrm{E}}^{\diamond} = \det \mathbf{C}_{\mathrm{E}}^{\diamond} (\mathbf{C}_{\mathrm{E}}^{\diamond})^{-1} = \begin{bmatrix} a_{\mathrm{E}}^{\diamond} & b_{\mathrm{E}}^{\diamond}/2 & d_{\mathrm{E}}^{\diamond}/2 \\ b_{\mathrm{E}}^{\diamond}/2 & a_{\mathrm{E}}^{\diamond} & e_{\mathrm{E}}^{\diamond}/2 \\ d_{\mathrm{E}}^{\diamond}/2 & e_{\mathrm{E}}^{\diamond}/2 & f_{\mathrm{E}}^{\diamond} \end{bmatrix} . \tag{3.44}$$

The coefficients of $\mathbf{C}_{\mathrm{E}}^{\diamond}$ are given by

$$\begin{cases} a_{\mathrm{E}}^{\diamond} &= k_{\mathrm{E}}[-l_m^2+(x_{1_m}+x_{1_s})^2+(x_{2_m}-x_{2_s})^2] \\ b_{\mathrm{E}}^{\diamond} &= 4k_{\mathrm{E}}(x_{1_m}x_{2_s}+x_{1_s}x_{2_m}) \\ c_{\mathrm{E}}^{\diamond} &= k_{\mathrm{E}}[-l_m^2+(x_{1_m}-x_{1_s})^2+(x_{2_m}+x_{2_s})^2] \\ d_{\mathrm{E}}^{\diamond} &= 4k_{\mathrm{E}}(x_{1_m}+x_{1_s}) \\ e_{\mathrm{E}}^{\diamond} &= 4k_{\mathrm{E}}(x_{2_m}+x_{2_s}) \\ f_{\mathrm{E}}^{\diamond} &= 4k_{\mathrm{E}} \end{cases} , \tag{3.45}$$

where

$$k_{\mathrm{E}}=-4l_m^2[-l_m^2+(x_{1_m}-x_{1_s})^2+(x_{2_m}-x_{2_s})^2] \, .$$

We recall that (x_{1_m}, x_{2_m}) and (x_{1_s}, x_{2_s}) are the Euclidean coordinates of the microphone and the source positions, respectively, and they represent the foci of an ellipse with major axis l_m. We observe that, if either the microphone or the source are located at the origin of the reference system (i.e., $[x_{1_m}, x_{2_m}] = [0, 0]$ or $[x_{1_s}, x_{2_s}] = [0, 0]$), the parameters in (3.45) simplify in such a way that $a_{\mathrm{E}}^{\diamond} = c_{\mathrm{E}}^{\diamond}$ and $b_{\mathrm{E}}^{\diamond} = 0$. This means that, when one of the two foci of the ellipse coincides with the origin, the dual conic is a circumference and the corresponding matrix $\mathbf{C}_{\mathrm{E}}^{\diamond}$ has the form of $\mathbf{C}_{\mathrm{C}}$ in (3.43). Notice that this property holds also for the case of the hyperbola, whose parameter vector is equivalent to that of the ellipse (see Section 3.2). For the sake of completeness, we can prove that also the dual of a parabola can be brought to a circumference. The conic matrix $\mathbf{C}_{\mathrm{P}}$ of a parabola is determined by the parameters in (3.30) and the corresponding dual matrix is

$$\mathbf{C}_{\mathrm{P}}^{\diamond} = \det \mathbf{C}_{\mathrm{P}}^{\diamond} (\mathbf{C}_{\mathrm{P}}^{\diamond})^{-1} = \begin{bmatrix} a_{\mathrm{P}}^{\diamond} & b_{\mathrm{P}}^{\diamond}/2 & d_{\mathrm{P}}^{\diamond}/2 \\ b_{\mathrm{P}}^{\diamond}/2 & a_{\mathrm{P}}^{\diamond} & e_{\mathrm{P}}^{\diamond}/2 \\ d_{\mathrm{P}}^{\diamond}/2 & e_{\mathrm{P}}^{\diamond}/2 & f_{\mathrm{P}}^{\diamond} \end{bmatrix} , \tag{3.46}$$

where

$$\begin{cases} a_{\mathrm{P}}^{\diamond} &= k_{\mathrm{P}}[-(x_{1_s}+x_{1_c})\sin\theta_{s'}-(x_{2_s}-x_{2_c})\cos\theta_{s'}] \\ b_{\mathrm{P}}^{\diamond} &= 2k_{\mathrm{P}}(x_{1_s}\cos\theta_{s'}-x_{2_s}\sin\theta_{s'}) \\ c_{\mathrm{P}}^{\diamond} &= k_{\mathrm{P}}[(x_{1_s}-x_{1_c})\sin\theta_{s'}+(x_{2_s}+x_{2_c})\cos\theta_{s'}] \\ d_{\mathrm{P}}^{\diamond} &= -2k_{\mathrm{P}}\sin\theta_{s'} \\ e_{\mathrm{P}}^{\diamond} &= 2k_{\mathrm{P}}\cos\theta_{s'} \\ f_{\mathrm{P}}^{\diamond} &= 0 \end{cases} \tag{3.47}$$

and

$$k_{\mathrm{P}} = -(x_{1_s} - x_{1_c}) \sin\theta_{s'} + (x_{2_s} - x_{2_c}) \cos\theta_{s'} \ .$$

With reference to (3.30), the pair (x_{1_s}, x_{2_s}) refers to the Euclidean coordinates of the source position, which corresponds to the focus of the parabola. It is clear that we obtain a simplification in the parameters of the dual parabola (3.47), under the assumption that the focus coincides with the origin of the reference system (i.e. $[x_{1_s}, x_{2_s}] = [0, 0]$). In particular, once again we get $a_{\mathrm{P}}^{\diamond} = c_{\mathrm{P}}^{\diamond}$ and $b_{\mathrm{P}}^{\diamond} = 0$. Hence, the dual of a parabola with the focus at the origin is a circumference and therefore $\mathbf{C}_{\mathrm{P}}^{\diamond}$ has the same structure of $\mathbf{C}_{\mathrm{C}}$ in (3.43).

We now exploit the particular form of a constraint described by a circumference. The conic matrix related to the nth measurement has the form

$$\mathbf{C}_n = \begin{bmatrix} a_n & 0 & d_n/2 \\ 0 & c_n & e_n/2 \\ d_n/2 & e_n/2 & f_n \end{bmatrix} \ . \tag{3.48}$$

Inserting (3.48) into (3.35), after some manipulations we obtain the following form of the cost function

$$J(\mathbf{v}) = \sum_{n=1}^{N} \left[a_n(v_1^2 + v_2^2) + d_n v_1 v_3 + e_n v_2 v_3 + f_n v_3^2 \right]^2 \ . \tag{3.49}$$

We now perform a cutting of the parameter space with the plane $v_3 = 1$, yielding the modified version of the cost function

$$J_3(\mathbf{v}) \triangleq J(\mathbf{v})|_{v_3=1} = \sum_{n=1}^{N} \left[a_n(v_1^2 + v_2^2) + d_n v_1 + e_n v_2 + f_n \right]^2 \ .$$

Notice that this operation does not exclude meaningful solutions having $v_3 = 0$. In fact, if the primitive to be estimated is a microphone or source position, $v_3 = 0$ represents points at infinity, which are meaningless if all the acoustic objects are at finite distances. Conversely, if we aim at estimating a reflector line parameters, $v_3 = 0$ denotes lines passing through the origin. It is evident that there are no lines passing through one of the foci of the conic (i.e., the origin) and tangent to it. In the light of the previous considerations, we can therefore define the optimization

problem as

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} J_3(\mathbf{v}) \ , \tag{3.50}$$

whose solution $\hat{\mathbf{v}}$ represents the best estimate of the primitive of interest. Following the same approach proposed in [59], the optimization problem (3.50) can be reformulated as a quadratically constrained linear Least-Squares one. More specifically, introducing the simple substitution $w = v_1^2 + v_2^2$ and defining a vector of unknowns $\mathbf{w} = [w, v_1, v_2]$, (3.50) can be rewritten in matrix form as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left\{ \|\mathbf{A}\mathbf{w} - \mathbf{b}\|^2 : \mathbf{w}^T\mathbf{D}\mathbf{w} + 2\mathbf{f}^T\mathbf{w} = 0 \right\} \ , \tag{3.51}$$

where

$$\mathbf{A} = \begin{bmatrix} a_1 & d_1 & e_1 \\ \vdots & \vdots & \vdots \\ a_N & d_N & e_N \end{bmatrix} \ , \ \mathbf{b} = \begin{bmatrix} -f_1 \\ \vdots \\ -f_N \end{bmatrix}$$

and

$$\mathbf{D} = \mathrm{diag}(0, 1, 1) \ , \ \mathbf{f} = [-0.5 \ \ 0 \ \ 0]^T \ .$$

Assuming that $\mathbf{A}$ has full column rank, the minimum is found as

$$\hat{\mathbf{w}}(\lambda) = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D})^{-1}(\mathbf{A}^T\mathbf{b} - \lambda\mathbf{f}) \ ,$$

where λ is the unique solution of $\hat{\mathbf{w}}(\lambda)^T\mathbf{D}\hat{\mathbf{w}}(\lambda) + 2\mathbf{f}^T\hat{\mathbf{w}}(\lambda) = 0$ on the interval for which $\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}$ is positive definite [59]. From the optimum value $\hat{\mathbf{w}} = [\hat{w} \ \hat{v}_1 \ \hat{v}_2]^T$, the solution to the original problem in (3.50) is finally given by $\hat{\mathbf{v}} = [\hat{v}_1 \ \hat{v}_2 \ 1]^T$.

3.3.3 Discussion

The minimization strategies described in the previous paragraphs aim at minimizing a cost function that sums up a number of homogeneous quadratic constraints. Except for the GTRS approach (Section 3.3.2.4), which requires the conic matrices to have a specific form, all the strategies can be applied to any estimation problem that can be brought to the general form given by (3.35). In particular, all of them are expected to converge to the same result. Nevertheless, the peculiarities of each of the approaches make them best suited for specific (class of) problems. A comparison of the minimization strategies is shown in Table 3.1, which summarizes the pros and cons, and provides some hints for their application.

Table 3.1: A comparison of the minimization strategies

Min. strategy	**Pros**	**Cons**	**Suggested application**
Par. space cutting (sec. 3.3.2.1)	Exact	Time demanding	Array calibration, reflector localization
Unit sphere (sec. 3.3.2.2)	Fast convergence	Requires initialization	Source localization using TDOAs
Unit cylinder (sec. 3.3.2.3)	Fast convergence, direct geometrical interpretation	Requires initialization	Reflector localization
GTRS (sec. 3.3.2.4)	Exact, fast convergence	Requires conic matrices to have a specific form	Array calibration using TOAs, reflector localization using TOAs

In the next Section the different minimization strategies are applied within the context of acoustic scene analysis. In particular, the class of problems indicated in Table 3.1 are discussed and solved.

3.4 Acoustic scene analysis

With the term *acoustic scene analysis* we refer to a set of operations that aim at building a self-aware and environment-aware multichannel input/output audio system. To this end, we can identify the following sequence of operations:

1. array calibration,
2. source localization and characterization,
3. environment inference.

The self-awareness of the system is reached when all the microphone and loudspeaker arrays/clusters composing the audio system are completely characterized, i.e. when they are calibrated. For example, the microphones within an array are characterized by a position, a polar pattern, and, possibly, an orientation if the pattern is not omnidirectional. Loudspeaker arrays need a similar characterization. The environment-awareness is a more challenging task, since it aims at identifying and characterizing the acoustic entities that may interact with the audio system, such as acoustic sources, or obstacles that cause reverberations to appear (environment inference). The degree of awareness of the system with respect to the environment depends on the level of detail in describing such acoustic entities. As an example, an acoustic source can be fixed at a certain position, or moves following a trajectory; it can also present a proper radiation pattern. Furthermore, an acoustic obstacle, or reflector, is characterized by its shape and its reflective properties, which can vary with the frequency.

In this Section we limit our discussion to propose some solutions for acoustic scene analysis, putting the geometric constraints derived in Section 3.2 at good use for a number of estimation problems. Notice that not all of the constraints presented in Section 3.2 can be easily implemented in practice. This is the case of the constraint associated to the TDOAs relative to the indirect path. Although algorithms for disambiguating TDOA estimates in multi-path and multi-source environments have been proposed [61, 62], its application to reflective TDOAs requires high computation time, especially when multiple reflectors are present. Moreover, in our discussion we will not consider constraints related to DOAs associated to the direct path. The simplicity of such constraints (lines on which sources are bound to lie) makes them less interesting than other type of constraints. Moreover, source localization through

the triangulation of multiple DOAs is a common approach in the literature [63, 64], and it has been widely studied.

In the next paragraphs we devise algorithms for array calibration (in terms of position of the elements) from TOA measurements; source localization from TDOA measurements; and reflector localization from TOA and DOA measurements.

3.4.1 Array calibration

The location and pose of the microphone/loudspeaker arrays with respect to the environment and the position of the elements within each array need to be available or they need to be somehow estimated. This problem is referred in the literature as self-calibration. Calibration is a fairly common operation in 3D vision, as it is a preliminary step to any application that uses multiple cameras to extract 3D geometric information from the imaged scene [11]. Only in recent times the multi-channel audio processing community started investigating the problem of self-calibration of microphone arrays. We can categorize the existing algorithms of self-calibration in two classes: we refer to intra-calibration and inter-calibration techniques.

As far as intra-calibration is concerned, the goal is to accurately localize elements within the same array and it is required when space-time processing algorithms cannot rely on approximate measurements of the locations of the sensors. In [65] the authors approach the problem of source localization with non-calibrated microphone arrays with a two-step iterative algorithm: at alternate steps source and microphone locations are localized using the knowledge acquired at the previous step. In [66] the authors tackle the intra-calibration problem using the Multi-Dimensional Scaling (MDS) [67] framework. More specifically, given the tape measures of the distances between all possible pair of sensors, the locations of microphones are estimated with respect to a reference one. In [68] the authors use the same approach but the measurement of the distances between sensors is accomplished through the analysis of the complex coherence function between microphones when diffuse noise field is present in the environment. As a result, the calibration procedure turns out to be greatly simplified with respect to [66].

Inter-calibration is required when multiple arrays are present in the acoustic scene and the knowledge of their mutual positions is needed. Some solutions in this direction have recently appeared in the literature. In [69] the authors address the inter-calibration problem by jointly localizing with two microphone arrays an acous-

tic source. The mutual positions are found with tools borrowed by computer vision that jointly analyze the localization results for several locations of the acoustic source.

In [70] the authors combine intra and inter-calibrations using a hierarchical approach: first each array is "intra-calibrated" using diffuse-noise field as input signal, then arrays are "inter-calibrated" with an approach similar to [69]. We notice, however, that intra-calibration and inter-calibration need two different procedures, as different input signals are used in the two phases.

In this Section we describe a novel solution for the intra-calibration. Moreover, we will introduce the concept of array global calibration which refers to the case of estimating the pose of an array whose geometry is known in advance. Finally, inter-calibration of multiple arrays is achieved by extending the global calibration technique. The input measurements are the Times Of Arrival of signals emitted by an acoustic source located at multiple positions. Notice that the same procedure can be applied for the self-calibration of loudspeaker arrays, simply by switching the roles of emitters and sensors.

3.4.1.1 Intra-calibration using TOAs

In this paragraph we discuss the problem of determining the position of an array of microphones with respect to a calibration pattern (i.e. a set of sources in known positions) using the constraints derived from the measurement of Times Of Arrival. We refer to the case of intra-calibration, which means that the geometry of the array is unknown. As a consequence, we are bound to estimate the location of each microphone separately.

The scenario of the self-calibration problem is depicted in Figure 3.12. Sources in given positions $\boldsymbol{x}_{s_1}, \ldots, \boldsymbol{x}_{s_N}$ produce a synchronized and known signal. TOAs t_{mn} related to the direct path are extracted for each receiver positions $\boldsymbol{x}_{r_1}, \ldots, \boldsymbol{x}_{r_M}$. Our goal is to infer the receiver positions from the knowledge of t_{mn}. In order to do so, we recall from Section 3.2.1 that each measurement imposes a quadratic constraint on the receiver position in the form

$$\mathbf{x}^T \mathbf{C}_{mn} \mathbf{x} = 0 \,, \tag{3.52}$$

where $\mathbf{C}_{mn}$ is a conic matrix representing a circumference centered at $\boldsymbol{x}_{s_n}$ with radius equal to $t_{mn}c$; and $\mathbf{x}$ is the homogeneous representation of a point lying on the circumference. If we consider all constraints related to the mth microphone at

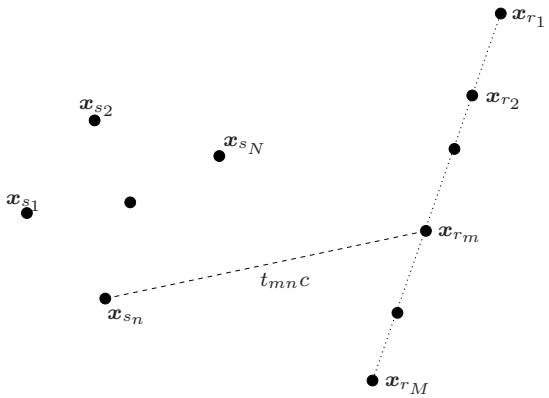


Figure 3.12: Sources in known positions $\boldsymbol{x}_{s_1}, \ldots, \boldsymbol{x}_{s_N}$ emit a known signal. Times Of Arrival t_{mn} related to the direct path are extracted for each receiver positions $\boldsymbol{x}_{r_1}, \ldots, \boldsymbol{x}_{r_M}$.

the same time we can collect them to obtain a cost function in the form of (3.35), i.e.

$$J(\mathbf{x}) = \sum_{n=1}^{N} \left(\mathbf{x}^T \mathbf{C}_{mn} \mathbf{x}\right)^2 \ .$$

The most natural choice for the minimization of this function is the GTRS approach proposed in Section 3.3.2.4, which is specific for constraints expressed by circumferences. Therefore, the microphone positions are estimated as

$$\hat{\mathbf{x}}_{r_m} = \arg\min_{\mathbf{x}} \ J(\mathbf{x})|_{x_3=1} \ .$$

3.4.1.2 Global calibration and inter-calibration using TOAs

When the internal geometry of the array (i.e., the mutual position of the microphones) is known in advance we can fruitfully take advantage of this information. Let us denote with $\boldsymbol{x}'_{r_m}$ the nominal position of the mth microphone (i.e. its position in the local reference system), while $\boldsymbol{x}_{r_m}$ is its position in the world coordinate system. We intend to estimate the rotation matrix and the translation vector that relate the world and the local coordinate systems. The homogeneous points $\mathbf{x}'_{r_m}$ and $\mathbf{x}_{r_m}$ corresponding to $\boldsymbol{x}'_{r_m}$ and $\boldsymbol{x}_{r_m}$, respectively, are related by an isometry, which is a projective transformation $\mathbf{H}$ that embeds the rotation matrix and the translation vector (see Section 2.2.4.1). In mathematical terms, this is written as

$$\mathbf{x}_{r_m} = \mathbf{H}\mathbf{x}'_{r_m} \ , \tag{3.53}$$

where

$$\mathbf{H} = \begin{bmatrix} \boldsymbol{R}(\theta) & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix} , \tag{3.54}$$

where $\boldsymbol{R}(\theta)$ is the matrix operating a rotation of the reference frame by an angle θ and $\boldsymbol{t}$ is the translation vector.

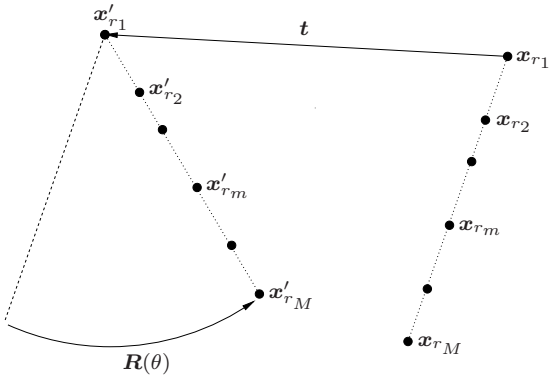


Figure 3.13: Self-calibration problem when the geometry of the array is known in advance: our goal is to estimate the rotation matrix $\mathbf{R}(\theta)$ and the translation vector $\mathbf{t}$ that bring the array from the nominal position to the actual one.

The problem has now become the estimation of the isometry that brings the local coordinate system to the global one, i.e. find the rotation matrix and the translation vector that moves $\mathbf{x}'_{r_m}$ to $\mathbf{x}_{r_m}$. Replacing (3.53) into (3.52) leads to a new form of the constraint:

$$\mathbf{x}'^T_{r_m} \mathbf{H}^T \mathbf{C}_{mn} \mathbf{H} \mathbf{x}'_{r_m} = 0 . \tag{3.55}$$

If the nominal positions of the microphones in the array have been measured, the only unknown in (3.55) is the isometry matrix $\mathbf{H}$. As neither $\mathbf{C}_{mn}$ nor $\mathbf{x}'_{r_m}$ are unknowns in (3.55), we can define a cost function collecting the squared residuals for all microphones in the same array:

$$J_{gl}(\theta, \boldsymbol{t}) = \sum_{m=1}^{M} \sum_{n=1}^{N} \left(\mathbf{x}'^T_{r_m} \mathbf{H}^T \mathbf{C}_{mn} \mathbf{H} \mathbf{x}'_{r_m} \right)^2 , \tag{3.56}$$

and the solution is given by

$$(\hat{\theta}, \hat{\boldsymbol{t}}) = \arg\min_{\theta, \boldsymbol{t}} J_{gl}(\theta, \boldsymbol{t}) . \tag{3.57}$$

The function in (3.56) does not present the same form of the generic cost function

(3.35), hence we cannot use any of the optimization techniques introduced in Section 3.3.2. Moreover, we notice that the cost function has a strongly nonlinear dependence on the rotation angle and the translation vector and therefore we can expect it to present several local minima. The minimization problem can be solved using, for example, the non-linear least-squares algorithm proposed in [60]. In order to guarantee that the iterative minimizer is not trapped in a local minimum, it needs to be initialized. A practical way to do so is to use the results from the intra-calibration. In particular, the initial translation is set to the estimated position of the first microphone, assumed that its nominal position is $[0, 0]^T$. The initial value for the rotation is then estimated finding the angle that minimizes the Euclidean distances between the estimated sensor positions and the rotated nominal ones. Finally, the position of the mth microphone is given by

$$\hat{\mathbf{x}}_{r_m} = \hat{\mathbf{H}}\mathbf{x}'_{r_m} \ ,$$

where $\hat{\mathbf{H}}$ is the homography estimated according to the minimization in (3.57).

If multiple arrays are present in the acoustic scene, the above equation is replicated for the two arrays, therefore we end up with the estimation of two rotation angles and two translation vectors. In this case, the two arrays can be inter-calibrated (according to the definition of inter-calibration given in [69]) remapping the world coordinate system with one of the two arrays.

3.4.1.3 Examples

In this Section we show some examples of array intra and global calibration on simulated and real data. The experimental setup is depicted in Figure 3.14. It is installed into a typical office-room, whose reverberation time is 0.7 s. The setup consists of a calibration grid (visible in the foreground) and two uniform linear microphone arrays (in the background), each accommodating 5 sensors distant 0.15 m. The calibration grid is rectangular with size $0.9\text{m}\times0.5\text{m}$, and defines a set of 25×5 potential positions of the loudspeaker. The origin of the grid is also the origin of the global reference frame, whose axes are parallel to those of the calibration pattern. The loudspeaker produces a sequence of white noise in the band $[0, 22$ kHz] and it is synchronized with the microphones. The impulse response from the loudspeaker to each microphone in the arrays is measured through a cross-correlation. The Times Of Arrival t_{mn} of the direct path are estimated by picking the first relevant maximum in the

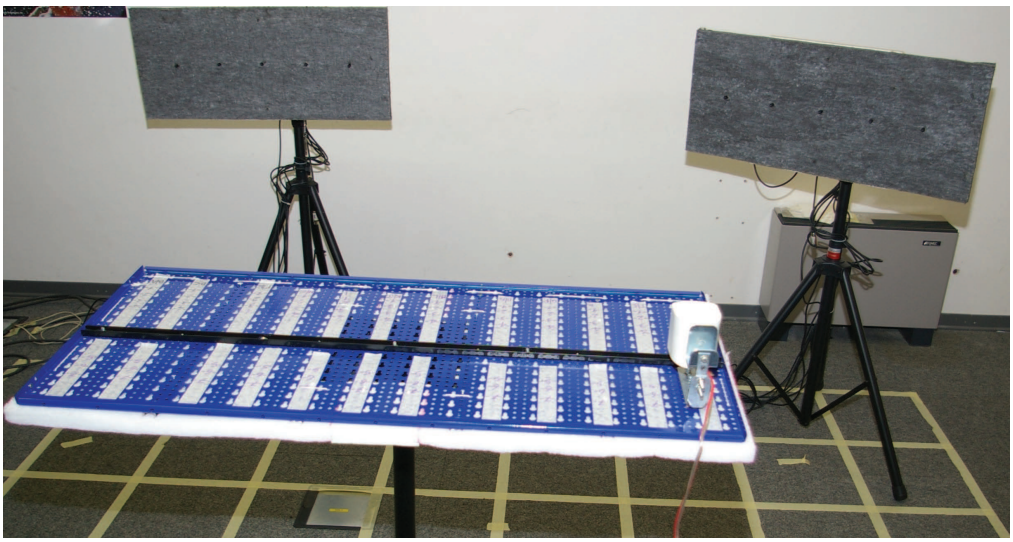


Figure 3.14: Snapshot of the acquisition system: the loudspeaker is placed on any position on the calibration pattern (in the foreground). The gray panels close to the walls, on the background, are the two arrays (each accommodating five sensors).

impulse responses. The measurements are then converted into geometric constraints as described in Section 3.2.1 and collected into the corresponding cost function.

Simulations We first consider a simulation of the setup described above. The positions of the arrays (in terms of rotation matrix and translation vector with respect to the calibration pattern) are given in Table 3.2. The acquisition of the

Table 3.2: Displacement of the arrays with respect to the reference frame

Array #	**Rotation angle**	**Translation vector**
1	$-\pi/6$	$[-1.8, 0.5]^T$
2	$\pi/6$	$[1.8, 0.5]^T$

signal is simulated for up to $N = 19$ positions of the loudspeaker and with a sampling frequency $F_s = 44.1$ kHz. For each position the measurement process is emulated by corrupting the theoretical TOAs with $K = 100$ realizations of additive white noise whose standard deviation error is 0.03 ms, corresponding to an error of 0.01 m on the distances of flight, assuming the standard sound speed $c = 343$ m/s. The effectiveness of the intra and inter-calibration algorithms is evaluated using the averaged Root

Mean Square Error

$$E = \sqrt{\frac{1}{MK}\sum_{k=1}^{K}\sum_{m=1}^{M}\|\hat{\mathbf{x}}_{m,k} - \mathbf{x}_m\|^2}\ ,$$

where $M = 10$ is the total number of microphones; $\mathbf{x}_m$ is the position of the mth microphone; and $\hat{\mathbf{x}}_{mk}$ is the estimation of the position of the mth microphone in the kth repetition. In order to verify the increase of robustness of the algorithms to a variable number of constraints, the error has been computed for N ranging from 3 to 19. Table 3.3 confirms that when a few measurements are available (from

Table 3.3: Average root mean square error of localization of microphones for a variable number of TOA measurements

TOA meas.	Intra-cal. RMSE	Global cal. RMSE
3	0.0522 m	0.0331 m
4	0.0215 m	0.0187 m
5	0.0208 m	0.0194 m
8	0.0126 m	0.0086 m
12	0.0109 m	0.0077 m
16	0.0096 m	0.0063 m
19	0.0091 m	0.0056 m

three to five) the global calibration process is much more efficient than the intra-calibration. When more information is available, the global calibration presents only a little advantage over the intra-calibration.

Experiments We now test the accuracy of intra and global array calibration in the real scenario of Figure 3.14. As far as global calibration is concerned, we test two different configurations. First, we assume that the nominal geometry of the array is preserved, i.e. the five sensors are collinear and the inter-distances are 0.15 m. Due to the construction imperfections, however, the arrays do not exactly match the above configuration, as they are not exactly collinear and uniformly distributed. As a consequence, the global calibration procedure is repeated when the nominal positions are assessed by the intra-calibration algorithm, thus performing a 2 steps global calibration. Figure 3.15 shows the averaged Root Mean Square Errors of intra and global calibration processes as a function of the number of observations, ranging from 3 to 19. We notice that the intra-calibration overcomes the global calibration algorithm when assuming perfect uniform linear arrays. The different behaviour with respect to the simulations can be interpreted as the consequence of non exact knowledge of the inner geometry of the arrays and an imperfect positioning of the

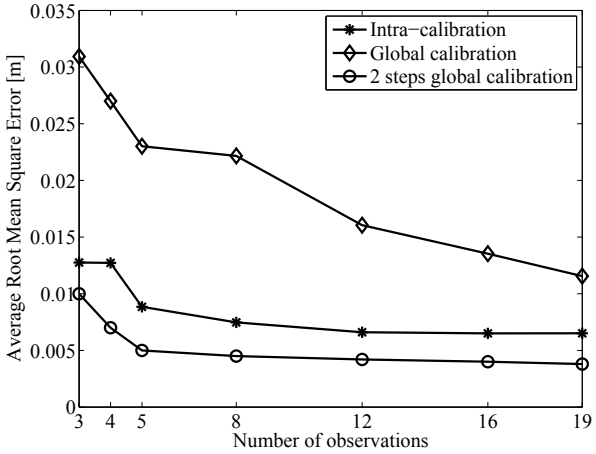


Figure 3.15: Error of intra and global calibration as a function of the number of measurement of the Times Of Arrival.

sensors. Indeed, when the a-priori information comes from intra-calibration results, the global calibration exhibits improved accuracy. We observe that the RMSE relative to the experimental results is smaller than in simulative conditions; the reason for this is that the error introduced in the simulations is greater than the error in the experiments.

3.4.2 Source localization using TDOAs

Source localization with microphone arrays from TDOA measurements is a problem that has been widely studied by the audio signal processing community. Most of the localization algorithms are based on the fact that, given a pair of sensors, the source is bound to lie on a hyperbola whose foci are on the sensors and whose distance between vertices is proportional to the measured TDOA. This knowledge is exploited in different ways. In [71] the authors propose a closed-form solution for the intersection of TDOA-related hyperbolas measured on an array of three microphones. In [72] the author proposes another closed-form algorithm for range computation that applies to an arbitrary array geometry. In [73] a two-stage weighted least-squares approach for source localization with a linear microphone array is proposed. Here the weighting procedure is either based on some a-priori information on the statistics of the measurement errors, or it starts from an approximate location of the source. Spherical Interpolation [74] and Spherical Intersection [75] are shown [2] to improve the localization accuracy over the above hyperbola-based methods, by intersecting

spheres (circles in 2D). In order to accomplish this task, the source-reference sensor distance is added as an extra unknown. In [2] a least squares reformulation of the Spherical Interpolation is proposed, with a more efficient implementation. A linear correction is also presented in order to improve the results. A similar least squares method is proposed in [76]. An alternate approach is the linear intersection algorithm proposed in [77]. In [73] the authors attenuate the non-linearity problem by using the geometry of linear uniform arrays, and by reformulating the hyperbolic constraint in terms of eccentricity and foci of the hyperbolas. Recently, in [59] the authors summarized the most relevant approaches to source localization and provided a set of exact and approximate solutions to the related estimation problems. In this Section we will show how source localization can be efficiently formulated combining projective constraints, leading to some advantages with respect to the state-of-the-art solutions. In particular, since this approach does not require the definition of a reference sensor, it makes it possible to realize localization systems composed by multiple clusters of microphones. The clusters do not need to be synchronized each other, thus making simpler and cheaper the required hardware.

3.4.2.1 Localization algorithm

Here we discuss the proposed source localization algorithm. In order to emphasize the advantage of the absence of a reference sensor, we distinguish between two opposite scenarios, as depicted in Figure 3.16. In particular, Figure 3.16-(a) shows a distri-

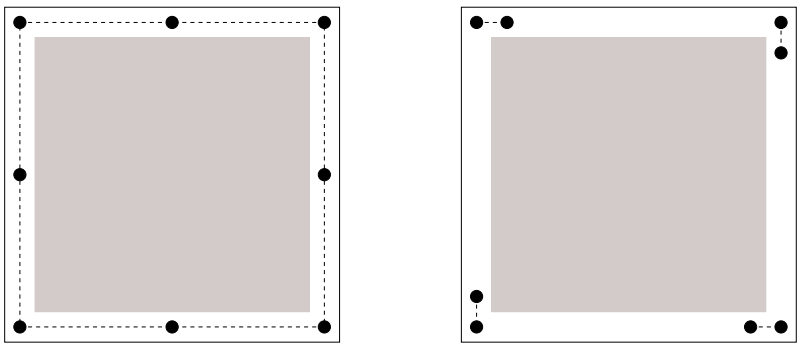


Figure 3.16: Two different source localization scenarios. In (a) the microphones are all synchronized and one of them is selected as the reference. The sensors surrounds the potential source positions denoted by the gray area. In (b) the microphones are arranged in synchronized pairs, distributed along the border of the room. The dashed lines depict the synchronization of the microphones.

bution of microphones installed along the borders of a room in which a source has

to be localized, whose potential positions are denoted by the gray area. The state-of-the-art algorithms typically operate in this situation, where all the sensors are synchronized and one of them is selected as a reference. The TDOAs are measured between microphone pairs that include the reference sensor. A set of $N+1$ sensors leads to the measurements of N TDOAs, denoted with $t_1, \dots t_N$. Figure 3.16-(b) depicts the same room in which the microphones are arranged in distributed pairs. In this case, only the sensors within a same pair are synchronized, thus making meaningless the definition of a reference microphone. The TDOAs are measured between paired microphones. From the signals on $2N$ microphones we measure N TDOAs, namely $t_1, \dots t_N$.

The proposed localization algorithm is suitable for both the scenarios. However, the situation in Figure 3.16-(b) presents several advantages with respect to one in Figure 3.16-(a). First of all, since the synchronism is required only between pair of microphones, the acquisitions can be made with a set of two-channel (stereo) audio interfaces, instead of using expensive multi-channel devices. Moreover, this solution makes it possible to place the paired sensors at a distance that allows a reliable estimation of the corresponding TDOA. It is well known that best performances are given by microphone pairs separated by a distance in the range from 5 cm to 20 cm. Higher distances may cause a loss in the correlation of the microphone signals, especially in reverberant rooms [78], thus making harder the correct estimation of the TDOA.

As shown in Section 3.2.2, the TDOA t_n measured on a pair of sensors leads to an hyperbolic constraints, which is expressed as

$$\mathbf{x}^T \mathbf{C}_n \mathbf{x} = 0 \ , \tag{3.58}$$

where $\mathbf{C}_n$ is a conic matrix representing an hyperbola with foci at the microphone positions and major axis equal to $t_n c$. All the individual constraints can be collected in a cost function in the form of (3.35):

$$J(\mathbf{x}) = \sum_{n=1}^{N} \left(\mathbf{x}^T \mathbf{C}_n \mathbf{x} \right)^2 \ . \tag{3.59}$$

It is important to notice that the cost function is defined independently from the particular scenario, as it does not involve the definition of a reference microphone.

Source localization is typically performed in real-time, therefore it is convenient to

minimize (3.59) on the unit sphere by means of an iterative optimization algorithm. The source position is therefore estimated as

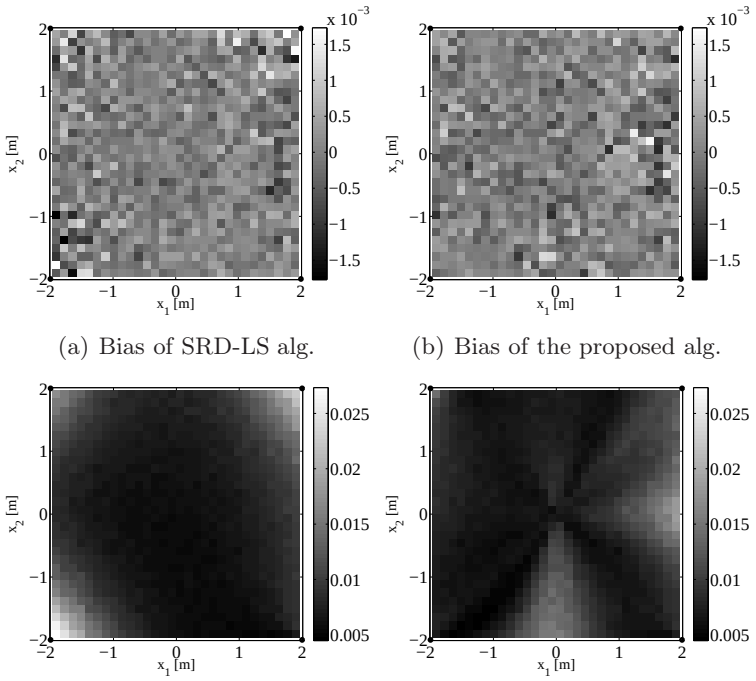
$$\hat{\mathbf{x}}_S = \arg\min_{\mathbf{x}} J(\mathbf{x}) \;\text{ subject to } \|\mathbf{x}\| = 1 \,. \tag{3.60}$$

The advantages of this approach are discussed in Section 3.3.2. Notice that the iterative search of the minimum requires an initialization value. A good starting point may be provided by linear methods such as the one presented in [76].

3.4.2.2 Examples

We now show some simulative examples in order to verify the feasibility of the source localization algorithm. The first example we consider is a comparison of the accuracy of the proposed methodology and of the state-of-the-art algorithm SRD-LS [59], both operating with 4 sensors located at the corners of a 2 m $\times$ 2 m square room. The SRD-LS algorithm has been chosen as a reference due to its superior localization accuracy (low bias and standard deviation of the estimation error) with respect to other standard methods, as shown in [59]. The localization is performed inside the square defined by the sensors, which has been regularly sampled with 900 test source positions. The theoretical range-differences (i.e. the TDOAs scaled by the speed of sound) have been corrupted with 500 realizations of a zero-mean Gaussian noise with standard deviation $\sigma = 1$ cm. The results are reported in Figures 3.17 and 3.18, relative to the x_1 and x_2 coordinates, respectively. In particular, from Figures 3.17-(a,b) and 3.18-(a,b), it can be noticed that the proposed algorithm present moderate bias on the estimation of both x_1 and x_2 coordinates, within the same range of values of SRD-LS technique. Similarly, Figures 3.17-(c,d) and 3.18-(c,d) shows a comparable behaviour of the two algorithms in terms of standard deviation. In fact, the average standard deviation of the estimation error is 81 mm for both the algorithms, on both coordinates. What mainly distinguishes SDR-LS from the proposed technique is the different distribution of the error in space.

The first example shows that the proposed localization technique has performances comparable to the state-of-the-art algorithms, in particular SRD-LS [59]. We now consider a scenario similar to one in Figure 3.16-(b), where the localization is performed by means of pairs of synchronized algorithms. This case is interesting to show the actual strength of the proposed approach. Here, the comparison with SDR-LS algorithm is not possible, since it is based on measurements relative to a reference

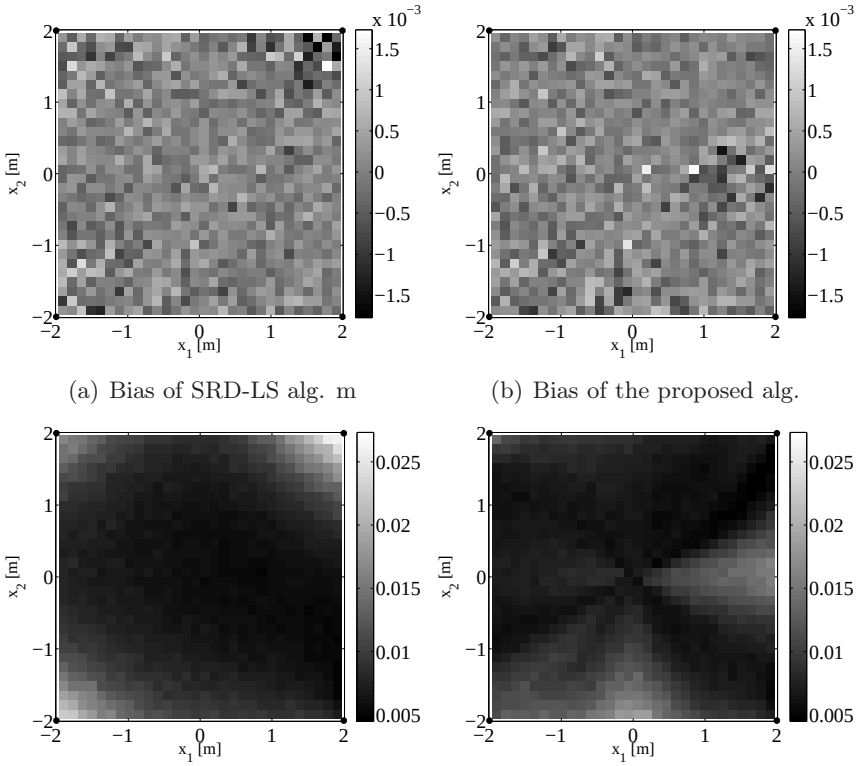


(a) Bias of SRD-LS alg. (b) Bias of the proposed alg.

(c) Std. dev. of SRD-LS alg. (d) Std. dev. of the proposed alg.

Figure 3.17: Comparison between the proposed source localization algorithm and the SRD-LS technique, operating with 4 synchronized microphones located at the corners of a 2 m $\times$ 2 m square room. The sensors are depicted as black dots; one of them is selected as a reference. Figures (a) and (b) show the bias on the x_1 coordinate; (c) and (d) show the standard deviation on the x_1 coordinate.

microphone. Therefore, we decided to make a comparison with the linear technique presented in [76] (referred as to GS), which allows to define an arbitrary number of reference sensors. The results are shown in Figure 3.19-(a,b,c,d), where the subfigures are organized as in the previous example. The standard deviation of the measurement error is kept fixed to $\sigma = 7$mm. Here we limit to show the behaviour for the x_1 axis, since that of x_2 is analogous to the first one. In this case, the proposed technique outperforms the linear one. The advantages over the first scenario are evident: as discussed before, the synchronization is required only among sensors belonging to the same pair; moreover, the TDOAs measured on closely-spaced sensors are more reliable than one estimated from distant microphones.



(a) Bias of SRD-LS alg. m

(b) Bias of the proposed alg.

(c) Std. dev. of SRD-LS alg.

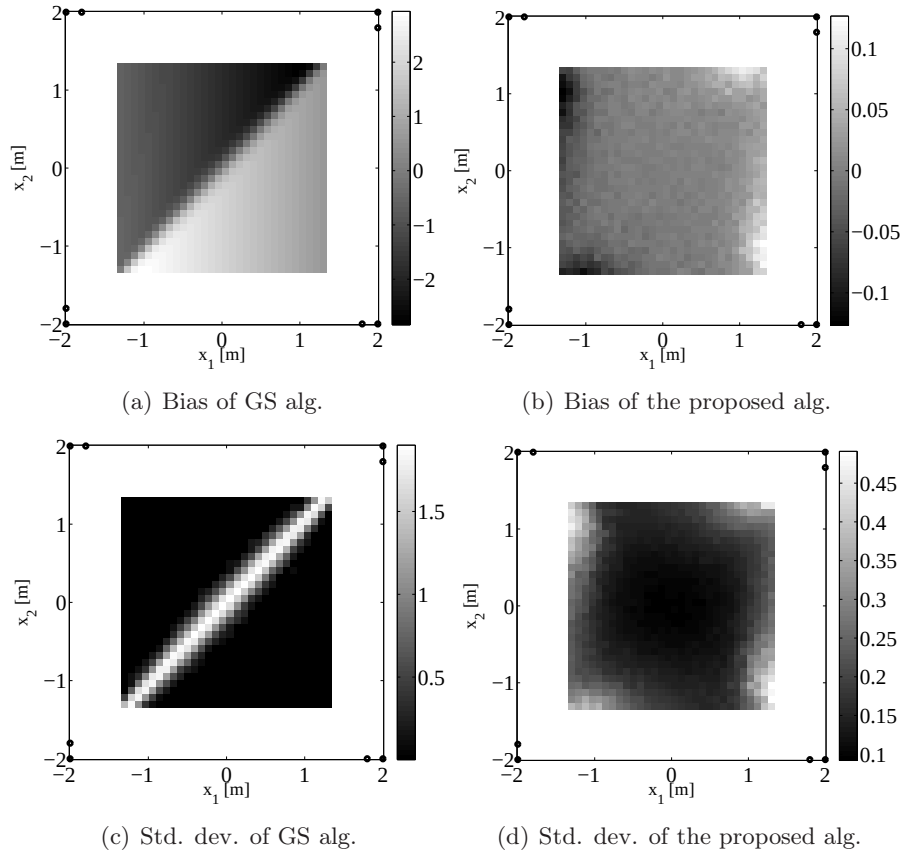
(d) Std. dev. of the proposed alg.

Figure 3.18: Comparison between the proposed source localization algorithm and the SRD-LS technique, operating with 4 synchronized microphones located at the corners of a 2 m × 2 m square room. The sensors are depicted as black dots; one of them is selected as a reference. Figures (a) and (b) show the bias on the x_2 coordinate; (c) and (d) show the standard deviation on the x_2 coordinate.

3.4.3 Reflector localization

The knowledge of the acoustic properties of the environment is crucial for many space-time processing applications. For example, in [79] source localization is approached using a maximum likelihood estimator whose data model incorporates the prediction of the early reflections. In [21] the knowledge of the room geometry is used to improve the rendering of soundfields through a loudspeaker array in a reverberant environment.

Techniques for the reconstruction of the geometry of the environment are common in computer vision. However, due to the different wavelengths of sound and optical waves, computer vision techniques return a geometry of the environment which contains too much detail for acoustic purposes and, moreover, reflectance proper-



(a) Bias of GS alg.

(b) Bias of the proposed alg.

(c) Std. dev. of GS alg.

(d) Std. dev. of the proposed alg.

Figure 3.19: Comparison between the proposed source localization algorithm and the GS technique, operating with 4 pairs of sensors located at the corners of a 2 m $\times$ 2 m square room. The sensors are depicted as black dots. Figures (a) and (b) show the bias on the x_1 coordinate; (c) and (d) show the standard deviation on the x_1 coordinate.

ties of materials can strongly differ between the acoustic and optical domains. For this reason the use of acoustic stimuli for the reconstruction of the geometry of the environment is highly desirable.

The inference of room geometry from acoustic measurements is a quite novel problem, therefore the amount of related literature is limited. In [80] the authors use a constrained room model and a ℓ_1 least-squares regularization to perform the estimation of the room geometry from the acoustic impulse responses. In [81] the authors present a technique for the estimation of the reflective surfaces from continuous signals in theaters and large auditory rooms, which is based on inverse mapping of the acoustic multi-path propagation problem. In [82] the problem is addressed using a single room impulse response. Although introducing a remarkable theoretical formu-

lation, this approach reveals to be much sensible to errors on the measured impulse response, making it inapplicable in real situations.

In this Section we introduce two algorithms for localizing a planar obstacle, exploiting the quadratic constraints derived in Section 3.2 arising from measurement related to the reflective paths. More specifically, the first algorithm is based on TOA measurements and requires the synchronization between source and microphones. The second method we propose is based on DOA measurements, which is suitable whenever is it not possible to synchronize the acquisition system with the acoustic source. At the end of the Section, both the approaches are generalized for the estimation of multiple planar reflectors.

As discussed in Chapter 1, reflector localization (and consequently the room geometry inference) are at the base of the application scenario considered for the testing of the methodologies proposed in this thesis. For this reason, the results relative to this topic are reported in Chapter 6, which is devoted to simulations and experiments concerning the specific application scenario.

3.4.3.1 Using TOAs

We consider a microphone array with sensors placed at $\boldsymbol{x}_{M_1}, \dots \boldsymbol{x}_{M_N}$. An acoustic source is located at $\boldsymbol{x}_S$ and, with no loss of generality, we assume the origin of the reference frame to be placed in that location. This scenario is depicted in Figure 3.20, where the a single planar reflector lies on the line $\mathbf{l}_R$. The image source $\boldsymbol{x}_{S'}$ is

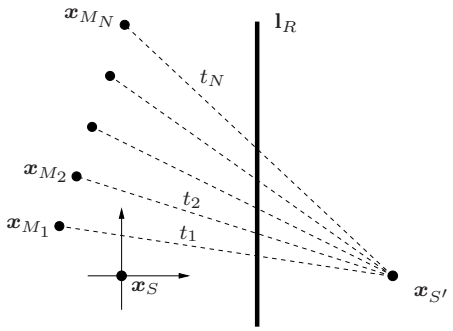


Figure 3.20: A microphone array with sensors at $\boldsymbol{x}_{M_1}, \dots \boldsymbol{x}_{M_N}$ is used to estimate the reflective TOAs $t_1, \dots t_N$ caused by the presence of a reflector lying on the line $\mathbf{l}_R$. The source location $\boldsymbol{x}_S$ coincides with the origin of the reference frame.

obtained by mirroring $\boldsymbol{x}_S$ over $\mathbf{l}_R$, and from the microphone signals we estimate the reflective TOAs $t_1, \dots t_N$. As discussed in Section 3.2.1, each measurement leads to

a quadratic constraint in the form

$$\mathbf{l}^T \mathbf{C}_n^{\diamond} \mathbf{l} = 0 \,, \tag{3.61}$$

where $\mathbf{C}_n^{\diamond}$ denotes the dual conic matrix of an ellipse with foci at $\boldsymbol{x}_S$ and $\boldsymbol{x}_{M_n}$ and major axis $t_n c$; and $\mathbf{l} = [l_1, l_2, l_3]^T$ is a line tangent to the ellipse. The constraints are then collected into the cost function

$$J(\mathbf{l}) = \sum_{n=1}^{N} \left(\mathbf{l}^T \mathbf{C}_n^{\diamond} \mathbf{l} \right)^2 \,.$$

Since the source is located at the origin, $\mathbf{C}_n^{\diamond}$ has the form of a circumference matrix (see Section 3.3.2.4). It is therefore convenient to cut $J(\mathbf{l})$ with the plane $l_3 = 1$, and write the cost function as the constrained least-squares problem (3.51). The reflector line is therefore estimated as

$$\hat{\mathbf{l}}_R = \arg\min_{\mathbf{l}} \; J(\mathbf{l})|_{l_3=1} \,. \tag{3.62}$$

3.4.3.2 Using DOAs

We now consider the case of DOA measurements. In this case, a microphone array is employed for estimating the DOAs related to the reflected paths. Without loss of generality, the array is centered at the origin of the reference system, therefore DOAs are measured with respect to this point. Multiple measurements can be performed by moving a source in a set of known positions $\boldsymbol{x}_{S_1}, \ldots \boldsymbol{x}_{S_N}$ in order to estimate the angles $\theta_1, \ldots \theta_N$, as depicted in Figure 3.21. As for the case of TOAs, each measurement is turned into a quadratic constraint having the form

$$\mathbf{l}^T \mathbf{C}_n^{\diamond} \mathbf{l} = 0 \,, \tag{3.63}$$

where now $\mathbf{C}_n^{\diamond}$ denotes the dual conic matrix of a parabola with focus at $\boldsymbol{x}_{S_n}$ and directrix $\mathbf{l}_n = [-\sin\theta_n, \cos\theta_n, 0]^T$. As usual, the constraints are combined into the cost function

$$J(\mathbf{l}) = \sum_{n=1}^{N} \left(\mathbf{l}^T \mathbf{C}_n^{\diamond} \mathbf{l} \right)^2 \,, \tag{3.64}$$

whose minimum provides an estimate of the reflector line. Since the computation time is not crucial for the estimation of the reflector line, in this case the most

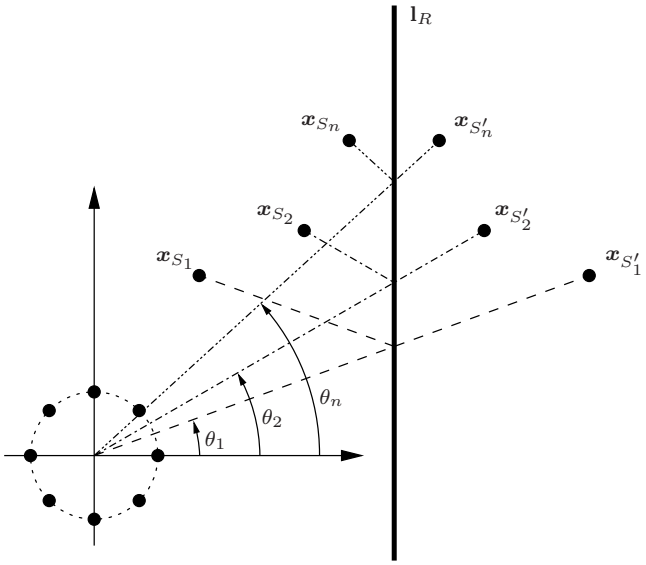


Figure 3.21: A circular microphone array, centred at the origin of the reference frame, estimates the reflective DOAs $\theta_1, \dots \theta_N$ generated by the reflections of a source moved at multiple positions $\boldsymbol{x}_{s_1}, \dots \boldsymbol{x}_{s_N}$. We assume the presence of a single reflector, lying on the line $\mathbf{l}_R$.

suitable approach for minimizing (3.64) is the one described in Section 3.3.2.1. In particular, $J(\mathbf{l})$ is cut with the planes $l_1 = 1$ and $l_2 = 2$, leading to the reduced cost functions $J_1(\mathbf{l}) = J(\mathbf{l})|_{l_1=1}$ and $J_2(\mathbf{l}) = J(\mathbf{l})|_{l_2=1}$. Denoting with $\hat{\mathbf{l}}_R^1$ and $\hat{\mathbf{l}}_R^2$ the global minima of $J_1(\mathbf{l})$ and $J_2(\mathbf{l})$, respectively, the estimated reflector line is finally given by

$$\hat{\mathbf{l}}_R = \min \left\{ \hat{\mathbf{l}}_R^1, \hat{\mathbf{l}}_R^2 \right\} . \tag{3.65}$$

3.4.3.3 Localization of multiple reflectors: Hough Transform

We consider here the problem of localizing multiple reflectors present at the same time in the acoustic scene. In this case, a single acoustic acquisition leads to the estimation of multiple measurements. More specifically, as far as TOAs are concerned, the impulse response measured at the nth microphone will present peaks at positions $t_{l,n}$, where $l = 1, \dots L$ is an index denoting different reflective paths. Similarly, for the case of DOAs, the nth source position generates L reflective paths corresponding to the angles $\theta_{l,n}$, $l = 1, \dots L$. For the sake of simplicity, we assume that the number of detectable reflections L remains constant for all the measurements. Notice that L does not correspond, in general, to the number of obstacles in the environment, since it accounts also for higher order reflections. However, higher order reflections

can be seen as generated by virtual walls, which can be treated just as real ones. At the end of the estimation process, potential virtual reflectors related to higher order reflections may be pruned, if needed, by means of geometrical considerations.

The problem now is to assign a label to each measurement, in order to group the TOAs or DOAs generated by the same reflector. We resort for this purpose to the generalized Hough transform [83], which is a feature extraction technique commonly used in computer vision to find out lines and other parametric curves in an image. The basic idea is to make a number of hypotheses on the reflector locations and then to verify, through a voting procedure, which of them are mostly coherent with the measurements. We define a grid of $I \times J$ test reflector lines parametrized by their orientation $\alpha_i,\ i = 1, \dots I$ in the range $[0, 2\pi]$; and distance from the origin $\rho_j,\ j = 1, \dots J$, in the range $[\rho_{\min}, \rho_{\max}]$. Therefore, the pair (α_i, ρ_j) identifies the line

$$x_1 \cos\alpha_i + x_2 \sin\alpha_i - \rho_j = 0 \ ,$$

whose parameter vector is $\tilde{\mathbf{l}}_{ij} = [\cos\alpha_i, \sin\alpha_i, -\rho_j]^T$. Mirroring the source located at $\mathbf{x}_S = [\boldsymbol{x}_S, 1]^T$ over the test reflector $\tilde{\mathbf{l}}_{ij}$, we obtain an hypothesis mirror source at

$$\tilde{\mathbf{x}} = \mathbf{H}_r \mathbf{x}_S \ , \tag{3.66}$$

where

$$\mathbf{H}_r = \begin{bmatrix} 1 - 2\sin^2\alpha_i & -2\cos\alpha_i \sin\alpha_i & 2\rho_j \cos\alpha_i \\ -2\cos\alpha_i \sin\alpha_i & 1 - 2\cos^2\alpha_i & 2\rho_j \sin\alpha_i \\ 0 & 0 & 1 \end{bmatrix} \tag{3.67}$$

is the homography describing the specular reflection (see Section 2.2.4.2 for details). The TOA associated to the test reflector $\tilde{\mathbf{l}}_{ij}$ and relative to the nth microphone in the array can be calculated as

$$\tilde{t}_{n,ij} = \frac{\|\tilde{\boldsymbol{x}} - \boldsymbol{x}_{M_n}\|}{c} \ , \tag{3.68}$$

where $\tilde{\boldsymbol{x}}$ is the Cartesian representation of $\tilde{\mathbf{x}}$; and the speed of sound c is assumed to be known. Similarly, the DOA associated to $\tilde{\mathbf{l}}_{ij}$ and relative to the source located

at $\mathbf{x}_{S_n} = [\boldsymbol{x}_{S_n}, 1]^T$ is obtained as

$$\tilde{\theta}_{n,ij} = \begin{cases} \arctan\left(\frac{\tilde{x}_{2_n}}{\tilde{x}_{1_n}}\right) , & \text{if} \quad \tilde{x}_{2_n} \geq 0 \\ \\ \arctan\left(\frac{\tilde{x}_{2_n}}{\tilde{x}_{1_n}}\right) + \pi , & \text{if} \quad \tilde{x}_{2_n} < 0 \end{cases} \tag{3.69}$$

where $(\tilde{x}_{1_n}, \tilde{x}_{2_n})$ are the coordinates of the hypothesis mirror source given by

$$\tilde{\mathbf{x}}_n = [\tilde{x}_{1_n}, \tilde{x}_{2_n}, 1]^T = \mathbf{H}_r \mathbf{x}_{S_n} .$$

At this point we can set up a voting procedure that estimates the likelihood of each test reflector. For this purpose we consider the functions $f_t(i,j)$ and $f_d(i,j)$ for TOAs and DOAs, respectively:

$$\begin{aligned} f_t(i,j) &= \sum_{n=1}^{N} \sum_{l=1}^{L} \eta_t(t_{l,n}, \tilde{t}_{n,ij}) , \\ f_d(i,j) &= \sum_{n=1}^{N} \sum_{l=1}^{L} \eta_d(\theta_{l,n}, \tilde{\theta}_{n,ij}) . \end{aligned} \tag{3.70}$$

As far as TOAs are considered, $\eta_t(t_{l,n}, \tilde{t}_{n,ij})$ is a binary function that evaluates if the measured TOA $t_{l,n}$ is compatible with the hypothesis $\tilde{t}_{n,ij}$ and it is defined as

$$\eta_t(t_{l,n}, \tilde{t}_{n,ij}) = \begin{cases} 1 & \text{if} \quad |t_{l,n} - \tilde{t}_{n,ij}| < \nu_t \\ 0 & \text{if} \quad |t_{l,n} - \tilde{t}_{n,ij}| \geq \nu_t \end{cases} ,$$

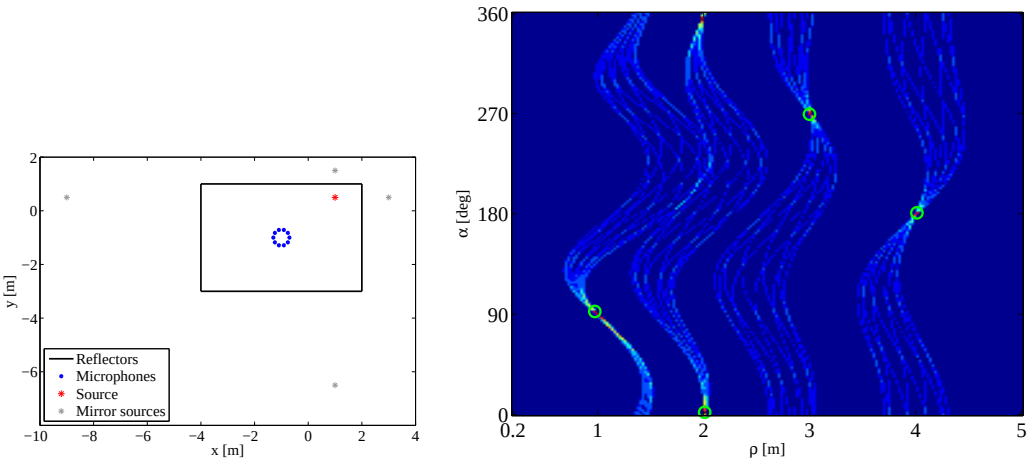
where ν_t is an acceptance threshold. In a similar way, the function $\eta_d(t_{l,n}, \tilde{t}_{n,ij})$ for DOAs is defined as

$$\eta_d(\theta_{l,n}, \tilde{\theta}_{n,ij}) = \begin{cases} 1 , & \text{if} \quad |\theta_{l,n}, \tilde{\theta}_{n,ij}|_\pi < \nu_d \\ 0 , & \text{if} \quad |\theta_{l,n}, \tilde{\theta}_{n,ij}|_\pi \geq \nu_d \end{cases} ,$$

where $|\theta_{l,n}, \tilde{\theta}_{n,ij}|_\pi$ denotes the unsigned angular distance in the range $[0, \pi]$ between $\theta_{l,n}$ and $\tilde{\theta}_{n,ij}$; and ν_d is a different acceptance threshold.

For the case of TOA measurements, a set of L candidate reflector lines (i.e., the reflectors that most likely have generated the acoustic measurements) are selected as

the hypothesis reflectors such that $f_t(i,j) > T_t$, where T_t is an acceptance threshold. If necessary, it is also possible to define a neighborhood suppression rule in order to retain a single candidate reflector in certain portion of the grid. An example of generalized Hough transform for the case of TOA measurements is shown in Figure 3.22-(b). This map is obtained from simulations on the simple scenario in Figure 3.22-(a), where reflections have been predicted up to the first order. The measurement process has been simulated by corrupting the calculated TOAs with zero-mean gaussian noise with standard deviation $\sigma_t = 0.045$ ms, corresponding to 2 samples at 44100 Hz in the discrete impulse responses. The candidate reflectors are denoted by the green circles superimposed to the Hough map.



(a) (b)

Figure 3.22: Generalized Hough transform for the case of TOA measurements. Figure (a) depicts the simulated scenario; the corresponding Hough map is shown in Figure (b).

The case of DOA measurements is a bit more involved. We discuss about this case through the example in Figure 3.23. The Hough map in Figure 3.23-(b) is calculated from simulations relative to the scenario in Figure 3.23-(a), considering reflections up to the first order. The DOAs have been corrupted with zero-mean white noise with standard deviation $\sigma_d = 1°$. We notice that the Hough transform is capable to precisely detect the orientation of candidate reflectors, but it fails in determining their distance. It is therefore convenient to estimate the orientation and the distance in two stages. More specifically, we first estimate the orientations from the highest

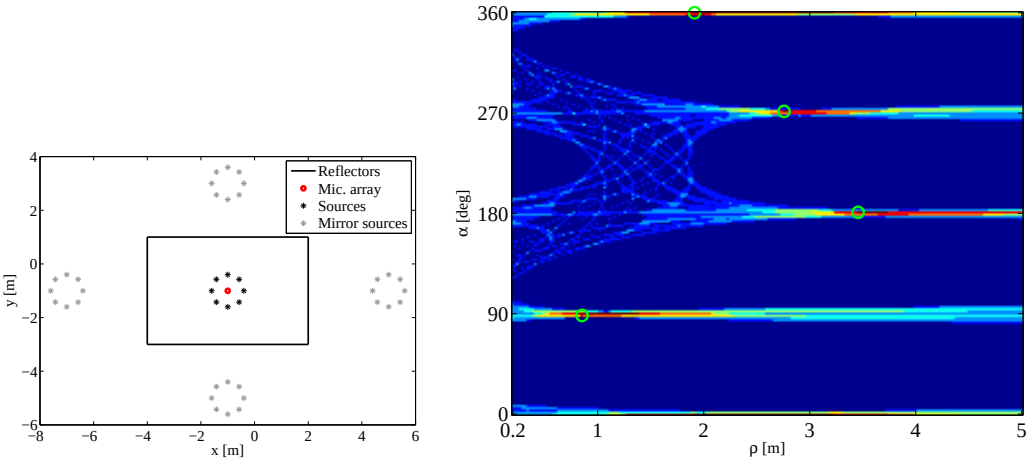
peaks of the function

$$a(i) = \sum_{j=1}^{J} f_d(i,j) \ , \tag{3.71}$$

located at $\hat{i}_1, \ldots \hat{i}_L$. The second steps concerns the estimation of the corresponding distance indexes, which are given by

$$\hat{j}_l = \arg\min_j \ f_d(\hat{i}_l, j) \ . \tag{3.72}$$

Finally, the parameters of the candidate reflectors are obtained as $\left\{\alpha(\hat{i}_l), \rho(\hat{j}_l)\right\}_{l=1}^{L}$.



(a) (b)

Figure 3.23: Generalized Hough transform for the case of DOA measurements. Figure (a) depicts the simulated scenario; the corresponding Hough map is shown in Figure (b).

The generalized Hough transform provides only a rough estimate of the reflector positions. The estimation can be refined by setting up L optimization problems in the form of (3.62) for the case of TOAs, or in the form of (3.65) for the case of DOAs. In order to do so, it is necessary to associate each measurement with one of the candidate reflectors. This labeling operation can be done as follows. Denoting with $\tilde{t}_{n,l}$ the TOA associated to the lth candidate reflector $\tilde{\mathbf{l}}_l$, the measured TOA t_n is assigned to that reflector if

$$|t_n - \tilde{t}_{n,l}| < G_t \ , \tag{3.73}$$

where G_t is an acceptance threshold. Similarly, the measured DOA θ_n is assigned to the lth candidate reflector if

$$|\theta_n, \tilde{\theta}_{n,l}|_\pi < G_d \ , \tag{3.74}$$

where $\tilde{\theta}_{n,l}$ is the DOA relative to the candidate reflector $\tilde{\mathbf{l}}_l$; and G_d is an acceptance threshold.

Notice that the procedure presented above is useful for detecting potential outliers in the set of measurements. Indeed, if (3.73) (for TOAs) or (3.74) (for DOAs) are not satisfied for any of the candidate reflectors, the nth measurement is considered as an outlier and therefore discarded.

3.4.4 Discussion

All the presented estimation algorithms rely on quadratic constraints based on a two-dimensional geometry. Notice that this fact does not constitute a limitation, since the underlying theory can be readily extended to the three-dimensional case. In particular, moving from 2D to 3D means working in the projective space $\mathbb{P}^3$, where a generic point is represented by means of 4 homogeneous coordinates, namely $\mathbf{X} = [x_1, x_2, x_3, x_4]^T$. In $\mathbb{P}^3$ there exists a duality between points and planes [11]. The homogeneous representation of a plane with equation $p_1x_1 + p_2x_2 + p_3x_3 + p_4 = 0$ is given by the vector $\mathbf{P} = [p_1, p_2, p_3, p_4]^T$, which is homogeneous and therefore represents a point in $\mathbb{P}^3$. Quadratic forms in $\mathbb{P}^3$ are called *quadrics*, which have the same duality properties of conics in $\mathbb{P}^2$. In particular, a quadric is described by $\mathbf{X}^T\mathbf{D}\mathbf{X} = 0$, where $\mathbf{D}$ is a symmetric 4×4 matrix containing the quadric coefficients [11]. Its dual representation is given by $\mathbf{P}^T\mathbf{D}^\diamond\mathbf{P}$ where $\mathbf{P}$ is the parameter vector of a plane tangent to the quadric, and $\mathbf{D}^\diamond = \det(\mathbf{D})\mathbf{D}^{-1}$ contains the coefficients of the dual quadric [11].

As shown in [84], TOA measurements related to reflective paths in a three-dimensional space lead to quadratic constraints in $\mathbb{P}^3$, i.e. to quadrics. In particular, the TOA related to the path linking a source at $\mathbf{X}_S$ and a microphone at $\mathbf{X}_M$ through a reflection, constrains the reflector (a plane in 3D) to be tangent to an ellipsoid. The foci of the ellipsoid are $\mathbf{X}_S$ and $\mathbf{X}_M$, while its major axis is proportional to the TOA. Following the same reasoning, it can be proved that TOA measurements on the direct paths lead to quadrics representing spheres; TDOA measurements on direct paths correspond to hyperboloids; and DOAs related to reflective paths lead to paraboloids. The combination of multiple constraints conducts to the definition of

a homogeneous cost function in $\mathbb{P}^3$, whose minimization leads to an estimate of the acoustic primitive of interest, which can be either a 3D point (microphone or source position) or a plane (on which a reflector lies). As an example, in [84] the position of a planar reflector is estimated as the common tangent to a set of ellipsoids originated from multiple TOA measurements.

4 Geometric wave field rendering

In this Chapter we describe a methodology for wave field rendering through an array of loudspeakers, based on the geometric decomposition of a complex sound field. Starting from the problem of rendering a virtual source in an anechoic chamber, we will incrementally discuss the steps needed to render also the virtual environment in which the virtual source is collocated. Adopting a *beam tracing* technique, it is possible to predict the effects of the virtual source in the virtual environment. More specifically, the acoustic propagation is modelled as a set of beams that originate from virtual image source positions, each one parametrized with a direction and an angular aperture. Each beam represents the region where the correspondent virtual image source is visible. As a consequence, the use of the beam tracing engine makes it possible to describe complex wave fields as the superposition of elementary beams. The virtual environment can therefore be rendered by synthesizing all the individual beams by means of a loudspeaker array, each properly delayed and attenuated. The rendering of a beam is performed through a *beam shaping* algorithm.

The beam tracing engine is developed in a two-dimensional space. This fact does not constitute a restriction since we aim at rendering the wave field in the plane surrounding the listener ears. Although the proposed rendering engine is easily estendible to the three-dimensional case, we prevent us from the rendering in a 3D space due to the disproportionate cost of the implementation, as already discussed in Section 1.2.

As a second step, we consider also the problem of sound field reproduction within a real room that, in most cases, is far from being anechoic. This problem is known in the literature as *room compensation*. Relevant works related to this topic can be found, for example, in [20] and [21]. More specifically, in [20] the authors address the problem through a particular multichannel adaptive filtering which considers also the physical acoustic propagation; in [21] the room compensation in simple environments is achieved by modelling the early reflections by means of the image source

method. Here we propose an alternate method which exploits the before mentioned beam tracing technique also for predicting the early reflections caused by the environment to be compensated for. Given the floor plan of the environment and the loudspeaker positions, the beam tracing engine determines the set of image loudspeakers along with their visibility in all the listening points. We notice that, if the information about the geometry of the room in which the rendering system operates is not available in advance, it may be inferred from suitable acoustic measurements, as described in Section 3.4.3.

The Chapter is organized as follows. Section 4.1 introduces the problem of rendering, taking into account all the aspects mentioned above; Section 4.2 presents the beam tracing algorithm employed for determining the set of acoustic beams to be rendered as well as the beams that generate the early reflections to be compensated; in Section 4.3 we describe the rendering methodology, showing some preliminary tests on room compensation.

4.1 Problem formulation

In this section we give a general description of the rendering problem. In particular we define the concept of a rendering system pointing out issues that we need to address to achieve reproduction of the desired wave field.

4.1.1 Scenario description

The goal of a *rendering system* is to accurately reproduce a desired wave field inside the *listening area*, by means of arbitrary distribution or array(s) of *loudspeakers*. The listening area is defined by a set of *control points*. A simple example of a rendering scenario is shown in Figure 4.1, where the loudspeakers (indicated with black crosses) are placed on a circle around the listening area, which is sampled by a regular distribution of control points.

Let us consider the situation of rendering an omnidirectional *virtual source*. In Figure 4.1 the virtual source is depicted by a black circle and the control points by dots. The listening area encloses the control points and it is denoted by shaded grey. This rendering scenario is interesting for walkthrough applications, where we intend to improve the immersivity of the acoustic scene.

We now consider a more challenging scenario in which we aim at rendering not only

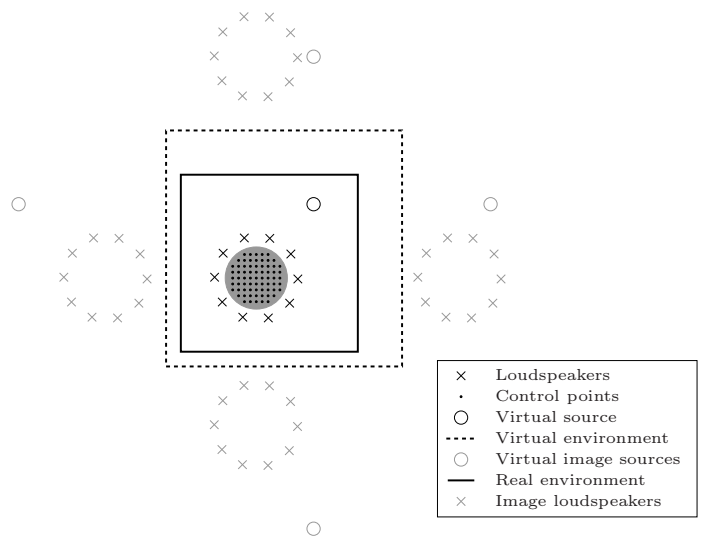


Figure 4.1: Geometry of a rendering system operating in a real environment and aiming at reproducing the reflections of a virtual environment.

the presence of the virtual source but also of a *virtual environment* that surrounds it. The environment causes reverberations to appear, and consequently we should perceive the reflective paths coming from the *reflectors* along with the direct path. In order to reproduce the effect of reverberations (at least in terms of early reflections) we need therefore, under the hypothesis of geometrical acoustics, to render also the *virtual image sources* obtained by mirroring the virtual source against the reflectors [22, 85, 86]. We aim at resorting to the superposition principle, i.e. we superimpose the effect of reflective sources to the direct one as adopted in [87]. A simple example is shown in Figure 4.1, where the dashed line defines the contour of a rectangular room. More specifically, each segment represents a planar reflector. The virtual image sources are represented by grey circles.

So far, however, we neglected the effect on the wave field reproduction of the *real environment* in which the rendering system is operating, i.e. we supposed a free-field (anechoic) propagation between each loudspeaker and each control point. We aim at devising a methodology for compensating, at least partially, the effect of reverberations coming from the real environment on the reproduced wave field. Once again, if the real environment is composed by planar reflectors the early reflections can be modelled by *image loudspeakers* whose positions are predicted with the image source technique. In Figure 4.1 the continuous line represents the real environment and the grey crosses denote image loudspeakers.

The example in Figure 4.1 is very simple. Indeed, the real and virtual environments present a rectangular geometry. In more complex environments (both real and virtual) image sources (virtual sources or image loudspeakers) are not visible from all points in space. Consider for example the environment shown in Figure 4.2. Here, the presence of a wall occludes the direct path between the acoustic source and a receiver (marked with a black point). We notice that in this example no reflective paths imping on the receiver. In fact, the *visibility* conditions cause the subdivision of rays originating from the (image) source into *beams*, which split and branch during propagation and interaction with the environment. Each beam has a limited region in which it is visible. We notice that this problem is common for both virtual sources in virtual environments and image loudspeakers in real environments, as they both can be occluded by virtual or real walls respectively. It is well known in the literature [86, 10] that the image source technique is not suitable for occluded environments, as a demanding visibility check between receivers (control points) and image sources (virtual sources or image loudspeakers) is required. A technique that addresses the above visibility issues is *beam tracing* [88, 89, 33, 90, 91, 10, 17, 92], which models the interaction of the wave field with the environment as the propagation, branching and reflection of beams, each characterized by the position of the image source, its orientation and its aperture.

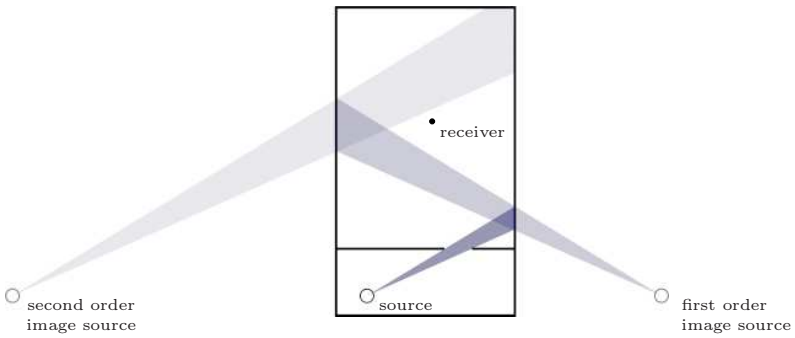


Figure 4.2: Occlusions limit regions in which the source and image sources are visible; the visibility conditions can be encoded by beams that split and branch during propagation and interaction with the environment.

As a result, beam tracing models the wave field as a tiling and superposition of acoustic beams. The Figure 4.3 shows the superposition of beams predicted by beam tracing in a densely occluded enclosure. For clarity of visualization the figure shows only beams up to the second order of reflection.

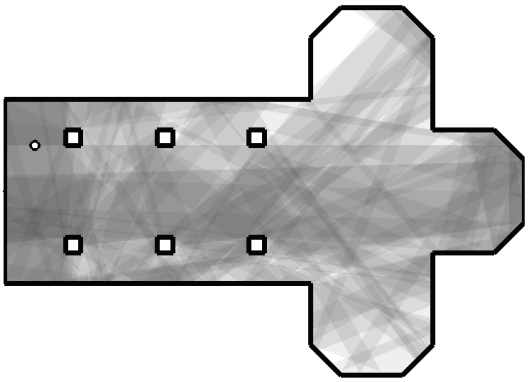


Figure 4.3: The superposition of beams predicted by beam tracing in a densely occluded environment; beams up to the second order of reflection are shown.

4.1.2 Requirements

In the light of the issues introduced in the previous paragraph, we now identify the requirements of a rendering system that reproduces the acoustics of a virtual environment when it is operating in a reverberant room. We can distinguish the following components:

1. *A propagation modelling technique that traces beams as they propagate in the virtual environment.* This operation can be efficiently performed by employing the beam tracing method described in Section 4.2;

2. *A beam shaping engine that allows us to render a beam in the listening area by means of a loudspeaker array.* This component is detailed in Section 4.3.1;

3. *A methodology for the compensation of the early reflections produced by the environment hosting the rendering system.* As we see in Section 4.3.4, the room compensation requires the knowledge of the positions of the image loudspeaker along with their visibility conditions. Once more, we resort for this task to beam tracing.

Let us consider the scenario of rendering a virtual source when both virtual and real environments are anechoic. In this situation we need only the beam shaping component. On the other hand, when we intend to reproduce the early reflections of a virtual environment we need to incorporate the propagation modelling technique to trace virtual image sources and render them. If, finally, our rendering system is

hosted in a reverberant real environment, we need also the compensation technique. As a consequence the components outlined above should be modularly used to achieve the desired reproduction result.

4.2 Beam tracing

As discussed in Section 4.1, visibility conditions cause the splitting of rays originating from an image source into beams. In Section 2.3.2 we saw that the mutual visibility between reflectors is conveniently encoded by visibility diagrams in the ray parameter space. With reference to Figure 4.4, we show how to use visibility diagrams to iteratively trace beams. Let us consider the reflection of a beam W_i onto the reflector $\mathcal{AB}$ delimited by the endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$. We first compute the reflected bundle of rays W'_i finding the image source $\boldsymbol{x}'_S$ (determined by mirroring the source $\boldsymbol{x}_S$ over the segment $\boldsymbol{x}_A\boldsymbol{x}_B$) and the rays $\mathbf{l}_1$ and $\mathbf{l}_2$ that limit W'_i, as in Figure 4.4-(a). The splitting process is accomplished in the ray space by intersecting the reflector's visibility diagram with the ray space representation of W'_i (i.e. the portion of the plane $\mathcal{S}'$ limited by $\mathbf{l}_1$ and $\mathbf{l}_2$), as shown in Figure 4.4-(b). The ray space representation of W'_i is made of the segments $\mathcal{W}_{i1}$, $\mathcal{W}_{i2}$ and $\mathcal{W}_{i3}$, each lying in a different visibility region. These segments represent the sub-beams originated from the splitting of W'_i. The corresponding beams in the geometric space are depicted in Figure 4.4-(c). One of them (W_{i3}) proceeds to infinity, the others (W_{i1} and W_{i2}) are blocked by reflectors and therefore they originate new beams. The recursive procedure stops when the preassigned order of reflection is reached or when the beams die out (i.e., when they are attenuated below a preassigned threshold of magnitude). It is well

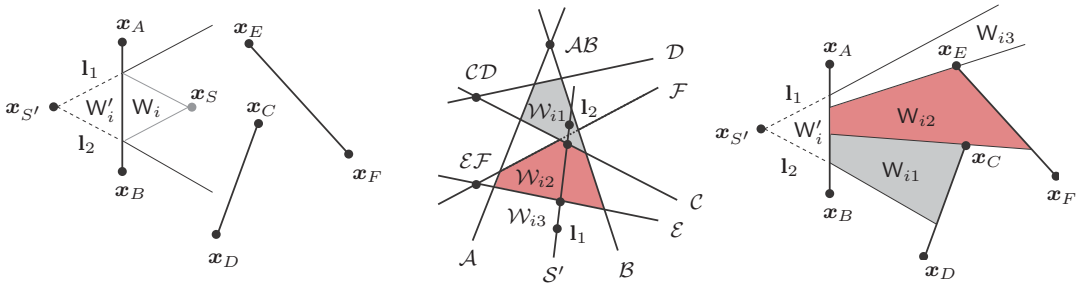


Figure 4.4: (a) Reflection of the beam in the geometric space; (b) Beam subdivision in the ray space; (c) New beams in the geometric space.

known in the literature [88, 10, 17, 92] that beams can be organized in a beam tree data structure that contains the branching relationship between acoustic beams and represent efficiently the visibility from the geometric space of the source position. Further references on the beam tracer and its implementation can be found in [18] and [93].

The knowledge of acoustic beams is necessary for the rendering stage whenever we want to reproduce the early reflections of a virtual environment and/or we want to compensate those of a real environment. In the first case the virtual environment causes virtual image sources to appear and the rendering system has to reproduce each one together with its beam-pattern (see Section 4.3.3). In the second case, in order to prevent that reverberations of the real environment damage substantially the quality of rendered sound field, early reflections have to be incorporated in the data model (see Section 4.3.4). To do this, the visibility between control points and image loudspeakers is extracted from the beam tree in the path tracing phase described in the following paragraph.

4.2.1 Path tracing

Once the receiver (control point) location $\boldsymbol{x}_r = (x_{1_r}, x_{2_r})$ is specified, a simple iterative procedure looks up the beam tree to find the paths from source to receiver. We use the beam parametrization shown in Figure 2.5-(f), Section 2.3, where the beam is parametrized with the vectors corresponding to its bounding lines oriented so that a point inside the beam is always on the right of those vectors. In order to test if a point is inside the beam, therefore, we only need to verify that it is on the right side of all vectors that parametrize the beam. Denoting the beam origin with $\boldsymbol{x}_{S'} = (x_{1_{S'}}, x_{2_{S'}})$, the length of the acoustic path is $d = \|\boldsymbol{x}_{S'} - \boldsymbol{x}_r\|$. The corresponding travelling time (delay) is $t = d/c$, where c indicates the speed of sound.

4.3 Soundfield Rendering

This Section focuses on the methodology for reproducing a desired wave field within a listening area by means of a loudspeaker array. First of all we describe the beam shaping engine used for rendering elementary narrow-band acoustic beams. After that the wideband extension is described. Then, we show how the rendering of a complex wave field is achieved through the superposition of individual beams.

Finally, the room compensation strategy is discussed.

4.3.1 Beam shaping

We now consider the rendering of an acoustic beam through a loudspeaker array. According to Figure 4.5, a set of ideal omnidirectional loudspeakers (i.e., point sources) are placed at arbitrary positions $\boldsymbol{x}_{p_1}, \dots \boldsymbol{x}_{p_M}$ in an anechoic room. We also define

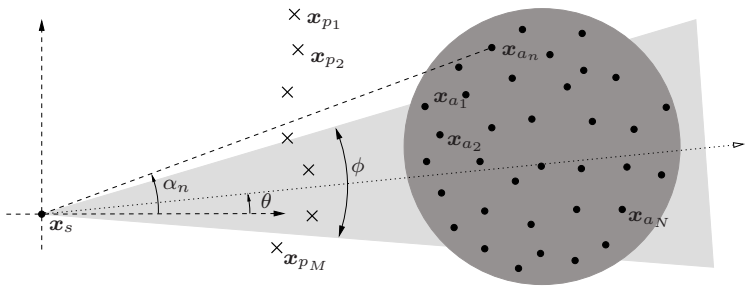


Figure 4.5: Geometry of the proposed rendering system.

a set of control points $\boldsymbol{x}_{a_1}, \dots \boldsymbol{x}_{a_N}$ within the listening area, depicted as the gray-shaded circle. The goal is to reproduce the acoustic beam generated by a virtual source located at $\boldsymbol{x}_s$ emitting towards the direction θ and with angular aperture ϕ. The Fourier transform $P_d(\omega, \boldsymbol{x}_{a_n})$ of the desired wave field at the nth control point is therefore given by

$$P_d(\omega, \boldsymbol{x}_{a_n}) = G_\omega(\boldsymbol{x}_{a_n}|\boldsymbol{x}_s)\Theta(\theta, \phi, \alpha_n)S(\omega) \ ,$$

where

$$G_\omega(\boldsymbol{x}_{a_n}|\boldsymbol{x}_s) = \frac{e^{-j\frac{\omega}{c}\|\boldsymbol{x}_s - \boldsymbol{x}_{a_n}\|}}{4\pi\|\boldsymbol{x}_s - \boldsymbol{x}_{a_n}\|}$$

is the Green's function (see (2.13)) from $\boldsymbol{x}_s$ to $\boldsymbol{x}_{a_n}$, ω being the angular frequency; $\Theta(\theta, \phi, \alpha_n)$ is an angular function describing the beam-pattern. With reference to Figure 4.5, α_n is the angle under which the virtual source is seen from $\boldsymbol{x}_{a_n}$; and $S(\omega)$ is the Fourier transform of the source signal. The goal of the beam shaping engine is to reproduce the effect of the virtual source at all the listening points by means of the loudspeakers. In other words, we aim at finding the vector of complex coefficients $\mathbf{h}_\omega$ applied to the loudspeakers that satisfies the following system of equations:

$$\mathbf{r}_\omega S(\omega) = \mathbf{G}_\omega \mathbf{h}_\omega S(\omega) \ , \tag{4.1}$$

where

$$\mathbf{G}_\omega = \begin{bmatrix} G_\omega(\boldsymbol{x}_{a_1}|\boldsymbol{x}_{p_1}) & \dots & G_\omega(\boldsymbol{x}_{a_1}|\boldsymbol{x}_{p_M}) \\ \vdots & \ddots & \vdots \\ G_\omega(\boldsymbol{x}_{a_N}|\boldsymbol{x}_{p_1}) & \dots & G_\omega(\boldsymbol{x}_{a_N}|\boldsymbol{x}_{p_M}) \end{bmatrix} \tag{4.2}$$

is the propagation matrix from each loudspeaker to each control point. The vector $\mathbf{r}_\omega$ encodes the desired response at all the control points, and it is therefore defined as

$$\mathbf{r}_\omega = \begin{bmatrix} G_\omega(\boldsymbol{x}_{a_1}|\boldsymbol{x}_s)\Theta(\theta,\phi,\alpha_1) \\ \vdots \\ G_\omega(\boldsymbol{x}_{a_N}|\boldsymbol{x}_s)\Theta(\theta,\phi,\alpha_N) \end{bmatrix} . \tag{4.3}$$

The system in (4.1) has, in general, no exact solution; the best solution in the least-squares sense is given by

$$\hat{\mathbf{h}}_\omega = \mathbf{G}_\omega^+ \mathbf{r}_\omega = (\mathbf{G}_\omega^H \mathbf{G}_\omega)^{-1} \mathbf{G}_\omega^H \mathbf{r}_\omega . \tag{4.4}$$

A smooth beam-pattern can be obtained by choosing $N >> M$ and specifying $\Theta(\theta,\phi,\alpha_n)$ as a Gaussian function [94]. Unfortunately, in some cases the matrix $\mathbf{G}_\omega^H \mathbf{G}_\omega$ is ill-conditioned, and a reconditioning step is needed in order to provide feasible values to the coefficients $\hat{\mathbf{h}}_\omega$. As noticed in [19], a SVD-based reconditioning technique reveals to be suitable for our purposes. With respect to Tikhonov regularization [95], the SVD-based reconditioning is independent from $\mathbf{r}_\omega$, and therefore it can be performed just with the knowledge of loudspeaker and control point locations, before defining the virtual source position. As a consequence, the matrix $\mathbf{G}_\omega^+$ can be stored in advance.

4.3.2 Wideband extension

The least-squares solution in (4.4) depends on the frequency ω. If the virtual source emits a narrow-band signal $s(t)$ centered at the frequency $\bar{\omega}$, the resulting coefficients $\hat{\mathbf{h}}_{\bar{\omega}} = [h_{\bar{\omega},1}, \dots h_{\bar{\omega},M}]^T$ can be directly applied to the loudspeakers to obtain the best approximation of the desired response. However, when the source signal is wide-band, a single-frequency solution is not sufficient for accurately reproducing the beam-pattern. In this case, we have to derive a digital filter to be applied to each loudspeaker, instead of a single complex coefficient.

Let us consider a wide-band source signal sampled at ω_S having frequency com-

ponents limited in the range $[\omega_{\min}, \omega_{\max}]$. We operate a uniform sampling of the frequency axis in $[0, \frac{\omega_S}{2}]$ using a sampling step of Δ_ω. Therefore, by posing $\omega_k = k\Delta_\omega$ with $k = 0, \ldots, L-1$, the frequency response of the loudspeaker filters can be obtained as

$$\hat{\mathbf{h}}_{\omega_k} = \begin{cases} \mathbf{G}^{+}_{\omega_k} \mathbf{r}_{\omega_k} & \text{if } \omega_k \in [\omega_{\min}, \omega_{\max}] \\ \mathbf{0} & \text{otherwise} \end{cases} .$$

The total number L of samples specifies the frequency resolution. As noticed in [94], in order to achieve an accurate wave field reproduction at all the frequencies, a proper choice is to select L such that $\Delta_\omega \leq 2\pi \cdot 40$ rad/s. In order to obtain a real-valued filter in the time domain, the response at the negative frequencies is derived as

$$\hat{\mathbf{h}}_{-\omega_k} = \hat{\mathbf{h}}^{*}_{\omega_k} ,$$

where * denotes the complex conjugation operation.

4.3.3 Rendering of the wave field as multiple beam shaping

We now consider the problem of rendering a virtual source along with the effect of the early reflections generated by the virtual environment surrounding it. As discussed in Section 4.1, the global wave field can be modelled by superposing the direct beam and the beams originated from the reflective paths. Once the planar geometry of the virtual environment and the position of the virtual source are specified, the beam tracing engine introduced in Section 4.2 determines the set of virtual image sources and the relative beams to be rendered. More specifically, the beam tracer specifies each beam with the virtual image source position, the direction and angular aperture. We notice that these parameters give a pure geometrical characterization of a beam. In order to reproduce the effect of a beam as a reflection, we need also to control two other parameters: the amplitude attenuation due to the reflective properties of the virtual environment and the emission delay. More specifically, the beam to be rendered has to be attenuated by a factor that depends on the reflection order of the virtual source and on the reflection coefficients of the virtual environment. Moreover, the beam emission has to be delayed in order to account for the distance between the virtual source and the listening area. The rendering of a complete wave field is finally performed by superposing the loudspeaker signals produced by the individual beams, each one properly scaled and delayed.

4.3.4 Room compensation

So far, we considered the ideal scenario of rendering a wave field in a anechoic room. When the sound reproduction system is operating into a real environment, the effect of the reverberations drastically affects the quality of the rendered wave field, as noticed in [21]. Here we discuss a methodology for performing the room compensation, i.e. a technique for attenuating the effect of the early reflections on the reproduced sound field. This can be done in a complete geometric fashion, exploiting the knowledge of the two-dimensional map of the environment. In order to do so, we employ, once again, the beam tracing engine presented in Section 4.2. In particular, the beam tracing allows us to:

1. determine the set $\{\boldsymbol{x}'_{p_{m,i}}\}_{i=1}^{Q_m}$ of the Q_m image loudspeakers associated to the mth loudspeaker, up to an arbitrary reflection order;

2. evaluate the visibility of the control point $\boldsymbol{x}_{a_n}$ from the image loudspeaker $\boldsymbol{x}'_{p_{m,i}}$; this operation is accomplished through the path-tracing described in Section 4.2.1.

The contribution of the mth loudspeaker to the nth control point can be modelled as follows:

$$G^{\star}_{\omega}(\boldsymbol{x}_{a_n}, \boldsymbol{x}_{p_M}) = G_{\omega}(\boldsymbol{x}_{a_n}|\boldsymbol{x}_{p_M}) + \sum_{i=1}^{Q_m} \beta_{m,i} V(\boldsymbol{x}_{a_n}, \boldsymbol{x}'_{p_{m,i}}) G_{\omega}(\boldsymbol{x}_{a_n}|\boldsymbol{x}'_{p_{m,i}}) \,, \tag{4.5}$$

where $V(\boldsymbol{x}_{a_n}, \boldsymbol{x}'_{p_{m,i}})$ is a binary function that maps the visibility of $\boldsymbol{x}_{a_n}$ from $\boldsymbol{x}'_{p_{m,i}}$; $\beta_{m,i}$ is the attenuation coefficient associated to the image loudspeaker at $\boldsymbol{x}'_{p_{m,i}}$. The value of $\beta_{m,i}$ depends on the reflective properties of the walls and on the reflection order of the image loudspeaker. In (4.5) we recognize a first term which corresponds to the free-field propagation (i.e., the Green's function), and a second term including the effect of all the visible reflective paths generated by the mth loudspeaker at the control point $\boldsymbol{x}_{a_n}$. The propagation matrix in (4.2) becomes

$$\mathbf{G}^{\star}_{\omega} = \begin{bmatrix} G^{\star}_{\omega}(\boldsymbol{x}_{a_1}, \boldsymbol{x}_{p_1}) & \dots & G^{\star}_{\omega}(\boldsymbol{x}_{a_1}, \boldsymbol{x}_{p_M}) \\ \vdots & \ddots & \vdots \\ G^{\star}_{\omega}(\boldsymbol{x}_{a_n}, \boldsymbol{x}_{p_1}) & \dots & G^{\star}_{\omega}(\boldsymbol{x}_{a_n}, \boldsymbol{x}_{p_M}) \end{bmatrix} \,, \tag{4.6}$$

and the corresponding least-squares solution is given by

$$\hat{\mathbf{h}_\omega^\star} = \mathbf{G}_\omega^{\star+}\mathbf{r}_\omega = (\mathbf{G}_\omega^{\star H}\mathbf{G}_\omega^\star)^{-1}\mathbf{G}_\omega^{\star H}\mathbf{r}_\omega \ .$$

It is important to notice how the total number of image loudspeakers, given by $\sum_{m=1}^{M} Q_m$, is determined by the maximum reflection order considered for the beam tracing step. In other words, this means that the proposed room compensation technique counteracts the effect of a limited number of reflections. An increase of the maximum reflection order is therefore expected to improve the compensation, at the expense of a slower and more memory consuming beam tracing step. A practical solution is to consider reflections up to the second or third order, which represents a good trade-off between the amount of compensation and the computation effort. As shown in the next paragraph, in fact, the effect of early reflections is mostly dampened considering image loudspeakers up to the second order of reflection.

4.3.5 Discussion

We now discuss the effect of room compensation in the listening area. For the sake of convenience, we consider a rendering system having a number of loudspeakers equal to the number of control points, i.e. $M = N$. In this case the propagation matrix $\mathbf{G}_\omega^\star$ is square. We also suppose that the reciprocal positions of the loudspeakers and of the control points guarantees that $\mathbf{G}_\omega^\star$ is non-singular. Under these assumptions, the system in (4.1) presents the exact solution

$$\mathbf{h}_\omega^\star = (\mathbf{G}_\omega^\star)^{-1}\mathbf{r}_\omega \ .$$

In order to characterize the effect of room compensation, let us consider the case of having a desired response $\mathbf{r}_\omega = [1 \dots 1]^T$. It is clear that, in this case, $\mathbf{G}_\omega^\star\mathbf{h}_\omega^\star = [1 \dots 1]^T$. This means that $\mathbf{h}_\omega^\star$ compensates for both the effects of the loudspeakers and of the early reflections. In other words, $\mathbf{h}_\omega^\star$ acts as an equalizer of the channels between each loudspeaker and each control point. When a generic desired response $\mathbf{r}_\omega$ is considered, we can interpret the filtering operation as composed of two steps: in a first stage the system performs a deconvolution operation in order to equalize the channels; after that, the system identifies the filters to synthesize the desired response.

In the more general case of $N > M$, the system in (4.1) admits only an approx-

imate solution (the best in the least-squares sense) $\hat{\mathbf{h}}_{\omega}^{\star}$. This means that an exact deconvolution is not possible. However, we expect the system to achieve the best room compensation in the least-squares sense.

For a better comprehension of the effects of room compensation, we show the results of a simple simulation. Consider the rendering of a beam inside a real environment composed of two parallel reflectors. The distance between the reflectors is 5 m and the reflection coefficient is 0.7. The acoustic propagation in the real environment is simulated with the beam tracer, modelling the reflections up to the 10^{th} order. Figure 4.6 shows the response of the beam shaping engine at a point $\boldsymbol{x}_{a_{\bar{n}}}$ in

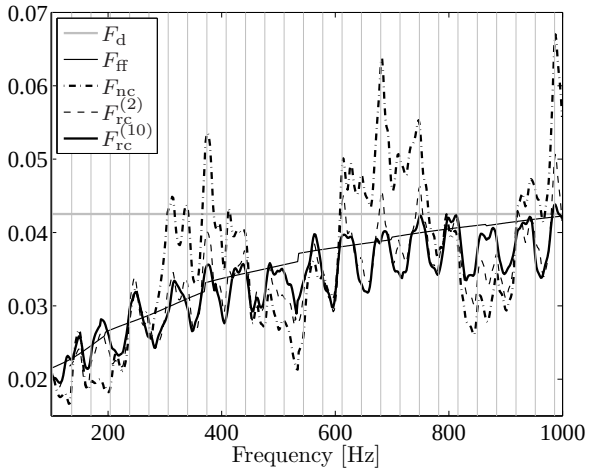


Figure 4.6: Frequency response of the beam shaping engine. The desired response F_{d} is compared with: the free-field response F_{ff}, the non-compensated response F_{nc}, the room-compensated response up to the 2^{nd} order $F_{\text{rc}}^{(2)}$ and the room-compensated response up to the 10^{th} order $F_{\text{rc}}^{(10)}$. The vertical lines denote the room resonant modes.

the center of the listening area. The desired response $F_{\text{d}}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$ is compared with:

- the free-field response, i.e. the frequency response $F_{\text{ff}}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$ of the beam shaping engine in an anechoic environment;
- the non-compensated response $F_{\text{nc}}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$, i.e. the frequency response of the beam shaping engine inside the real environment without compensating for it;
- the response of the room compensation $F_{\text{rc}}^{(2)}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$ performed considering reflections up to the 2^{nd} order;
- the response of the room compensation $F_{\text{rc}}^{(10)}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$ performed considering

reflections up to the 10^{th} order, i.e. the maximum compensation for this simulation.

We observe that the free-field response approaches the desired one as the frequency increases. The noticeable gap in the left part of the curve is due to the poor directivity at low frequencies. On the other hand, we notice that the environment acts on the non-compensated frequency response as a filter, which strongly alters the frequency content of the reproduced wave field. In particular, we notice that the most relevant peaks of $F_{\text{nc}}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$ coincide with a subset of the room resonant modes envisioned by the propagation theory, denoted in Figure 4.6 with vertical lines. Through the room compensation we aim at reducing the effect of the reverberations. When the room is compensated up to the 2^{nd} order, we observe a modest attenuation of the peaks. When the maximum compensation is performed (10^{th} order in this case), the frequency response $F_{\text{rc}}^{(10)}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$ follows the free-field response $F_{\text{ff}}(\omega, \boldsymbol{x}_{a_{\bar{n}}})$. The residual ripple around the free-field response remarks the fact that an exact deconvolution is not possible.

So far we considered the response in a single-point. However, the solution $\hat{\mathbf{h}}_{\omega}^{\star}$ is optimal in the least-squares sense, i.e. it guarantees that the residual $\|\mathbf{r}_{\omega} - \mathbf{G}_{\omega}^{\star}\hat{\mathbf{h}}_{\omega}^{\star}\|^2$ between the desired response and the actual response at the control points is minimum. In order to better appreciate the effect of room compensation, Figure 4.7 shows: the normalized residual[1] R_{ff} of the free-field response; the normalized residual R_{nc} of the non-compensated response; the normalized residual $R_{\text{rc}}^{(2)}$ of the room-compensated response up to the 2^{nd} reflection order; and the normalized residual $R_{\text{rc}}^{(10)}$ of the room-compensated response up to the 10^{th} order. Looking at the non-compensated response R_{nc}, we observe that the effect of early reflections becomes very evident. In fact, the environment acts as a comb filter on the free-field response, whose peaks match the room resonant frequencies, depicted as vertical lines in Figure 4.7. This phenomenon is less evident in the response at $\boldsymbol{x}_{a_{\bar{n}}}$, since some resonant modes may be cancelled out by the zeros of the impulse response at $\boldsymbol{x}_{a_{\bar{n}}}$. We also observe that the room compensation strongly dampens the resonant peaks, especially for the 10^{th} order compensation where the normalized error $R_{\text{rc}}^{(10)}$ approaches the free-field curve R_{ff}.

[1] The normalized residual is calculated as $\frac{\|\mathbf{r}_{\omega} - \mathbf{G}_{\omega}^{\star}\hat{\mathbf{h}}_{\omega}^{\star}\|^2}{\|\mathbf{r}_{\omega}\|^2}$.

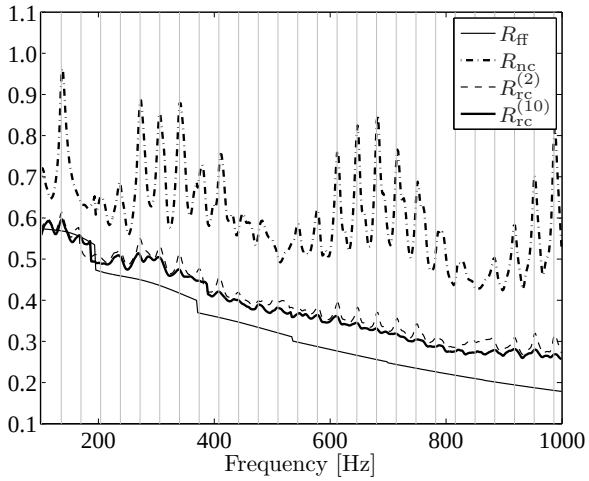


Figure 4.7: Normalized residual of beam shaping at the control points. The curves R_{ff}, R_{nc}, $R_{\mathrm{ff}}^{(2)}$ and $R_{\mathrm{ff}}^{(10)}$ depict the free-field response, the non-compensated response, the 2^{nd} room-compensated response and the 10^{th} room-compensated response, respectively. The vertical lines denote the room resonant modes.

5 Theoretical assessment methodologies and evaluation metrics

In this Chapter we introduce novel methodologies suitable for predicting the accuracy of arbitrary estimation algorithms, as well as for assessing the accuracy of wave field rendering techniques. Such methodologies are introduced in Sections 5.1 and 5.2, respectively, and they will be used for evaluating the analysis algorithms and the rendering technique proposed in Chapter 3 and 4, respectively. Finally, in Section 5.3 we present the evaluation metric used for evaluating the accuracy of the reflector localization techniques proposed in Section 3.4.3.

5.1 Error propagation analysis for estimation algorithms

In this Section we present a novel approach to error propagation analysis that allows us to predict the impact of measurement errors on an arbitrary minimization process. The proposed methodology is based on introductory concepts of Catastrophe Theory [96], and can be applied to a wide range of situations and cost functions. For our purposes, we adopt it for predicting the accuracy of some of the estimation algorithms presented in Chapter 3.

5.1.1 Mathematical derivation

Let $J(\boldsymbol{x}, \boldsymbol{c})$ be a generic cost function, with variables $\boldsymbol{x} = [x_1, \dots x_M]^T$ and parameters $\boldsymbol{c} = [c_1, \dots c_N]^T$. The parameters correspond to the experimental measurements or observations, and the variables represent the object of the estimation. As an example, for the case of source localization using TDOAs (see Section 3.4.2), the variables x_1 and x_2 denote the unknown source position; and the terms $c_1, \dots c_N$ are

given by the measured TDOA values. Let $\boldsymbol{x}_0$ be the correct value of the variable that we intend to determine, and $\boldsymbol{c}_0$ the related error-free measurements. In a real situation we are given noisy measurements $\bar{\boldsymbol{c}} = \boldsymbol{c}_0 + \delta\boldsymbol{c}$, $\delta\boldsymbol{c}$ being the measurement error. Consequently, the global minimum of the cost function moves from $(\boldsymbol{x}_0, \boldsymbol{c}_0)$ to $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{c}})$, where $\bar{\boldsymbol{x}} = \boldsymbol{x}_0 + \delta\boldsymbol{x}$. Assuming the error $\delta\boldsymbol{c}$ to be sufficiently small, we want to determine $\delta\boldsymbol{x}$ through a truncated Taylor expansion of $J(\boldsymbol{x}, \boldsymbol{c})$.

The second-order Taylor series expansion of $J(\boldsymbol{x}, \boldsymbol{c})$, centred at $(\boldsymbol{x}_0, \boldsymbol{c}_0)$, can be written as

$$\begin{aligned} J(\boldsymbol{x}, \boldsymbol{c}) \simeq\ & J|_{\boldsymbol{x}_0,\boldsymbol{c}_0} + (\nabla_{\boldsymbol{x}} J)^T|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\boldsymbol{x} - \boldsymbol{x}_0) + (\nabla_{\boldsymbol{c}} J)^T|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\boldsymbol{c} - \boldsymbol{c}_0) + \\ & + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\boldsymbol{x} - \boldsymbol{x}_0) + \\ & + \frac{1}{2}(\boldsymbol{c} - \boldsymbol{c}_0)^T \boldsymbol{H}_{\boldsymbol{c},\boldsymbol{c}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\boldsymbol{c} - \boldsymbol{c}_0) + \\ & + (\boldsymbol{c} - \boldsymbol{c}_0)^T \boldsymbol{H}_{\boldsymbol{c},\boldsymbol{x}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\boldsymbol{x} - \boldsymbol{x}_0) \,, \end{aligned} \tag{5.1}$$

where

$$\nabla_{\boldsymbol{x}} J = [J_{x_1}, \dots J_{x_M}]^T \,, \qquad \nabla_{\boldsymbol{c}} J = [J_{c_1}, \dots J_{c_N}]^T \,,$$

$$J_{x_i} = \frac{\partial J}{\partial x_i} \,, \qquad J_{c_j} = \frac{\partial J}{\partial c_j} \,,$$

and

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J) = \begin{bmatrix} J_{x_1x_1} & \cdots & J_{x_1x_M} \\ \vdots & \ddots & \vdots \\ J_{x_Mx_1} & \cdots & J_{x_Mx_M} \end{bmatrix} \,, \qquad \boldsymbol{H}_{\boldsymbol{c},\boldsymbol{c}}(J) = \begin{bmatrix} J_{c_1c_1} & \cdots & J_{c_1c_N} \\ \vdots & \ddots & \vdots \\ J_{c_Nc_1} & \cdots & J_{c_Nc_N} \end{bmatrix} \,,$$

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J) = \begin{bmatrix} J_{x_1c_1} & \cdots & J_{x_1c_N} \\ \vdots & \ddots & \vdots \\ J_{x_Mc_1} & \cdots & J_{x_Mc_N} \end{bmatrix} \,,$$

with

$$J_{x_ix_j} = \frac{\partial^2 J}{\partial x_i \partial x_j} \,, \qquad J_{x_ic_j} = \frac{\partial^2 J}{\partial x_i \partial c_j} \,, \qquad J_{c_ic_j} = \frac{\partial^2 J}{\partial c_i \partial c_j} \,.$$

Notice that $(\nabla_{\boldsymbol{x}} J)^T|_{\boldsymbol{x}_0,\boldsymbol{c}_0} = 0$, as the function with the correct parameters $\boldsymbol{c}_0$ has a minimum in $\boldsymbol{x}_0$.

Now we study $\nabla_{\boldsymbol{x}} J(\boldsymbol{x}, \boldsymbol{c})|_{\bar{\boldsymbol{x}},\bar{\boldsymbol{c}}} = 0$ to search for the new minimum $\bar{\boldsymbol{x}}$. Deriving the

Taylor series expansion in (5.1) we obtain

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\bar{\boldsymbol{x}} - \boldsymbol{x}_0) + \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}(\bar{\boldsymbol{c}} - \boldsymbol{c}_0) = 0 \; .$$

Observing that $\bar{\boldsymbol{x}} - \boldsymbol{x}_0 = \delta\boldsymbol{x}$ and $\bar{\boldsymbol{c}} - \boldsymbol{c}_0 = \delta\boldsymbol{c}$, the above equation becomes

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}\delta\boldsymbol{x} + \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}\delta\boldsymbol{c} = 0 \; ,$$

and finally we can write

$$\delta\boldsymbol{x} = -\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J)|^{-1}_{\boldsymbol{x}_0,\boldsymbol{c}_0} \cdot \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}\delta\boldsymbol{c} \; . \tag{5.2}$$

5.1.2 Discussion

The above analysis relies on the hypothesis that we can truncate the Taylor series expansion of J to the second order. This is, of course, true if $\delta\boldsymbol{x}$ and $\delta\boldsymbol{c}$ can be assumed as being sufficiently small. In order to provide a theoretical upper bound for $\delta\boldsymbol{c}$, we need a more in-depth study of the Taylor expansion, for example, through the analysis of the Lagrange remainder of the Taylor expansion.

It is also important to notice that (5.2) does not envision the possibility of estimation bias, as $\delta\boldsymbol{x}$ and $\delta\boldsymbol{c}$ are linearly related. As a consequence, the proposed error propagation analysis is applicable whenever the estimation bias is negligible.

Finally, (5.2) is valid under the hypothesis that $\det(\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J))|_{\boldsymbol{x}_0,\boldsymbol{c}_0} \neq 0$. Mathematically this condition means that J should have an isolated non-degenerate minimum at $(\boldsymbol{x}_0, \boldsymbol{c}_0)$.

5.1.3 Statistical error analysis

In a real scenario we cannot assume the measurement noise $\delta\boldsymbol{c}$ to be known. However, some statistical information could be available or could be estimated from data. It is therefore important to find a relation between statistical descriptors of the noise $\delta\boldsymbol{c}$ and of $\delta\boldsymbol{x}$. In this paragraph, we provide examples for the case of zero-mean Gaussian error on $\delta\boldsymbol{c}$. We define $\boldsymbol{A} = -\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J)|^{-1}_{\boldsymbol{x}_0,\boldsymbol{c}_0} \cdot \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J)|_{\boldsymbol{x}_0,\boldsymbol{c}_0}$, such that (5.2) can be rewritten as

$$\delta\boldsymbol{x} = \boldsymbol{A}\delta\boldsymbol{c} \; .$$

The relationship between the covariance matrix $\Sigma_{\boldsymbol{x}}$ of the localization error and the covariance matrix $\Sigma_{\boldsymbol{c}}$ of the measurements is

$$\Sigma_{\boldsymbol{x}} = \boldsymbol{A}\Sigma_{\boldsymbol{c}}\boldsymbol{A}^T \,, \tag{5.3}$$

where

$$\Sigma_{\boldsymbol{c}} = \begin{bmatrix} \sigma_{c_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{c_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{c_N}^2 \end{bmatrix}, \qquad \Sigma_{\boldsymbol{x}} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_M} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2x_M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_Mx_1} & \sigma_{x_Mx_2} & \cdots & \sigma_{x_M}^2 \end{bmatrix}$$

under the assumption of statistical independence of measurement errors.

5.1.4 Error propagation analysis and Cramer-Rao Lower Bound

We now investigate the link between the proposed error propagation analysis and the standard analysis tool given by the Cramer-Rao Lower Bound (CRLB). More specifically, we aim at showing that CRLB is a particular case of the error propagation analysis applied to a Maximum-Likelihood (ML) cost function. It is well known in the literature (see for example [97, 23, 98]) that ML estimation techniques asymptotically attain the CRLB. As a consequence, for the presented theory to be valid, the results of the error propagation analysis performed onto a ML cost function must match the CRLB of that problem.

In order to demonstrate this fact, we first derive the ML cost function of a generic estimation problem. We notice that the noise-free measurements can be modelled as

$$\boldsymbol{c}_0 = \boldsymbol{f}(\boldsymbol{x}_0) \triangleq \begin{bmatrix} f_1(\boldsymbol{x}_0) \\ f_2(\boldsymbol{x}_0) \\ \vdots \\ f_N(\boldsymbol{x}_0) \end{bmatrix} \,,$$

where $\boldsymbol{f} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ is a multidimensional function that maps the variables to be estimated, $\boldsymbol{x}_0$, to the observations $\boldsymbol{c}_0$. In real situations, the measurement are corrupted with some measurement error. We denote the noisy measurements with

$$\bar{\boldsymbol{c}} = \boldsymbol{f}(\boldsymbol{x}_0) + \delta\boldsymbol{c} \,.$$

The error $\delta \boldsymbol{c} \in \mathbb{R}^N$ is commonly modelled [2] as a multivariate zero-mean Gaussian distribution with covariance matrix $\Sigma_{\boldsymbol{c}}$, i.e. $\delta \boldsymbol{c} \sim \mathcal{N}(0, \Sigma_{\boldsymbol{c}})$. The probability that the measurements $\bar{\boldsymbol{c}} = [\bar{c}_1, \dots \bar{c}_N]^T$ are generated from a generic point $\boldsymbol{x} = [x_1, \dots x_M]^T$ is therefore given by

$$p(\bar{\boldsymbol{c}}, \boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma_{\boldsymbol{c}})}} e^{-\frac{1}{2}\left\{[\bar{\boldsymbol{c}} - \boldsymbol{f}(\boldsymbol{x})]^T \Sigma_{\boldsymbol{c}} [\bar{\boldsymbol{c}} - \boldsymbol{f}(\boldsymbol{x})]\right\}} , \tag{5.4}$$

which corresponds to the *likelihood function* [23]. The ML estimator is defined as [23]

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \mathcal{L}(\bar{\boldsymbol{c}}, \boldsymbol{x}) ,$$

where $\mathcal{L}(\bar{\boldsymbol{c}}, \boldsymbol{x})$ is the *log-likelihood* function [23], which is given by

$$\mathcal{L}(\bar{\boldsymbol{c}}, \boldsymbol{x}) = \ln p(\bar{\boldsymbol{c}}, \boldsymbol{x}) = -\ln \sqrt{(2\pi)^N \det(\Sigma_{\boldsymbol{c}})} - \frac{1}{2}\left\{[\bar{\boldsymbol{c}} - \boldsymbol{f}(\boldsymbol{x})]^T \Sigma_{\boldsymbol{c}} [\bar{\boldsymbol{c}} - \boldsymbol{f}(\boldsymbol{x})]\right\} . \tag{5.5}$$

The first term of the right-hand side of (5.5) is constant, therefore the estimation problem can be rewritten as

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \left(-\frac{1}{2}\left\{[\bar{\boldsymbol{c}} - \boldsymbol{f}(\boldsymbol{x})]^T \Sigma_{\boldsymbol{c}} [\bar{\boldsymbol{c}} - \boldsymbol{f}(\boldsymbol{x})]\right\}\right) ,$$

or equivalently as

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} J_{\mathrm{ML}}(\boldsymbol{x}, \boldsymbol{c} = \bar{\boldsymbol{c}}) ,$$

where

$$J_{\mathrm{ML}}(\boldsymbol{x}, \boldsymbol{c}) = \frac{1}{2}\left\{[\boldsymbol{c} - \boldsymbol{f}(\boldsymbol{x})]^T \Sigma_{\boldsymbol{c}} [\boldsymbol{c} - \boldsymbol{f}(\boldsymbol{x})]\right\} \tag{5.6}$$

denotes the ML cost function.

The CRLB of the estimated variables covariance is given by the inverse of the Fisher Information Matrix [23], whose $[i, j]$ element is

$$[\boldsymbol{I}(\boldsymbol{x})]_{ij} \triangleq -E\left[\frac{\partial^2 \mathcal{L}(\bar{\boldsymbol{c}}, \boldsymbol{x})}{\partial x_i \partial x_j}\right] .$$

For a Gaussian distribution, the Fisher Information Matrix becomes [2]

$$\boldsymbol{I}(\boldsymbol{x}) = \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T \Sigma_{\boldsymbol{c}}^{-1} \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right] , \tag{5.7}$$

where

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1} \quad \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_2} \quad \cdots \quad \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_M} \right], \qquad \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_i} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial x_i} \\ \vdots \\ \frac{\partial f_N(\boldsymbol{x})}{\partial x_i} \end{bmatrix}.$$

We now apply the proposed error propagation analysis to the ML cost function (5.6). The covariance matrix $\Sigma_{\boldsymbol{x}}$ of the estimation error can be computed using (5.3). In this case we have that

$$\Sigma_{\mathrm{ML},\boldsymbol{x}} = \boldsymbol{A}_{\mathrm{ML}} \Sigma_{\boldsymbol{c}} \boldsymbol{A}_{\mathrm{ML}}^T \tag{5.8}$$

where $\boldsymbol{A}_{\mathrm{ML}} = -\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})|^{-1}_{\boldsymbol{x}_0,\boldsymbol{c}_0} \cdot \boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J_{\mathrm{ML}})|_{\boldsymbol{x}_0,\boldsymbol{c}_0}$. The $[i,j]$ element of $\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})$ is given by

$$\begin{aligned} [\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})]_{ij} &= \frac{\partial^2 J_{\mathrm{ML}}(\boldsymbol{x},\boldsymbol{c})}{\partial x_i \partial x_j} = \\ &= \frac{\partial}{\partial x_i} \left\{ - \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_j} \right]^T \Sigma_{\boldsymbol{c}}^{-1} [\boldsymbol{c} - \boldsymbol{f}(\boldsymbol{x})] - [\boldsymbol{c} - \boldsymbol{f}(\boldsymbol{x})]^T \Sigma_{\boldsymbol{c}}^{-1} \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_j} \right] \right\}. \end{aligned}$$

The covariance matrix $\Sigma_{\boldsymbol{c}}$ is symmetric by definition, and therefore the above expression can be simplified as

$$\begin{aligned} [\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})]_{ij} &= -2 \frac{\partial}{\partial x_i} \left\{ \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_j} \right]^T \Sigma_{\boldsymbol{c}}^{-1} [\boldsymbol{c} - \boldsymbol{f}(\boldsymbol{x})] \right\} = \\ &= -2 \left\{ \left[\frac{\partial^2 \boldsymbol{f}(\boldsymbol{x})}{\partial x_i \partial x_j} \right]^T \Sigma_{\boldsymbol{c}}^{-1} [\boldsymbol{c} - \boldsymbol{f}(\boldsymbol{x})] - \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_j} \right]^T \Sigma_{\boldsymbol{c}}^{-1} \left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i} \right] \right\}. \end{aligned} \tag{5.9}$$

In a similar way we compute the matrix $\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J_{\mathrm{ML}})$, whose $[i,j]$ element is

$$\begin{aligned}
[\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J_{\mathrm{ML}})]_{ij} &= \frac{\partial^2 J_{\mathrm{ML}}(\boldsymbol{x},\boldsymbol{c})}{\partial x_i \partial c_j} = \\
&= \frac{\partial}{\partial x_i}\left\{\left[\frac{\partial \boldsymbol{c}}{\partial c_j}\right]^T \Sigma_{\boldsymbol{c}}^{-1}\left[\boldsymbol{c}-\boldsymbol{f}(\boldsymbol{x})\right] + \left[\boldsymbol{c}-\boldsymbol{f}(\boldsymbol{x})\right]^T \Sigma_{\boldsymbol{c}}^{-1}\left[\frac{\partial \boldsymbol{c}}{\partial c_j}\right]\right\} = \\
&= 2\frac{\partial}{\partial x_i}\left\{\left[\boldsymbol{c}-\boldsymbol{f}(\boldsymbol{x})\right]^T \Sigma_{\boldsymbol{c}}^{-1}\left[\frac{\partial \boldsymbol{c}}{\partial c_j}\right]\right\} = \\
&= -2\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i}\right]^T \Sigma_{\boldsymbol{c}}^{-1}\left[\frac{\partial \boldsymbol{c}}{\partial c_j}\right] . \qquad (5.10)
\end{aligned}$$

Equations (5.9) and (5.10) have to be evaluated at $(\boldsymbol{x}=\boldsymbol{x}_0, \boldsymbol{c}=\boldsymbol{c}_0)$. We observe that $\boldsymbol{f}(\boldsymbol{x}_0)=\boldsymbol{c}_0$, and therefore $[\boldsymbol{c}-\boldsymbol{f}(\boldsymbol{x})]|_{\boldsymbol{x}_0,\boldsymbol{c}_0}=\boldsymbol{0}$. As a consequence, we obtain

$$[\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})|_{\boldsymbol{x}_0,\boldsymbol{c}_0}]_{ij} = 2\left\{\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_j}\right]^T \Sigma_{\boldsymbol{c}}^{-1}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_i}\right]\right\}\Bigg|_{\boldsymbol{x}_0,\boldsymbol{c}_0} ,$$

which leads to

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})|_{\boldsymbol{x}_0,\boldsymbol{c}_0} = 2\left\{\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T \Sigma_{\boldsymbol{c}}^{-1}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]\right\}\Bigg|_{\boldsymbol{x}_0,\boldsymbol{c}_0} . \qquad (5.11)$$

In the right-hand side of (5.11) we recognize the Fisher Information Matrix (5.7), and therefore we can finally write

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{x}}(J_{\mathrm{ML}})|_{\boldsymbol{x}_0,\boldsymbol{c}_0} = 2\,\boldsymbol{I}(\boldsymbol{x})|_{\boldsymbol{x}_0,\boldsymbol{c}_0} = 2\boldsymbol{I}(\boldsymbol{x}_0) . \qquad (5.12)$$

From (5.10) we readily obtain

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J_{\mathrm{ML}}) = -2\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T \Sigma_{\boldsymbol{c}}^{-1} ,$$

and therefore

$$\boldsymbol{H}_{\boldsymbol{x},\boldsymbol{c}}(J_{\mathrm{ML}})|_{\boldsymbol{x}_0,\boldsymbol{c}_0} = -2\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T\Bigg|_{\boldsymbol{x}_0,\boldsymbol{c}_0} \Sigma_{\boldsymbol{c}}^{-1} . \qquad (5.13)$$

Inserting (5.12) and (5.13) into (5.8) we obtain

$$\begin{aligned}\Sigma_{\mathrm{ML},\boldsymbol{x}} &= \left\{-[2\boldsymbol{I}(\boldsymbol{x}_0)]^{-1}\right\}\left\{-2\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T\Bigg|_{\boldsymbol{x}_0,\boldsymbol{c}_0}\Sigma_{\boldsymbol{c}}^{-1}\right\}\Sigma_{\boldsymbol{c}}\times \\ &\quad\times\left\{-2\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T\Bigg|_{\boldsymbol{x}_0,\boldsymbol{c}_0}\Sigma_{\boldsymbol{c}}^{-1}\right\}\left\{-[2\boldsymbol{I}(\boldsymbol{x}_0)]^{-T}\right\}\ . \end{aligned} \tag{5.14}$$

Since the Fisher Information Matrix is symmetric [23], its inverse is symmetric as well and therefore $[\boldsymbol{I}(\boldsymbol{x}_0)]^{-T} = [\boldsymbol{I}(\boldsymbol{x}_0)]^{-1}$. Consequently, after some passages (5.14) simplifies to

$$\Sigma_{\mathrm{ML},\boldsymbol{x}} = [\boldsymbol{I}(\boldsymbol{x}_0)]^{-1}\left\{\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T\Sigma_{\boldsymbol{c}}^{-1}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]\right\}\Bigg|_{\boldsymbol{x}_0,\boldsymbol{c}_0}[\boldsymbol{I}(\boldsymbol{x}_0)]^{-1}\ . \tag{5.15}$$

In (5.15) we recognize again the Fisher Information Matrix, and therefore we finally obtain

$$\Sigma_{\mathrm{ML},\boldsymbol{x}} = [\boldsymbol{I}(\boldsymbol{x}_0)]^{-1}\boldsymbol{I}(\boldsymbol{x}_0)[\boldsymbol{I}(\boldsymbol{x}_0)]^{-1} = [\boldsymbol{I}(\boldsymbol{x}_0)]^{-1}\ . \tag{5.16}$$

Equation (5.16) proves that the error propagation analysis applied to the ML cost function (5.6) is equivalent to the CRLB of the corresponding estimation problem, which is defined as the inverse of the Fisher Information Matrix evaluated at $\boldsymbol{x} = \boldsymbol{x}_0$.

5.2 Measuring the accuracy of wave field rendering

Sound field reproduction methodologies, such as Wave Field Synthesis [6] and Higher-Order Ambisonics [7] are generally studied and developed under specific assumptions. The typical scenario consists of a distribution of ideal loudspeakers mounted in a completely anechoic acoustic environment. The accuracy of these systems is therefore predicted in a theoretical fashion by modeling the wave propagation inside the same ideal scenario (see for example [21, 16, 99, 100, 101]). When the rendering methodologies are implemented on a real reproduction system, however, the non-ideality of both loudspeakers and environment may alter the quality of the rendered wave field. In order to evaluate this degradation, a methodology for accurately measuring the reproduced wave field inside the listening area is required.

This task is not trivial, because it implies a space-time sampling of the wave field within a wide region. If the sampling is performed with a single microphone moving inside the listening area, the number of acquisitions required to obtain a sufficiently dense reconstruction of the wave field makes this solution impracticable. On the other hand, if a regular grid composed of many microphones is used for sampling the wave field, their presence significantly alters the measured wave field. Furthermore, one has to be aware of the risk that the error introduced by the measurement methodology exceeds the degradation of the rendered wave field caused by the non-ideality of the reproduction system.

In this Section we consider a well-established technique for the measurement of a two-dimensional wave field. As presented in [24] the wave field is sampled over a circle by means of a pair of rotating microphones (an omnidirectional one and a figure-of-eight one). The wave field inside and outside the circle is extrapolated exploiting the Circular Harmonic Decomposition [102]. In order to compare target, theoretical and measured wave fields using synthetic parameters, we define some evaluation metrics. More specifically, we first introduce two MSE metrics based on the difference between the theoretical and the measured wave fields. Although useful, these metrics are not able to distinguish at which propagation modes the error between the measured and theoretical wave field is introduced. In order to circumvent this problem, we also define two novel evaluation metrics that rely on a modal analysis based on the two-dimensional discrete Fourier transform.

5.2.1 Problem formulation

A generic rendering system, shown in Fig. 5.1, is composed by an arbitrary distribution of M loudspeakers located at $\boldsymbol{x}_{p_1}, \dots \boldsymbol{x}_{p_M}$. The goal of this system is to

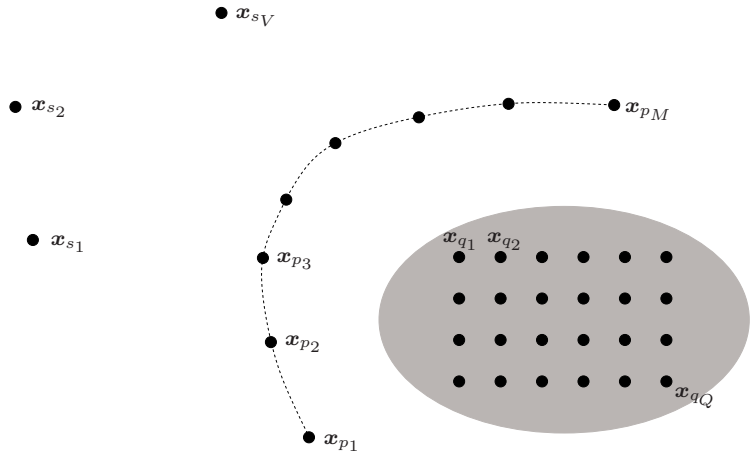


Figure 5.1: A very general model of a rendering system.

reproduce the wave field generated by a set of V virtual point sources located at $\boldsymbol{x}_{s_1}, \dots \boldsymbol{x}_{s_V}$ by means of the loudspeakers. The rendering system is designed for reproducing the target wave field inside the listening area, denoted by the gray-shaded region shown in Fig. 5.1. We define an evaluation region composed of a set of Q points $\boldsymbol{x}_{q_1}, \dots \boldsymbol{x}_{q_Q}$ regularly distributed inside the listening area. Let $S_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})$ and $S_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})$ be the Fourier transforms of the target and theoretical wave fields at the point $\boldsymbol{x}_{q_i}$, respectively. Using the wave field measuring technique described in Section 5.2.2 we acquire the measured wave field $S_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i})$.

Our goal is to introduce some methodologies to evaluate the differences between $S_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})$, $S_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})$ and $S_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i})$.

5.2.2 Measurement of the wavefield

In this Section we summarize the methodology for the measurement of the wave field presented in [24]. Adopting a polar coordinate system (radius ρ, angle ϕ), the time-domain Fourier transform of the sound pressure can be written as an angular

Fourier series, whose coefficients can be further decomposed into circular harmonics

$$\begin{aligned} P(\omega,\rho,\phi) &= \sum_{\mu=-\infty}^{\infty} \mathring{P}_\mu(\omega,\rho)e^{j\mu\phi} \\ &= \sum_{\mu=-\infty}^{\infty} C_\mu(\omega)J_\mu(k\rho)e^{j\mu\phi} \ , \end{aligned} \tag{5.17}$$

where $C_\mu(\omega)$ represents the μth circular harmonic at frequency ω; $k = \omega/c$ is the wave number, c being the sound speed; J_μ is the Bessel function of the first kind and order μ. We notice that the circular harmonic coefficients depend only on the angular mode μ and the frequency ω. The knowledge of the sound pressure $P(\omega,\rho_0,\phi)$ on a circle with radius ρ_0 is therefore, in principle, sufficient to determine the wave field $P(\omega,\rho,\phi)$ inside and outside the circle, by adopting the following equalization:

$$C_\mu(\omega) = \frac{1}{J_\mu(k\rho_0)}\mathring{P}_\mu(\omega,\rho_0) \ .$$

However, the zeros of the Bessel functions at the denominator of (5.18) make the direct implementation of this procedure impracticable. This issue is faced in [24] with the knowledge of the pressure and its gradient on the circle ρ_0. The combination of the signals acquired by the two sensors leads to a modified version of (5.18), whose associated equalization function is free of zeros [24]. Moreover, the measurement of the wave field over the circle with radius ρ_0 in infinite positions is not feasible. A spatial sampling of $P(\omega,\phi)$ over the circle is therefore adopted in [24] through the rotating rig shown in Figure 5.2. The radius of the circle described by the arm is $\rho_0 = 0.74$ cm. The position of the arm is controlled by a stepper motor that, for every complete rotation, stops at Z intermediate positions, under the assumption that the wave field is stationary during the rotation of the arm. This solution allows

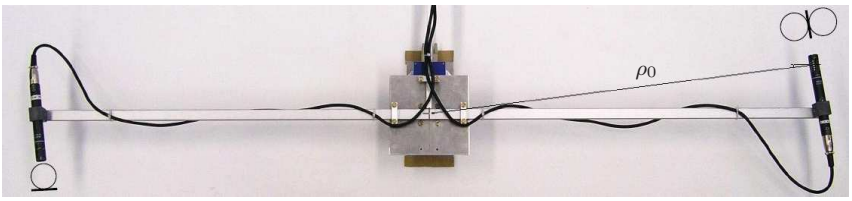


Figure 5.2: The virtual array. An omnidirectional and a figure-of-eight microphones are used to emulate a cardioid microphone, as described in [24].

to sample the measurement circle with high angular resolution (up to $Z = 4000$ positions can be used), keeping thus moderate the effect of modal aliasing introduced by the spatial sampling. Finally, we notice that the wave field $P(\omega, \rho_0, \phi)$ is measured on a circle with radius ρ_0, while the listening points $\boldsymbol{x}_{q_i}$ are located in arbitrary positions. Therefore an interpolation or extrapolation as described by (5.17) and (5.18) is required to map $P(\omega, \rho, \phi)$ on the listening points $\boldsymbol{x}_{q_i}$. The final result of the measurement procedure is the measured wave field $S_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i})$.

5.2.3 Evaluation methodology

In this Section we introduce the metrics used to evaluate the accuracy of the rendering technique. The evaluation methodology proposed here distinguishes between the target, theoretical and the measured wave field. The target wave field is the desired outcome of the reproduction (e.g. a plane wave). The theoretical wave field results from an approximation of the target wave field with a finite number of ideal loudspeakers and a given rendering technique. The measured wave field is the one resulting from acoustic measurements in a real-world rendering situation. First we introduce an evaluation metric that synthetically describes the accuracy of the theoretical and measured wave fields with respect to the target one. Later on we define a new evaluation metric that gives further insight into the evaluation of the accuracy, by performing a modal analysis on target, theoretical and measured wave fields.

5.2.3.1 Normalized mean square error metrics

As a first step we normalize the target wave field $S_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})$ in order to make its energy unity. The resulting normalization coefficient $W_{\mathrm{des}}(\omega)$ is used also to normalize the theoretical wave field. The resulting normalized wave fields are

$$\bar{S}_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i}) = \frac{S_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})}{W_{\mathrm{des}}(\omega)} , \qquad \bar{S}_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i}) = \frac{S_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})}{W_{\mathrm{des}}(\omega)} ,$$

where

$$W_{\mathrm{des}}(\omega) = \sqrt{\frac{1}{Q} \sum_{i=1}^{Q} |S_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})|^2} . \tag{5.18}$$

This way, the target wave field is forced to have unitary energy in the evaluation region, independently on the set of virtual sources to be reproduced. As a consequence, the results related to different distributions of virtual sources can be directly

compared.

Using the normalization in (5.18), we derive an evaluation metric based on the mean square error between $\bar{S}_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})$ and $\bar{S}_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})$. This is a common approach for the theoretical evaluation of rendering techniques [103, 100, 104]. In particular, the quantity

$$E_{\mathrm{T}}(\omega) = \frac{1}{Q} \sum_{i=1}^{Q} |\bar{S}_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i}) - \bar{S}_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})|^2 \, . \tag{5.19}$$

corresponds to the Normalized Mean Square Error (NMSE) as defined in [100]. The metric in (5.19) synthetically quantifies the accuracy obtained by the theoretical wave field in reproducing the target one.

As far as the measured wave field is concerned, one could define an evaluation metric analogous to the one introduced for the theoretical evaluation. However, the measured wave field cannot be directly compared with the target wave field because its amplitude depends on the microphone gains and the loudspeaker volumes. Therefore, a normalization of the measured wave field is in order. More specifically, we propose to normalize it in such a way that the global energy within the evaluation region equals the energy of the theoretical wave field. In other words

$$\bar{S}_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i}) = \frac{S_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i}) \sqrt{\frac{1}{Q} \sum_{i=1}^{Q} |\bar{S}_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})|^2}}{\sqrt{\frac{1}{Q} \sum_{i=1}^{Q} |S_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})|^2}} \, .$$

This way, the NMSE between the theoretical and measured wave fields becomes

$$E_{\mathrm{M}}(\omega) = \frac{1}{Q} \sum_{i=1}^{Q} |\bar{S}_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i}) - \bar{S}_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i})|^2 \, . \tag{5.20}$$

5.2.3.2 Modal analysis

We notice that the evaluation metrics in (5.19) and (5.20) do not give insight on the distribution of the error for different wave fronts in the sound field. In order to address this issue, we perform a two-dimensional discrete Fourier transform on $\Re\{\bar{S}_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})\}$, $\Re\{\bar{S}_{\mathrm{T}}(\omega, \boldsymbol{x}_{q_i})\}$ and $\Re\{\bar{S}_{\mathrm{M}}(\omega, \boldsymbol{x}_{q_i})\}$ to obtain $\hat{S}^{\mathrm{c}}_{\mathrm{des}}(k_x, k_y)$, $\hat{S}^{\mathrm{c}}_{\mathrm{T}}(k_x, k_y)$ and $\hat{S}^{\mathrm{c}}_{\mathrm{M}}(k_x, k_y)$, respectively. The variables (k_x, k_y) are the spatial frequency bin coordinates. We omit the variable ω in the notation for the sake of compactness. A planar wavefront propagating in the environment is compactly described by the

wavenumber k, inversely proportional to the wavelength, and by its direction of propagation φ. After some arithmetic we find that k and φ are related to k_x and k_y by

$$k = \frac{2\pi}{D}\sqrt{k_x^2 + k_y^2} \tag{5.21}$$

$$\varphi = \arctan k_y/k_x \ , \tag{5.22}$$

where D is the diameter of the area under analysis, in meters. For the sake of clarity, Figure 5.3 shows an example of target wave field on the left-hand side and the corresponding two-dimensional discrete Fourier transform on the right-hand side. We obtain therefore from $\hat{S}^{\mathrm{c}}_{\mathrm{des}}(k_x, k_y)$ the function $\hat{S}^{\mathrm{p}}_{\mathrm{des}}(k, \varphi)$ using the change of

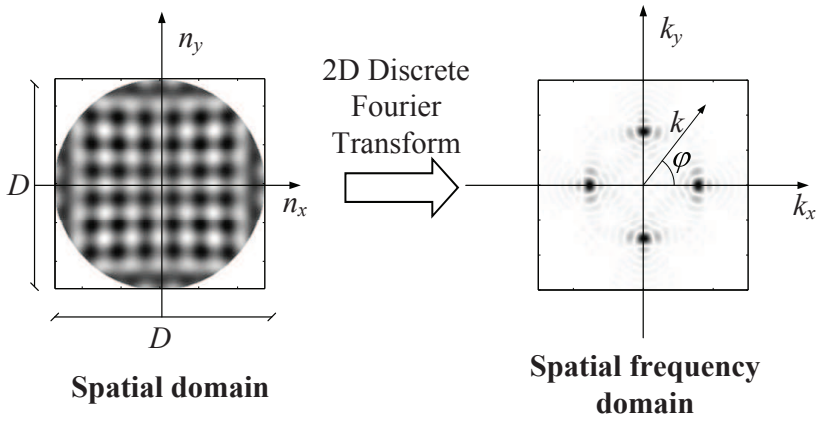


Figure 5.3: An example of target wave field and the corresponding two-dimensional Fourier transform.

variables in (5.21) and (5.22). We derive from $\hat{S}^{\mathrm{p}}_{\mathrm{des}}(k, \varphi)$ two functions, $\hat{S}'_{\mathrm{des}}(k)$ and $\hat{S}''_{\mathrm{des}}(\varphi)$. The function $\hat{S}'_{\mathrm{des}}(k)$ aims at finding the distribution of the wavenumbers of the propagating wavefronts, regardless of their propagation direction and it is defined as

$$\hat{S}'_{\mathrm{des}}(k) = \sqrt{\sum_{\varphi} |\hat{S}_{\mathrm{des}}(k, \varphi)|^2} \ . \tag{5.23}$$

In other words $\hat{S}'_{\mathrm{des}}(k)$ sums all the wavefronts in $\hat{S}^{\mathrm{p}}_{\mathrm{des}}(k, \varphi)$, which are characterized by the same wavenumber. Under the hypothesis that the loudspeaker array is rendering a monochromatic signal, however, we obtain a wavenumber distribution characterized by a dominant wavenumber at which almost all the energy of the wavefront is contained. In order to distinguish wavefronts with the same wavenum-

ber but with different directions of propagation, we define the function $\hat{S}''_{\mathrm{des}}(\varphi)$. The dominant wavenumber is found as

$$\hat{k} = \arg\max_{k} \hat{S}'_{\mathrm{des}}(k) \; . \tag{5.24}$$

The metric $\hat{S}''_{\mathrm{des}}(\varphi)$ assesses the distribution of the directions of propagation for the dominant wave number

$$\hat{S}''_{\mathrm{des}}(\varphi) = \hat{S}^{\mathrm{p}}_{\mathrm{des}}(\hat{k}, \varphi) \; . \tag{5.25}$$

Notice that, due to the symmetry properties of the spatial Fourier transform of $\Re\{\bar{S}_{\mathrm{des}}(\omega, \boldsymbol{x}_{q_i})\}$, the function $\hat{S}''_{\mathrm{des}}(\varphi)$ is periodic with period π. This means that $\hat{S}''_{\mathrm{des}}(\varphi)$ is fully determined for $\varphi \in [0, \pi]$ or $\varphi \in [-\pi, \pi]$. In other words, $\hat{S}''_{\mathrm{des}}(\varphi)$ is not able to discriminate between inward and outward propagating wave components.

Adopting the same procedure described above, we derive $\hat{S}'_{\mathrm{T}}(k)$ and $\hat{S}''_{\mathrm{T}}(\varphi)$ from $\hat{S}_{\mathrm{T}}(k_x, k_y)$; and $\hat{S}'_{\mathrm{M}}(k)$ and $\hat{S}''_{\mathrm{M}}(\varphi)$ from $\hat{S}_{\mathrm{M}}(k_x, k_y)$. The comparison of the two functions allows us to analyze the differences for each wavenumber and, limited to the dominant wavenumber, for each direction of propagation between theoretical, measured and target wave fields.

5.3 Evaluation metric for reflector localization algorithms

In order to asses the accuracy of reflector localization, we adopt a polar representation of reflector lines. With reference to Figure 5.4, a line is defined by its distance ρ and angle α with respect to the origin, and therefore its parameter vector is $\mathbf{l} = [\cos\alpha,\ \sin\alpha,\ -\rho]^T$. Denoting with (α, ρ) and $(\hat{\alpha}, \hat{\rho})$ the actual and estimated

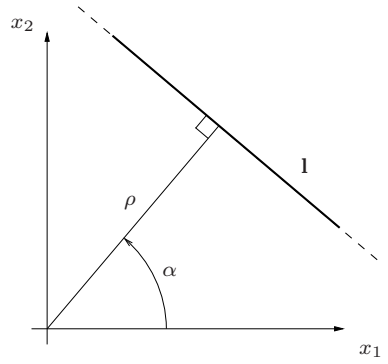


Figure 5.4: Polar representation of a line.

reflector line parameters, we define the following metrics:

- distance error $\varepsilon_\rho = \rho - \hat{\rho}$;
- angular error $\varepsilon_\alpha = \langle \alpha, \hat{\alpha} \rangle_\pi$, where $\langle \alpha, \hat{\alpha} \rangle_\pi$ denotes the signed angular distance in the range $[0, \pi]$ between α and $\hat{\alpha}$.

6 Simulations and experiments

In this Chapter we evaluate the accuracy of the most relevant algorithms presented in this thesis, considering the application scenario introduced in Chapter 1. In particular, we consider a geometric wave field rendering system operating in a reverberant room. As discussed in Section 4.3.4, this system requires the knowledge of the geometry of the real environment in which it operates. More precisely, the position of the walls are used for determining the location of the image loudspeakers that appear from acoustic reflections. This information is then employed for modelling the acoustic propagation from the loudspeakers to the listening area, making possible the compensation of the real environment. For this reasons, simulations and experiments are focused on the evaluation of

- room geometry inference techniques introduced in Section 3.4.3;
- the wave field rendering methodology presented in Chapter 4.

This chapter is organized as follows. In Section 6.1 we test the effectiveness of room geometry inference algorithms based on TOA and DOA measurements. In particular, as far as simulations are concerned, we consider both standard Monte-Carlo simulations and the error propagation analysis described in Section 5.1. As for experiments, we test the accuracy of the algorithms in localizing one or more acoustic reflectors. In Section 6.2 we first evaluate the rendering technique by simulating anechoic and reverberant environments. After that, we analyse the accuracy of the rendering working on a real reproduction system. To do so, we consider measured wave fields obtained adopting the wave field measuring methodology outlined in Section 5.2.2.

6.1 Inference of the environment

In this Section we evaluate the effectiveness of the reflector localization algorithms based on TOA and DOA measurements, described in Sections 3.4.3.1 and 3.4.3.2, respectively. On the base of the evaluation metric proposed in Section 5.3, we present a set of simulative and experimental results.

6.1.1 Inference using TOAs: simulations

Here we test, by means of simulations, the reflector localization algorithm described in Section 3.4.3.1. Such method relies on TOA measurements, which generate elliptic constraints acting on the reflector line. We first compare the accuracy of two different minimization strategies, namely the minimization on the unit cylinder (see Section 3.3.2.3) and the exact solution provided by the GTRS approach (see Section 3.3.2.4). After that we compare the results of Monte-Carlo simulations with those of the theoretical error propagation analysis devised in Section 5.1, proving that the latter can be proficiently used for predicting the accuracy of the reflector localization technique.

6.1.1.1 Simulation setup

All the simulations are conducted with reference to the setup of Fig. 6.1. A set of $N = 5$ microphones located at $\boldsymbol{x}_{M_1}, \dots \boldsymbol{x}_{M_5}$ are uniformly spaced on a circle of radius 30 cm centered in the origin of the reference frame (corresponding to the acoustic source position $\boldsymbol{x}_S$). The simulations are performed on a set of 9000 test reflector lines defined by their distance ρ and angle α with respect to the origin, as shown in Figure 6.1. The test reflector lines are defined by distances in the range $[1 \text{ m}, 4 \text{ m}]$ and angles in the range $[0, 2\pi]$.

6.1.1.2 Exact and iterative methods

The localization of a reflector is based on the minimization of the cost function in (3.62), which collects multiple elliptical constraints arising from TOA measurements. As discussed in Section 3.3.2, when the acoustic source is located at the origin of the reference frame, there exist two suitable optimization strategies for minimizing the cost function, namely the GTRS approach and the iterative minimization on the

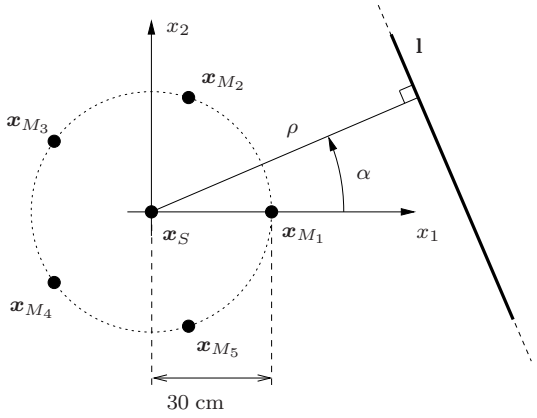
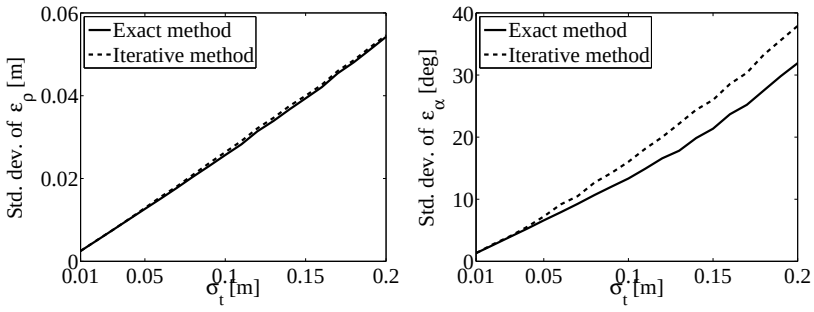


Figure 6.1: Simulation setup for reflector localization using TOAs.

unit cylinder. While the former provides an exact solution, the latter may provide a bad estimate if local minima are present.

In this paragraph, using the setup in Figure 6.1, we compare the results of the iterative methods with the groundtruth provided by GTRS. In order to simulate the measurement process, for each reflector position, reflective TOAs are calculated and then corrupted by 1000 realizations of independent identically distributed zero-mean Gaussian noise with standard deviation σ_t/c, where $c = 340$ m/s is the speed of sound. The accuracy is evaluated by considering the standard deviation of the distance error ε_ρ and of the angular error ε_α. Figures. 6.2-(a) and 6.2-(b) show the results as a function of σ_t, respectively, averaged over all the tested locations and repetitions. As far as the distance error is concerned, the iterative and the exact



(a) Average std. dev. of ε_ρ (b) Average std. dev. of ε_α

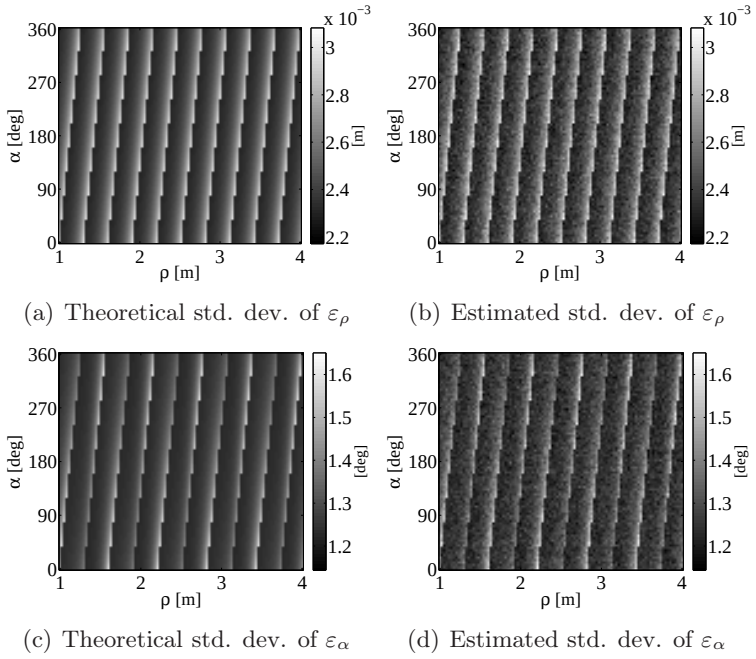
Figure 6.2: Comparison between the iterative and the exact solutions.

solutions turn out to exhibit almost identical errors, which are proportional to the

standard deviation σ_t of the measurement error. As for the angular error, for values of σ_t below 0.05 m, the two approaches have the same results, but for higher values of σ_t, the iterative method is affected by larger errors. This is due to the presence of multiple local minima in the cost function. For large measurement errors, the risk of encountering local minima increases as the cost function becomes less smooth. Although this phenomenon occurs occasionally, its impact on the standard deviation of the angular error is quite noticeable. The exact solution is therefore preferable over the iterative one, especially for large measurement errors.

6.1.1.3 Monte-Carlo simulations and error propagation analysis

We now compare the results of Monte-Carlo simulations with ones predicted through error propagation analysis. We keep fixed the standard deviation of the measurement noise to $\sigma_t = 0.01$ m. The standard deviation of the error predicted with the analytic method is compared with the results of the simulations conducted on the same testing reflector positions, adopting the GTRS approach. The results depicted in Figure 6.3 show the distance error for theoretical (a) and simulated (b) analysis,



(a) Theoretical std. dev. of ε_ρ (b) Estimated std. dev. of ε_ρ

(c) Theoretical std. dev. of ε_α (d) Estimated std. dev. of ε_α

Figure 6.3: Comparison between the theoretical standard deviation of the error (predicted with the error propagation analysis) and the simulation results.

respectively. Similarly, Figures 6.3-(c) and 6.3-(d) show the theoretical and simulated results relative to the angular error. The comparison between Monte-Carlo simulations and the theoretical prediction of the accuracy confirms the validity of the error propagation analysis. In fact, the results of the simulations accurately match the theoretical ones: they present the same mean error of the expected values (2.5 mm for the distance and 1.3° for the angle). The patterns of local maxima (i.e. diagonal white lines) correspond to configurations where two or more reflective paths are (almost) collinear, thus producing ellipses having the same point of tangency on the reflector. A simple graphical explanation is given in Figure 6.4. In particular, Figure 6.4-(a) depicts a situation in which two ellipses are generated from distinct reflection points. Figure 6.4-(b) shows the case in which two reflective paths are collinear, thus sharing the same reflection point. In this situation, it is evident that two measurements carry similar information, thus reducing the robustness of the estimation.

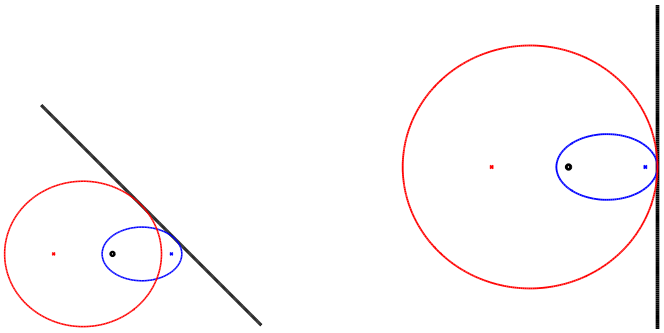
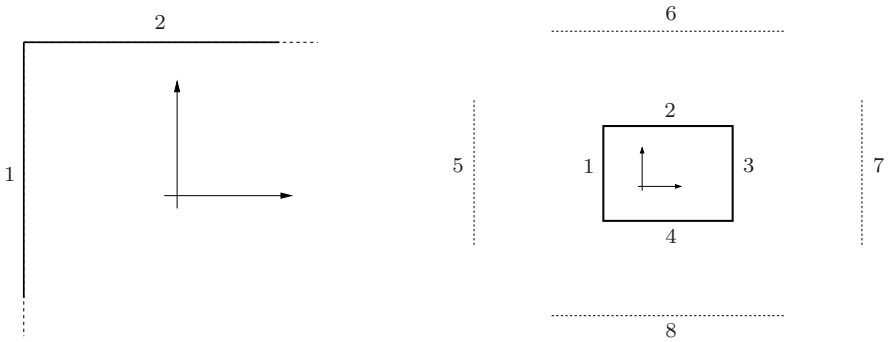


Figure 6.4: The reciprocal positions of reflector (black line), source (black circle) and microphones (colored dots) determine different configurations of the ellipses generated from reflective TOAs. In (a) the reflective paths generate two ellipses tangent to the reflector in distinct points. In (b) the reflective paths are so that the tangency points are coincident.

6.1.2 Inference using TOAs: experiments

In this Section we present experimental results for the localization of the reflectors of two environments. Experiments have been conducted in an acoustically dry room in which wood panels have been placed to alter the impulse response. The first experiment concerns the problem of localizing two reflectors that are not mutually visible. In a later stage we consider the case of mutually visible reflectors.

In these experiments we used a loudspeaker that was moved in 24 different positions on a circle and a microphone located at the center of the circle, coincident with the origin of the reference frame[1]. The radius of the circle is 0.16 m. The sampling frequency is $F_s = 44.1$ kHz and the excitation signal is a white noise in the bandwidth $[0, 5$ kHz$]$. TOAs relative to different reflectors have been disambiguated and labeled by means of the generalized Hough transform described in Section 3.4.3.3, and the corresponding estimation problems have been solved using the GTRS approach. The geometry of the system for the first environment is depicted in Figure



(a) First test environment. (b) Second test environment.

Figure 6.5: Geometry of the environments used for the experiments. The first environment (a) is composed of two reflectors which are not mutually visible. The second environment (b) is composed of four reflectors; parallel reflectors are mutually visible causing multiple reflections which generate "virtual walls", denoted by dashed lines.

6.5-(a). Numbers next to the walls refer to indexes in Table 6.1, where localization results for both reflectors are shown according to the criteria defined above. In the

Table 6.1: Experimental results for the environment in Figure 6.5-(a).

Index	$(\rho$ [m] , α [°])	$(\hat{\rho}$ [m] , $\hat{\alpha}$ [°])	ε_ρ [m]	ε_α [°]
1	(1.805 , 180)	(1.81 , 178)	−0.005	2
2	(2.24 , 90)	(2.24 , 89)	0	1

[1]Notice that, from the geometrical point of view, the roles of loudspeakers and microphones can be interchanged. As a consequence, a single microphone located at the origin of the reference frame leads to ellipses with a focus in that point, thus making possible the use of the GTRS approach for solving the estimation problem.

second experiment we have tested the system in the environment depicted in Figure 6.5-(b). Dashed lines refer to "virtual reflectors" originated from second-order reflections. Table 6.2 shows the corresponding localization results using the metrics defined in Section 5.3. We notice that even "virtual" reflectors are localized with a good accuracy.

Table 6.2: Experimental results for the environment in Figure 6.5-(b).

Index	(ρ [m] , α [°])	($\hat{\rho}$ [m] , $\hat{\alpha}$ [°])	ε_ρ [m]	ε_α [°]
1	(1.18 , 270)	(1.17 , 270)	0.01	0
2	(1.37 , 180)	(1.38 , 179)	−0.01	1
3	(1.68 , 90)	(1.68 , 89)	0	1
4	(2.08 , 360)	(2.08 , 358)	0	2
5	(2.82 , 270)	(2.84 , 271)	−0.02	−1
6	(2.82 , 90)	(2.85 , 85)	−0.03	5
7	(3.44 , 180)	(3.45 , 181)	−0.01	−1
8	(3.44 , 360)	(3.44 , 357)	0	3

6.1.3 Inference using DOAs: simulations

We now test the accuracy of the reflector localization algorithm described in Section 3.4.3.2. Such method converts DOA measurements into quadratic constraints which have the form of parabolas tangent to the reflector line. More specifically, each DOA generates a parabola with focus at the source position, and whose directrix coincides with the DOA line. As shown in Section 3.3.2.4, if the source is kept fixed at the origin of the reference frame, the dual of the parabola arising from each DOA measurement takes the form of a circumference. Therefore, in principle, the GTRS approach could be used to obtain the exact solution of the estimation problem. However, this setup is not easy to implement, since it would require to move the microphone array around the source. The converse (i.e. moving the source around a fixed microphone array) turns out to be much more practical and easier to be automatized. Although in this case the GTRS method can not be applied, the exact solution can still be obtained by means of cutting of the search space with two planes orthogonal to the parameter axes, as discussed in Section 3.3.2.1.

6.1.3.1 Simulation setup

We now describe the simulation setup adopted for the simulations. According to Figure 6.6, the source is moved at $N = 10$ positions $x_{S_1}, \dots x_{S_N}$ uniformly distributed

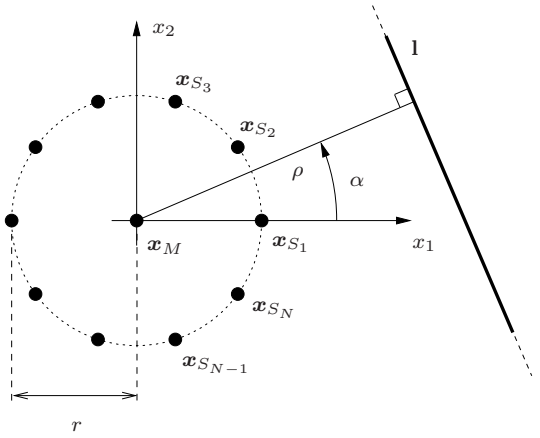


Figure 6.6: Simulation setup for reflector localization using TOAs.

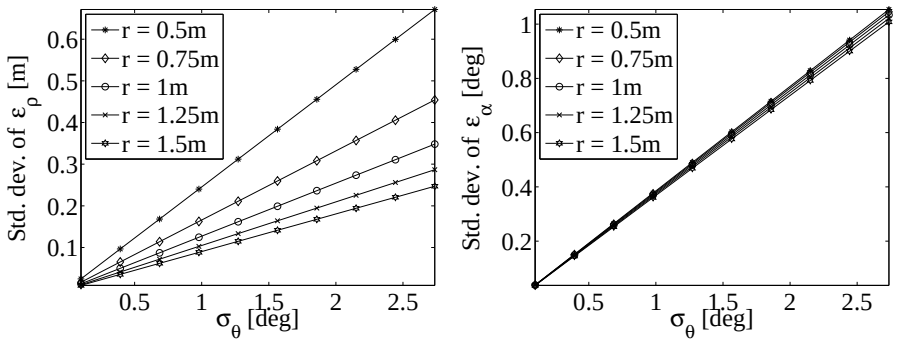
on a circle centred at $\boldsymbol{x}_M = [0, 0]^T$ and with radius r. The origin of the reference system coincides with the centre $\boldsymbol{x}_M$ of the microphone array, and reflective DOAs are calculated with respect to this point. As for the case of TOAs described in Section 6.1.1.1, the simulations are performed on a set of 9000 test reflector lines $\mathbf{l} = [\cos\alpha,\ \sin\alpha,\ -\rho]^T$ defined by their distance $\rho \in [1\text{ m}, 4\text{ m}]$ and angle $\alpha \in [0, 2\pi]$ with respect to the origin.

6.1.3.2 Error propagation analysis

The results presented in Section 6.1.1.3 prove the effectiveness of the theoretical error propagation analysis in predicting the impact of the measurement error on the estimated reflector lines. For this reason, it is reasonable to limit the evaluation of the DOA-based localization algorithm to a theoretical analysis, applying the error propagation analysis to the setup of Figure 6.6. This allows us to avoid Monte-Carlo simulations, which in this case are extremely time-demanding. In fact, as discussed in Section 3.3.2.4, the minimization process based on the cutting of the parameter space is very expensive in terms of computation effort. This makes Monte-Carlo simulations impracticable, which require a high number of repetitions for each test.

As a first test, we evaluate the accuracy of the estimation varying the radius r of the circle on which the sources are disposed. The error on DOAs is assumed to be Gaussian, zero-mean and independently distributed on each measurement with standard deviation σ_θ. The results are shown in Figure 6.7. In particular, Figure 6.7-(a) shows the average standard deviation of the distance error ε_ρ as a function

of σ_θ. These results are obtained averaging the standard deviation of ε_ρ for all the

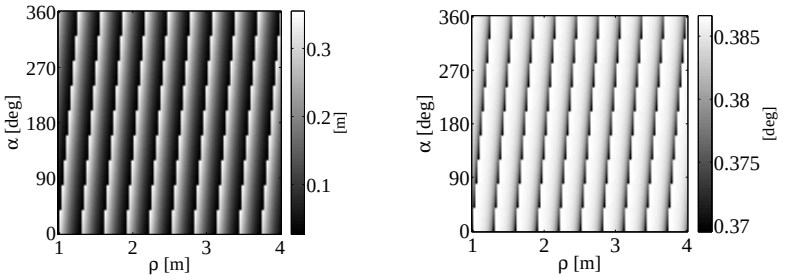


(a) Average std. dev. of ε_ρ (b) Average std. dev. of ε_α

Figure 6.7: Accuracy of reflector localization as a function of the standard deviation σ_θ of the measurement error.

9000 test reflector positions. Similarly, Figure 6.7-(b) shows the average standard deviation of the angular error ε_ρ as a function of σ_θ. First of all, we observe that the localization technique based on DOAs suffers from noticeable errors (up to 0.6 m for $r = 0.25$ m and $\sigma_\theta = 3°$) in estimating the distance of the reflectors. On the other hand, the angular error is kept moderate (slightly above 1° in the worst case). This behaviour is not surprising, since by definition DOA measurements provide information only about the angle of emission of indefinitely distant sources. The distance error is reduced when the source occupies positions on larger circles. For instance, when $r = 1.5$ m, the distance error reduces to 0.2 m for $\sigma_\theta = 3°$. As far as the angular error ε_α is concerned, it reveals to be almost independent from the size of the source circle.

As a second test, we fix the radius of the circle to $r = 0.75$ m and the measurement error to $\sigma_\theta = 1°$. The predicted standard deviation of ε_ρ and ε_α for all the 9000 test positions are shown in Figures 6.8-(a) and 6.8-(b), respectively. As for the case of TOAs shown in Figure 6.3, the estimation error exhibits a periodic behaviour in the parameter space (ρ, α). In this case however, unlike for TOAs, the spread between the minimum and the maximum values of the distance error is more significant. While in the best case the standard deviation of ε_ρ maintains below 3 cm, the patterns of local maxima in Figure 6.8-(a) are characterized by a standard deviation of 35 cm. This happens whenever one ore more DOA measurements lead to degenerate parabolas, i.e. when one or more sources lie on the same line determined by the measured DOA. On the other side, the standard deviation of the angular error maintains moderate



(a) Theoretical std. dev. of ε_ρ (b) Theoretical std. dev. of ε_α

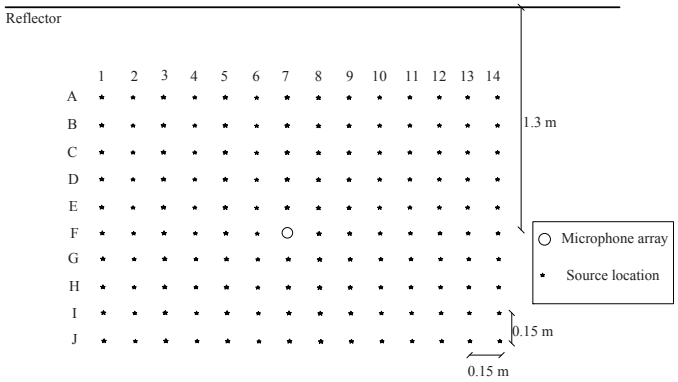
Figure 6.8: Standard deviation of the error predicted with the error propagation analysis.

with values around 0.38°. Moreover, it is interesting to notice that the patterns of local minima of the angular error (i.e, the diagonal black lines in Figure 6.8-(b)) correspond to the patterns of local maxima of the distance error (i.e. the diagonal white lines in Figure 6.8-(a)). This means that a lack in the estimation of the reflector distance is partially recovered by a better estimation of the reflector angle.

6.1.4 Inference using DOAs: experiments

In the experiments we tested the reflector localization accuracy as a function of the number of DOA measurements used in the estimation. The experimental setup is depicted in Figure 6.9. The experiments have been conducted using the microphone array[2] shown in Figure 6.9-(b), which is composed by 10 sensors disposed on a rigid cylindrical baffle with radius 0.04 m. For further reference on the microphone array, see [105]. The signal acquired by the sensors has been sampled at $F_s = 44.1$ kHz. In all the experiments we considered the presence of a single reflector in the environment. For this purpose, we have placed a planar reflective surface in a low-reverberating chamber. Figure 6.9-(a) shows 140 potential positions of the source, disposed on a grid. However, due to the reciprocal configuration of the array and the reflector, not all the positions on the grid generate a reflective path. For this reason, we have placed a loudspeaker at 88 positions out of 140. The loudspeaker emitted a white noise in the band [1 kHz , 10 kHz]. We tested the accuracy of the reflector localization as a function of the number of constraints N used in the minimization, N being between 3 and 24. For each value of N, we tested all the possible combinations of DOA measurements. The results are depicted in Figures 6.10-(a) and 6.10-(b),

[2]The array has been provided by the LMS laboratory at the Technische Fakultät of the Friedrich-Alexander University of Erlangen-Nuremberg.



(a) Setup of the experiment: the grid marks all the potential positions of the loudspeaker. The microphone array is at the center of the grid. The reflector is distant 1.3 m from the center of the microphone array.

(b) The circular microphone array used in the experiments [105].

Figure 6.9: Setup of the experiments.

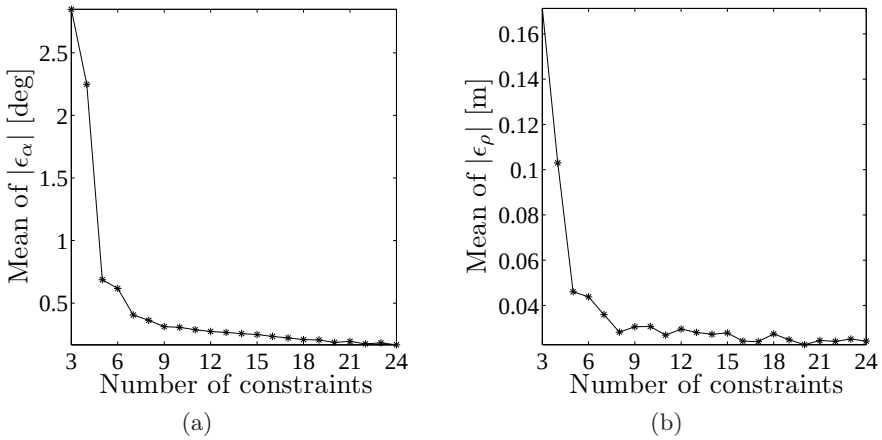


Figure 6.10: Mean angular (a) and distance (b) errors vs number of DOAs.

which show the average of $|\varepsilon_\rho|$ and of $|\varepsilon_\alpha|$, respectively. We notice that even with three constraints, i.e. three source positions, the reflector is well localized.

In order to show the stability of the estimation, we show also the standard deviation of ε_ρ and ε_α as a function of the number of DOA measurements. We notice that as the number of constraints increases the standard deviation decreases, as expected. Nonetheless, when $N \geq 4$ measurements are used, the solution is almost independent from the specific set of measures used.

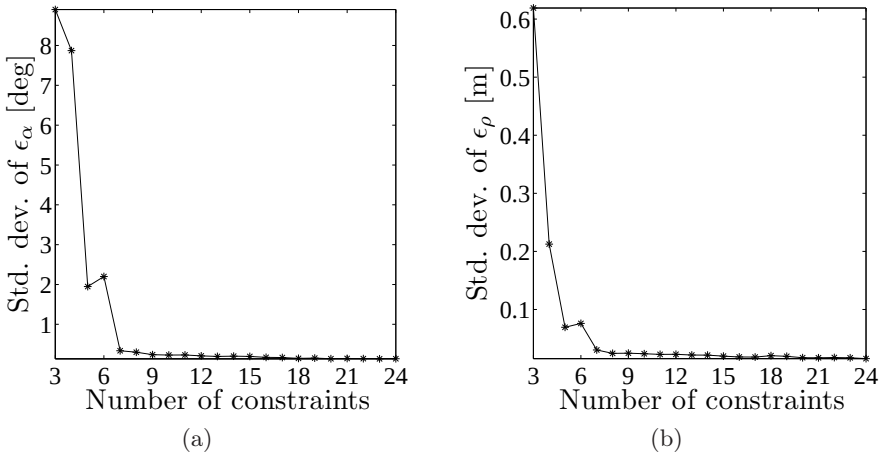


Figure 6.11: Std. dev. of angular (a) and distance (b) errors vs number of DOAs.

6.2 Wave field rendering

This Section is dedicated to evaluating the rendering technique proposed in Chapter 4. In the first part of the Section, we aim at testing the rendering methodology in a theoretical fashion. To do so, the reproduced wave field is first predicted by modelling the propagation between each loudspeaker and the listening area, taking into account the effect of the environment hosting the rendering system. The accuracy of the reproduced wave field is then evaluated through a comparison with the desired wave field, adopting the NMSE and modal metrics introduced in Section 5.2. Since the room estimation methodologies revealed a great accurateness (see the results in the previous Section), here we omit to test the robustness of room-compensation against errors in the geometry of the environment. In all the simulations, we assume the knowledge of the floor-map.

In the second part of the Section we consider a real reproduction system consisting in an arrangement of 48 high-quality loudspeakers mounted on a circular array. Such system is tested with both the proposed rendering technique and the Wave Field Synthesis [6] technique. The rendered wave fields are measured by means of the Circular Harmonic Decomposition, as discussed in Section 5.2.2. The comparison of the results of the two rendering techniques makes it possible, on one hand, to verify the suitability of the measuring methodology. On the other hand, it allows us to investigate the differences between the proposed technique and a well-established rendering methodology such as WFS, even in a real scenario.

6.2.1 Environment-aware rendering: a theoretical evaluation

In this Section we evaluate the accuracy of the proposed rendering methodology in two different scenarios. In a first stage, we test the beam shaping engine with a linear array of loudspeakers emitting a single beam. After that, we consider a circular loudspeaker array for the rendering of a virtual source along with its effects on three different virtual environments. In both the scenarios we simulate the presence of an L-shaped real environment.

The terminology employed throughout this Section was introduced and explained in Chapter 4. For the sake of clarity, in the following we recall the main definitions:

- *virtual source*: a source to be rendered by the loudspeakers;
- *real environment*: the environment in which the rendering system operates;
- *beam shaper*: the system devoted to synthesizing beams by means of the loudspeakers;
- *virtual environment*: the environment hosting the virtual source. Its effect is reproduced by overlapping multiple beams;
- *room compensation*: the process of predicting and compensating the impact of the real environment on the rendered wave field.

6.2.1.1 Simulation scenarios

Scenario 1: the first scenario is shown in Figure 6.12. The loudspeakers are arranged on a linear array with aperture l and the listening area covers a rectangle with dimensions 3 m$\times$2.5 m. The real environment models an L-shaped room, whose reflection coefficient is set to 0.7 for all the reflectors. The number M of loudspeakers varies for the different tests shown in Section 6.2.1.3. The beam shaper synthesizes a single virtual source located at a distance d from the array and emitting a beam directed as θ and with angular aperture ϕ. The virtual source is positioned so that the center of the beam always passes through the array center. The listening area is sampled with $N = 1000$ control points. A more dense grid of $Q = 10000$ evaluation points is used for evaluating the NMSE and modal performances of the beam shaper.

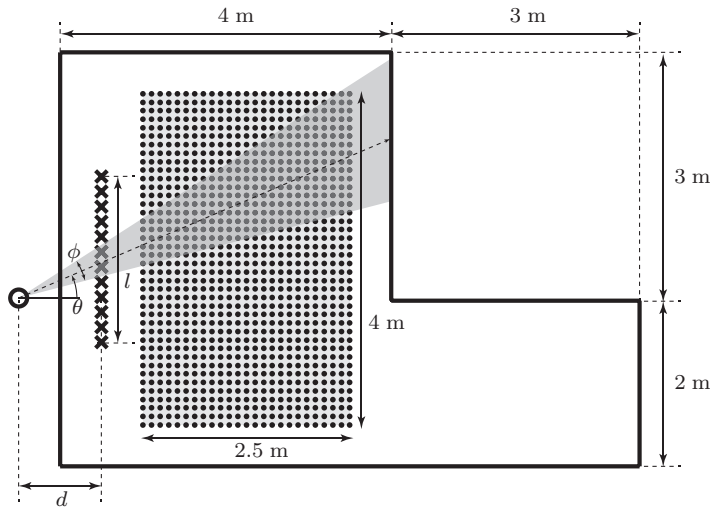


Figure 6.12: Setup for the evaluation of the beam shaping engine. The loudspeakers are arranged on a linear array with aperture l and the listening area is rectangular. The real environment models an L-shaped room. The virtual source, depicted with the black circle, is positioned at a distance d from the array and emits a beam passing through the array centre and directed as θ, with angular aperture ϕ.

Scenario 2: the second scenario is depicted in Figure 6.13. The rendering system is composed of M loudspeakers disposed on a circumference with radius r_a and a listening area covering a circle with radius r_l inside the array. The control points are regularly spaced within the listening area. The rendering system is operating in the same L-shaped real environment of the first scenario. The listening area is regularly sampled with $N = 1000$ control points, while the evaluation is performed on a regular distribution of $Q = 10000$ evaluation points.

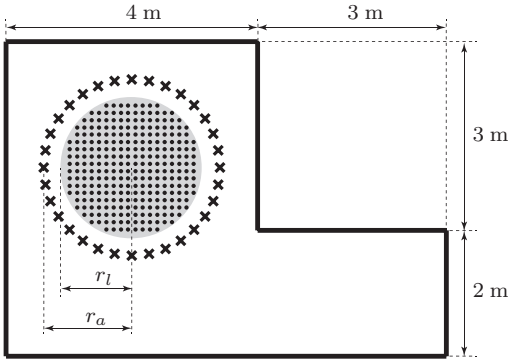


Figure 6.13: Setup for the evaluation of the rendering engine. The loudspeakers are disposed on a circle with radius r_a and the listening area covers a circle with radius r_l inside the array. The real environment models an L-shaped room.

6.2.1.2 Evaluation metrics

The accuracy of the rendered wave fields is evaluated by means of the NMSE and modal metrics introduced in Section 5.2.

The NMSE metric provides a global information about the accuracy of the reproduced wave fields. It is calculated by means of (5.19), and we distinguish the following quantities:

- E_{rc} $\rightarrow$ NMSE associated to wave fields synthesized performing the room compensation;
- E_{nc} $\rightarrow$ NMSE of wave fields rendered without compensating for the real environment;
- E_{ff} $\rightarrow$ NMSE of wave fields reproduced in the free-field, i.e. when the real environment is anechoic and no compensation is required.

The modal metrics allow us to examine the wave field propagation. In particular, the function $\hat{S}'(k)$ defined in (5.23) is used for determining the distribution of the wavenumbers k in the reproduced wave fields; and the function $\hat{S}''(\varphi)$ gives insight on the propagation directions φ of the rendered wave fields. We consider the following quantities:

- $\hat{S}'_{\text{des}}$, $\hat{S}''_{\text{des}}$ $\rightarrow$ modal analysis of the desired wave field;
- $\hat{S}'_{\text{rc}}$, $\hat{S}''_{\text{rc}}$ $\rightarrow$ modal analysis for room-compensated rendering;
- $\hat{S}'_{\text{nc}}$, $\hat{S}''_{\text{nc}}$ $\rightarrow$ modal analysis for non-compensated rendering;
- $\hat{S}'_{\text{ff}}$, $\hat{S}''_{\text{ff}}$ $\rightarrow$ modal analysis for rendering in the free-field.

6.2.1.3 Beam shaping with a linear array

In this paragraph we consider the first scenario, described in Section 6.2.1.1. We evaluate the effectiveness of the beam shaper engine for different values of the parameters of the array (number of loudspeakers) and of the beam (distance, direction and aperture). For all the following tests, the room compensation is performed up to the 3^{rd} order of reflection, while the reflections of the real environment are modelled up to the 15^{th} order.

Test 1 Let us consider the rendering of a beam with an array of $M = 32$ loudspeakers and aperture $l = 2$ m. The desired beam-pattern has parameters $d = 5$ m, $\theta = 0\,°, \phi = 10°$. A snapshot of the desired wave field at 1 kHz is shown in Figure 6.14-

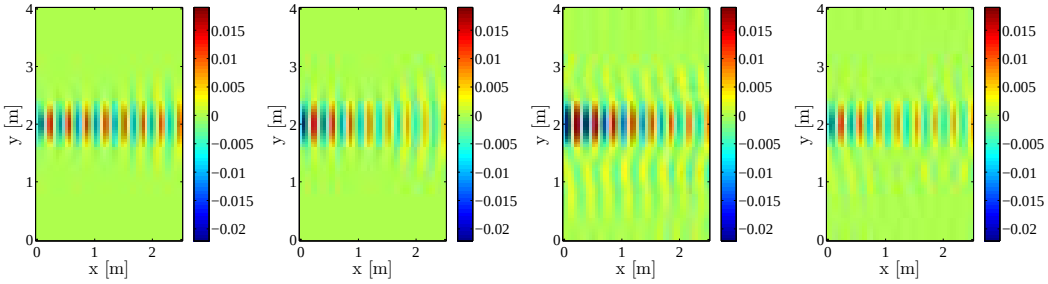


Figure 6.14: Example of beam shaping with a linear array of 32 loudspeakers, at 1 kHz. The desired beam-pattern (a) has parameters $d = 5$ m, $\theta = 0\,°, \phi = 10°$. The free-field, non-compensated and room-compensated wave fields are shown in (b), (c) and (d), respectively.

(a). The corresponding rendered wave fields are shown in Figures 6.14-(b), 6.14-(c) and 6.14-(d), which show, respectively: the wave field rendered in the free-field; the wave field reproduced without compensating for the real environment; the wave field reproduced performing the compensation of the real environment. It can be noticed that the free-field and room-compensated rendered wave fields are similar to the desired one. On the other hand, the non-compensated reproduction suffers from the effect of the real environment, which produces, in this case, visible artifacts in the bottom part of the listening area. This deterioration is probably due to a reflection coming from the lower part of the L-shaped environment (see Figure 6.12).

We now focus on the NMSE performances of the system reproducing the beam under analysis. Figure 6.15 shows the NMSE values in the range [100 Hz, 4 kHz]. The vertical grey lines denote the natural modes associated to the two parallel walls orthogonal to the beam. These resonant modes are expected to be the most likely to be excited in the case under analysis, since for $\theta = 0°$ the beam is oriented exactly towards the wall in front of the loudspeaker array (see Figure 6.12). Indeed, the performances of the non-compensated rendering reveals to be highly affected in correspondence of a subset of the resonant modes. On the other hand, the deconvolution provided by room compensation strongly attenuates the effect of reverberations. In fact, the NMSE related to the room compensation approaches the curve relative to the free-field NMSE.

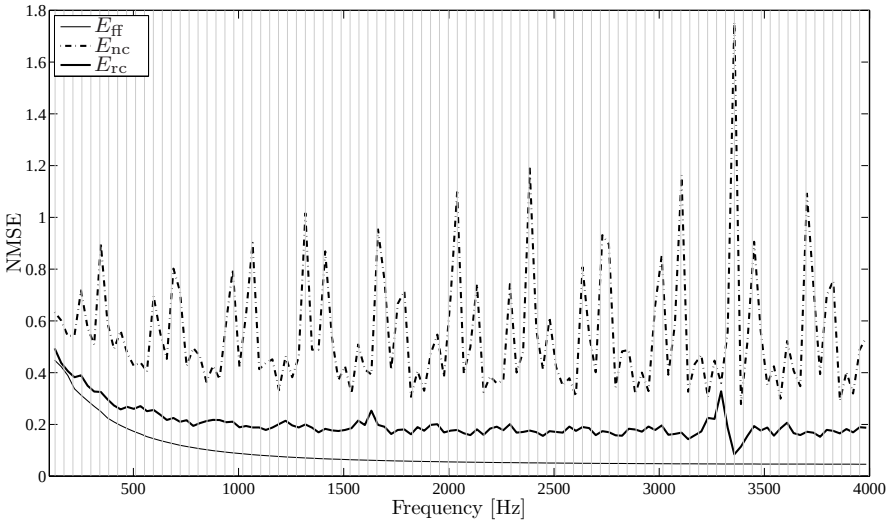
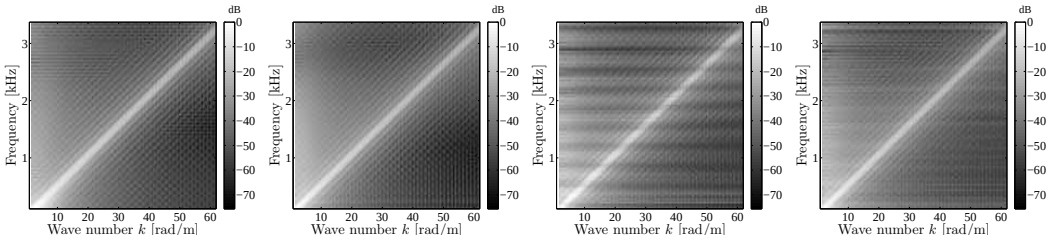


Figure 6.15: NMSE values of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of the beam shaper. The array has an aperture $l = 2$ m and it is composed of $M = 32$ loudspeakers. The desired beam-pattern has parameters $d = 5$ m, $\theta = 0\ ^\circ$, $\phi = 10^\circ$.

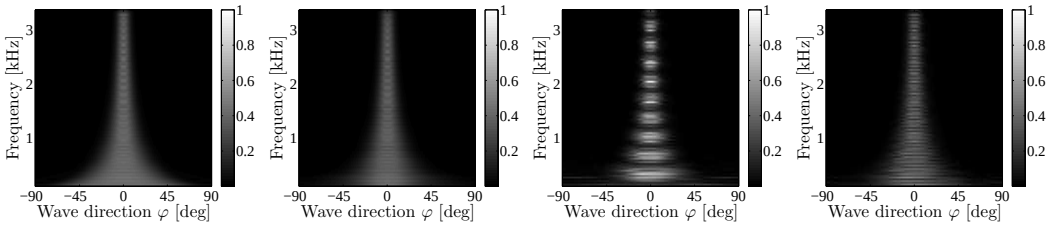
We now analyse the modal performances relative to the same test. The distribution of the wavenumbers is depicted in Figure 6.16, which shows the function $\hat{S}'(k)$ for all frequencies in the range [100 Hz, 2.5 kHz]. We notice that Figure 6.16-(c), which



(a) S'_{des} (b) S'_{ff} (c) S'_{nc} (d) S'_{rc}

Figure 6.16: Analysis of the wave number distribution for the rendering of a beam-pattern with parameters $d = 5$ m, $\theta = 0\ ^\circ$, $\phi = 10^\circ$.

refers to the function S'_{nc}, presents a set of parallel horizontal stripes. This fact is due, as expected, to the dominant natural modes. In fact, such horizontal stripes are in correspondence of the frequencies at which the curve E_{nc} in Figure 6.15 exhibits the highest peaks. The boosting of the resonant modes is almost completely eliminated by room-compensation. Indeed, the function S'_{rc} in Figure 6.16-(d) approaches the



(a) S''_{des} (b) S''_{ff} (c) S''_{nc} (d) S''_{rc}

Figure 6.17: Propagation direction analysis for the rendering of a beam-pattern with parameters $d = 5$ m, $\theta = 0\,^\circ$, $\phi = 10^\circ$.

free-field behaviour S'_{ff} and the desired behaviour S'_{des} in Figure 6.16-(a). As far as the propagation direction is concerned, the results are shown in Figure 6.17. The results are organized as in Figure 6.16, i.e. the function S''_{des} is plotted in Figure 6.17-(a); S''_{ff} in (b); S''_{nc} in (c); and S''_{rc} in (d). We notice that the function S''_{nc} presents local maxima in correspondence of the resonant frequencies. Moreover, at low frequencies we observe the presence of undesired propagation modes which are not present in the desired (S''_{des}), free-field (S''_{ff}) and room-compensated (S''_{rc}) responses.

Test2 We now consider a second simulation to evaluate the effectiveness of the beam shaper for different number of loudspeakers, keeping constant the aperture to $l = 2$ m. The beam parameters are again set to $d = 5$ m, $\theta = 0\,^\circ$, $\phi = 10^\circ$. Figure 6.18 shows the NMSE results of beam shaping at 4 kHz. We observe that a good

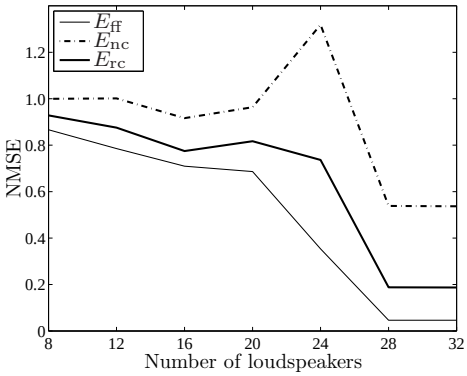


Figure 6.18: NMSE values, at 4 kHz, of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of the beam shaper as a function of the number of loudspeakers. The array has an aperture $l = 2$ m and the desired beam parameters are $d = 5$ m, $\theta = 0\,^\circ$, $\phi = 10^\circ$.

reproduction of the beam-pattern requires a sufficient number of loudspeakers, i.e.

$M \geq 28$. This problem is due to the spatial aliasing artefacts introduced by an incorrect spacing between the loudspeakers. For a constant aperture of the linear array, aliasing can be avoided increasing the number of loudspeakers.

In order to better characterize the phenomenon of spatial aliasing, we perform a modal analysis of the reproduced wave fields. As noticed in [106], spatial aliasing in rendering applications can be interpreted as the effect of undesired propagation modes. For this reason, we focus on analysing the distribution of the propagation directions. In Figure 6.19 the function $\hat{S}''_{\text{rc}}$ relative to the reproduced wave fields with $M = 12$, $M = 20$, and $M = 28$ loudspeakers is compared with the function $\hat{S}''_{\text{des}}$ relative to the desired wave field at 4 kHz. We observe that, with $M = 12$,

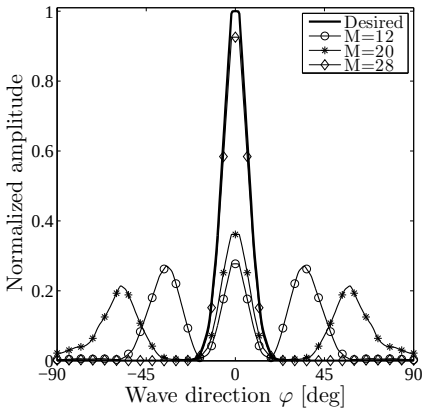
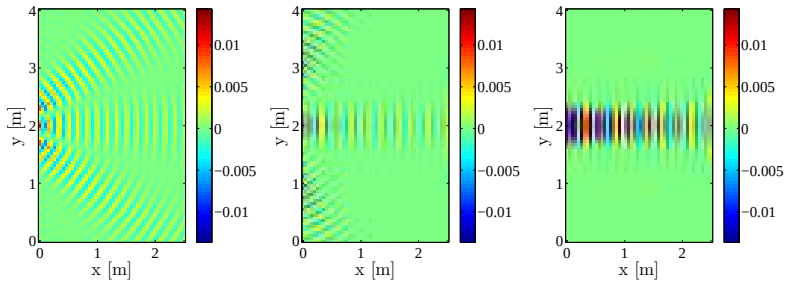


Figure 6.19: Spatial aliasing: analysis of the direction of propagation. The rendering system reproduces a beam at 4 kHz with $M = 12, 20, 28$ loudspeakers. The corresponding functions $\hat{S}''_{\text{rc}}$ are compared with the desired one $\hat{S}''_{\text{des}}$.

the beam-pattern is not correctly synthesized since unwanted propagation modes at directions $\pm 40°$ are present. Moreover, we notice an equal repartition of the energy of the rendered wave field at $-40°$, $0°$ and $40°$, causing an attenuation of the reproduction in the desired direction. Similar considerations hold for $M = 20$. In this case, however, the desired propagation direction $0°$ gains energy with respect to the undesired ones, which are now moved to $\pm 60°$. When $M = 28$ aliasing artifacts disappear, and the function $\hat{S}''_{\text{rc}}$ approaches the desired one $\hat{S}''_{\text{des}}$, exhibiting only the desired propagation direction. A visual confirmation of the analysis reported above is given by Figure 6.20, which shows the rendered wave fields at 4 kHz for $M = 12$ (a), $M = 20$ (b), and $M = 28$ (c).

The analysis of the distribution of the wavenumbers by means of the function $\hat{S}'$



(a) $M = 12$ (b) $M = 20$ (c) $M = 28$

Figure 6.20: Spatial aliasing: rendering a beam with a variable number M of loudspeakers.

is omitted, since it leads to results similar to those presented above.

Test 3 We now analyze the performances of the beam shaping engine varying the beam parameters. The number of loudspeakers and the array aperture are kept fixed to $M = 32$ and $l = 2$ m, respectively. In a first stage, we consider a virtual source positioned at a distance $d = 5$ m from the array emitting a beam with aperture $\phi = 10°$; the beam direction varies in the range $[0°, 50°]$. Figure 6.21 shows the results of the NMSE analysis at 1 kHz. We notice that the free-field and room-

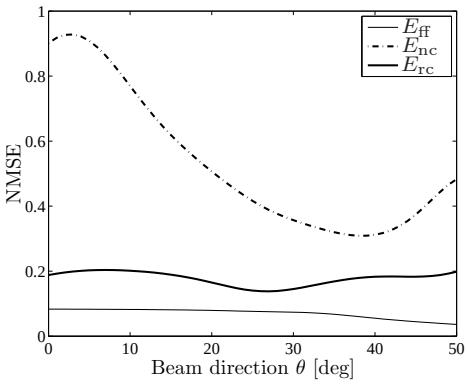


Figure 6.21: NMSE values, at 1 kHz, of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of the beam shaper as a function of the beam direction θ. The other beam parameters are fixed to $d = 5$ m and $\phi = 10°$. The array has an aperture $l = 2$ m and it is composed of $M = 32$ loudspeakers.

compensated responses are nearly independent from the beam direction. On the other hand, looking at the non-compensated response, we observe that the effect of the environment strongly depends on the emission direction. Once again, the room compensation technique reveals to be effective for attenuating the effect of

reverberations. We now analyse the modal behaviour of the rendered wave fields. Figure 6.22 shows the function $\hat{S}''$ for two different beam directions, namely $\theta = 9°$

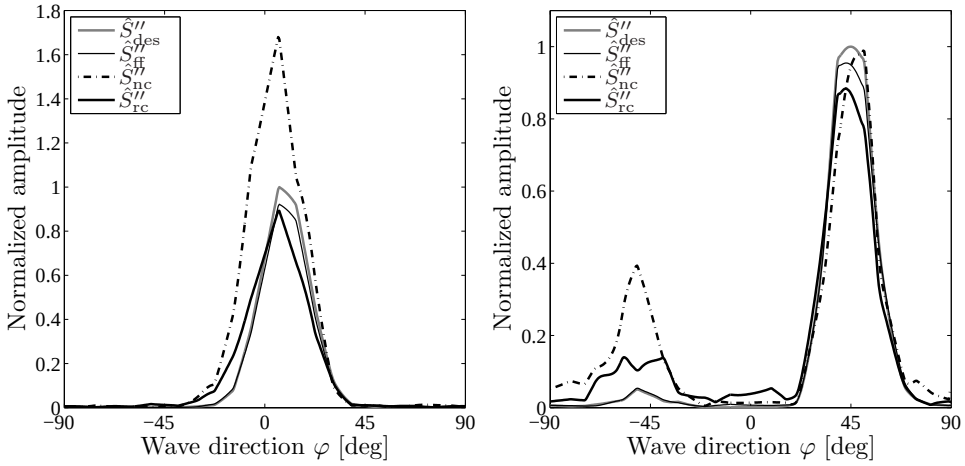


Figure 6.22: Analysis of the propagation direction for the rendering of beams oriented towards $\theta = 9°$ (a) and $\theta = 46°$ (b).

(a) and $\theta = 46°$ (b). In the first case, the real environment introduces only an error on the amplitude of the rendered wave field, as it can be noticed looking the curve $\hat{S}''_{\text{nc}}$ in Figure 6.22-(a). In the case shown in Figure 6.22-(b), the environment reacts generating a reflection from the direction around $-45°$. This happens due to the major steering of the beam with respect to the previous case. It is interesting to observe that the room-compensation is able to dampen both the types of distortion introduced by the environment. Indeed, the curves of $\hat{S}''_{\text{rc}}$ approach those of $\hat{S}''_{\text{des}}$ and $\hat{S}''_{\text{ff}}$.

We now keep fixed the orientation of the beam, and we consider a virtual source positioned at various distances from the array, in the range $[0.1\text{ m}, 5\text{ m}]$. The virtual source emits a beam directed as $\theta = 0°$ with aperture $\phi = 5°$. Figure 6.23 shows the NMSE results at 1 kHz. We notice, once again, that the effect of reverberations seriously damages the performances of the beam shaper if the real environment is not compensated. Regarding the free-field and compensated responses, we notice that their respective NMSEs depend on the distance of the virtual source from the array. More specifically, we observe that the accuracy increases for virtual sources located far from the array. This happens since the portion of listening area illuminated by the beam decreases as the source moves close to the array. In other words, when the

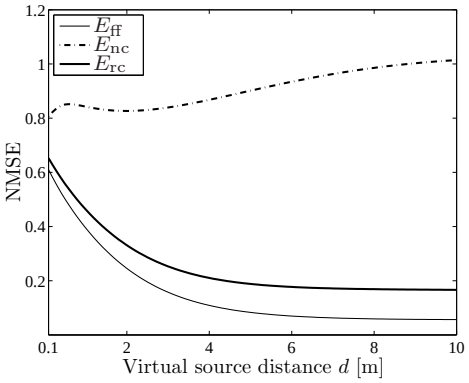


Figure 6.23: NMSE values, at 1 kHz, of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of the beam shaper as a function of the distance d of the virtual source from the array. The other beam parameters are fixed to $\theta = 0°$ and $\phi = 10°$. The array has an aperture $l = 2$ m and it is composed of $M = 32$ loudspeakers.

source is very close to the array, the beam shaper is asked to render a very sharp beam. Better performances may be obtained by increasing the array aperture. Also in this case, the degradations introduced by the environment have different natures, as it can be observed from the propagation direction analysis shown in Figure 6.24. For virtual sources located close to the array, the environment alters the propagation direction of the non-compensated rendered wave field. An example is shown in Figure 6.24, which is relative to a virtual source located at a distance $d = 1.8$ m from the array. Here, the non-compensated wave field exhibits two propagating components at $\pm 15°$, while the desired one is centred at $0°$. On the other hand, far sources determine only an amplitude error on the reproduction of non-compensated wave fields. This can be noticed in Figure 6.24-(b), which is relative to a virtual source located at a distance $d = 10$ m behind the array.

6.2.1.4 Wave field rendering with a circular array

We consider here the rendering of a virtual source along with the effect of three different virtual environments. The simulations are conducted considering the second scenario described in Section 6.2.1.1, setting $M = 32$, $r_a = 1.4$ m and $r_l = 1.1$ m. The reflections of the real environment are modelled up to the 15^{th} order, whereas room compensation is performed up to the 3^{rd} order of reflection. The virtual source and the virtual environments considered for the experiments are shown in Figure 6.25, which depicts also their mutual positions with respect to the real environment.

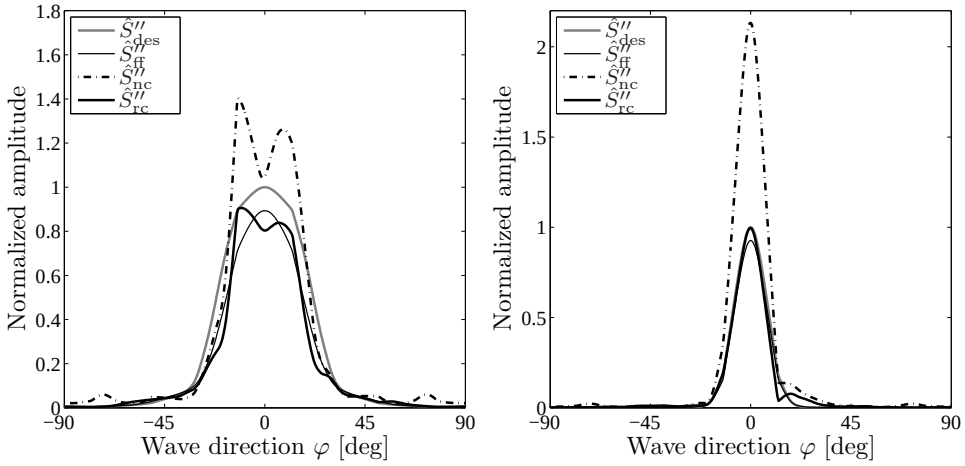


Figure 6.24: Analysis of the propagation direction for the rendering of beams originated from a virtual source close ($d = 1.8$ m) to the array (a) and far ($d = 10$ m)from the array (b).

The circular geometry of the loudspeaker array and of the listening area allows us to asses the maximum alias-free frequency f_A. In fact, as noticed in [21], for such geometry the number M of loudspeaker necessary to achieve an alias-free reproduction has to fulfill the relationship $(M-1)/2 \geq kr_l$, where $k = 2\pi f/c$ is the wave number, f being the frequency. For the rendering system used in simulations ($M = 32$ and $r_l = 1.1$ m) we obtain $f_A \approx 800$ Hz.

In the first simulation we consider the virtual environment in Figure 6.25-(a), that models a 5 m side square room with reflection coefficient 0.7. We simulate the rendering of the virtual source along with the beams associated to reflections up to the 2^{nd} order. Figure 6.26 shows the NMSE results in the range [100 Hz, 2.2 kHz]. As expected, looking at the NMSE of the free-field and room-compensated responses, we observe that the wave field is rendered with good accuracy up to the maximum alias-free frequency f_A. Above this frequency value, the quality of rendering decreases because of the spatial aliasing artifacts. On the other hand, the quality of the rendered wave field is poor when no compensation is performed. In fact, the NMSE of the non-compensated response highlights, as usual, the effect of the real environment on the listening area. As far as the modal analysis is concerned, we focus on the analysis of the directions of propagation. In particular, Figures 6.27-(a), 6.27-(b) and 6.27-(c) show the quantities $|\hat{S}''_{\mathrm{ff}} - \hat{S}''_{\mathrm{des}}|$, $|\hat{S}''_{\mathrm{nc}} - \hat{S}''_{\mathrm{des}}|$ and $|\hat{S}''_{\mathrm{rc}} - \hat{S}''_{\mathrm{des}}|$,

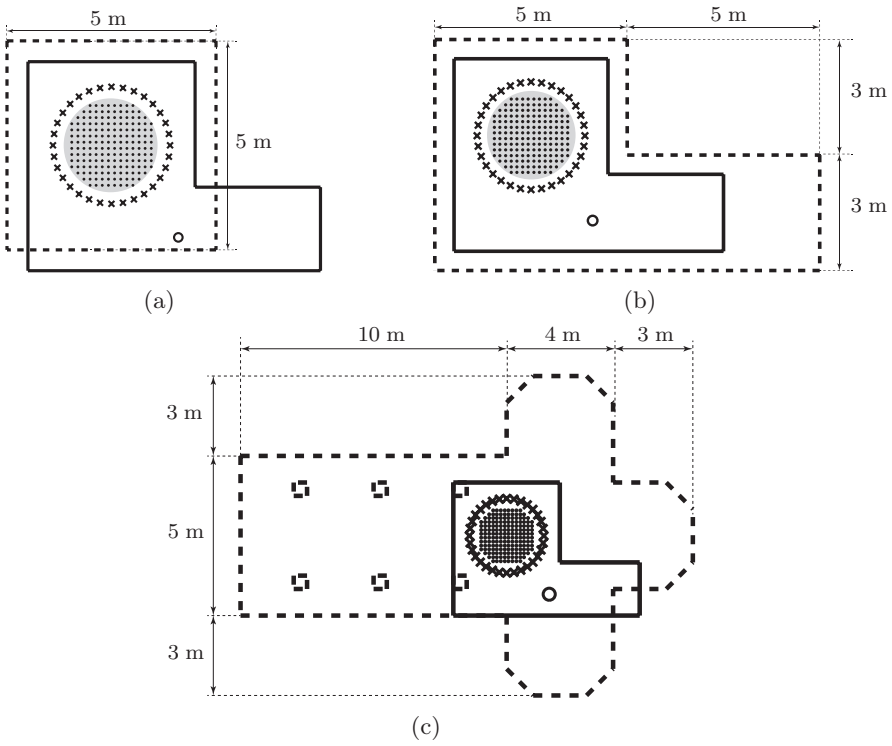


Figure 6.25: The rendering system operates in an L-shaped real environment and renders a virtual source (depicted as a black circle) along with its effect on three different virtual environments. The first virtual environment (a) models a square room; the second virtual environment (b) represents an L-shaped room slightly bigger than the real environment; the third virtual environment (c) models a small church.

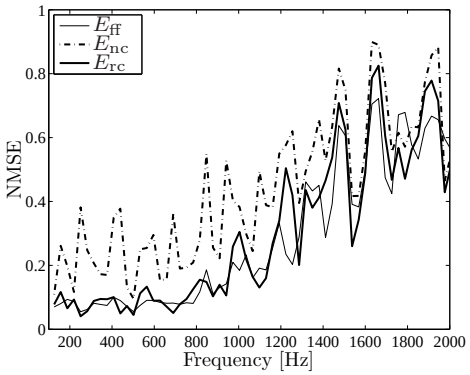
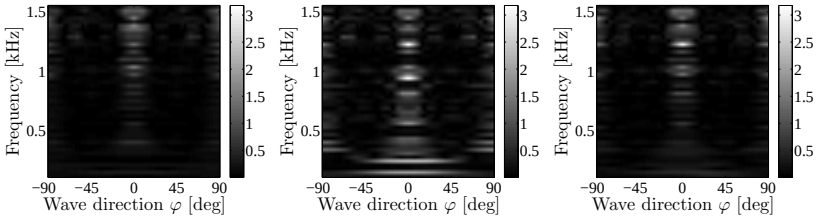


Figure 6.26: NMSE values of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of rendering the virtual source and the virtual environment in Figure 6.25-(a).

respectively. When no compensation of the real environment is performed, the reverberation interferes with the reproduced wave field introducing several undesired

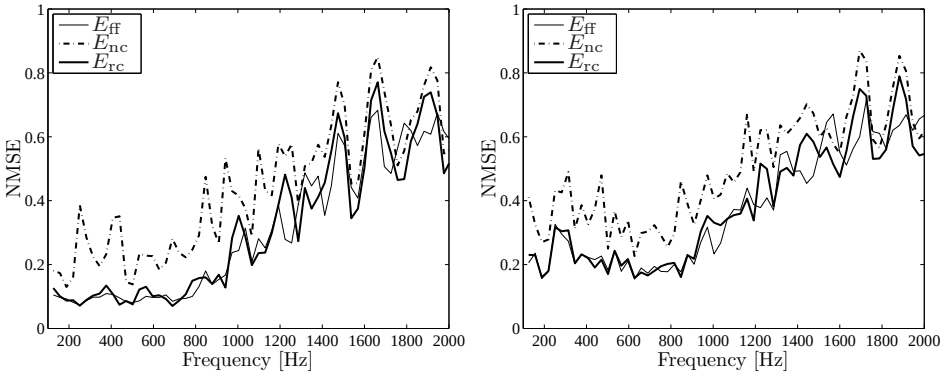


(a) $|\hat{S}''_{\mathrm{ff}} - \hat{S}''_{\mathrm{des}}|$ (b) $|\hat{S}''_{\mathrm{nc}} - \hat{S}''_{\mathrm{des}}|$ (c) $|\hat{S}''_{\mathrm{rc}} - \hat{S}''_{\mathrm{des}}|$

Figure 6.27: Propagation direction analysis relative to the rendering of the virtual environment in Figure 6.25-(a).

propagation modes, as noticed in Figure 6.27-(b). On the other hand, the free-field and room-compensated wave fields share a similar behaviour, presenting only few isolated undesired propagation modes, as it can be observed from the local maxima in Figures 6.27-(a) and 6.27-(c).

The same simulation is repeated using the virtual environments in Figures 6.25-(b) and 6.25-(c), representing an L-shaped room (slightly bigger than the real environment) and a small church, respectively. The reflection coefficient is set to 0.7 for both the virtual environments. The corresponding NMSE results are shown in Figures 6.28-(a) and 6.28-(b), respectively. For these tests, virtual sources are rendered along with the beams associated to acoustic reflections up to the 5^{th} order. The

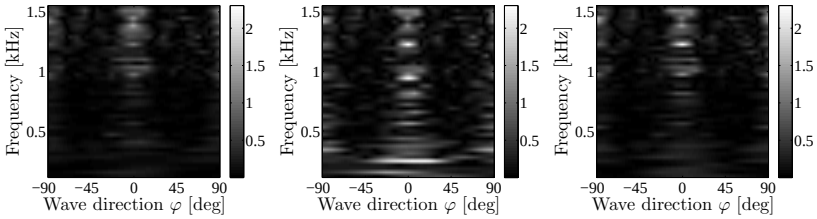


(a) (b)

Figure 6.28: NMSE values of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of rendering the virtual source and the virtual environment in Figures 6.25-(b) (left picture)) and 6.25-(c) (right picture).

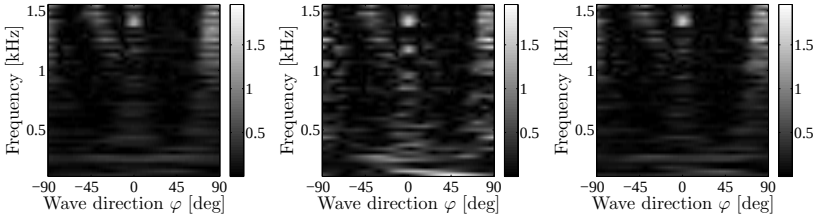
results of the propagation direction analysis for the two test environments are shown

in Figures 6.29 and 6.30, respectively, which are organized as in Figure 6.27. We notice that the results present the same behaviour of the first test, revealing the effectiveness of the free-field and room-compensated rendering independently from the characteristics of the virtual environment.



(a) (b) (c)

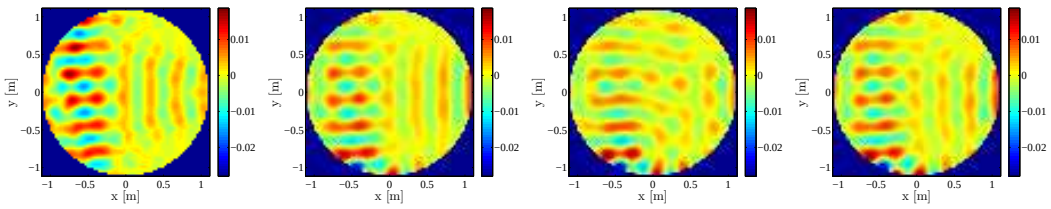
Figure 6.29: Propagation direction analysis relative to the rendering of the virtual environment in Figure 6.25-(b).



(a) (b) (c)

Figure 6.30: Propagation direction analysis relative to the rendering of the virtual environment in Figure 6.25-(c).

An example of the desired, free-field, non-compensated and room-compensated wave fields is shown in Figure 6.31. More specifically, Figure 6.31-(a) depicts the real part of the desired wave field, at 1 kHz, relative to the rendering of a virtual source inside a small church (Figure 6.25-(c)). Looking at Figures 6.31-(b) and 6.31-(d), we notice that the free-field and room-compensated wave fields accurately reproduce the desired one. On the other hand, when no compensation is performed, the reverberation causes a degradation in the reproduced wave field (Figure 6.31-(c)).



(a) (b) (c) (d)

Figure 6.31: The real part of the desired wave field (a) at 1 kHz produced considering the virtual source and the virtual environment in Figure 6.25-(c) is compared with: the real part of the free-field wave field (b), the real part of the non-compensated wave field (c), and the real part of the room-compensated wave field (d).

6.2.2 Experimental evaluation

In this Section we show the results of an experimental evaluation of the proposed rendering methodology. The experiments have been performed on the loudspeaker array installed at the LMS laboratory at the Friedrich-Alexander University of Erlangen-Nuremberg. This system arranges 48 high quality loudspeakers mounted on a circular array operating in a low-reverberating chamber. It has been extensively employed in many works [87, 107, 108] that contributed to make Wave Field Synthesis [6] one of the standards for sound field reproduction over wide regions.

The experiments presented in this Section have a twofold goal. On one hand, they serve to confirm the suitability of the wave field measuring technique introduced in Section 5.2.2. On the other and, they allow us to perform a comparative analysis between the proposed rendering technique and the Wave Field Synthesis method. For the sake of compactness, in the following we will refer to the proposed technique as to Geometric Rendering.

6.2.2.1 Experimental setup

A picture of the experimental setup is depicted in Figure 6.32, which shows the rendering system installed at the LMS laboratory at the University of Erlangen-Nuremberg. The loudspeaker array accommodates 48 high-quality emitters disposed

Figure 6.32: Setup for the experimental evaluation of the proposed rendering technique.

on a circumference with radius 1.5 m. The virtual microphone array (described in Section 5.2.2 - Figure 5.2) is located at the centre of the loudspeaker array. It samples the rendered wave field in 200 intermediate angular positions on the circle described by the arm, which has a radius of 0.74 m. For each position, the signal is acquired for $T = 2$ s. Signals are emitted and acquired at the frequency of $F_s = 44.1$ kHz.

6.2.2.2 Results

In this paragraph we report the results of the experimental evaluation of Geometric Rendering, comparing them with those of Wave Field Synthesis. For the sake of clarity, we recall the definition of the terms used in the following, which has been introduced in Section 5.2.2:

- *target wave field*: it refers to the wave field to be rendered;
- *theoretical wave field*: it refers to the theoretical prediction of the rendered wave field;
- *measured wave field*: it refers to the wave field reproduced by the real system, measured by means of the technique based on the Circular Harmonic Decomposition introduced in Section 5.2.2.

We will show two set of experiments. In a first stage, we evaluate the accuracy of the measurement and of the reproduction of the wave field produced by the loudspeaker array when it simulates the presence of a virtual point source. In a second stage, we consider the rendering of a virtual source along with the acoustics of a virtual environment.

Rendering of a virtual point source Simple source models such as plane waves and point sources are commonly used for comparing the accuracy of different rendering systems [101, 99, 109]. In order to compare Geometric Rendering (GR) and Wave Field Synthesis (WFS) we considered point sources, which are simply rendered with both the techniques. As far as GR is concerned, a point source can be interpreted as a beam with angular aperture $\phi = 2\pi$ that originates at the source position[3], and therefore it can be rendered, as usual, by means of the beam shaping engine. As far WFS is concerned, the loudspeaker signals are computed in a closed-form [16].

[3]Notice that, in this case, the beam direction θ is meaningless.

We evaluated the accuracy of the two rendering systems for three distances ($d = 3$ m, $d = 6$ m and $d = 10$ m) of the source from the array centre. The source emits monochromatic signals at 500 Hz, 1000 Hz and 1500 Hz. Table 6.3 summarizes the NMSE results for all the source distances and frequencies. In particular, it reports the NMSE values relative to the theoretical wave fields (denoted with $E_{\mathrm{T}}^{\mathrm{GR}}$ and $E_{\mathrm{T}}^{\mathrm{WFS}}$ for GR and WFS, respectively), and the NMSE values relative to the measured wave fields ($E_{\mathrm{M}}^{\mathrm{GR}}$ and $E_{\mathrm{M}}^{\mathrm{WFS}}$). Best performances are marked in bold.

Table 6.3: NMSE results of the experimental analysis related to the rendering of point sources.

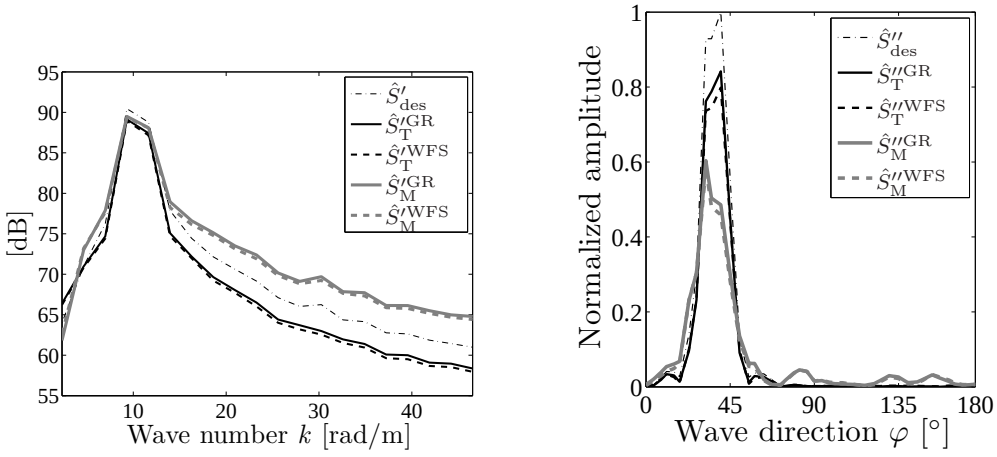
d	f	$E_{\mathrm{T}}^{\mathrm{GR}}$	$E_{\mathrm{T}}^{\mathrm{WFS}}$	$E_{\mathrm{M}}^{\mathrm{GR}}$	$E_{\mathrm{M}}^{\mathrm{WFS}}$
3 m	500 Hz	**0.049**	0.055	0.253	**0.252**
	1000 Hz	**0.058**	0.088	**0.462**	0.567
	1500 Hz	**0.248**	0.627	**0.534**	1.088
6 m	500 Hz	**0.073**	0.074	**0.207**	0.256
	1000 Hz	**0.083**	0.097	**0.364**	0.376
	1500 Hz	**0.270**	0.475	**0.490**	0.745
10 m	500 Hz	**0.083**	0.090	0.205	**0.199**
	1000 Hz	**0.092**	0.112	**0.345**	0.352
	1500 Hz	**0.277**	0.456	**0.530**	0.806

We first consider the theoretical performances, whose NMSE values are denoted with $E_{\mathrm{T}}^{\mathrm{GR}}$ and $E_{\mathrm{T}}^{\mathrm{WFS}}$ in Table 6.3. We notice that the performances of the two rendering methods vary with respect to the distance of the virtual source and with respect to the frequency. More specifically, the two rendering methods have almost the same behaviour at 500 Hz, although GR performs slightly better with respect to WFS. As it is predictable, due to the spatial Nyquist criterion, when the frequency increases the performances of both methods tends to degrade. The spatial Nyquist frequency for this loudspeaker array is around $f_A = 900$ Hz, which is derived from the fact that $f_A \approx \frac{c(M-1)}{4\pi r}$, $M = 48$ being the number of loudspeakers; $r = 1.5$ m the radius of the listening area; and c the speed of sound. For details on the spatial Nyquist frequency see Section 6.2.1.4 or [21]. Looking at the NMSE values, it can be noticed that the degradation amount is not very relevant for 1000 Hz (i.e., when the frequency is slightly above the Nyquist), but it becomes significant for 1500 Hz. Consider, for example, the source at 6 m rendered with GR: the error $E_{\mathrm{T}}^{\mathrm{WFS}}$ is 0.073 at 500 Hz, 0.083 at 1000 Hz and 0.27 at 1500 Hz. Moreover, comparing the behaviours of WFS and GR, it can be noticed that GR tends to limit the degradation introduced by spatial aliasing. For example, considering again the source at 6 m, at 1500 Hz we

have that $E_{\mathrm{T}}^{\mathrm{GR}} = 0.27$, while $E_{\mathrm{T}}^{\mathrm{WFS}} = 0.475$. Analogous considerations can be made for all the source distances and for all the errors types, and they are valid for both WFS and GR.

The experimental results, denoted with $E_{\mathrm{T}}^{\mathrm{GR}}$ and $E_{\mathrm{T}}^{\mathrm{WFS}}$ in Table 6.3, basically confirm the results relative to the theoretical comparison. In particular, GR generally performs better than WFS, especially at high frequency. As it can be expected, the NMSE values are always greater than the correspondent theoretical values. Many factors may contribute to degrade the rendered wave fields, such as the non-ideality of the loudspeakers; a certain amount of reverberations; and noise in the measurements, which can be amplified by the wave field measuring technique.

In order to investigate, at least partially, the reasons of the degradation introduced in the measured wave fields, we resort to a modal analysis. To this end, we selected a case for which spatial aliasing is absent. Figure 6.33. shows the wavenumber (a) and wave direction (b) analyses for the source located at a distance $d = 6$ m from the centre of the array. We observe that the experimental wave number and wave



(a) Wave-number distribution

(b) Wave direction distribution

Figure 6.33: Wave-number and wave direction distributions for the target, theoretical and measured wave fields of GR and WFS for the first experiment for $d = 6$ m.

direction distribution functions well fit the respective target distributions. We notice that, however, that the measured wave fields (relative to both GR and WFS) are slightly degraded by waves propagating from directions around $\Phi = 80°$, $\Phi = 130°$ and $\Phi = 150°$. This is probably due to some spurious reverberations introduced by the environment, which is not completely anechoic.

Rendering of a source in a virtual environment For the second experiment we limit our evaluation to the GR technique. Here we aim at rendering the acoustics of a virtual environment. This is accomplished by spatially overlapping the acoustic beams generated by the image sources, following the approach proposed in Section 4.3.3. We aim at rendering the same virtual environments used for the theoretical evaluation proposed in Section 6.2.1.4, in particular:

- a 5 m $\times$ 5 m square room;
- an L-shaped room;
- a small church.

The mutual arrangements of the loudspeaker array and each virtual environment are shown in Figures 6.34-(a), 6.34-(b) and 6.34-(c). The array is depicted as black crosses disposed on a circle; the listening points are denoted by the black points within the array; and the virtual source is depicted by a black circle.

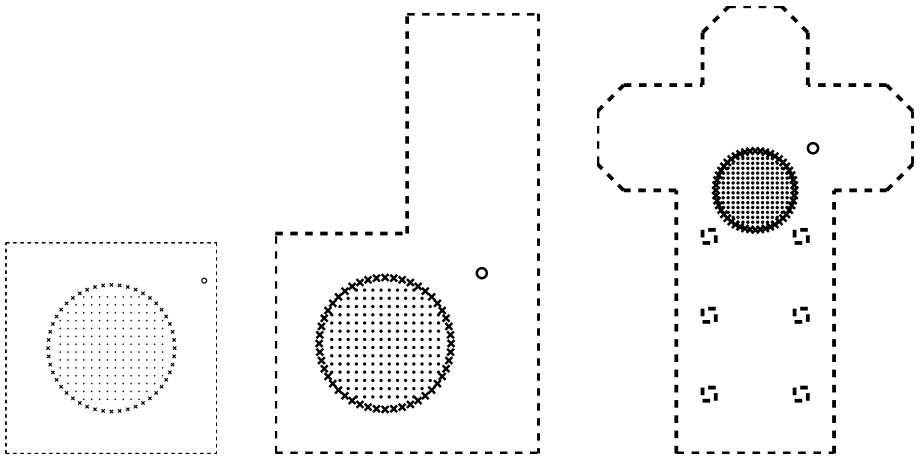


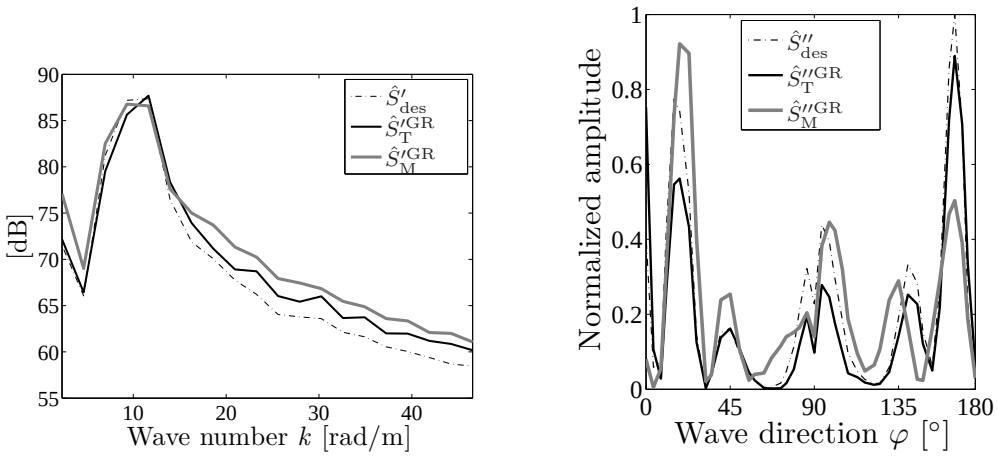
Figure 6.34: Test environment used in the second experiment.

Table 6.4 shows the results of the NMSE evaluation. The results confirm the effectiveness of GR in reproducing a virtual source along with the effects of different virtual environments. The theoretical performances are comparable with those of the rendering of a single point source. We notice that, also in this case, a degradation of the experimental results with respect to the theoretical ones is present. As before, we investigate the origin of such degradation by means of a modal analysis. Figure

Table 6.4: NMSE results of the experimental analysis related to the rendering of a source in three different virtual environments: a square room (a), an L-shaped room (b), a small church (c).

Env.	f	$E_{\mathrm{T}}^{\mathrm{GR}}$	$E_{\mathrm{M}}^{\mathrm{GR}}$
a	500 Hz	0.073	0.283
	1000 Hz	0.069	0.202
	1500 Hz	0.238	0.681
b	500 Hz	0.075	0.203
	1000 Hz	0.080	0.292
	1500 Hz	0.251	0.498
c	500 Hz	0.050	0.286
	1000 Hz	0.053	0.445
	1500 Hz	0.336	0.552

6.35 shows the wave number (a) and wave direction (b) distributions relative to the rendering of a source in the L-shaped virtual environment, at 500 Hz. We notice



(a) Wave-number distribution

(b) Wave direction distribution

Figure 6.35: Wave-number and wave direction distributions for the target, theoretical and measured wave fields of GR for the second experiment.

that the wave number distributions of theoretical and measured wave fields generally well fit the target one. Also in this experiment, the measurement methodology allows us to appreciate the presence of some spurious propagating wave fronts in the range [60° , 90°], which probably correspond to reverberations generated by the reproduction room.

Discussion As stated at the beginning of this Section, the experiments we have presented can be used to assess the suitability of the wave field measuring technique introduced in Section 5.2.2. This is possible because the experiments have been conducted in a controlled environment. For this reason, we expect the experimental results to match, as much as possible, the theoretical ones. This was indeed verified by the results of the proposed experiments. In fact, although a certain amount of degradation in the rendered wave fields is inevitably present (due to the non-ideality of the loudspeakers and of the environment, as well as of the acquisition system), the experimental results exhibit a behaviour similar to the theoretical ones. In particular:

- the experimental results shown in Tables 6.3 and 6.4 present the same trend of the corresponding theoretical results;
- the measuring technique was able to capture the differences between GR and WFS. More specifically, Table 6.3 shows how the theoretical better performances of GR with respect of WFS are captured also in the experimental results.

Moreover, the measurement technique revealed to be capable of detecting the subtle details of the reproduced wave fields. In particular, through a modal analysis of the measured wave fields, we noticed the presence of spurious reverberation caused by the non-completely anechoic environment.

7 Conclusions and future research directions

In this thesis we have presented novel solutions for the analysis and the synthesis of acoustic wave fields. All the proposed methodologies have been developed following a geometric approach, which allowed us to describe the propagation in terms of acoustic rays, and to model complex wave fields as the superposition of elementary acoustic beams.

As far as wave field analysis is concerned, we have proposed a number of techniques that make use of projective geometry for reformulating standard estimation problems such as source localization and array calibration; and to formulate novel problems, such as the estimation of the position of acoustic reflectors present in an environment. Furthermore, we have proposed an unified theory for dealing with all the types of estimation problem. In particular, we have seen how acoustic measurements (TOAs, TDOAs, DOAs) can be turned into homogeneous quadratic constraints (i.e., conic sections) acting on the primitive of interest leading to the formulation of a cost function that shares the same form for all the problems. We have also discussed a number of minimization strategies for that cost function, proposing both exact and iterative solutions. While the former are suitable for achieving the maximum accuracy in the estimation (useful in reflector localization tasks), the latter are characterized by a shorter computation time, thus being more suitable for real-time applications such as source localization.

As far as wave field synthesis is concerned, we have proposed a geometric approach for the rendering of complex wave fields with loudspeaker arrays in a reverberant environment. Through an efficient beam tracing engine, the reverberation of the real environment is predicted and included into the data model as the contribution of a number of image (wall-reflected) loudspeakers. This way, image loudspeakers actually contribute to the rendering and operate room-compensation in a completely

geometric manner. To do so, the system requires the knowledge of the geometry of the environment, which can be estimated in advance by means of reflector localization algorithms based on TOA and DOA measurements. The same beam tracing engine is then used for modelling the wave field to be rendered. More specifically, given the geometry of a virtual environment, the beam tracer is used for determining the set of the acoustic beams generated by the reflections of an acoustic source over the virtual walls. Each beam is reproduced by means of the loudspeakers through a suitable beam shaping technique.

For the evaluation of the proposed techniques we have introduced novel methodologies and metrics. In particular, we have proven the existence of a linear relationship between the measurement error and the estimation error. This fact allowed us to verify in a very efficient way the effectiveness of the proposed estimation algorithms. Moreover, this technique is applicable to all the estimation problems that are based on the minimization of a cost function. For the evaluation of the geometric rendering technique, we have developed a methodology that is valid for both simulative and experimental analyses. In particular, standard MSE-based metrics have been complemented with a novel modal analysis conducted in the wave number domain, which allowed us to give insight on the distribution of the error for different wave fronts in the sound field. In order to measure wave fields reproduced by real loudspeaker systems, we have resorted to an extrapolation technique. The wave field is first sampled on a circle by means of a pair of rotating microphones, and then extrapolated through circular harmonic decomposition.

Simulative and experimental results proved the effectiveness of both analysis and synthesis methodologies. In particular, reflector localization techniques turned out to be accurate and robust even in real scenarios, enabling the reconstruction of the geometry of different environments. The geometric approach to sound field synthesis revealed to be accurate in rendering the effect of a single source as well as the effect of a source located inside a virtual environment. Geometric room compensation revealed to be effective in canceling out most of the reverberations, enabling a free-field behaviour in an arbitrary environment.

To conclude, we mention a set of possible evolutions of the work presented in this thesis. As far as wave field analysis is concerned, it could be interesting a further exploitation of the common formulation of all the constraints arising from different acoustic measurements. As an example, let us consider a scenario in which multiple microphone arrays or clusters provide DOA and TOA measurements relative to the

reflective paths. This way, it could be possible to build an unique cost function embedding DOA and TOA constraints. Moreover, assuming a priori knowledge on the covariance of the measurement error it may be possible to assign different weights to each type of constraints, such that the estimation error is minimized. In the context of wave field analysis, another extension could be that of exploiting different acoustic measurements along with the knowledge of the room geometry. This way, we could devise novel estimation problems, such as "behind the corner" source localization and tracking. As far as wave field rendering is concerned, it may be desirable to include also other propagating phenomena in the reverberation model, such as diffusion and diffraction. Assuming the knowledge of diffusive and diffractive properties of the walls, room compensation could be made even more effective if complemented by this information. Furthermore, the effect of a source within a virtual environment may be rendered in an even more realistic fashion modelling also diffusive and diffractive propagations.

Bibliography

[1] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*, Digital signal processing. Springer, 2001.

[2] Y.Huang, J.Benesty, and G.Elko, *Audio Signal Processing for Next Multimedia Communication Systems*, Kluwer Academic Publishers, 2004.

[3] Ce Wang, S. Griebel, and M. Brandstein, "Robust automatic video-conferencing with multiple cameras and microphones," in *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, 2000, vol. 3, pp. 1585 –1588 vol.3.

[4] D.N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 5, pp. 499 – 508, sept. 2004.

[5] Zicheng Liu, Zhengyou Zhang, Li-Wei He, and Phil Chou, "Energy-based sound source localization and gain normalization for ad hoc microphone arrays," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, april 2007, vol. 2, pp. II–761 –II–764.

[6] A. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J.Acoust.Soc.Am.*, vol. 93, pp. 2764–2778, 1993.

[7] R. Nicol, "Sound spatialization by higher order ambisonics," in *Proc. 2nd Int. Symp. on Ambisonics and Spherical Acoustics*, 2010.

[8] A.A. Kaufman and A.L. Levshin, *Acoustic and elastic wave fields in geophysics*, Number v. 2 in Acoustic and Elastic Wave Fields in Geophysics. Elsevier, 2002.

[9] A. Krokstad, S Strom, and S. Sorsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, pp. 118–125, 1968.

[10] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West, "A beam tracing approach to acoustic modeling for interactive virtual environments," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1998, SIGGRAPH '98, pp. 21–32, ACM.

[11] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2003.

[12] J.Stolfi, "Oriented projective geometry," in *Proc. of Symposium on Computational Geometry*, 1987, pp. 76–85.

[13] J.Semple and T.Kneebone, *Algebraic projective geometry*, Oxford University Press, 1952.

[14] Mei Han and T. Kanade, "Multiple motion scene reconstruction with uncalibrated cameras," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 7, pp. 884 – 894, july 2003.

[15] P.J. La Riviere, Junguo Bian, and P.A. Vargas, "Penalized-likelihood sinogram restoration for computed tomography," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 8, pp. 1022 –1036, aug. 2006.

[16] Sascha Spors, Rudolf Rabenstein, and Jens Ahrens, "The theory of wave field synthesis revisited," in *Proc. 124th Convention of the Audio Engineering Society, Amsterdam, May 17-20*, 2008.

[17] Thomas Funkhouser, Patrick Min, and Ingrid Carlbom, "Real-time acoustic modeling for distributed virtual environments," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM, SIGGRAPH '99*, Los Angeles, California, USA, 1999, pp. 365–374, ACM Press/Addison-Wesley Publishing Co.

[18] Fabio Antonacci, Marco Foco, Augusto Sarti, and Stefano Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 812–824, May 2008.

[19] Fabio Antonacci, Alberto Calatroni, Antonio Canclini, Andrea Galbiati, Augusto Sarti, and Stefano Tubaro, "Rendering of an acoustic beam through an array of loudspeakers," in *Proceedings of the 12th International Conference on Digital Audio Effects, DAFx-09*, Como, Italy, September 2009.

[20] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 354–369, July 2007.

[21] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.

[22] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[23] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*, Prentice Hall signal processing series. Prentice-Hall PTR, 1993.

[24] A. Kuntz and R. Rabenstein, "Cardioid pattern optimization for a virtual circular microphone array," in *proc. of EAA Symposium on Auralization*, 2009.

[25] S.Thrun, "Affine structure from sound," in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, Cambridge, MA, 2005, MIT Press.

[26] J.Hu C.C. Wang; C.H. Lin, "Probabilistic structure from sound," *Advanced Robotics*, vol. 23, no. 12-13, pp. 1687–1702, 2009.

[27] L.L. Beranek and Acoustical Society of America, *Acoustics*, Electrical and electronic engineering. Published by the American Institute of Physics for the Acoustical Society of America, 1986.

[28] D.T. Blackstock, *Fundamentals of physical acoustics*, Wiley-Interscience. Wiley, 2000.

[29] P.M.C. Morse and H. Feshbach, *Methods of theoretical physics*, Number v. 1 in International series in pure and applied physics. McGraw-Hill, 1953.

[30] T.D. Rossing, *Springer handbook of acoustics*, Springer Handbook of. Springer, 2007.

[31] V. Troyan and Y. Kiselev, *Statistical Methods of Geophysical Data Processing*, World Scientific, 2010.

[32] A.S. Monin, *Theoretical geophysical fluid dynamics*, Environmental fluid mechanics. Kluwer Academic Publishers, 1990.

[33] Michael Monks, Byong Mok Oh, and Julie Dorsey, "Acoustic simulation and visualization using a new unified beam tracing and image source approach," in *Audio Engineering Society Convention 101*, 1996.

[34] "Adaptive algorithms for the identification of sparse impulse responses," in *Topics in Acoustic Echo and Noise Control*, Eberhard Hansler and Gerhard Schmidt, Eds., Signals and Communication Technology, pp. 125–153. Springer Berlin Heidelberg, 2006.

[35] K. Crammer and D.D. Lee, "Room impulse response estimation using sparse online prediction and absolute loss," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, may 2006, vol. 3, p. III.

[36] Yuanqing Lin and Daniel D. Lee, "Bayesian regularization and nonnegative deconvolution for room impulse response estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 839–847, 2006.

[37] Jingdong Chen, Jacob Benesty, and Yiteng Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 170–170, January 2006.

[38] Yuanqing Lin, Daniel D. Lee, and Lawrence K. Saul, "Nonnegative deconvolution for time of arrival estimation," in *in ICASSP*, 2004, pp. 377–380.

[39] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 4, pp. 320 – 327, aug 1976.

[40] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *Signal Processing, IEEE Transactions on*, vol. 41, no. 2, pp. 525 –533, feb 1993.

[41] P. Stoica and R.L. Moses, *Introduction to spectral analysis*, Prentice Hall, 1997.

[42] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, no. 4, pp. 661–675, 1971.

[43] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408 – 1418, aug. 1969.

[44] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276 – 280, mar 1986.

[45] G. Bienvenu, "Influence of the spatial coherence of the background noise on high resolution passive methods," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.*, apr 1979, vol. 4, pp. 306 – 309.

[46] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984 –995, jul 1989.

[47] H. Teutsch and W. Kellermann, "Eb-esprit: 2d localization of multiple wideband acoustic sources using eigen-beams," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, march 2005, vol. 3, pp. iii/89 – iii/92 Vol. 3.

[48] M. R. Azimi-Sadjadi, A. Pezeshkib, L. L. Scharfb, and M. Hohil, "Wideband DOA estimation algorithms for multiple target detection and tracking using unattended acoustic sensors," in *Proc. of Spie - The International Society For Optical Engineering*, 2004.

[49] Liqun Qi and Kok Lay Teo, "Multivariate polynomial minimization and its application in signal processing," *Journal of Global Optimization*, vol. 26, pp. 419–433, 2003, 10.1023/A:1024778309049.

[50] P Dreesen and B De Moor, "Polynomial optimization problems are eigenvalue problems," *Electrical Engineering*, pp. 49–68, 2009.

[51] Martin Mevissen and Masakazu Kojima, "Sdp relaxations for quadratic optimization problems derived from polynomial optimization problems," *APJOR*, vol. 27, no. 1, pp. 15–38, 2010.

[52] Bernard Hanzon and Dorina Jibetean, "Global minimization of a multivariate polynomial using matrix methods," *Journal of Global Optimization*, vol. 27, pp. 1–23, 2003, 10.1023/A:1024664432540.

[53] Victor Blanco and Justo Puerto, "Some algebraic methods for solving multiobjective polynomial integer programs," *Journal of Symbolic Computation*, vol. 46, no. 5, pp. 511 – 533, 2011, Groebner Bases and Applications.

[54] Zhi-Quan Luo and Shuzhong Zhang, "A semidefinite relaxation scheme for multivariate quartic polynomial optimization with quadratic constraints," *SIAM J. on Optimization*, vol. 20, pp. 1716–1736, January 2010.

[55] Simai He, Zhening Li, and Shuzhong Zhang, "General constrained polynomial optimization: an approximation approach," 2009.

[56] Masakazu Kojima, "Sums of squares relaxations of polynomial semidefinite programs," in *of Mathematical and computing Sciences, Tokyo Institute of Technology, Meguro, Tokyo*, 2003, pp. 152–8552.

[57] Jean B. Lasserre, "Convergent sdp-relaxations in polynomial optimization with sparsity," *SIAM Journal on Optimization*, vol. 17, pp. 822–843.

[58] D. Du, P.M. Pardalos, and W. Wu, *Mathematical theory of optimization*, Nonconvex optimization and its applications. Kluwer Academic, 2001.

[59] A. Beck, P. Stoica, and Jian Li, "Exact and approximate solutions of source localization problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770 –1778, May 2008.

[60] J. Lagarias, J. Reeds, M. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 1, pp. 112–147, 1998.

[61] J. Scheuing and Bin Yang, "Disambiguation of tdoa estimates in multi-path multi-source environments (datemm)," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, may 2006, vol. 4, p. IV.

[62] C.M. Zannini, A. Cirillo, R. Parisi, and A. Uncini, "Improved tdoa disambiguation techniques for sound source localization in reverberant environments," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 30 2010-june 2 2010, pp. 2666 –2669.

[63] Jingjing Jia, Mingjie Liu, and Xiaofeng Li, "Acoustic passive localization algorithm based on wireless sensor networks," in *Mechatronics and Automation, 2009. ICMA 2009. International Conference on*, aug. 2009, pp. 1145 –1149.

[64] Jingjing Jia, Mingjie Liu, and Xiaofeng Li, "Acoustic localization algorithm using wireless sensor networks," in *Intelligent Computation Technology and Automation, 2009. ICICTA '09. Second International Conference on*, oct. 2009, vol. 2, pp. 434 –437.

[65] A. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations - a maximum likelihood approach," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 1958–1966, Dec. 1989.

[66] Stanley T. Birchfield and Amarnag Subramanya, "Microphone array position calibration by basis-point classical multidimensional scaling," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1025–1034, September 2005.

[67] T. Cox and M. Cox, *Multidimensional Scaling*, Chapman and Hall, 2001.

[68] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 666–670, Mar. 2008.

[69] A. Redondi, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Geometric calibration of distributed microphone arrays," in *in proc. of IEEE International Workshop on Multimedia Signal Processing*, Oct. 2009, pp. 1–5.

[70] M. Hennecke, T. Plotz, G. Fink, J. Schmalenstroer, and R. Hab-Umbach, "A hierarchical approach to unsupervised shape calibration of microphone array networks," in *IEEE/SP 15th Workshop on Statistical Signal Processing, SSP '09.*, Sept. 2009, pp. 257–260.

[71] A.Tobias, "Acoustic emission source location in two dimensions by an array of three transducers," in *Non-destructive testing*, 1976.

[72] S.S Reddi, "An exact solution to range computation with time delay information for arbitrary array geometries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 485–486, 1993.

[73] N. Liu, Z. Xu, and B.M. Sadler, "Low-complexity hyperbolic source localization with a linear sensor array," *IEEE Signal Processing Letters*, vol. 15, pp. 865–868, 2008.

[74] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1661–1669, 1987.

[75] H. Schau and A. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1223–1225, 1987.

[76] M. Gillette and H. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Processing Letters*, vol. 15, pp. 1–4, 2008.

[77] M.S. Brandstein, J.E. Adcock, and H.F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *Speech and Audio Processing, IEEE Transactions on*, vol. 5, no. 1, pp. 45 –50, jan 1997.

[78] E A P Habets and P C W Sommen, "Optimal microphone placement for source localization using time delay estimation," *Proc Workshop Circuits Systems and Signal Processing ProRISC*, pp. 284–287, 2002.

[79] F. Ribeiro, C. Zhang, D. Florêncio, and D.Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1781–1792, 2010.

[80] D. Ba, F. Ribeiro, C. Zhang, and D. Florencio, "ℓ_1 regularized room modeling with compact microphone arrays," in *proc. of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2010.

[81] S. Tervo and T. Korhonen, "Estimation of reflective surfaces from continuous signals," in *proc. of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2010.

[82] Ivan Dokmanic, Yue M. Lu, and Martin Vetterli, "Can one hear the shape of a room: The 2-d polygonal case," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 321 –324.

[83] Richard O. Duda and Peter E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, pp. 11–15, January 1972.

[84] E. Nastasia, F. Antonacci, A. Sarti, and S. Tubaro, "Robust inference of room geometry from acoustic measurements using the hough transform," in *Proceedings of the 19th European Signal Processing Conference*, 2011.

[85] Jeffrey Borish, "Extension of the image model to arbitrary polyhedra," *The Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, 1984.

[86] U.R. Kristiansen, A. Krokstad, and T. Follestad, "Extending the image method to higher-order reflections," *Applied Acoustics*, vol. 38, no. 2-4, pp. 195 – 206, 1993.

[87] "Wave field synthesis techniques for spatial sound reproduction," in *Topics in Acoustic Echo and Noise Control*, Eberhard Hnsler and Gerhard Schmidt, Eds., Signals and Communication Technology, pp. 517–545. Springer Berlin Heidelberg, 2006.

[88] Paul S. Heckbert and Pat Hanrahan, "Beam tracing polygonal objects," in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques, ACM, SIGGRAPH '84*, Minneapolis, Minnesota, USA, 1984, vol. 18, pp. 119–127.

[89] Norm Dadoun, David G. Kirkpatrick, and John P. Walsh, "The geometry of beam tracing," in *Proceedings of the first annual symposium on Computational geometry, ACM, SCG '85*, Sedona, Arizona, USA, June 1985, pp. 55–61.

[90] U. Stephenson and U. Kristiansen, "Pyramidal beam tracing and time dependent radiosity," in *Proceedings of the 15th International Congress on Acoustics*, Trondheim, Norway, June 1995, pp. 657–660.

[91] John P. Walsh and Norman Dadoun, "What are we waiting for? the development of godot, ii," *Proceedings of the 103rd Meeting of the Acoustical Society of America*, vol. 71, no. S1, pp. S5, April 1982.

[92] Samuli Laine, Samuel Siltanen, Tapio Lokki, and Lauri Savioja, "Accelerated beam tracing algorithm," *Applied Acoustics*, vol. 70, no. 1, pp. 172 – 181, 2009.

[93] D. Markovic, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Visibility-based beam tracing for soundfield rendering," in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, oct. 2010, pp. 40 –45.

[94] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA '09*, New Paltz, New York, USA, October 2009, pp. 313 –316.

[95] Albert Tarantola, *InverseProblemTheory*, SIAM - Society for Industrial and Applied Mathematics, 2004.

[96] R. Gilmore, *Catastrophe theory for scientists and engineers*, Dover Publications, Inc., 1993, Reprint of the 1981 original.

[97] S.I. Amari, H. Nagaoka, and D. Harada, *Methods of Information Geometry*, Translations of Mathematical Monographs. American Mathematical Society, 2007.

[98] E.L. Lehmann and G. Casella, *Theory of point estimation*, Springer texts in statistics. Springer, 1998.

[99] Sascha Spors and Jens Ahrens, "A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling," in *Audio Engineering Society Convention 125*, 10 2008.

[100] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, 2010.

[101] Jerome Daniel, Sebastien Moreau, and Rozenn Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, 3 2003.

[102] T.D. Abhayapala and A. Gupta, "Non-spherical microphone array structures for 3D beamforming and spherical harmonic analysis," in *proc. of the 11th International Workshop on Acoustic Echo and Noise Control*, 2008.

[103] A. Gupta and T.D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1149 –1159, july 2011.

[104] P. Peretti, L. Romoli, S. Cecchi, L. Palestini, and F. Piazza, "Phase approximation of linear geometry driving functions for sound field synthesis," in *proc. of the 2010 European Signal Processing Conference*, 2010.

[105] H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoust. Soc. Am.*, vol. 120, pp. 2724–2736, 2006.

[106] Sascha Spors, "Spatial aliasing artifacts produced by linear loudspeaker arrays used for wave field synthesis," *Second IEEEEURASIP Inter national Symposium on Control Communica tions and Signal Processing*, , no. 1, pp. 1–4, 2006.

[107] S. Spors, H. Teutsch, and R. Rabenstein, "High-quality acoustic rendering with wave field synthesis," in *Vision, Modeling, and Visualization*, 2002, pp. 101–108.

[108] Thomas Sporer, Jan Plogsties, and Sandra Brix, "Carrouso - an european approach to 3d-audio," in *Audio Engineering Society Convention 110*, 5 2001.

[109] J.H. Chang, J.W. Choi, and Y.H. Kim, "Reproduction of a plane-wave sound field based on wave number domain focusing: Comparison with ambisonics and wfs," *The 8th International conference on Motion and Vibration Control*, 08 2006.