Sound source localization with data and model uncertainties using the EM and Evidential EM algorithms

Thèse présentée pour l’obtention du grade de Docteur de l’UTC

Soutenue le 09 décembre 2014
Spécialité : Information Technologies and Systems
Université de Technologie de Compiègne

Thesis

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Area of specialization: Information Technologies and Systems

by

Xun WANG

Sound Source Localization with Data and Model Uncertainties Using the EM and Evidential EM Algorithms

Heudiasyc Laboratory, UMR CNRS 7253
Roberval Laboratory, UMR CNRS 7337

Defended on December 9th, 2014.
Thesis Committee:

President: Thierry Denœux Université de Technologie de Compiègne
Reviewers: Alain Appriou ONERA
Jean-Marc Girault Université François Rabelais de Tours
Examiners: Charles Pézerat Université du Maine
Olivier Strauss Université Montpellier II
Supervisors: Jérôme Antoni Université de Lyon
Benjamin Quost Université de Technologie de Compiègne
Jean-Daniel Chazot Université de Technologie de Compiègne
Abstract

This work addresses the problem of multiple sound source localization for both deterministic and random signals measured by an array of microphones. The problem is solved in a statistical framework via maximum likelihood. The pressure measured by a microphone is interpreted as a mixture of latent signals emitted by the sources; then, both the sound source locations and strengths can be estimated using an expectation-maximization (EM) algorithm. In this thesis, two kinds of uncertainties are also considered: on the microphone locations and on the wavenumber. These uncertainties are transposed to the data in the belief functions framework. Then, the source locations and strengths can be estimated using a variant of the EM algorithm, known as Evidential EM (E2M) algorithm.

The first part of this work begins with the deterministic signal model without consideration of uncertainty. The EM algorithm is then used to estimate the source locations and strengths: the update equations for the model parameters are provided. Furthermore, experimental results are presented and compared with the beamforming and the statistically optimized near-field holography (SONAH), which demonstrate the advantage of the EM algorithm.

The second part raises the issue of model uncertainty and shows how the uncertainties on microphone locations and wavenumber can be taken into account at the data level. In this case, the notion of the likelihood is extended to the uncertain data. Then, the E2M algorithm is used to solve the sound source estimation problem. In both the simulation and real experiment, the E2M algorithm proves to be more robust in the presence of model and data uncertainty.

The third part of this work considers the case of random signals, in which the amplitude is modeled by a Gaussian random variable. Both the certain and uncertain cases are investigated. In the former case, the EM algorithm is employed to estimate the sound sources. In the latter case, microphone location and wavenumber uncertainties are quantified similarly to the second part of the thesis. Finally, the source locations and the variance of the random amplitudes are estimated using the E2M algorithm.

**Key words:** sound source localization, holography, beamforming, acoustic imaging,
statistical inference from imprecise data, belief functions, Evidential EM algorithm.
Résumé

Ce travail de thèse se penche sur le problème de la localisation de sources acoustiques à partir de signaux déterministes et aléatoires mesurés par un réseau de microphones. Le problème est résolu dans un cadre statistique, par estimation via la méthode du maximum de vraisemblance. La pression mesurée par un microphone est interprétée comme étant un mélange de signaux latents émis par les sources. Les positions et les amplitudes des sources acoustiques sont estimées en utilisant l’algorithme espérance-maximisation (EM). Dans cette thèse, deux types d’incertitude sont également pris en compte : les positions des microphones et le nombre d’onde sont supposés mal connus. Ces incertitudes sont transposées aux données dans le cadre théorique des fonctions de croyance. Ensuite, les positions et les amplitudes des sources acoustiques peuvent être estimées en utilisant l’algorithme E2M, qui est une variante de l’algorithme EM pour les données incertaines.

La première partie des travaux considère le modèle de signal déterministe sans prise en compte de l’incertitude. L’algorithme EM est utilisé pour estimer les positions et les amplitudes des sources. En outre, les résultats expérimentaux sont présentés et comparés avec le beamforming et la holographie optimisée statistiquement en champ proche (SONAH), ce qui démontre l’avantage de l’algorithme EM.

La deuxième partie considère le problème de l’incertitude du modèle et montre comment les incertitudes sur les positions des microphones et le nombre d’onde peuvent être quantifiées sur les données. Dans ce cas, la fonction de vraisemblance est étendue aux données incertaines. Ensuite, l’algorithme E2M est utilisé pour estimer les sources acoustiques. Finalement, les expériences réalisées sur les données réelles et simulées montrent que les algorithmes EM et E2M donnent des résultats similaires lorsque les données sont certaines, mais que le dernier est plus robuste en présence d’incertitudes sur les paramètres du modèle.

La troisième partie des travaux présente le cas de signaux aléatoires, dont l’amplitude est considérée comme une variable aléatoire gaussienne. Dans le modèle sans incertitude, l’algorithme EM est utilisé pour estimer les sources acoustiques. Dans le modèle incertain, les incertitudes sur les positions des microphones et le nombre d’onde sont transposées aux données comme dans la deuxième partie. Enfin, les positions et les variances des amplitudes

**Mots clés:** localisation de sources acoustiques, holographie, beamforming, imagerie acoustique, inférence statistique à partir de données imprécises, fonctions de croyance, algorithme E2M.
Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Jérôme Antoni, Dr. Benjamin Quost and Dr. Jean-Daniel Chazot. I would never have been able to finish my dissertation without their patient guidance, advice, encouragement and careful modification of the manuscript they have provided throughout my Ph.D. study and research.

I would like to thank other members of staff at the laboratories Heudiasyc and Roberval, for the assistance they provided at all levels of the research work. In particular, I would like to thank the technicians at Roberval Jean-Marc Gherbezza and Félix Foucart for their helps during the experiments.

I would also like to acknowledge the financial support by Collegium UTC/CNRS for providing the funding which allows me to undertake this research and to attend the conferences.

Finally, I would like to thank my family. They were always supporting me and encouraging me with their best wishes.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>List of Abbreviations and Symbols</td>
<td>xv</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>I Sound Sources Localization From Certain Measurements</td>
<td>3</td>
</tr>
<tr>
<td>2 State of Art</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Background and Basic Description of Sound Source Localization Model</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Review of Different Source Localization Methods</td>
<td>6</td>
</tr>
<tr>
<td>2.2.1 Beamforming</td>
<td>7</td>
</tr>
<tr>
<td>2.2.2 NAH</td>
<td>8</td>
</tr>
<tr>
<td>2.2.3 SONAH</td>
<td>9</td>
</tr>
<tr>
<td>2.3 EM Algorithm</td>
<td>10</td>
</tr>
<tr>
<td>2.3.1 Introduction of the EM Algorithm</td>
<td>11</td>
</tr>
<tr>
<td>2.3.2 Review of Existing Works on Source Estimation Via the EM Algorithm</td>
<td>13</td>
</tr>
<tr>
<td>3 Sound Sources Localization Under the Deterministic Amplitude Assumption</td>
<td>15</td>
</tr>
<tr>
<td>3.1 Model</td>
<td>15</td>
</tr>
<tr>
<td>3.2 Model Estimation Via the EM Algorithm</td>
<td>16</td>
</tr>
<tr>
<td>3.3 Comparison of Maximum Likelihood and Beamforming</td>
<td>19</td>
</tr>
<tr>
<td>3.4 Experiments</td>
<td>20</td>
</tr>
<tr>
<td>3.5 Conclusion</td>
<td>22</td>
</tr>
</tbody>
</table>
## II Sound Source Localization in Presence of Uncertainty

### 4 State of Art: Uncertainties in Sound Source Localization Model

4.1 Uncertainties in the Model

4.1.1 Microphone Location Uncertainty

4.1.2 Medium Uncertainty

4.1.3 Sound Source Localization Model with Uncertainties

4.2 Fuzzy Sets, Belief Functions, and the Evidential EM Algorithm

5 Sound Source Estimation in Presence of Uncertainty: Deterministic Amplitude Case

5.1 Model

5.2 Uncertainty Representation and Propagation

5.2.1 Uncertain Data

5.2.2 Likelihood Function of Uncertain Data

5.3 Model Estimation Using the Evidential EM Algorithm

5.4 Experiments

5.4.1 Simulated Data

5.4.2 Real Experiment

5.5 Conclusion

### III Sound Source Localization for Random Signals

6 Sound Source Estimation Under the Random Amplitude Assumption

6.1 Sound Source Model With Random Amplitude

6.2 Model Estimation Via the EM Algorithm

6.3 Experiments

6.4 Conclusion

7 Sound Source Localization in Presence of Uncertainty: Random Amplitude Case

7.1 Model
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2</td>
<td>Uncertainty Representation and Propagation</td>
<td>88</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Uncertain Data</td>
<td>88</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Likelihood Function of Uncertain Data</td>
<td>89</td>
</tr>
<tr>
<td>7.3</td>
<td>Model Estimation Via the Evidential EM Algorithm</td>
<td>89</td>
</tr>
<tr>
<td>7.4</td>
<td>Experiments</td>
<td>93</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Simulated Data</td>
<td>94</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Real Experiment</td>
<td>95</td>
</tr>
<tr>
<td>7.5</td>
<td>Conclusion</td>
<td>97</td>
</tr>
<tr>
<td>8</td>
<td>Conclusion and Perspectives</td>
<td>101</td>
</tr>
<tr>
<td>8.1</td>
<td>Conclusion</td>
<td>101</td>
</tr>
<tr>
<td>8.2</td>
<td>Perspectives</td>
<td>102</td>
</tr>
</tbody>
</table>

## References | 103 |

## Appendix A | Complex Gaussian Random Variables | 109 |

## Appendix B | Output Variance Approximated By Taylor Expansion | 111 |

## Appendix C | Computation of the Gradient Matrix in Section 5.2 | 113 |

## Appendix D | Computation of the Log-Likelihood of the Uncertain Data | 119 |

## Appendix E | Computation of the Gradient Matrix in Section 7.2 | 121 |
List of Figures

2.1 Spherical waves emitted by a sound source $r$ .......................... 6

3.1 Sound source estimation problem (2D case). The microphone locations are represented by red crosses, the sound sources by blue crosses. The origin is assumed here to be the center of the array of microphones. .............. 16

3.2 Experimental setup and microphone distribution. .......................... 20

3.3 Comparison between EM, beamforming and SONAH at $f=525\text{Hz}$. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. .... 23

3.4 Comparison between EM, beamforming and SONAH at $f=1325\text{Hz}$. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. .... 24

3.5 Comparison between EM, beamforming and SONAH at $f=5025\text{Hz}$. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. .... 25

3.6 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S=2$ (c) EM estimates, $S=6$ (d) EM estimates, $S=10$. $f=525\text{Hz}$ .... 26

3.7 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S=2$ (c) EM estimates, $S=6$ (d) EM estimates, $S=10$. $f=1325\text{Hz}$. .... 27

3.8 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S=2$ (c) EM estimates, $S=6$ (d) EM estimates, $S=10$. $f=5025\text{Hz}$. .... 28

5.1 Sound source estimation problem (2D case) with uncertain meta-parameters (microphone locations and wavenumber). .......................... 53
5.2 2D sound source localization example. The sound sources are represented by blue crosses. The actual microphone locations are represented by red crosses. The real location of 5th microphone is uncertain, taking possible values on the pink line segment. .......................... 55

5.3 Log-likelihood of \( p_l \) (a); generalized log-likelihood \( pl_l \) (b-d) for different x-coordinates of \( S_2 \) and x-coordinates actual locations of microphone \( M_5 \). The values of \( \sigma_1^2 \) in \( pl_l \) are 0.003, 0.01 and 0.05 in (b), (c) and (d) respectively. In all cases, microphone \( M_5 \) is assumed to have a x-coordinate \( m_{x,5} = 2 \). ... 56

5.4 Source location estimates obtained using the EM (blue circles) and E2M (red crosses) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. \( \sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.02 \text{m}, \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 15 \times \frac{\pi}{180}, \sigma_k = 0.5 \text{m}^{-1} \). .......................... 63

5.5 Source location estimates obtained using the EM (blue circles) and E2M (red crosses) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. \( \sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.04 \text{m}, \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 20 \times \frac{\pi}{180}, \sigma_k = 0.5 \text{m}^{-1} \). .......................... 64

5.6 MSE and 95% confidence intervals for the sound source location estimates. Case (a): \( \sigma_k = 0, \sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.01 \text{l m}, \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 5 \times \frac{\pi}{180}, \text{l} = 0,1,2,3,4 \). Case (b): \( \sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 0, \sigma_k = 0,0.25,0.5,0.75,1 \text{m}^{-1} \). .......................... 65

5.7 Source location estimates obtained using EM (blue points) and E2M (red crosses), and corresponding 95% confidence ellipses on the x-y plane. Black crosses represent the actual source locations. .......................... 66

5.8 Source location estimates obtained using EM (left) and E2M (right), and corresponding 95% confidence ellipsoids in 3D space. Black crosses represent the source locations. .......................... 66

6.1 Sound source estimation problem (2D case) in the far-field case. ............... 76

6.2 Comparison between EM, beamforming and SONAH at f=400Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. ... 78

6.3 Comparison between EM, beamforming and SONAH at f=1200Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. ... 79
6.4 Comparison between EM, beamforming and SONAH at f=2200Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. 80
6.5 Comparison between EM, beamforming and SONAH at f=4300Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane. 81
6.6 Sound field on the microphone plane using (a) measurement (b) EM estimates, S = 2 (c) EM estimates, S = 6 (d) EM estimates, S = 10. f=400Hz. 82
6.7 Sound field on the microphone plane using (a) measurement (b) EM estimates, S = 2 (c) EM estimates, S = 6 (d) EM estimates, S = 10. f=1200Hz. 83
6.8 Sound field on the microphone plane using (a) measurement (b) EM estimates, S = 2 (c) EM estimates, S = 6 (d) EM estimates, S = 10. f=2200Hz. 84
6.9 Sound field on the microphone plane using (a) measurement (b) EM estimates, S = 2 (c) EM estimates, S = 6 (d) EM estimates, S = 10. f=4300Hz. 85

7.1 Random amplitude signal source location estimates obtained using the EM (blue) and E2M (red) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. \( \sigma_x = \sigma_y = \sigma_z = 0.02 \text{m}, \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 10 \frac{\pi}{180}, \sigma_k = 0.5 \text{m}^{-1} \). 96
7.2 Random amplitude signal source location estimates obtained using the EM (blue) and E2M (red) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. \( \sigma_x = \sigma_y = \sigma_z = 0.03 \text{m}, \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 15 \frac{\pi}{180}, \sigma_k = 0.5 \text{m}^{-1} \). 97
7.3 MSE and 95% confidence intervals for the (random amplitude) source location estimates. Case (a): \( \sigma_k = 0, \sigma_x = \sigma_y = \sigma_z = 0.01 \text{m}, \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 5l \frac{\pi}{180}, l = 0, 1, 2, 3, 4 \). Case (b): \( \sigma_x = \sigma_y = \sigma_z = 0 = \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 0, \sigma_k = 0, 0.25, 0.5, 0.75, 1 \text{m}^{-1} \). 98
7.4 Random amplitude source location estimates obtained using EM (blue points) and E2M (red crosses), and corresponding 95% confidence ellipsoids on the x-y plane. Black crosses represent the actual source locations. 98
7.5 Random amplitude source location estimates obtained using EM (left) and E2M (right), and corresponding 95% confidence ellipsoids in 3D space. Black crosses represent the source locations. 99
List of Abbreviations and Symbols

Symbols

\[ \mathbb{E} \quad \text{Mathematical expectation} \]
\[ \text{Cov} \quad \text{Covariance or covariance matrix} \]
\[ \mathbf{I}_M \quad M\text{-dimensional identity matrix} \]
\[ x^c \quad \text{conjugate of complex number } x \]
\[ \mathbf{A}^H \quad \text{conjugate transpose of matrix } \mathbf{A} \]
\[ \mathbf{A}^T \quad \text{transpose of matrix } \mathbf{A} \]
\[ \det(\mathbf{A}) \quad \text{Determinant of matrix } \mathbf{A} \]
\[ \text{tr}(\mathbf{A}) \quad \text{Trace of matrix } \mathbf{A} \]
\[ \mathcal{N}(\mu, \Sigma) \quad \text{Gaussian distribution with mean } \mu \text{ and covariance matrix } \Sigma \]
\[ \phi(\cdot|\mu, \Sigma) \quad \text{Probability density function of Gaussian random vector with mean } \mu \text{ and covariance matrix } \Sigma \]
\[ \text{Re}(x) \quad \text{Real part of complex number } x \]
\[ \text{Im}(x) \quad \text{Imaginary part of complex number } x \]
\[ \text{diag}(a_1, \cdots, a_n) \quad \text{Diagonal matrix with diagonal entries } a_1, \cdots, a_n \]

Abbreviations

\[ \text{CRLB} \quad \text{Cramer-Rao lower bound} \]
\[ \text{DoA} \quad \text{direction of arrival} \]
\[ \text{EM} \quad \text{expectation-maximization (algorithm)} \]
\[ \text{E2M} \quad \text{evidential expectation-maximization (algorithm)} \]
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIM</td>
<td>Fisher Information Matrix</td>
</tr>
<tr>
<td>LS</td>
<td>least square</td>
</tr>
<tr>
<td>ML</td>
<td>maximum likelihood</td>
</tr>
<tr>
<td>MLE</td>
<td>maximum likelihood estimation (or estimate, estimator)</td>
</tr>
<tr>
<td>NAH</td>
<td>near-field acoustical holography</td>
</tr>
<tr>
<td>PDF</td>
<td>probability density function</td>
</tr>
<tr>
<td>SONAH</td>
<td>statistically optimized near-field acoustical holography</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

One of the crucial issues in acoustics engineering is to reduce the noise emitted by complex devices. In the automobile industry, for example, due to the requirements expressed by the users (in terms of acoustic comfort) as well as the legislation set by the European community to limit noise pollution, acoustic quality is now accepted as a decisive criterion to judge the performance of a vehicle, as well as its safeness or its fuel consumption. The first step of this task is to understand the acoustic behavior of the device in order to focus the efforts on the main sources and the most annoying frequency bands. However, the noise sources emitted by a machine are multiple and structurally complex. Thus, the need for tools allowing visualization, quantification and identification of noise sources has become significant in acoustics engineering.

In this thesis, we investigate the problem of multiple sound source localization. The sound signals are subdivided into two classes: deterministic and random signals. The former assumes that the amplitude is a constant to be estimated, while in the latter case the amplitude is modeled as a random variable. The task of this work is to estimate the sound source locations and amplitudes for deterministic signals, or the sound source locations and variances of amplitudes for random signals, using sound pressures measured by an array of microphones.

Part I of this thesis studies the sound source localization problem for deterministic signals in a maximum likelihood framework. The sound pressure measured by a microphone is interpreted as a mixture of signals emitted by the sources, and the contributions of this measurement from various sources are assumed as latent variables. In this case, the EM algorithm is used to solve the estimation problem.

However, contrary to the ideal experimental setting, different kinds of uncertainties may pervade the sound propagation and measurement process. For example, sometimes the mi-
microphone array has to be placed on a vibrating object so that the microphone locations are never precisely measured. Moreover, the sound propagation medium can be uncertain as well. For instance, the sound speed, thus the wavenumber, may vary significantly due to changes of temperature. Therefore, in Part II, we take into account these two kinds of uncertainties in the estimation process. The uncertainties are first transposed to the data, via first-order approximations. The resulting uncertain pressures are then represented using contour functions, which are mathematical tools proposed within the framework of belief functions. Then, parameter estimation can be carried out using the Evidential EM (E2M) algorithm, which is a variant of the EM algorithm to perform maximum likelihood estimation from uncertain data.

Finally, in Part III, the source localization problem for random signals is investigated. First, under the assumption of certain measurements, the source locations and variances of random amplitudes are estimated using the EM algorithm, as in Part I. Then, the uncertainties of the microphone locations and the wavenumber are considered. The parameters are estimated using the E2M algorithm as in Part II.

The simulated and real experiments are presented for both deterministic and random signals. In the certain case, the EM algorithm is compared with beamforming and SONAH, which are popular methods for identifying the sound sources. The results show that the EM algorithm can clearly localize the sources while beamforming and SONAH are restricted to the studied frequency. Moreover, when the data are pervaded with uncertainties, the EM and E2M algorithm are compared to each other. The experimental results show that E2M algorithm performs better than the EM in the sense of estimation error.
Part I

Sound Sources Localization From Certain Measurements
Chapter 2

State of Art

2.1 Background and Basic Description of Sound Source Localization Model

In this thesis, we investigate the sound source localization problem in a statistical framework. The sound pressure radiated by point sources is described by the following model in the frequency domain:

\[ p_t = G(r)A + n_r. \]  

(2.1)

Here, \( p_t = (p_{t1}, \cdots, p_{tM})^T \) corresponds to the \( M \)-dimensional vector of pressures measured in the \( t \)-th snapshot \( (t = 1, \cdots, T) \), and \( G \) is an \( M \)-by-\( S \) matrix describing the sound propagation process. The \( (m,s) \) element of this matrix is the propagation operator from the \( m \)-th microphone to the \( s \)-th source: \( G(r_s | r'_m) = \frac{\mu |r_s - r'_m|}{4\pi |r_s - r'_m|} \), where \( r_s \) and \( r'_m \) respectively stand for the sound source and microphone locations, and \( k \) is the wavenumber. The measurement noise \( n_t = (n_{t1}, \cdots, n_{tM})^T \) is assumed to be complex Gaussian\(^1\) distributed with 0-mean and covariance matrix \( \Sigma \). The vector \( A = (A_1, \cdots, A_S)^T \) contains the amplitudes of the sound sources, that will be assumed as deterministic in Chapters 3-5, and modeled by a random variable in Chapters 6-7. In the former case, \( p_t \) is called deterministic amplitude signal, in the latter, it is referred to as random amplitude signal.

\(^1\)Please see Appendix A for complete definition of complex-valued Gaussian random variable.
2.2 Review of Different Source Localization Methods

Beamforming [10, 44, 68], Near-field Acoustical Holography (NAH) [52, 71], and Statistically Optimized Near-field Acoustical Holography (SONAH) [35, 39, 64] are the most widely used sound source localization methods. These approaches are based on pressure measurements obtained from an array of microphones. Beamforming estimates the Direction of Arrival (DoA) of the plane waves in the far-field case, or the point source location in the near-field case, by maximization of the Delay-and-Sum beamformer. However, beamforming is still restricted in frequency range due to its limited dynamic range and minimum resolvable source separation. On the other hand, NAH is used to retro-propagate the sound pressure over a surface near the sound sources and ensures a higher spatial resolution by also taking into account the evanescent waves. SONAH is similar to NAH. The pressure field is indeed expanded over the same plane wave basis in the two methods, but the analytical formalism of SONAH avoids the errors caused by the use of discrete spatial Fourier transform in NAH. However, NAH and SONAH both reconstruct the sound field on a specific surface, so that non-parallel point sources cannot be localized in a 3D space. In a recent paper [3], Antoni presented a unified approach of these different acoustic imaging methods. Based on a Bayesian framework, a super-resolution of the reconstructed pressure field is made possible by taking into account a prior information over the source distribution.
2.2 Review of Different Source Localization Methods

2.2.1 Beamforming

The conventional (or Bartlett) beamforming is based on the assumption that the signals are emitted by a single source. In this section, only the point source case is introduced—corresponding results for plane wave case can be found in [44].

The Delay-and-Sum beamformer applied to the measured pressures $\mathbf{p}_t$ is defined as

$$B_t = \mathbf{w}^H \mathbf{p}_t = \sum_{m=1}^{M} w_m^* p_{mt}$$

and the beamformer power output is

$$|B_t|^2 = \mathbf{w}^H \mathbf{p}_t \mathbf{p}_t^H \mathbf{w}. \quad (2.2)$$

The idea behind each beamformer is to steer the measured pressures to find the source position. The steering vector $\mathbf{w}$ depends on the estimated source location $\mathbf{r}$, and its expression depends on the kind of sources considered (plane waves or point sources). The estimated position of the source is then given by the maximum of the chosen beamformer. Note that this method considers one source at a time; thus a secondary source is related to a secondary local maximum of the beamformer. The choice of the steering vector $\mathbf{w}$ and the method to find the source position is now detailed, for a pressure model (Eq. (2.1)) with a single source ($S = 1$). In the far-field case, $\mathbf{p}_t$ can be approximated by

$$\mathbf{p}_t \approx \left( e^{j|\mathbf{r} - \mathbf{r}_1|}, \ldots, e^{j|\mathbf{r} - \mathbf{r}_M|} \right)^T \mathbf{A} + \mathbf{n}_t = A e^{j|\mathbf{r}|} \mathbf{g}(\mathbf{r}) + \mathbf{n}_t, \quad (2.4)$$

with the following operator $\mathbf{g}$

$$\mathbf{g}(\mathbf{r}) = \left( e^{-j|\mathbf{r} - \mathbf{r}_1|}, \ldots, e^{-j|\mathbf{r} - \mathbf{r}_M|} \right)^T. \quad (2.5)$$

As shown in Figure 2.1, the operator $\mathbf{g}(\mathbf{r})$ represents the phase delay between the incoming wave at each microphone and the phase reference at the origin.

By assuming that the noise covariance is diagonal: $\mathbb{E}(\mathbf{n}_t^T \mathbf{n}_t) = \sigma^2 \mathbf{I}_M$, the steering vector is then given by

$$\max_{\mathbf{w}} \mathbb{E} \left( \mathbf{w}^H \mathbf{p}_t \mathbf{p}_t^H \mathbf{w} \right) = \max_{\mathbf{w}} \left( \mathbb{E} \left( |\mathbf{A}|^2 |\mathbf{w}^H \mathbf{g}|^2 + \sigma^2 |\mathbf{w}|^2 \right) \right). \quad (2.6)$$

To obtain a non-trivial solution, the steering vector $\mathbf{w}$ is constrained to $|\mathbf{w}| = 1$. The steering
vector maximizing the beamformer power output, denoted as $w_{BF}$, can then be obtained by:

$$w_{BF} = \frac{g}{\sqrt{g^Hg}}. \quad (2.7)$$

Substituting Eq. (2.7) back into Eq. (2.3), the source location estimated by the beamformer power output is

$$\hat{r} = \arg \max_r \frac{p_r^Hgg^Hp_t}{g^Hg} = \arg \max_r p_r^Hgg^Hp_t. \quad (2.8)$$

On the other hand, the Delay-and-Sum beamformer

$$B_t = w_{BF}^H p_t = \frac{1}{M} \sum_{m=1}^{M} e^{jk(|r| - |r_m|)} p_{mt} \quad (2.9)$$

can be interpreted as an operator that “steers” with a specific phase delay applied at each microphone before the summation. When this delay matches the real delay related to the time of arrival of the wave at each microphone, then the Delay-and-Sum reaches its maximum.

### 2.2.2 NAH

The NAH method is used to back-propagate the sound pressure over a surface near the sound sources and ensures a higher spatial resolution by also taking into account the evanescent field. At a given frequency, the sound pressure field on the microphone array plane, denoted as $p_a(x,y)$, is retro-propagated to the source parallel plane. It includes three steps:

- The transform of the spatial pressure field in the wavenumber domain by a 2D Spatial Fourier Transform:

$$q_a(k_x, k_y) = \frac{1}{(2\pi)^2} \int \int p_a(x,y)e^{-j(k_x x + k_y y)} \, dx \, dy. \quad (2.10)$$

Here $q_a$ denote the spatial pressure in the wavenumber domain in the array plane.

- The back-propagation of the sound pressure in the wavenumber domain to the new defined plane:

$$q_d(k_x, k_y) = q_a(k_x, k_y)e^{jk_z(z_d - z_a)}, \quad (2.11)$$

\(^2\)Evanescent wave: a near-field wave with an intensity that exhibits exponential decay without absorption as a function of the distance from the boundary at which the wave was formed. Mathematically, it is a wave vector where one or more of the vector’s components has an imaginary value.
where \( z_d \) and \( z_a \) are the \( z \)-coordinates of the new defined plane and the microphone plane, and \( q_d \) denotes the spatial pressure in the wavenumber domain in the new defined plane. In order to optimize the spatial resolution, evanescent waves \((k^2 < k_x^2 + k_y^2)\) have to be included.

- The transform of the sound pressure back to the spatial domain by an Inverse 2D Spatial Fourier Transform:

\[
p_d(x,y) = \int \int q_d(k_x,k_y)e^{i(k_x x + k_y y)} \, dk_x dk_y. \quad (2.12)
\]

Would the whole microphone plane be exactly measurable, the sound field could be perfectly reconstructed using NAH. However, only a finite number of points on the microphone plane is measured, therefore with an unavoidable noise. Therefore, Eqs. (2.10) and (2.12) have to be replaced by the Discrete Fourier Transform (DFT) over a finite domain and caution needs to be taken to avoid the retropropagation of background noise.

### 2.2.3 SONAH

In NAH, the use of a spatial DFT and a retropropagation operator in the wavenumber domain is computationally efficient. However, the discretization also causes some errors [6, 64]. SONAH performs the plane to plane transformation directly in the spatial domain instead of in the wavenumber domain.

The sound pressure at any position in the source plane \( \mathbf{r} \) can be predicted as a weighted sum of \( M \) measurements at position \( \mathbf{r}'_m \) in the microphone array plane

\[
p(\mathbf{r}) = \sum_{m=1}^{M} c_m(\mathbf{r}) p(\mathbf{r}'_m) = \mathbf{p}^T \mathbf{c}(\mathbf{r}). \quad (2.13)
\]

The transfer vector \( \mathbf{c}(\mathbf{r}) \) does not depend on the measurements, but only on the position \( \mathbf{r} \), and is found as the Least Square (LS) solution of the linear equations

\[
\mathbf{B} \mathbf{c}(\mathbf{r}) = \alpha(\mathbf{r}), \quad (2.14)
\]

where \( \mathbf{B} \) and \( \alpha(\mathbf{r}) \) are defined by the elementary wave functions \( \Psi_n(\mathbf{r}) \) such that

\[
\mathbf{B} = \begin{pmatrix}
\Psi_1(\mathbf{r}_1) & \cdots & \Psi_1(\mathbf{r}_M) \\
\vdots & \ddots & \vdots \\
\Psi_N(\mathbf{r}_1) & \cdots & \Psi_N(\mathbf{r}_M)
\end{pmatrix}, \quad \alpha = (\mathbf{r}) \begin{pmatrix}
\Psi_1(\mathbf{r}) \\
\vdots \\
\Psi_N(\mathbf{r})
\end{pmatrix}. \quad (2.15)
\]
As in the classical NAH, the pressure field is expanded over a plane wave basis with the following elementary functions

\[ \Psi_n(r) = e^{-j(k_x x + k_y y + k_z z)}, n = 1, \ldots, N, N \to \infty, \]  

where

\[ k_{z,n} = \begin{cases} \sqrt{k^2 - k_{x,n}^2 - k_{y,n}^2}, & \text{if } \sqrt{k_{x,n}^2 + k_{y,n}^2} \leq k \\ -j \sqrt{k_{x,n}^2 + k_{y,n}^2 - k^2}, & \text{otherwise} \end{cases}, \]  

The LS solution of (2.14) can be expressed by

\[ c(r) = (B^H B + \varepsilon I_M)^{-1} B^H \alpha(r), \]  

where \( \varepsilon \) is the Tikhonov regularization parameter and \( I_M \) is the \( M \)-dimensional identity matrix. From Eq. (2.15), it follows that the elements of \( B^H B \) and \( B^H \alpha(r) \) are given by

\[ [B^H B]_{ij} = \sum_n \Psi_n^*(r_i) \Psi_n^*(r_j), \quad [B^H \alpha(r)]_i = \sum_n \Psi_n^*(r_i) \Psi_n^*(r). \]  

By letting the sampling spacing in the \( (k_x, k_y) \) plane go to zero \( (N \to \infty) \), the summations in Eq. (2.19) become integrals over the \( (k_x, k_y) \) plane [35]. This limit method can avoid the wrap-around problem caused by the discrete representation in the wavenumber domain.

### 2.3 EM Algorithm

In this thesis, we solve the sound source localization problem in a parametric framework: the model is estimated through the sound source locations \( \mathbf{r} \) and amplitudes \( \mathbf{A} \) in Eq. (2.1). More specifically, we propose a maximum likelihood approach for this source estimation problem. In the case of a single source, the estimates of these parameters are obtained by maximizing the likelihood function of the measured pressure \( \mathbf{p} \). For multiple sources, however, computing the MLE of the parameters becomes difficult due to the number of parameters to be estimated and the form of the likelihood function. To overcome this problem, an iterative method, which is called the Expectation-Maximization (EM) algorithm, is proposed to find the maximum likelihood estimates (MLE) of the parameters. The pressure measured by a microphone is interpreted as a mixture of signals emitted by the various sources. Then, the MLE of the various sources may be obtained using the EM algorithm. In this section, we first briefly introduce the EM algorithm, and then we review some existing
works on parameter estimation of superimposed signals based on this procedure.

2.3.1 Introduction of the EM Algorithm

The EM algorithm [19] is an iterative method allowing to deal with incomplete (or missing) data: such data correspond to partial or imprecise information. Incomplete data can occur because of nonresponse: no information is provided for several items or a whole unit. For example, researchers test the service life of a group of products, such as light bulbs. Since the testing time is very long, they can not stay in the laboratory at any time. Alternatively, they may have periodic checks (e.g., every two or three hours) for the number of failure products. Thus, the precise service life of bulbs is missed. Instead, only incomplete data in the form of intervals representing imprecise observation are obtained. Sometimes incomplete data are caused by researchers themselves, when data collection is done improperly or inapplicable to the researchers’ need. For example, in the multiple sources localization problem, the measured pressure \( p_i \) is a precise observation but does not offer enough information for the parameter estimation. Therefore, we see it as an incomplete data and define the contributions of the various sound sources to these measured pressures as a complete data.

Let \( x \) denote the incomplete or observed data with PDF \( f_X(x|\Phi) \) and \( y \) stand for the complete (unknown) data with PDF \( f_Y(y|\Phi) \), where \( \Phi \) is the parameter to estimate. The ML approach consists in estimating the parameter \( \Phi \) by maximizing the log-likelihood function of the observed data, or observed log-likelihood:

\[
L(\Phi|x) = \log f_X(x|\Phi),
\]

over all the possible values of parameter \( \Phi \). The conditional probability of the incomplete data variable \( y \) given \( x \) is

\[
f(y|x, \Phi) = \frac{f_Y(y|\Phi)}{f_X(x|\Phi)} = \frac{f_Y(y|\Phi)}{\int_y f_Y(y|\Phi)dy}.
\]

Here \( y(x) = \{y : x(y) = x\} \) stands for all the possible values of the complete data \( y \) that may correspond to the imprecise observation (incomplete data) \( x \).

For the purpose of maximizing Eq. (2.20), the EM algorithm proceeds with the complete log-likelihood by iterating back and forth between two steps. In the \( l \)-th iteration:
• the **E-step** consists in computing the conditional expectation

\[ Q(\Phi | \Phi') = \mathbb{E}(\log f_Y(y|\Phi)|x, \Phi'), \tag{2.22} \]

knowing the parameter \( \Phi' \) estimated at the previous iteration;

• the **M-step** consists in determining \( \Phi^{l+1} \) by maximizing \( Q(\Phi|\Phi') \):

\[ \Phi^{l+1} = \arg\max_{\Phi} Q(\Phi|\Phi'). \tag{2.23} \]

Let \( \Phi' \) denote the parameter estimated at \( l \)-th step and take log and conditional expectation \( E(\bullet|x, \Phi') \) at Eq. (2.21), we obtain

\[ \mathbb{E}(\log f(y|x, \Phi)|x, \Phi') = \mathbb{E}(\log f_Y(y|\Phi)|x, \Phi') - L(\Phi|x). \tag{2.24} \]

Let \( H(\Phi|\Phi') = \mathbb{E}(\log f(y|x, \Phi)|x, \Phi') \), then

\[ L(\Phi|x) = Q(\Phi|\Phi') - H(\Phi|\Phi'). \tag{2.25} \]

This algorithm guarantees that the likelihood increases at each iteration. Indeed, \( L(\Phi^{l+1}) \geq L(\Phi') \) because

• in the M-step \( Q(\Phi^{l+1}|\Phi') \geq Q(\Phi'|\Phi') \), and

• by Jensen’s inequality,

\[ H(\Phi^{l+1}|\Phi') - H(\Phi'|\Phi') = \mathbb{E} \left( \log \frac{f(y|x, \Phi^{l+1})}{f(y|x, \Phi')} |x, \Phi' \right) \]

\[ \leq \log \mathbb{E} \left( \frac{f(y|x, \Phi^{l+1})}{f(y|x, \Phi')} |x, \Phi' \right) = \log \int f(y|x, \Phi^{l+1}) dy = 0. \tag{2.26} \]

By the Monotone Convergence Theorem, the increasing sequence \( L(\Phi) \), which is bounded above, converges to its limit \( L^* \). Wu [72] studied the convergence properties of \( L(\Phi) \) and concluded that the limit of the observed log-likelihood \( L^* \) is a stationary value of \( L(\Phi) \) if \( Q(\Phi|\Phi') \) is continuous in both \( \Phi \) and \( \Phi' \), which is a very weak condition that should be satisfied in most practical situations. Moreover, Theorem 3 of [72] gives a discriminant for local maxima of \( \Phi \). However, it is typically hard to verify. Generally speaking, the convergence to a stationary point, a local maximum or a global maximum depends on the
choice of the starting points. Therefore, it is well-advised to run the EM algorithm with different starting points that are representative of the parameter space and finally retain the result with highest log-likelihood among the various EM estimates.

2.3.2 Review of Existing Works on Source Estimation Via the EM Algorithm

Feder and Weinstein [26–28] were the first to use the EM algorithm to investigate the source parameter estimation problem in the superimposed signals (or multiple sources signals) model. In particular, a wide range of models involving superimposed signals are generalized in [28]:

\[ x(t) = \sum_{k=1}^{K} s_k(t|\Phi_k) + n(t), \]  

(2.27)

where \( \Phi_k \) is the unknown parameter vector associated with the \( k \)-th signal component \( s_k \) and \( n(t) \) represents the additive noise. In this model, each measured signal \( x(t) \) is a mixture of various signal components or contributions. The purpose is to estimate the parameters of each signal component separately. The authors proposed to introduce latent variables

\[ y_k(t) = s_k(t|\Phi_k) + n_k(t), k = 1, \ldots, K \]  

(2.28)

to represent each (unknown) contribution to the measurement; note that the noise \( n(t) \) has been arbitrarily decomposed into components \( n_k(t) \) as well. Obviously, should these contributions be known, estimating the parameters of the model would be easy. In this case, applying the EM algorithm consists in iterating the following steps. In the E-step, the expected source contributions given a current model estimate \( \Phi' \) are estimated:

\[ Q(\Phi|\Phi') = E \left( y_k(t)|x(t), \Phi' \right), k = 1, \ldots, K \]  

(2.29)

for deterministic signals or

\[ Q(\Phi|\Phi') = E \left( y_k(t)y_k^H(t)|x(t), \Phi' \right), k = 1, \ldots, K \]  

(2.30)

for random signals. Then, the M-step amounts to updating the model parameters according to the source contributions newly computed. Feder and Weinstein [28] also applied this algorithm to the multi-path time delay problem and multiple source location estimation problem. We have to note that in the latter, plane waves with specific angles of arrival were
Some researchers have investigated different kinds of superimposed signals using the EM algorithm. Cirpan and Cekli [11, 12] and Kabaoglu et al. [42, 43] used a similar EM approach to study the localization of near-field sources. Each of these four contributions addresses a specific case, namely: deterministic signals in 2D space and 3D space, and Gaussian random signals in 2D and 3D space. However, note that the near-field was just taken into account with the Fresnel approximation of the source-microphone distance in the time delay (without the $1/r$ spatial attenuation in Eq. (2.1)), and that even in the deterministic case [11, 42] the signal strength depends on the snapshots (unlike in Eq. (2.1), where the strength $A$ is a constant parameter). Lu et al. [46] and Yan et al. [73] studied the source localization of wideband signals, in which the basic assumption of the propagation operator is more or less same as in [11, 12, 42, 43]. Moreover, [46] solved the problem with a diagonal but non-identity noise covariance matrix, and [73] analyzed the computational complexity of the algorithm. Sheng and Hu [61] and Meng and Xiao [47] investigated the multiple source localization problem via acoustic energy measurement using the EM algorithm. In these two papers, only the strength and the attenuation of the sound ($A$ and $1/r$) are considered, but without regard to the phase (the term $e^{jkr}$ in Eq. (2.1)). Frenkel and Feder [31] applied the EM algorithm to the multiple target tracking problem. At each time interval, the sonar sends a signal that is reflected from the moving targets and received by the sensors. The purpose of the algorithm is to estimate the locations and velocities of the multiple moving targets.

In the next chapter, we propose a solution of the multiple sound sources localization problem presented in Section 1.1, using the EM algorithm. The methodology is very similar to that described in Ref. [28], but the assumptions underlying our model differ from all the works mentioned above.
Chapter 3

Sound Sources Localization Under the Deterministic Amplitude Assumption

3.1 Model

In this chapter, we consider the sound source localization problem (2.1) under the assumption that the amplitude $A$ is deterministic.

The traditional MLEs of $r$ and $A$ are obtained by maximizing the log-likelihood

$$\log L(r, A|p) = -\sum_{t=1}^{T} |p_t - GA|^2.$$ (3.1)

When multiple sources are in presence ($S \geq 2$), computing ML estimates necessitates to solve a difficult optimization problem: indeed, $4S$ parameters need to be estimated (for each source, the 3D coordinates $r_s$ and the source amplitude $A_s$). However, by using the EM algorithm, this optimization problem can be largely simplified.

For this purpose, the key is to interpret the signal (2.1) as a mixture of several contributions, as suggested by Feder and Weinstein [28]. Fig. 3.1 shows a 2D example of the problem. In this example, each measurement made by one of the five microphones includes two components, each one representing the contribution from one of the two sources. Let us rewrite Eq. (2.1) in the form of superimposed signals:

$$p_t = G(r)A + n_t = \sum_{s=1}^{S} G_s(r_s)A_s + n_t.$$ (3.2)

In this equation, $G_s A_s$ stands for the contribution from the $s$-th source. The vector $G_s$, which is a function of $r_s$, is the $s$-th column of the matrix $G$, i.e., $(G(r_s|r_1'), \cdots, G(r_s|r_M'))^T$. In this
chapter, the strengths $A_s, s = 1, \cdots, S$ are deterministic. The noise $n_t$ is assumed to follow an $M$-dimensional Gaussian distribution with 0-mean and covariance matrix $\Sigma = \sigma^2 I_M$.

### 3.2 Model Estimation Via the EM Algorithm

For each snapshot, the latent variables are assumed to be the contributions of the sources to the measurement $p_t$:

$$c_{st} = G_s A_s + n_{st}, \ s = 1, \cdots, S,$$

where the noise vector $n_{st}$ is obtained by arbitrarily decomposing the total noise $n_t$ into $S$ components, i.e., $\sum_{s=1}^{S} n_{st} = n_t$. Thus, $p_t$ (the observed incomplete pressure) is related to the $c_{st}$ (the latent contributions of the sources) by

$$p_t = \sum_{s=1}^{S} c_{st}.$$  

Let us further assume that $n_{st}, s = 1, \cdots, S$ are mutually independent Gaussian random variables with 0-mean and covariance matrix $\Sigma_s = \frac{1}{3} \Sigma$. Then, the complete data log-likelihood is

$$L(r, A|c) = - \sum_{t=1}^{T} \sum_{s=1}^{S} |c_{st} - G_s A_s|^2.$$
Given the parameters \( r^l, A^l \) estimated in the M-step of the \( l \)-th iteration, the E-step in the \((l + 1)\)-th iteration consists in computing

\[
Q(\Phi|\Phi^l) = \mathbb{E}(L(r, A|c)|p_t, r^l, A^l)
\]

\[
= d_1 - \sum_{t=1}^{T} \sum_{s=1}^{S} \left[ -\mathbb{E}(c_{st}^H|p_t, r^l, A^l)G_sA_s - (G_sA_s)^H\mathbb{E}(c_{st}|p_t, r^l, A^l) + |G_sA_s|^2 \right]
\]

\[
= d_2 - \sum_{t=1}^{T} \sum_{s=1}^{S} \left[ \hat{c}_{st}^l - G_sA_s \right]^2,
\]

where

\[
\hat{c}_{st}^l = \mathbb{E}(c_{st}|p_t, r^l, A^l)
\]

and \( d_1 \) and \( d_2 \) are constants independent of \( r \) and \( A \). Computing \( Q(\Phi|\Phi^l) \) thus amounts to determining the expected source contributions (3.7). For this purpose, let us remind the following theorem [56].

**Theorem 1.** Let \( X \) and \( Y \) be \( n \)-dimensional Gaussian random vectors with expectation \( \mu_X \) and \( \mu_Y \), and with covariance matrix \( \Sigma_{XX} \) and \( \Sigma_{YY} \). Let \( \Sigma_{XY} = \text{Cov}(X, Y) \) and \( \Sigma_{YX} = \text{Cov}(Y, X) \), the conditional probability density function is Gaussian:

\[
f_{Y|X}(\cdot|X) \sim \phi \left( \cdot | \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \right).
\]

Here \( (p_t, c_{st})^T \) are jointly Gaussian with expectation \( \left( \sum_{s=1}^{S} G_sA_s, G_sA_s \right)^T \) and covariance matrix \( \frac{1}{S} \Sigma + \frac{1}{S} \Sigma \). Thus, Theorem 1 makes it possible to compute Eq. (3.7)

\[
\hat{c}_{st}^l = \mathbb{E}(c_{st}|p_t, r^l, A^l) = G_s^lA_s^l + \frac{1}{S} \left( p_t - \sum_{s=1}^{S} G_s^lA_s^l \right).
\]

The M-step consists in computing new parameter estimates by maximizing (3.6):

\[
r_{s}^{l+1}, A_{s}^{l+1} = \arg\min_{r_s, A_s} \sum_{t=1}^{T} \left| \hat{c}_{st}^l - G_sA_s \right|^2.
\]

(3.10)

The computation of Eq. (3.10) can be simplified as follows. For any sound source location \( r_s \), the update equation for the source strength \( A_s \) is given by

\[
A_{s}^{l+1} = \arg\min_{A_s} \left| \hat{c}_{st}^l - G_sA_s \right|^2 = \frac{G_s^Hc_{st}^l}{G_s^H G_s}.
\]

(3.11)
in which

\[ \hat{c}_s^l = \frac{1}{T} \sum_{t=1}^{T} \hat{c}_{st}^l. \]  

(3.12)

Then, by substituting Eq. (3.11) back into Eq. (3.10), the source location estimate is given by

\[
\hat{r}_s^{l+1} = \arg \min_{r_s} \left| \hat{c}_s^l - \frac{G_s^H G_s}{G_s^H G_s} \hat{c}_s^l \right|^2
\]

\[
= \arg \min_{r_s} \left( \hat{c}_s^l \right)^H \left( I_M - \frac{G_s G_s^H}{G_s^H G_s} \right)^H \left( I_M - \frac{G_s G_s^H}{G_s^H G_s} \right) \hat{c}_s^l
\]

\[
= \arg \min_{r_s} \left( \hat{c}_s^l \right)^H \left( I_M - \frac{G_s G_s^H}{G_s^H G_s} \right) \hat{c}_s^l
\]

\[
= \arg \max_{r_s} \left( \hat{c}_s^l \right)^H \frac{G_s G_s^H}{G_s^H G_s} \hat{c}_s^l. 
\]

(3.13)

We can see that Eq. (3.13) is a 3-parameter optimization problem. If we assume further that all the sources are on a given plane, it boils down to a 2-parameter optimization problem, which is much easier to solve than maximizing Eq. (3.1).

The strategy for estimating the parameters in the sound source localization model using the EM algorithm is summarized by Algorithm 1:

**Algorithm 1** EM algorithm for sound sources localization in the case of deterministic amplitude

For \( l = 0 \), pick starting values for the parameters \( r^0, A^0 \). For \( l \geq 1 \):

repeat

  estimate the source contribution \( \hat{c}_s^l \) from Eqs. (3.9) and (3.12) for \( s = 1, \ldots, S \);

  estimate the new source location \( \hat{r}_s^{l+1} \), for \( s = 1, \ldots, S \), with Eq. (3.13);

  estimate the new source strength \( \hat{A}_s^{l+1} \) for \( s = 1, \ldots, S \) by substituting the new source location \( \hat{r}_s^{l+1} \) back to Eq. (3.11);

until the relative increase of the observed data log-likelihood (3.1) is less than a given threshold \( \kappa \):

\[
\frac{L(\hat{r}_s^{l+1}, \hat{A}_s^{l+1}|p) - L(\hat{r}_s^l, \hat{A}_s^l|p)}{L(\hat{r}_s^l, \hat{A}_s^l|p)} < \kappa.
\]

(3.14)
3.3 Comparison of Maximum Likelihood and Beamforming

In this section, we compare the ML approach detailed above with beamforming. First, we consider an equivalent form of the beamforming strategy presented in Section 2.2.1: in the beamformer power output (2.3), the pressure model $\mathbf{p}_r = \mathbf{G}(\mathbf{r})\mathbf{A} + \mathbf{n}_r$ is taken without the far-field approximation (2.4). Beamforming is then reformulated as

$$\max_{\mathbf{w}} E \left( \mathbf{w}^H \mathbf{p}_r \mathbf{p}_r^H \mathbf{w} \right) = \max_{\mathbf{w}} \left( A^2 \left| \mathbf{w}^H \mathbf{G} \right|^2 + \left| \mathbf{w} \right|^2 \right).$$

(3.15)

Therefore, the non-trivial solution ($\left| \mathbf{w} \right| = 1$) of $\mathbf{w}$ becomes

$$\mathbf{w}_{BF} = \frac{\mathbf{G}}{\sqrt{\mathbf{G}^H \mathbf{G}}}.$$ 

(3.16)

and the sound source location obtained via beamforming is

$$\hat{\mathbf{r}} = \arg \max_{\mathbf{r}} \frac{\mathbf{p}_r^H \mathbf{G} \mathbf{G}^H \mathbf{p}_r}{\mathbf{G}^H \mathbf{G}}.$$ 

(3.17)

On the other hand, the traditional MLE under the assumption of a Gaussian noise with 0-mean and covariance matrix $\sigma^2 \mathbf{I}_M$ is obtained by solving

$$\hat{\mathbf{r}}, \hat{\mathbf{A}} = \arg \min_{\mathbf{r}, \mathbf{A}} \left| \mathbf{p}_r - \mathbf{G}(\mathbf{r})\mathbf{A} \right|^2,$$

(3.18)

for $\mathbf{r}$ and $\mathbf{A}$, which finally leads to the same location estimates as Eq. (3.17).

For multiple sources, beamforming solves the problem in the same way as the single source case. An $\mathbf{r} - \left| \mathbf{B}_1 \right|^2$ colormap is usually employed to find the sources. The locations corresponding to the $S$ maxima on the colormap are generally retained as the locations of the $S$ sources to estimate. Note, however, that this method is theoretically inappropriate. By contrast, the ML approach is well defined once the number of sources is specified. The update equations of the source amplitudes are

$$\hat{\mathbf{r}} = \arg \max_{\mathbf{r}} \mathbf{p}_r^H \mathbf{G} \left( \mathbf{G}^H \mathbf{G} \right)^{-1} \mathbf{G}^H \mathbf{p}_r,$$

(3.19)

and the strength estimates are

$$\hat{\mathbf{A}} = \left( \hat{\mathbf{G}}^H \hat{\mathbf{G}} \right)^{-1} \hat{\mathbf{G}}^H \mathbf{p}_r,$$

(3.20)

in which $\hat{\mathbf{G}} = \mathbf{G}(\hat{\mathbf{r}})$. However, as presented at the beginning of this chapter, Eq. (3.19) is
a complicated multiple parameters optimization problem and difficult to solve, unless the problem is simplified using latent variables and solved by the EM algorithm as in Section 3.2. The update equation of the sound source location estimate Eq. (3.13) indicates that each iteration of the EM algorithm actually performs a beamforming projection for each source.

### 3.4 Experiments

This section illustrates the proposed method by experiments realized on real data. The experimental setup is shown in Figure 3.2 (a). We set 60 microphones on an array (on the plane $z = 0$, the center of the array being at the origin), the distribution of which is displayed in Figure 3.2 (b). In addition, two sound sources are placed in the half space $\{z : z > 0\}$ at $r_1 = (-0.239m, -0.112m, 0.314m)$ and $r_2 = (0.172m, -0.012m, 0.314m)$. In this experiment, multi-sine signals with a wide frequency range (100 - 6000 Hz) are played during 60 seconds. The signal is divided into 60 segments and then transformed in the frequency domain using a Discrete Fourier Transform, so that 60 snapshots are obtained in the frequency domain.

Note that in this experiment, the loudspeakers are not perfect point sources. Therefore, estimating a model with two point sources may not yield accurate estimates of the pressures emitted by the two vibrating membranes. A higher number of sources could indeed be more efficient to model these radiating surfaces, for instance by layers of monopole or dipoles.
Here we thus compare three different models, with $S = 2$, $S = 6$ and $S = 10$ point sources, respectively. For each model, in order to overcome the problem of local maxima mentioned in Section 2.3.1, the procedure is run 100 times with initial parameter values selected at random. The initial location of each source is generated at random according to a 2D uniform distribution having a 60 cm $\times$ 60 cm support centered on the loudspeaker. Half of the sources are related to the first loudspeaker and another half to the second. The initial source amplitudes are taken randomly in a range of $\pm 3\text{dB}$ around $(A_1^*, A_2^*)^T$. Here $(A_1^*, A_2^*)^T$ is the LS solution of $p_t = G(r_1, r_2)A$ in dB, in which $r_1$ and $r_2$ denote the actual locations of the two sources. Half of the initial source strengths are obtained from a complex-valued uniform distribution with support in the range of $A_1^* \pm 3\text{dB}$, and another half are from another uniform distribution with support $A_2^* \pm 3\text{dB}$. Next, we compare the EM results with beamforming and SONAH with Tikhonov regularization (using the optimal Tikhonov parameter, cf. [3]). Figures 3.3, 3.4 and 3.5 display the estimation results at different frequencies. In each figure, (a) shows the 100 EM estimates obtained on the given $z$-plane (in all three methods, the $z$-coordinates of the sources are assumed to be known in advance) using blue points, and the retained estimate with highest likelihood using red crosses in the model of $S = 2$. In Figure 3.3 (a), at low frequency (lower than 1000 Hz), the EM estimates with different initial values are almost all identical. By contrast, Figures 3.4 (a) and 3.5 (a) show that at relatively high frequencies (greater than 1000 Hz), different initial values lead to different EM estimates. However, the strategy consisting in retaining the estimate with highest likelihood is able to pick up an accurate source location estimate. Subfigures (b)-(d) display the sound field on the sources plane obtained via the EM estimates under the assumption of $S = 2$, $S = 6$ and $S = 10$ respectively. In the latter two cases, although the assumed number of sources is more than the actual source number $S = 2$, the EM sound field clearly indicates the two sound sources. Moreover, Subfigures (e) and (f) show the similar sound field obtained via beamforming and SONAH. In the present case, both can separate the two sources at low frequencies but do not work at high frequencies, while the EM results are robust at all frequencies.

Given the EM estimates $\hat{\mathbf{r}}_s$ and $\hat{A}_s$ for $s = 1, \ldots, S$, we can reconstruct the sound pressure at any position $\mathbf{r}$ on the microphone plane by

$$p(\mathbf{r}) = \sum_{s=1}^{S} G(\hat{\mathbf{r}}_s | \mathbf{r}) \hat{A}_s. \tag{3.21}$$

Figures 3.6, 3.7 and 3.8 compare the measured sound field on the microphone plane with the reconstructed sound field obtained with the EM estimates at three different frequencies.
We may observe that the quality of the reconstruction increases with the number of sources assumed in the model.

3.5 Conclusion

In this chapter, we proposed to use the maximum likelihood approach to solve the sound source localization problem for deterministic signals. In the single source case, the MLE is identical to that obtained via beamforming. In the case of multiple sources, beamforming is however not valid anymore. The MLE becomes difficult to compute due to the complexity of the optimization problem. Then, the iterative EM algorithm makes it possible to simplify the model estimation problem. In this chapter, we also performed a two-source experiment to validate the proposed method. We made a comparison between our own approach, beamforming and SONAH, which clearly showed the advantage of the EM method.
3.5 Conclusion

Fig. 3.3 Comparison between EM, beamforming and SONAH at f=525Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
Fig. 3.4 Comparison between EM, beamforming and SONAH at f=1325Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
3.5 Conclusion

Fig. 3.5 Comparison between EM, beamforming and SONAH at f=5025Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
Fig. 3.6 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f = 525$ Hz.
Fig. 3.7 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f = 1325\text{Hz}$. 
Fig. 3.8 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f=5025$Hz.
Part II

Sound Source Localization in Presence of Uncertainty
Chapter 4

State of Art: Uncertainties in Sound Source Localization Model

4.1 Uncertainties in the Model

Part I of this thesis addresses the issue of solving the multiple sound source localization problem in an ideal experimental environment. However, different kinds of uncertainties may pervade the sound propagation and measurement process. In order to avoid confusion, we consider randomness and uncertainty to be of different natures: the former stands for irreducible random errors (e.g., measurement noise), while the latter represents systematic errors due to incomplete knowledge of the experimental environment or the measurement process. For example, repeated measurements of the period of a pendulum using a stopwatch may yield different results because it is difficult to start and stop the stopwatch exactly at the same point in the pendulum's swing. This measurement error is generally considered as a random error and described by a 0-mean random variable. However, if the watch is slow due to mechanical flaws, this will lead to a systematic error. This kind of uncertainty has to be quantified and distinguished from random errors. Uncertainty is frequently left aside, being identified as one of the causes for randomness. Obviously, this strategy may not be reasonable. The conventional approach, which consists in using a large number of repeated measurements to decrease the random error, may still lead to inaccurate results in presence of uncertainty. Our claim is that uncertainty should be quantified and taken into account in the estimation process, rather than left aside.

Two kinds uncertainties have been known for a while in the sources localization problem: imprecise microphones (or sensors) locations [2, 40] and uncertain sound propagation environment [23, 29, 30, 36, 62, 65, 66]. In Section 4.1.1, these works are introduced to
illustrate the importance of these two sources of uncertainty.

### 4.1.1 Microphone Location Uncertainty

In a real experimental setting, one may argue that microphone locations are never totally certain, being obtained either by hand measurement or automatically by using auto-calibration schemes. Sometimes the microphone array has to be placed on a vibrating object so that the microphone locations are constantly changed: this is the case, for instance, when measuring the sound radiating from a powerful engine.

Ajdler et al. [2] discussed the impact of uncertain sensor position in the single sound source localization problem. In their work, they used the MLE method to localize the single sound source and assessed the Cramer-Rao Lower Bound (CRLB) of the estimators. In their simulation experiment, white Gaussian noise with different standard deviations was added to the exact microphone positions before the parameter estimation process. Without surprise, as the standard deviation increases, the variance of the estimator moves away from the CRLB. For higher errors on the microphone positions, the algorithm does not converge anymore. Luckily, the experiments show that placing the microphones further away from each other and using more microphones result in an increase of the estimation precision.

Jacoby et al. [40] investigated the sensor location uncertainty in the single source localization problem in shallow water environment. The signal model is

$$
p = A T(\theta_0, x, z) q(r_0, z_0) + n,
$$

where $p$ is the $M$-dimensional measurement, $\theta_0$, $r_0$, $z_0$ and $A$ are respectively the bearing, range, depth and amplitude of the sound source, $n$ is an additive Gaussian noise with 0-mean and covariance matrix $\sigma^2 I_M$, $x = (x_1, \cdots, x_M)^T$ and $z = (z_1, \cdots, z_M)^T$ are 2D coordinate vectors of the sensors; and where the $M \times N$ matrix $T(\theta_0, x, z)$ and the $N$-dimensional vector $q(r_0, z_0)$ are given by the general terms

$$
T_{mn}(\theta_0, x, z) = \phi_n(z_m) e^{-j k_n x_m \sin \theta_0},
$$

and

$$
q_n(r_0, z_0) = \phi_n(z_0) e^{j k_n r_0 \sqrt{k_n r_0}}.
$$

Here $\phi_n, n = 1, \cdots, N$ are modal depth eigenfunctions and $k_n, n = 1, \cdots, N$ are the corresponding horizontal wavenumbers. Let $\alpha = (\theta_0, r_0, z_0, A)$ denote the four source parameters,
the sensor location parameters and $\eta = [\alpha^T, \beta^T]^T$ the overall parameter vector. Then, [40] used the CRLB to analyze the inherent limitations associated with the joint source and sensor location estimation problem. The CRLB for the overall parameter vector $\eta$ is expressed by

$$CRLB^{-1} = J = \frac{2}{\sigma^2} \text{Re} \left[ \frac{\partial (b^T T q)^H}{\partial \eta} \frac{\partial b^T T q}{\partial \eta} \right],$$

where the Fisher Information Matrix (FIM) $J$ can be written in the partitioned form

$$J = \begin{pmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{pmatrix},$$

with $J_{\alpha\alpha}$ and $J_{\beta\beta}$ being the FIMs for the source and sensor parameter vectors, respectively. In the case of known sensor locations, the CRLB for the source parameter $\alpha$ is obtained by

$$CRLB_{\alpha}^{-1} = J_{\alpha\alpha}. \quad (4.6)$$

When both the source and sensor location parameters are unknown, the inverse CRLB for the source parameters becomes

$$CRLB_{\alpha}^{-1} = J_{\alpha\alpha} - J_{\alpha\beta} J_{\beta\beta}^{-1} J_{\beta\alpha} = J_{\alpha\alpha} - \Delta J_{\alpha\alpha},$$

where $\Delta J_{\alpha\alpha}$ may be seen as a loss in Fisher information brought to the source parameters due to a lack of knowledge of the sensor locations. This loss of information increases the CRLB of the source location estimation as well. When there is no prior information on the sensor locations, the FIM of the overall parameter estimator $J$ proves to be singular. Then, three kinds of partial prior information are considered in the paper: i) prior knowledge on the vertical sensor location, ii) on the horizontal sensor location and iii) on the distance between adjacent sensors. The simulation results demonstrate that when the sensor location uncertainty exists (missing partial sensor location information), the CRLB is higher than when there is no uncertainty, but the loss is still tolerable.

### 4.1.2 Medium Uncertainty

The uncertainty of the medium in which the sound propagates may also affect the source location estimation. The wavenumber $k = \frac{2\pi f}{c}$ in Eq. (2.1) depends on the sound speed $c$, which may vary significantly due to changes of the temperature in the medium. For example, when localizing the sound sources of a working engine, the heat-radiation cannot
be neglected although it is sometimes difficult to quantify. In the ocean environment, due to limited knowledge of the physical properties of the medium, sound source localization is often limited as well.

Hamson and Heitmeyer [36] investigated the environmental effects on the source localization in the shallow water model (4.1). Let \( \mathbf{p} \) and \( \mathbf{p}_r \) be the received pressure field and replica field, \( (r_0, z_0) \) and \( (r'_0, z'_0) \) be the real and hypothetical source positions varying over the required surface. Then they defined an “ambiguity function” by

\[
A(r_0, z_0, r'_0, z'_0) = \frac{\left| \sum_{m=1}^{M} \mathbf{p}(r_0, z_0, z_m) \mathbf{p}_r(r'_0, z'_0, z_m) H \right|^2}{\sum_{m=1}^{M} |\mathbf{p}(r_0, z_0, z_m)|^2 \sum_{m=1}^{M} |\mathbf{p}_r(r'_0, z'_0, z_m)|^2},
\]

such that \( A = 1 \) for the correctly located source position at \( r'_0 = r_0, z'_0 = z_0 \). In the case of no uncertainty, the \( r'_0, z'_0 \) ambiguity function clearly indicates a peak located at the actual source position \( (r_0, z_0) \). In [36], several environmental parameters were then considered: sound speed, water depth and water bottom (the latter two are reflected by the sound speed, density and attenuation in the model). When these uncertainties are introduced in the simulation experiments, some sidelobes appear in the \( r'_0, z'_0 \) ambiguity function, which may lead to a wrong source location estimation.

Tabrikian and Krolik [65] studied the effect of uncertain environmental parameters in the shallow-water model (4.1) as well. In this paper, the sound speed parameter is assumed to be uniformly distributed in a given range of values. The surface and bottom sound speeds are \( 1500 \pm 2.5 m/s \) and \( 1480 \pm 2.5 m/s \) respectively, and the bottom depth is measured as \( 102.5 \pm 2.5 m \). The sub-bottom and lower halfspace sound speeds are less precise: \( 1600 \pm 50 m/s \) and \( 1750 \pm 100 m/s \) respectively. Furthermore, the higher modes \( N \) in the shallow-water model (4.1) are associated with rays that have a higher number of reflections from the boundaries and larger group delays in the medium, so that these modes are affected by medium variations more than the lower-order modes. This means that the available information on the source location is weaker in the higher modes than in the lower modes. In the proposed model, assuming some high-order modes carrying no useful information concerning the source location may lead to a suboptimal but computationally efficient ML solution.

Shorey and Nolte [62] also assumed the uncertainty of the sound speed in the shallow-water model to be uniformly distributed. The various sound speed measurements at hand indicate a wide variation of values in the water column sound speed. In the shallow wa-
ter environment considered in [62] (water depth is 73 m), the sound speed decreases from 
$1522 \pm 2 m/s$ at the surface ($0^+ m$) to $1486 \pm 2 m/s$ at the bottom ($73^- m$), and the bottom 
half sound speed ($73^+ m$) appears to bear a higher level of uncertainty, being $1550 \pm 50 m/s$.

Tabrikian and Krolik [66] went on to investigate the single source problem in shallow water as in [65]. Here, Eq. (4.1) is rewritten as

$$p_m = A \sum_{n=1}^{N} \phi_n(z_m, \Gamma) \phi_n(z_0, \Gamma) \frac{e^{jk_n(\Gamma)(r_n-x_m \sin \theta_0)}}{\sqrt{k_n(\Gamma)r_0}} + n_m, \ m = 1, \ldots, M,$$

(4.9)

where $\phi_n(\cdot, \Gamma)$ and $k_n(\Gamma)$ are the modal eigenfunction and horizontal wavenumber depending on the uncertain environmental parameter vector $\Gamma$ including the bathymetry, geo-acoustic properties of the bottom, and the sound speed. However, the effect of errors in the modal horizontal wavenumber $k_n(\Gamma)$ is typically more significant than the errors in the modal depth eigenfunctions $\phi_n(\cdot, \Gamma)$ since small errors in $k_n(\Gamma)$ can result in large modal phase perturbations. Therefore, the environmental uncertainty is represented by considering the wavenumber $\Gamma$ as a random variable and neglecting the mismatch in the modal depth eigenfunctions. The wavenumber can also be approximated by its sample mean and sample covariance matrix, in which the sample $k(\Gamma_l), l = 1, \ldots, L$ may be obtained by Monte Carlo realizations of the uncertain environment $\Gamma_l$.

Finette [29] proposed to include the environmental (sound speed) uncertainty into the ocean acoustic model. In the far-field, the sound pressure $p(r,z)$ at range $r$ and depth $z$ is approximated by

$$p(r,z) \approx \sqrt{\frac{2}{\pi k_0 r}} e^{j(k_0 r - \frac{\pi}{2})} \psi(r,z),$$

(4.10)

where the deterministic narrow angle parabolic equation $\psi(r,z)$ satisfies the partial differential equation

$$2jk_0 \frac{\partial \psi(r,z)}{\partial r} + \frac{\partial^2 \psi(r,z)}{\partial z^2} + k_0^2 (n^2(r,z) - 1) \psi(r,z) = 0,$$

(4.11)

with index of refraction $n(r,z) = \frac{c_0}{c(r,z)}$ and wavenumber $k_0 = \frac{\omega}{c_0}$. An incomplete knowledge of the sound speed is then considered by expressing

$$c(r,z) \Rightarrow c(r,z; \theta) = \bar{c}(r,z) + \delta c(r,z; \theta),$$

(4.12)

in which $\bar{c}(r,z)$ is a deterministic component and $\delta c(r,z; \theta)$ a stochastic contribution, $\theta$ represents a (random) realization associated with uncertainty. Then (4.11) can be generalized
to the following stochastic differential equation

$$
2j k_0 \frac{\partial \psi(r,z;\theta)}{\partial r} + \frac{\partial^2 \psi(r,z;\theta)}{\partial z^2} + k_0^2 \left( \frac{c_0^2}{c^2(r,z)} - 1 - \frac{2c_0^2 \delta c(r,z;\theta)}{c^3(r,z)} \right) \psi(r,z;\theta) = 0.
$$

(4.13)

A general solution of Eq. (4.13) may be obtained via polynomial chaos expansion of $\psi(r,z;\theta)$ and $\delta c(r,z;\theta)$ (the details are omitted here); then, the sound pressure can be estimated, in presence of uncertainty, by

$$
p(r,z;\theta) \approx \sqrt{\frac{2}{\pi k_0 r}} e^{j(k_0 r - \xi)} \psi(r,z;\theta).
$$

(4.14)

Finette [30] further assumed that the sound speed field $c(r,z;\theta)$ can be modeled by a Gaussian random variable: $\bar{c} + \sigma \xi$, where $\bar{c}$ and $\sigma$ are the mean value and standard deviation of the sound speed, and $\xi$ follows a standard Gaussian distribution. Under this assumption, an analytic solution of (4.13) may be obtained.

### 4.1.3 Sound Source Localization Model with Uncertainties

As presented in Sections 4.1.1 and 4.1.2, the microphone locations and wavenumber are two major sources of uncertainty (sound speed, temperature or other environmental uncertainties may be reflected by the wavenumber). Therefore, in the second part of this thesis, the microphone locations $r_m', m = 1, \cdots, M$ and wavenumber $k$ in the sound source localization model (2.1) are not assumed to be precisely known anymore. In Part II, the microphone locations and wavenumber are considered as meta-parameters, which means parameters that are generally considered to be known and on which the model implicitly depends. We propose to transpose these uncertainties to the pressures measured by the microphones. Then, we use belief functions to represent the resulting uncertain pressures. In Section 4.2, fuzzy sets theory and the theory of belief functions are introduced. Then, some existing methods taking the uncertainty into account in the acoustical model are reviewed in Section 4.3.

### 4.2 Fuzzy Sets, Belief Functions, and the Evidential EM Algorithm

In this section, a short introduction of fuzzy sets is given first. Then, we introduce the theory of belief function and the evidential EM algorithm, which are the main tools used to cope
with uncertain data in this thesis.

4.2.1 Fuzzy Sets

Zadeh [75] introduced fuzzy sets to extend the classical notion of set. In classical set theory, an element either belongs or does not belong to a set. By contrast, a fuzzy set assumes the degree of membership of an element to a set by a value in the real unit interval [0,1]. Let \( \Omega \) be a space of points: a fuzzy number \( \tilde{A} \in \Omega \) is characterized by a membership function \( \mu_{\tilde{A}}(\omega) \), which associates with each point \( \omega \in \Omega \) a real number in the interval [0,1]. The value of \( \mu_{\tilde{A}}(\omega) \) represents the “grade of membership” of \( \omega \) in \( \tilde{A} \). Obviously, if a membership takes values that are either 0 or 1, its corresponding fuzzy set is a classical set. Some important concepts of classical sets, such as union, intersection, complement and convexity, can all be generalized in fuzzy sets [75].

The \( \alpha \)-cut of a fuzzy set \( \tilde{A} \) with membership function \( \mu_{\tilde{A}}(\omega) \) is defined as

\[
\tilde{A}^\alpha = \{ \omega | \mu_{\tilde{A}}(\omega) \geq \alpha \},
\]

for \( \alpha \in [0,1] \). If the membership function \( \mu_{\tilde{A}}(\omega) \) is concave, each \( \alpha \)-cut \( \tilde{A}^\alpha \) is an interval.

4.2.2 Belief Functions

Belief Functions on Discrete Domains

Belief functions were originally proposed by Dempster [16] in order to represent imprecise and uncertain observations using multi-valued functions, also known as random sets or set-valued mappings. This seminal work was then further extended and developed by [60, 63]. Belief functions were successfully applied to statistical inference [17, 21, 22], signal and image processing [45], and multisensor data function [4, 25].

Let \( X \) be a variable taking values in a finite domain \( \Omega \). Partial knowledge about \( X \) may be represented by a mass function \( m : 2^\Omega \rightarrow [0,1] \), where \( 2^\Omega \) stands for the power set of \( \Omega \), such that \( \sum_{A \subseteq \Omega} m(A) = 1 \). Any subset \( A \) of \( \Omega \) such that \( m(A) > 0 \) is called a focal element of \( m \). Then, \( m(A) \) represents the degree of belief that the actual value taken by \( X \) is in \( A \). Two particular cases are of interest:

- when \( m(A) = 1 \) for some \( A \subseteq \Omega \), the mass function \( m \) is said to be categorical (if furthermore \( A = \Omega \), then \( m \) is vacuous);
• if $|A| = 1$ for all $A \subseteq \Omega$ such that $m(A) > 0$, then $m$ is called a Bayesian belief mass; it is then formally equivalent to a probability distribution over $\Omega$.

A mass function $m$ may also be represented by its associated belief and plausibility functions, defined for all $A \subseteq \Omega$ by:

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad Pl(A) = \sum_{B \setminus A \neq \emptyset} m(B).$$

(4.16)

We can interpret $Bel(A)$ as the degree to which the evidence supports $A$, while $Pl(A)$ can be interpreted as an upper bound on the degree of belief that could be assigned to $A$ if further evidence was available. The property $Bel(A) \leq Pl(A)$ always holds, for all $A \subseteq \Omega$. The belief function $Bel(A)$ and plausibility function $Pl(A)$ can thus be seen as lower and upper bounds on the possibility that the hypothesis $A$ could be true according to the evidence expressed by the mass function $m$. Furthermore, if $m$ is Bayesian, then $Bel(A) = Pl(A) = m(A)$ for all $A \subseteq \Omega$. Eventually, the function $pl : \Omega \rightarrow [0, 1]$ such that $pl(\omega) = Pl(\{\omega\})$ is called the contour function [60] associated to $m^\Omega$.

**Dempster’s Rule**

Dempster’s rule of combination makes it possible to pool two pieces of independent evidence. Let $m_1$ and $m_2$ be two mass functions induced by independent items of evidence. They can be combined using Dempster’s rule to form a new mass function:

$$m_1 \oplus m_2(A) = \frac{1}{1 - \kappa} \sum_{B \setminus C = A} m_1(B)m_2(C),$$

(4.17)

for all $A \subseteq \Omega$ and $A \neq 0$, and $m_1 \oplus m_2(\emptyset) = 0$, where

$$\kappa = \sum_{B \setminus C = \emptyset} m_1(B)m_2(C)$$

(4.18)

is the degree of conflict between $m_1$ and $m_2$. If $\kappa = 1$, there is a logical contradiction between two pieces of evidence and they cannot be combined.

**Example 1.** Let $m_1$ and $m_2$ be two categorical mass functions such that $m_1(A_1) = 1$, $m_2(A_2) = 1$ and $A_1 \cap A_2 \neq \emptyset$. Then $m_1 \oplus m_2$ is also categorical such that $m_1 \oplus m_2(A_1 \cap A_2) = 1$ and $m_1 \oplus m_2(B) = 0$ if $B \neq A_1 \cap A_2$. This example matches one’s intuitive expectation: for two pieces of trustworthy evidences, we may narrow the field of evidence by their intersection.
For a Bayesian mass function \( m_1 \), its contour function is the PDF \( f_1 \) defined by \( f_1(\omega) = m_1(\{\omega\}) \) for all \( \omega \in \Omega \). Combining \( m_1 \) with an arbitrary mass function \( m_2 \) with contour function \( pl_2 \) yields a Bayesian mass function \( m_1 \oplus m_2 \) with contour function

\[
f_1 \oplus pl_2(\omega) = \frac{f_1(\omega)pl_2(\omega)}{\sum_{\omega \in \Omega} f_1(\omega)pl_2(\omega)}.
\]

(4.19)

If \( m_2 \) is categorical and such that \( m_2(A) = 1 \), then \( f_1 \oplus pl_2 \) is the probability distribution obtained by conditioning \( f_1 \) with respect to \( A \).

**Cognitive Independence**

Let \( X_n \) be a variable defined on a finite frame of discernment \( \Omega_n \), and let \( m^\Omega \) be a mass function on the product frame \( \Omega = \Omega_1 \times \cdots \times \Omega_N \), expressing evidence on \( X = \{X_n, n = 1, \cdots, N\} \). The marginalization of \( m^\Omega \) on \( \Omega_n \), denoted as \( m^{\Omega_1 \Omega_n} \), is defined as

\[
m^{\Omega_1 \Omega_n}(A) = \sum_{C: \Omega_n \subseteq A} m^\Omega(C),
\]

(4.20)

for all \( A \subseteq \Omega_n \). Here \( C \downarrow \Omega_n \) denotes the projection of \( C \subseteq \Omega \) onto \( \Omega_n \).

Now, let \( m_n^\Omega \) be a mass function defined on \( \Omega_n \) and let \( Pl_n^\Omega \) be its associated plausibility function. The variables \( X_n \) are said to be **cognitively independent** if the following equality holds:

\[
Pl^\Omega(A_1 \times \cdots \times A_N) = \prod_{n=1}^{N} Pl^\Omega_n(A_n).
\]

(4.21)

Here \( Pl^\Omega \) is the plausibility function associated to \( m^\Omega \). This property means that new evidence on one variable does not affect our belief in the other variables.

**Belief Functions on the Real Line**

In the case of a continuous domain, e.g., \( \Omega = \mathbb{R} \), the notion of mass function is replaced by that of mass density function. A mass density is defined as a function \( m \) from the set of closed real intervals to \([0, +\infty)\) such that \( m([u, v]) = g(u, v) \) for all \( u \leq v \), where \( g \) is a two-dimensional probability density function with support in \( \{(u, v) \in \mathbb{R}^2 : u \leq v\} \). The intervals \([u, v]\) such that \( m([u, v]) > 0 \) are called **focal intervals** of \( m \). The contour function \( pl \) corresponding to \( m \) is defined by the integral:

\[
pl(x) = \int_{-\infty}^{x} \int_{x}^{+\infty} g(u, v) dv du.
\]

(4.22)
One important special case of mass density functions are Bayesian mass functions, for which focal intervals are reduced to points. Then, the two-dimensional PDF has the following form: $g(u, v) = f(u)\delta(u - v)$, where $f$ is a univariate PDF and $\delta$ is the Dirac delta function. In this case, the contour function $p_l(x)$ has the same form than the traditional PDF. If we assume further that $f$ is Gaussian, then $p_l(x)$ is called a Gaussian contour function.

As in the discrete case, let $m_1$ be a Bayesian mass density function defined by the univariate PDF $f_1$ and let $m_2$ be an arbitrary mass density function with contour function $p_{l_2}$. Then, the combination of $m_1$ and $m_2$ is Bayesian, and has the following PDF:

$$f_1 \oplus p_{l_2}(x) = \frac{f_1(x)p_{l_2}(x)}{\int_{-\infty}^{\infty} f_1(x)p_{l_2}(x)dx}.$$  

(4.23)

### 4.2.3 Statistical Inference From Uncertain Data

#### Likelihood of Uncertain Data

Let us consider a parameter estimation problem, where the data at hand are uncertain and represented by contour functions. Here, we present how the notion of likelihood may be extended to such uncertain data.

Assume that $y$ is an imprecise observation, but it is known for sure that $y \in A$ for some $A \subseteq \Omega_Y$. The likelihood function given such imprecise data is now:

$$L(\Phi|A) = f_Y(A|\Phi) = \sum_{y \in A} f_Y(y|\Phi).$$  

(4.24)

More generally, our knowledge of $y$ may be not only imprecise, but also uncertain; it can be described by a mass function $m$ on $\Omega_Y$ with focal elements $A_1, \cdots, A_r$ and corresponding masses $m(A_1), \cdots, m(A_r)$. To extend the likelihood function, we may simply compute the weighted sum of the terms $L(\Phi|A_i)$ with coefficients $m(A_i)$, which leads to the following expression:

$$L(\Phi|m) = \sum_{i=1}^{r} m(A_i)L(\Phi|A_i).$$  

(4.25)
Note that the likelihood function of uncertain data (4.25) can be rewritten as

\[
L(\Phi|m) = \sum_{i=1}^r m(A_i) \sum_{y \in A_i} f_Y(y|\Phi) \\
= \sum_{y \in \Omega_Y} f_Y(y|\Phi) \sum_{A_i \ni y} m(A_i) \\
= \sum_{y \in \Omega_Y} f_Y(y|\Phi)p_l(y). \tag{4.26}
\]

The likelihood function \( L(\Phi|m) \) thus only depends on \( m \) through its associated contour function \( p_l \). For this reason, we will write indifferently \( L(\Phi|m) \) or \( L(\Phi|p_l) \). Furthermore, we may observe that \( L(\Phi|m) \) is identical with the expectation of \( p_l(Y) \) given \( \Phi \):

\[
L(\Phi|p_l) = \sum_{y \in \Omega_Y} f_Y(y|\Phi)p_l(y) = \mathbb{E}[p_l(Y)]. \tag{4.27}
\]

The above definitions in the discrete case can be straightforwardly transposed to the continuous case. Let \( Y \) be a continuous random vector taking values in \( \Omega_Y \) with PDF \( f_Y(y|\Phi) \) and let \( p_l : \Omega_Y \to [0,1] \) be the contour function of a continuous mass function \( m \) on \( \Omega_Y \). The likelihood function given \( p_l \) can be defined as:

\[
L(\Phi|p_l) = \mathbb{E}[p_l(Y)] = \int_{\Omega_Y} f_Y(y|\Phi)p_l(y)dy, \tag{4.28}
\]

this integral is assumed to exist and to be nonzero.

**Evidential EM Algorithm**

We address here the problem of estimating the parameters of the model via maximum-likelihood when the uncertain data at hand are represented by contour functions. Obviously, the EM algorithm introduced in Section 2.3.1 cannot be used to maximize the generalized likelihood anymore, since the data are uncertain. An extension of the EM algorithm, known as the Evidential EM (E2M) algorithm, was proposed in [21] to address this case. We briefly remind here the principles of this algorithm; for an extensive presentation, the reader is invited to refer to [21].

The E2M algorithm iterates alternatively between two steps. In the E-step of iteration \( l \), the expectation of the log-likelihood of the complete but uncertain data is computed:

\[
Q(\Phi|\Phi^l) = E \left[ \log L(\Phi|y)|p_l(y), \Phi^l \right]. \tag{4.29}
\]

Note that this expectation is now computed with respect to both the current \( f \) and \( \Phi^l \) of the
parameter vector and the contour function \( pl \) through which the uncertain sample is known. As explained in [21], the PDF used in this expectation is defined by

\[
p_Y(y|pl(y), \Phi^l) = \frac{p_Y(y|\Phi^l)pl(y)}{L(\Phi^l|pl(y))}.
\]

Then, Eq. (4.29) is computed using the definition of the mathematical expectation:

\[
Q(\Phi|\Phi^l) = \int_{\Omega_Y} \log L(\Phi|y)p_Y(y|pl(y), \Phi^l)dy.
\]

The M-step of the E2M algorithm is unchanged and requires the maximization of \( Q(\Phi|\Phi^l) \) with respect to \( \Phi \). As in the EM algorithm, the E2M algorithm alternately repeats the E- and M-steps defined above until the relative increase of the observed-data likelihood becomes smaller than a given threshold.

The evidential EM algorithm inherits the monotonicity property of the EM algorithm, which guarantees the convergence of the likelihood function.

**Theorem 2.** Any sequence \( L(\Phi^l|m) \) for \( l = 0, 1, 2, \cdots \) of likelihood values obtained using the evidential EM algorithm is non-decreasing, i.e., it verifies

\[
L(\Phi^{l+1}|pl) \geq L(\Phi^l|pl),
\]

for all \( l \).

### 4.3 Uncertainty Quantification Methods

Some researchers have investigated taking uncertainties into account in acoustical problems, mainly using propagation of variance [8, 48] or error bars [14], probabilistic method [41] and more particularly Bayesian approaches [5, 23, 57, 62, 77, 78], interval and fuzzy sets analysis [1, 9, 24] and belief functions [34, 58].

For estimating the uncertainty on a variable of interest, the statistical moment technique via Taylor series expansion may be used. Let \( \Theta = (\theta_1, \cdots, \theta_U) \) be the vector of uncertain inputs and \( g(\Theta) \) denote an output function whose uncertainty depends on \( \Theta \); its expectation and variance (quantifying the level of uncertainty) may be approximated by Taylor expansion at \( \Theta \) [7, 48, 50]:

\[
\overline{g(\Theta)} \approx g(\Theta) + \frac{1}{2} \sum_{u=1}^{U} \sum_{v=1}^{U} \frac{\partial^2 g(\Theta)}{\partial \theta_u \partial \theta_v} \sigma^2_{\theta_u \theta_v}.
\]
4.3 Uncertainty Quantification Methods

and

\[ \sigma_2^2 \approx \sum_{u=1}^{U} \sum_{v=1}^{U} \frac{\partial g(\Theta)}{\partial \theta_{u}} \frac{\partial g(\Theta)}{\partial \theta_{v}} \sigma_{\theta_{u} \theta_{v}}, \]  

(4.34)

where \( \sigma_{\theta_{u} \theta_{v}} \) is the covariance of \( \theta_{u} \) and \( \theta_{v} \). Mu et al. [48] investigated parameter uncertainty (for shear moduli, bulk modulus and permeability) in the predicted acoustic response in ocean sediment. Using Eqs. (4.33) and (4.34), the mean and variance of the output sound speed are obtained. Bussow and Petersson [8] studied how energy transformation is affected by the coupling loss factors. In a system composed of \( N \) subsystems with respect energies \( \{E_n\}_{n=1}^{N} \), the subsystem energy vector \( \mathbf{E} = (E_1, \cdots , E_N)^T \) depends on a coupling (energy-flux) matrix \( \mathbf{C} \) by \( \mathbf{E} = \mathbf{C}^{-1} \mathbf{P} \), where \( \mathbf{P} \) and \( \omega \) stand for power input and angular frequency respectively. However, the coupling matrix \( \mathbf{C} \) is uncertain due to some input quantities \( \theta_{u}, \ u = 1, \cdots, U \) (for example, an uncertain coupling length between two subsystems). The uncertainties in the input quantities lead to uncertainties in the resulting energies, which may be characterized by variances. Finally, the energy variance \( \sigma_{E_n}^2 \) and normalized standard deviation \( \frac{\sigma_{E_n}}{E_n} \) may be estimated by Eq. (4.34).

Congedo et al. [14] used error bars to model the uncertainties in the problem of dense gas flow. Assume that the uncertain input parameters are \( \tilde{\xi} = (\tilde{\xi}_1, \cdots , \tilde{\xi}_N) \), the extend of variation of which is represented by \( \Delta \tilde{\xi}_n \): each input \( \tilde{\xi}_n \) is thus associated with an interval \( [\tilde{\xi}_n - \Delta \tilde{\xi}_n, \tilde{\xi}_n + \Delta \tilde{\xi}_n] \), \( n = 1, \cdots, N \). This work proposed to find optimal uncertainty bars on input data ensuring a given statistic properties for the output of interest by solving an optimization problem.

James and Dowling [41] investigated how the sound pressure is affected by the uncertain wavenumber and source strength. They assumed that the sound amplitude \( A \) and wavenumber \( k \) follow PDFs \( f_A \) and \( f_k \). In a one-dimensional space with coordinate \( x \), the PDF for a given sound pressure is thus \( f_P(R, I; x) \), where \( R \) and \( I \) are the real and imaginary parts of the pressure related to the amplitude \( A \) and the wavenumber \( k \) by

\[ A = \sqrt{R^2 + I^2}, \quad \tan \theta = \frac{I}{R}, \quad k_n = \frac{\theta - 2\pi n}{x}, \quad n = 0, \pm 1, \pm 2, \cdots. \]

(4.35)

Here \( \theta \) is the phase and \( k_n \) are possible wavenumbers at location \( x \). At the initial point \( x = 0 \), we have \( \theta = 0 \), and therefore the PDF of the sound pressure \( f_P \) is

\[ f_P(I, R; x = 0) = \frac{1}{A} f_A(A) \delta(\theta) = \frac{1}{\sqrt{R^2 + I^2}} f_A \left( \sqrt{R^2 + I^2} \right) \delta \left( \tan^{-1} \left( \frac{I}{R} \right) \right). \]

(4.36)
When $x > 0$, the PDF obtained after propagation of the uncertainty is then

$$f_p(I, R; x) = \frac{f_A(\sqrt{R^2 + I^2})}{x\sqrt{R^2 + I^2}} \sum_n f_k \left( \frac{\tan^{-1}(I/R) - 2\pi n}{x} \right)$$

(4.37)

for $I > 0$, and $2\pi n$ in $f_k$ is replaced by $2\pi(n - 1/2)$ when $I < 0$. However, in two or more spatial dimensions, the complexity prevents from finding a closed form solution. Then, either expensive numerical techniques have to be employed, or special cases have to be considered. In [41], the authors state that “uncertain wave propagation involving one or more independent spatial dimensions may not be adequately described by an expected value and a variance” (e.g., as done in [8, 48]). However, the proposed technique still cannot deal with complex models: the measurement error and the inverse problem addressed in our thesis were not considered, which is a much more complicated problem.

Zhang et al. used a Bayesian method to estimate the model parameter and quantify the model error (e.g., the systematic error due to linearization of the model) simultaneously, in the finite element method [78] and force reconstruction problem [77]. Let $\xi$ be the model parameter to be estimated, $\gamma$ be the parameter that characterizes the modeling error, $X$ be the measured output, and $f(\xi)$ and $f(\gamma)$ be the prior information of both parameters respectively. The joint posterior PDF of $\xi$ and $\gamma$ is

$$f(\xi, \gamma|X) \propto f(X|\xi, \gamma)f(\xi)f(\gamma),$$

(4.38)

where $f(X|\xi, \gamma)$ stands for the PDF of the output function. Maximum a posterior (MaP) estimates of both parameters can be obtained by maximizing $f(\xi, \gamma|X)$. The sample of this (high dimensional) posterior distribution may be obtained via a Markov Chain Monte Carlo (MCMC) method, so that marginal posterior of each parameter [78] and force reconstruction [77] can be obtained.

Becker et al. [5], whose theory is based on [49], studied the sensitivity analysis of a nonlinear finite element model. Denote the model to be emulated as an unknown function $g(x)$, where $x$ is a vector of $d$ model inputs, resulting in a model output $y$. Assume that the prior distribution of $g(x)$ is Gaussian with mean $w\phi(x)^T = w_0 + w_1x_1 + \cdots + w_dx_d$ and variance $\sigma^2$, denoted as $f(g(x)|w, \sigma^2)$. On the other hand, using a set of training data, which consist of $n$ input-output pairs, the posterior distribution of $w$ and $\sigma^2$, denoted as $f(w, \sigma^2|y)$, may be estimated. A joint distribution of $g(x)$, $w$ and $\sigma^2$ can be expressed as

$$f(g(x), w, \sigma^2|y) \propto f(g(x)|w, \sigma^2, y)f(w, \sigma^2|y).$$

(4.39)
By integrating (4.39), the posterior distribution of \( g(x) \) is obtained:

\[
f(g(x)|y) = \int \int f(g(x), w, \sigma^2|y)dw\sigma^2.
\]  (4.40)

Then, several quantities, e.g., posterior mean and variance, can be inferred to quantify the model sensitivity to the uncertainties.

Dosso [23] considered the ocean (single) sound source localization problem using a Bayesian approach. Let \( m \) and \( d \) represent vectors of model parameters (environment properties \( u \) and source location \( r \)) and data (measured acoustic fields) respectively. The posterior probability of the parameters given the measured field may then be expressed as

\[
f(m|d) = \frac{f(d|m)f(m)}{f(d)},
\]  (4.41)

where the prior distribution \( f(m) \) in this paper may be given by either an uniform or a Gaussian distribution. The sound source location is expressed by a jointly marginal probability over source location \( r \), which is obtained by the integral of \( P(m|d) \) over environmental properties parameter \( u \). The integral may be computed via Gibbs’ sampling.

Richardson and Nolte [57] and Shorey and Nolte [62] derived posterior source position PDF in an uncertain ocean environment. Assume that the signal consists of signal \( s \) and additive Gaussian noise \( n \) components, which is expressed as

\[ p = s(S, \Theta, \Psi) + n, \]  (4.42)

where \( S \) and \( \Theta \) denote the source position parameter and the transmitted pressure waveform parameter, and \( \Psi \) is the uncertain environment parameter. The posterior probability of the (single) source location is

\[
f(S|p) = \frac{f(S)}{f(p)} \int_\Psi f(p|S, \Psi)f(\Psi|S)d\Psi
\]

\[
= C f(S) \int_\Psi (E + 1)^{-1}e^{-\frac{|R|^2}{2E}}f(\Psi|S)d\Psi,
\]  (4.43)

where \( C \) is a normalization constant and \( E \) and \( R \) are functions of \( S \) and \( \Psi \).

Adhikari et al. [1] and Chowdhury et al. [9] proposed using fuzzy sets to quantify the parametric uncertainties in the structural dynamics and discussed uncertainty propagation. Erdogan and Bakir [24] described the measurement noise and model parameters by fuzzy numbers in a finite element model, then estimating the parameters by minimizing an objective function. The uncertainty propagation problem can be expressed in a general way
as
\[
\hat{y} = f(\hat{A}_1, \cdots, \hat{A}_N) \tag{4.44}
\]
where \( f \) is a smooth function of the fuzzy parameters \( (\hat{A}_1, \cdots, \hat{A}_N) \) and \( \hat{y} \) is the fuzzy output. For performing the uncertainty propagation, each fuzzy variable is considered as an interval variable for each \( \alpha \)-cut of the membership function of a fuzzy set in [24]. More specifically, for a fuzzy set \( \hat{A} \), its membership function \( \mu_{\hat{A}}(x) \) is assumed to be concave; therefore, its \( \alpha \)-cut \( \hat{A}^\alpha \) is an interval. Then the \( \alpha \)-cut of the fuzzy output variable can be computed using interval arithmetic
\[
y^\alpha = f(\hat{A}_1^\alpha, \cdots, \hat{A}_N^\alpha). \tag{4.45}
\]
By selecting a few \( \alpha \)-cuts and computing (4.15) and (4.45) respectively, the output fuzzy variable \( \hat{y} \) can be approximated. However, we may observe that applying interval arithmetic generally increases the length of the intervals. For example, the definition of the sum of two intervals \([a_1, b_1]\) and \([a_2, b_2]\) is
\[
[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2], \tag{4.46}
\]
which indicates that the length of the sum interval is equal to the sum of the lengths of the two intervals. Therefore, this approach may produce a meaningless large bound of the output. Furthermore, this approach is sometimes infeasible: \( f(A_1^\alpha, \cdots, A_N^\alpha) \) may not be computed for a complex function \( f \) because the interval space is not linear and the extension of classical operators such as multiplication and division to intervals is still an open question [69]. An alternative is to use global optimization techniques, yet this strategy is prone to high computational costs. It was suggested in [1, 9] to use a high-dimensional model representation to limit the computational complexity.

Gogu et al. [34] considered the propagation of uncertainty using the theory of belief functions. This paper represents the input variables by mass functions and gives a degree of imprecision on the output variable in terms of belief and plausibility functions. For convex and linear output functions, a solution can be obtained. However, for complex output functions (e.g., the complex-valued sound propagation function including a periodic phase term), the output uncertainty is difficult to compute using this method.

Riley [58] also quantified the model uncertainties using belief functions. This paper considered a finite model set \( \mathbf{M} = \{M_1, \cdots, M_N\} \) and quantified the model uncertainties as a mass function \( m \) on \( 2^\mathbf{M} \). Then, one may integrate the information of model uncertainties
to the output response of interest $y$ by its mass function

$$m(y) = \sum_{A_i \subseteq M} m(y|A_i)m(A_i). \quad (4.47)$$

The belief and plausibility functions on the response of interest $y$ can also be computed:

$$Bel(y) = \sum_{A_i \subseteq M} Bel(y|A_i)m(A_i), \quad (4.48)$$

$$Pl(y) = \sum_{A_i \subseteq M} Pl(y|A_i)m(A_i). \quad (4.49)$$

Note that most of the works mentioned above aim at propagating the uncertainty directly through the model. However, sound source localization is an inverse problem, in which imprecise measurements may have a significant impact on the accuracy of the parameter estimates. Other works [23, 57, 62] give a posteriori PDF of the source location, but these three papers only consider the single source problem and the models become very complicated when a few uncertainties are considered at the same time. In this thesis, we solve the multiple source localization problem and make it possible to take into account the uncertainties in the sound propagation and measurement process. More particularly, we propose to transpose the uncertainties to the data in order to estimate the model parameters via maximum likelihood.
Chapter 5

Sound Source Estimation in Presence of Uncertainty: Deterministic Amplitude Case

5.1 Model

In this chapter, we consider the sound source localization problem (2.1) under the assumption that the strength (or amplitude) $A$ is deterministic as in Chapter 3. However, we now assume that the microphone locations $r'_m$ and wavenumber $k$ are subject to uncertainties. For this reason, the propagation operator from the $s$-th source to the $m$-th microphone

$$G(r_s | r'_m) = \frac{\frac{\pi}{a} (r_s - r'_m)}{4\pi |r_s - r'_m|}$$ (5.1)

is also ill-known. As introduced in Section 4.3, there are several methods for quantifying the model uncertainties. In this thesis, the uncertainties are taken into account using the theory of belief functions. In the next section, we use a first order Taylor approximation method to quantify the uncertainty of measurement due to microphone location and wavenumber uncertainties. The data used in the parameter estimation are then represented by contour functions which include the information of uncertainty. Finally, the locations and strengths of sound sources are estimated using the E2M algorithm.
5.2 Uncertainty Representation and Propagation

5.2.1 Uncertain Data

From now on, we will denote by $\theta_1, \cdots, \theta_Q$ the $Q$ meta-parameters subject to uncertainties (e.g., microphone locations and wavenumber). This section investigates how uncertainties can be represented in the sound source model. First, it should be stressed out that the sound propagation model depends on the meta-parameters only through the observed data (here, the measured pressure vector $p_t$). Thus, rather than integrating the uncertainty on these meta-parameters directly in the model, we propose to transpose it to the observed data. This choice is motivated by the efficiency and the versatility of the resulting procedure. Indeed, many kinds of uncertain meta-parameters can then be taken into account, the only requirement being to identify is how these uncertainties impact the observed data. Furthermore, the propagation of the meta-parameter uncertainty to the measured pressures being separated from the estimation process, it will be done only once in the overall process.

Let the contour function $p_{l|p_t}$ specify the set of plausible values for the pressure measured at time $t$ (for convenience, from now on $p_{l|p_t}$ is simplified as $p_l$). That is, instead of an unique value corresponding to the pressure actually measured, $p_l$ associates degrees of plausibility to each of the possible pressures. Example 2 shows how meta-parameter uncertainty may be transferred to the data using uniform contour functions (equivalently interval-valued pressures).

**Example 2** (Uniform contour function). Assume that the covariance matrix of the contour function $p_l$ is diagonal (the uncertainties on the measurement from various microphones are cognitively independent): $\Sigma^* = \text{diag}(\sigma_1^2, \cdots, \sigma_M^2)$. Then, the entries of $\Sigma^*$ can be estimated by first order Taylor approximation (B.3):

$$
\sigma_m^* = \sum_{q=1}^{Q} \left[ \text{Re} \left( \frac{\partial p_{nt}}{\partial \theta_q} \right) \sigma_q \right]^2 + \sum_{q=1}^{Q} \left[ \text{Im} \left( \frac{\partial p_{nt}}{\partial \theta_q} \right) \sigma_q \right]^2, m = 1, \cdots, M. \quad (5.2)
$$

Now the uncertain knowledge of the sound pressure $p_{nt}$ (contour function) is represented by an interval, whose real and imaginary parts are $[a_{mp, \text{re}}, b_{mp, \text{re}}]$ and $[a_{mp, \text{im}}, b_{mp, \text{im}}]$ respectively. Let the center of the contour function be the measured pressure itself: $\mu_{mp} = p_{nt}$. Then the
left- and right-endpoints of the intervals \(a_{mp}^{re}, b_{mp}^{re}, a_{mp}^{im}, b_{mp}^{im}\) must satisfy that

\[
\begin{align*}
\frac{a_{mp}^{re} + b_{mp}^{re}}{2} &= \text{Re}(\mu_{mp}^{*}), \\
\frac{a_{mp}^{im} + b_{mp}^{im}}{2} &= \text{Im}(\mu_{mp}^{*}), \\
\frac{(b_{mp}^{re} - a_{mp}^{re})^2}{12} &= (\sigma_{m}^{re})^2, \\
\frac{(b_{mp}^{im} - a_{mp}^{im})^2}{12} &= (\sigma_{m}^{im})^2,
\end{align*}
\]

(5.3)

where the variance of the uniform variables of the real and imaginary parts are estimated by (B.3):

\[
(\sigma_{m}^{re})^2 = \sum_{q=1}^{Q} \left[ \text{Re}\left( \frac{\partial p_{mt}}{\partial \theta_q} \right) \sigma_q \right]^2
\]

(5.4)

and

\[
(\sigma_{m}^{im})^2 = \sum_{q=1}^{Q} \left[ \text{Im}\left( \frac{\partial p_{mt}}{\partial \theta_q} \right) \sigma_q \right]^2,
\]

(5.5)

and therefore

\[
\begin{align*}
a_{mt}^{re} &= \text{Re}(\mu_{mt}^{*}) - \sqrt{3}\sigma_{m}^{re}, \\
b_{mt}^{re} &= \text{Re}(\mu_{mt}^{*}) + \sqrt{3}\sigma_{m}^{re}, \\
a_{mt}^{im} &= \text{Im}(\mu_{mt}^{*}) + \sqrt{3}\sigma_{m}^{im}, \\
b_{mt}^{im} &= \text{Im}(\mu_{mt}^{*}) - \sqrt{3}\sigma_{m}^{im}.
\end{align*}
\]

(5.6)

Thus, the uncertainty of the measurement can equivalently be represented by an uniform contour function

\[
p_{l_{p_{mt}}}(p_{mt}) = 1_{[a_{mt}^{re}, b_{mt}^{re}]}(\text{Re}(p_{mt}))1_{[a_{mt}^{im}, b_{mt}^{im}]}(\text{Im}(p_{mt})),
\]

(5.7)

where \(1_{A}(x)\) is the characteristic function, defined by \(1_{A}(x) = 1\) if \(x \in A\) and 0 otherwise. Finally, the contour function of the measurement \(p = (p_1, \cdots, p_T)\) is a characteristic function on a 2MP-dimensional rectangle:

\[
p_{l_{p}}(p) = \prod_{i=1}^{T} \prod_{m=1}^{M} 1_{[a_{mt}^{re}, b_{mt}^{re}]}(\text{Re}(p_{mt}))1_{[a_{mt}^{im}, b_{mt}^{im}]}(\text{Im}(p_{mt})).
\]

(5.8)

In the remaining of Section 5.2, we will more particularly consider Gaussian contour
functions: the imprecise knowledge of the measured pressure is thus quantified by a Gaussian PDF, the expectation of which is given by the measured pressure: \( \mu^*_f = p_f \). Its covariance matrix \( \Sigma^* \) directly reflects the level of uncertainty to which the measurement is subject. This covariance matrix can be estimated from the meta-parameter uncertainties via (B.3):

\[
\Sigma^* \approx \nabla p_f(\mu) \Sigma_\Theta \nabla p_f(\mu)_H, \tag{5.9}
\]

where \( \nabla p_f(\mu) \) is the gradient with respect to \( \Theta \), and \( \mu_\Theta \) and \( \Sigma_\Theta \) are the expectation and covariance matrix of \( \Theta \), respectively. Next, we introduce a practical example of uncertainty representation in the case of Gaussian contour functions.

**Example 3** (Imprecise microphone locations and wavenumber). This example considers a two-source localization problem where the microphone locations are uncertain. Due to motions of the array, the actual position of its center and its orientation may change. We also assume that the wavenumber is subject to uncertainties, due to the presence of a heater which changes the temperature. Figure 5.1 gives a simplified representation of this problem in a two-dimensional case.

Assume the array center \( \mathbf{r}_0 = (x_0, y_0, z_0) \) and the rotation angles \( \theta_1, \theta_2 \) and \( \theta_3 \). Then, the actual coordinates of the microphones can be computed by

\[
r'_m = R_x(\theta_1) R_y(\theta_2) R_z(\theta_3) (r^*_m - \mathbf{r}_0), \text{ for } m = 1, \ldots, M, \tag{5.10}
\]

where the rotation matrices \( R_x, R_y, \) and \( R_z \) are defined by

\[
R_x(\theta_1) = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \theta_1 & -\sin \theta_1 \\
0 & \sin \theta_1 & \cos \theta_1 \\
\end{pmatrix}, \tag{5.11}
\]

\[
R_y(\theta_2) = \begin{pmatrix}
\cos \theta_2 & 0 & -\sin \theta_2 \\
0 & 1 & 0 \\
\sin \theta_2 & 0 & \cos \theta_2 \\
\end{pmatrix}, \tag{5.12}
\]

\[
R_z(\theta_3) = \begin{pmatrix}
\cos \theta_3 & -\sin \theta_3 & 0 \\
\sin \theta_3 & \cos \theta_3 & 0 \\
0 & 0 & 1 \\
\end{pmatrix}. \tag{5.13}
\]

Obviously, these parameters cannot be measured at each time. However, it is reason-
able to assume that partial information on the parameter uncertainties is available in the
form of variances $\sigma_{b_1}^2$, $\sigma_{b_2}^2$, $\sigma_{b_3}^2$, $\sigma_{x_0}^2$, $\sigma_{y_0}^2$, $\sigma_k^2$. Furthermore, these sources of uncertainty are assumed as mutually cognitively independent, and having zero mean (except for the wavenumber which has mean $k_0$). Thus, it is proposed to model the meta-parameter uncertainty by a multivariate Gaussian contour function $p_l(p_i)$ with expectation $\mu^*_i = p_i$, and whose covariance matrix can be estimated via (5.9):

$$\Sigma^* \approx \nabla p_l(\mu_\theta) \text{diag}(\sigma_{b_1}^2, \sigma_{b_2}^2, \sigma_{b_3}^2, \sigma_{x_0}^2, \sigma_{y_0}^2, \sigma_k^2) \nabla p_l(\mu_\theta)^H.$$  (5.14)

Here the gradient vector $\nabla p_l(\mu_\theta)$ is obtained by (C.42). Details of its computation can be found in Appendix C.

Fig. 5.1 Sound source estimation problem (2D case) with uncertain meta-parameters (microphone locations and wavenumber).

5.2.2 Likelihood Function of Uncertain Data

In this section, we present how the likelihood function of the model parameters may be computed from the uncertain data, which take the form of multivariate contour functions $p_l$. In this case, the uncertain data log-likelihood function, which is the counterpart to the certain data log-likelihood (3.1), is

$$\log L(r, A| p_l) = \log \int f(p_l|r, A) p_l(p_i) dp_i.$$  (5.15)
Assume further that each $p_i$ is a Gaussian contour function with mean $\mu_i^*$ and covariance matrix $\Sigma^*$. The uncertain data log-likelihood then becomes

$$\log L(r, A|p_i) = -(\mu_i^* - GA)^H (\sigma^2 I_M + \Sigma^*)^{-1} (\mu_i^* - GA).$$  \hspace{1cm} (5.16)

The detailed computation of (5.16) may be found in Appendix D. Now let us consider an example of uncertain data log-likelihood, which shows the advantage of taking microphone uncertainty into account.

**Example 4 (Log-likelihood of uncertain data).** This example considers a simple 2D example to show the advantage of taking into account data uncertainty. Figure 5.2 displays two sources with actual coordinates $r_1 = (-1,2)$ and $r_2 = (1,2)$; the coordinates of $S_1$ and the y-coordinate of $S_2$ are known, so that only the x-coordinate $r_{x,2}$ of $S_2$ needs to be estimated. Also, five microphones are placed on the y-axis, with assumed x-coordinates $(-2,-1,0,1,2)$. The first four x-coordinates are certain, but the fifth one, written hereafter $m_{x,5}$, is subject to noise, and may actually take any value in the interval $[1,3]$ (represented by the pink line segment in Figure 5.2). For various positions of microphone $M_5$, the pressures are generated by Eq. (2.1), in which the strengths are $A = (1,1)^T$, the frequency $f = 1000$ Hz and the standard deviation of noise $\sigma = 0.05$. Then, the log-likelihood $\log L(r_{x,2}|p_i)$ is computed via Eq. (3.1), assuming that the microphone position is $(2,0)$. Alternatively, the generalized log-likelihood (5.16) $\log L(r_{x,2}|p_i)$ is computed, where the Gaussian contour function $p_i$ has mean $\mu_i^*$ taking value as the generated pressure and covariance matrix $\Sigma^* = \text{diag}(0,0,0,0,\sigma_s^2)$, with $\sigma_s^2 \in \{0.003, 0.01, 0.05\}$.

Figure 5.3(a) displays the level curves of $\log L(r_{x,2}|p_i)$ as a function of $r_{x,2}$ and $m_{x,5}$. When the actual x-coordinate of the fifth microphone is equal to its assumed value ($m_{x,5} = 2$), the log-likelihood of $p_i$ reaches its maximum for the actual sound source x-coordinate ($r_{x,2} = 1$): in this case, a classical ML estimation makes it possible to correctly locate the sound source. However, when the information of the microphone location is wrong (i.e., $m_{x,5} \neq 2$), this strategy may result in a wrong source location estimate, due to local maxima in the likelihood function.

Figures 5.3(b-d) show the level curves of $\log L(r_{x,2}|p_i)$ as a function of $r_{x,2}$ and $m_{x,5}$, for $\sigma_s^2 = 0.003$, $\sigma_s^2 = 0.01$, and $\sigma_s^2 = 0.05$, respectively. As the level of uncertainty $\sigma_s^2$ on the assumed microphone location $m_{x,5}$ increases, it may be seen that the local maxima observed for the classical log-likelihood function disappear. The explanation is that the weight of the information provided by microphone $M_5$ in the estimation process then decreases. Thus, the estimation process relies principally on the information provided by the other microphones,
the likelihood function being “smoothed” with respect to the variations of $m_{x,5}$. As a result, maximizing the generalized likelihood makes it possible to correctly estimate the actual sound source location, even if the information provided by $M_5$ is wrong.

![Diagram](image)

**Fig. 5.2** 2D sound source localization example. The sound sources are represented by blue crosses. The actual microphone locations are represented by red crosses. The real location of 5th microphone is uncertain, taking possible values on the pink line segment.

### 5.3 Model Estimation Using the Evidential EM Algorithm

This section details how the ML approach used in the precise case can be extended to cope with uncertain sound pressures. The uncertain data at hand are represented by contour functions $p_l$. The purpose of the maximum likelihood approach is to find the optimal parameter estimates by maximizing the likelihood function of the incomplete uncertain data $L(r, A|p_l)$. By introducing the contributions of the various sources $c_t = (c_{1t}, \cdots, c_{St})$ as latent variables, with

$$c_{st} = G_s A_s + n_{st},$$

the problem can be equivalently solved using the E2M algorithm introduced in Section 4.2.3. Here $n_{st}, s = 1, \cdots, S$ are mutually independent Gaussian random variables, with 0-mean and covariance $\Sigma_s = \frac{\sigma^2}{S} I_M$. The E-step and M-step of the E2M algorithm for this sound source localization model are presented as follows.
Fig. 5.3 Log-likelihood of $p_t$ (a); generalized log-likelihood $pl_t$ (b-d) for different x-coordinates of $S_2$ and x-coordinates actual locations of microphone $M_5$. The values of $\sigma^2$ in $pl_t$ are 0.003, 0.01 and 0.05 in (b), (c) and (d) respectively. In all cases, microphone $M_5$ is assumed to have a x-coordinate $m_{x,5} = 2$.

E-step

First, let us notice that:

$$Q(\Phi|\Phi') = \mathbb{E} \left[ \log L(r, A|c_{st})|pl_t, r', A' \right]$$

$$= d - \sum_{t=1}^{T} \sum_{s=1}^{S} \left[ \mathbb{E}(c_{st}|pl_t, r', A') - G_sA_s \right]^2,$$

where $d$ is a constant independent of parameters $r$ and $A$. This expectation depends on the uncertain data only through the expected latent pressures $\hat{c}_{st} = \mathbb{E}(c_{st}|pl_t)$. To proceed with these latter, we first need to explicit the conditional PDF of the observed pressures given the uncertain data:

$$f(p_t|pl_t) = \frac{pl_t(p_t)f(p_t)}{\int pl_t(p_t)f(p_t)dp_t},$$

where $f(p_t) = \phi(p_t|GA, \Sigma)$ and $\phi(\cdot|\mu, \Sigma)$ denotes the PDF of the multivariate Gaussian distribution with expectation $\mu$ and covariance matrix $\Sigma$. By Theorem 1, the conditional
probability of the latent contributions given the measured pressure is

\[ f(c_{st}|p_t) = \phi \left( c_{st}|G_sA_s + \frac{1}{S}(p_t - GA), \left( \frac{1}{S} - \frac{1}{S^2} \right) \Sigma \right). \]  

(5.20)

Thus, the conditional PDF of the latent contributions \( c_{st} \) given the uncertain pressures \( pl_t \) is

\[ f(c_{st}|pl_t) = \int f(c_{st}|p_t)f(p_t|pl_t(p_t))dp_t. \]  

(5.21)

Finally, the expectation of the latent contributions \( c_{st} \) given the uncertain measured pressures \( pl_t \) is

\[
\hat{c}_{st} = \int c_{st}f(c_{st}|pl_t)dc_{st} = \int \int c_{st}f(c_{st}|p_t)dc_{st}f(p_t|pl_t(p_t))dp_t \\
= \frac{\int \left[ G_sA_s + \frac{1}{S}(p_t - GA) \right] pl_t(p_t)f(p_t)dp_t}{\int pl_t(p_t)f(p_t)dp_t} \\
= G_sA_s - \frac{1}{S}GA + \frac{1}{S} \frac{\int p_tpl_t(p_t)f(p_t|r^t,A^t)}{\int pl_t(p_t)f(p_t)dp_t}. 
\]  

(5.22)

Therefore, in the \( l \)-th iteration of the E2M algorithm, the E-step amounts to computing the expected latent pressures

\[
\hat{c}_{st}^l = G_s^lA_s^l - \frac{1}{S}G_s^lA_s + \frac{1}{S} \frac{\int p_tpl_t(p_t)f(p_t|r^t,A^t)}{\int pl_t(p_t)f(p_t)dp_t}. 
\]  

(5.23)

Example 5 (Interval-typed contour function continued). We assume that the uncertain measurements are represented by uniform contour functions \( pl_t \) (see Eq. (5.8) in Example 2):

\[
pl_t(p_t) = \prod_{m=1}^{M} I_{[a_{ml}^m,b_{ml}^m]}(\text{Re}(p_{mt}))I_{[d_{ml}^m,b_{ml}^m]}(\text{Im}(p_{mt})). 
\]  

(5.24)

Let \( \Sigma = \text{diag}(\sigma_1^2, \cdots, \sigma_M^2) \) denote the diagonal covariance matrix of the additive noise. Then, the expected latent pressures (5.23) can be computed as:

\[
\mathbb{E}(c_{st}|pl(p_t)) = G_s^lA_s^l + \frac{1}{Sl^2},
\]  

(5.25)
where \( l_1 \) and \( l_2 \) are obtained by:

\[
l_1 = \int p_l(p_t)f(p_t|r',A')dp_t = \prod_{m=1}^{M} \int p_{l,p_m}(p_{mt})f(p_{mt}|r',A')dp_{mt}
\]

\[
= \prod_{m=1}^{M} \int 1_{[r_{mp},r_{mp}]}(\text{Re}(p_{mt})) \phi \left( \text{Re}(p_{mt})|\text{Re} \left( [G'A']_m \right), \frac{\sigma_m^2}{2} \right) d\text{Re}(p_{mt})
\]

\[
= \prod_{m=1}^{M} \int 1_{[r_{mp},r_{mp}]}(\text{Im}(p_{mt})) \phi \left( \text{Im}(p_{mt})|\text{Im} \left( [G'A']_m \right), \frac{\sigma_m^2}{2} \right) d\text{Im}(p_{mt})
\]

\[
= \prod_{m=1}^{M} \frac{\sigma_m^2}{2} \left[ F \left( \frac{b_{mp} - \text{Re} \left( [G'A']_m \right)}{\sigma_m/\sqrt{2}} \right) - F \left( \frac{a_{mp} - \text{Re} \left( [G'A']_m \right)}{\sigma_m/\sqrt{2}} \right) \right]
\]

\[
= \prod_{m=1}^{M} \frac{\sigma_m^2}{4\pi} \left[ e^{-\frac{(a_{mp} - \text{Re} \left( [G'A']_m \right))^2}{\sigma_m^2}} - e^{-\frac{(b_{mp} - \text{Re} \left( [G'A']_m \right))^2}{\sigma_m^2}} \right]
\]

\[
= \prod_{m=1}^{M} \frac{\sigma_m^2}{4\pi} \left[ e^{-\frac{(a_{mp} - \text{Im} \left( [G'A']_m \right))^2}{\sigma_m^2}} - e^{-\frac{(b_{mp} - \text{Im} \left( [G'A']_m \right))^2}{\sigma_m^2}} \right].
\]

(5.26)

where \( F \) is the cumulative distribution function of standard Gaussian distribution, and

\[
l_2 = \int \left( p_t - G'A' \right) p_l(p_t)f(p_t|r',A')dp_t
\]

\[
= \prod_{m=1}^{M} \int \left( p_{mt} - [G'A']_m \right) p_{l,p_m}(p_{mt})f(p_{mt}|r',A')dp_{mt}
\]

\[
= \prod_{m=1}^{M} \int 1_{[r_{mp},r_{mp}]}(\text{Re}(p_{mt})) \text{Re} \left( p_{mt} - [G'A']_m \right) \phi \left( \text{Re}(p_{mt})|\text{Re} \left( [G'A']_m \right), \frac{\sigma_m^2}{2} \right) d\text{Re}(p_{mt})
\]

\[
= \prod_{m=1}^{M} \int 1_{[r_{mp},r_{mp}]}(\text{Im}(p_{mt})) \text{Im} \left( p_{mt} - [G'A']_m \right) \phi \left( \text{Im}(p_{mt})|\text{Im} \left( [G'A']_m \right), \frac{\sigma_m^2}{2} \right) d\text{Im}(p_{mt})
\]

\[
= \prod_{m=1}^{M} \frac{\sigma_m^2}{2} \left[ F \left( \frac{b_{mp} - \text{Im} \left( [G'A']_m \right)}{\sigma_m/\sqrt{2}} \right) - F \left( \frac{a_{mp} - \text{Im} \left( [G'A']_m \right)}{\sigma_m/\sqrt{2}} \right) \right]
\]

\[
= \prod_{m=1}^{M} \frac{\sigma_m^2}{4\pi} \left[ e^{-\frac{(a_{mp} - \text{Re} \left( [G'A']_m \right))^2}{\sigma_m^2}} - e^{-\frac{(b_{mp} - \text{Re} \left( [G'A']_m \right))^2}{\sigma_m^2}} \right]
\]

\[
= \prod_{m=1}^{M} \frac{\sigma_m^2}{4\pi} \left[ e^{-\frac{(a_{mp} - \text{Im} \left( [G'A']_m \right))^2}{\sigma_m^2}} - e^{-\frac{(b_{mp} - \text{Im} \left( [G'A']_m \right))^2}{\sigma_m^2}} \right].
\]

(5.27)

In the above example, the uncertain information of the microphone positions \( p_{l,p_m}, m = 1, \cdots, M \) are assumed to be cognitively independent, i.e., \( \Sigma^* \) is diagonal. However, in some situations, this assumption may not be reasonable. In the case of Gaussian contour functions, however, the computations lead to a closed form for the expected latent contributions, even when \( \Sigma^* \) is non-diagonal. In the remaining of this section, we consider such Gaussian contour functions \( p_l(p_t) = \phi(p_t|\mu^*_t, \Sigma^*) \). First, we remind here a well-known theorem of multivariate Gaussian distributions [51] and give its corollary.

**Theorem 3.** The product of two PDFs (denoted as \( \phi(x|\mu_1, \Sigma_1) \) and \( \phi(x|\mu_2, \Sigma_2) \), respec-
tively) of random variable \(X\) is
\[
\phi(x|\mu_3, \Sigma_3)\phi(\mu_1|\mu_2, \Sigma_1 + \Sigma_2),
\]
where \(\mu_3 = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)\) and \(\Sigma_3 = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\).

**Corollary 5.3.1.** Assume that \(X\) follows a complex Gaussian distribution with PDF \(f_X(x) = \phi(x|\mu_1, \Sigma_1)\), and that it is imprecisely observed through the Gaussian contour function \(p_X(x) = \phi(\cdot|\mu_2, \Sigma_2)\). Then, the PDF of the random variable \(X\) given the Gaussian contour function \(p_X\) is
\[
f_{X|p_X}(x|p_X(x)) = \phi(x|\mu_3, \Sigma_3),
\]
where \(\mu_3 = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)\) and \(\Sigma_3 = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\).

**Proof**

By Theorem 3,
\[
f_X(x)p_X(x) = \phi(x|\mu_3, \Sigma_3)\phi(\mu_1|\mu_2, \Sigma_1 + \Sigma_2),
\]
then the likelihood function can be computed by
\[
L(x) = \int p_X(x)f_X(x)dx = \phi(\mu_1|\mu_2, \Sigma_1 + \Sigma_2),
\]
so we finally have
\[
f_{X|p_X}(x|p_X(x)) = \frac{f_X(x)p_X(x)}{L(x)} = \phi(x|\mu_3, \Sigma_3). \quad \square
\]

Then, from Corollary 5.3.1, the PDF of the measured pressure vector \(p_t\) given the uncertain measurement \(l_t\) is obtained by
\[
f(p_t|l_t) = \phi\left(p_t|\left\{\frac{I_M}{\sigma^2} + (\Sigma^*)^{-1}\right\}^{-1}\left(\Sigma^*)^{-1}\mu^* + \frac{I_M}{\sigma^2}GA\right), \left\{\frac{I_M}{\sigma^2} + (\Sigma^*)^{-1}\right\}^{-1}\right),\]
(5.33)

This makes it possible to obtain the following expression for the expectations of the latent contributions of the sources:
\[
\hat{e}_{st}^l = G_s^l A_s^l - \frac{1}{S}G_s^l A^l + \frac{1}{S}\left(\frac{I_M}{\sigma^2} + (\Sigma^*)^{-1}\right)^{-1}\left(\Sigma^*)^{-1}\mu^* + \frac{I_M}{\sigma^2}G^l A^l\right).
\]
(5.34)
M-step

The M-step consists in computing new $\mathbf{f}$s for the parameters so as to maximize Eq. (5.18); in this case, it amounts to solving

$$
r_{s}^{l+1}, A_{s}^{l+1} = \arg \min_{r_{s}, A_{s}} \sum_{t=1}^{T} \left| \mathbf{c}_{st}^{l} - \mathbf{G}_{s}A_{s} \right|^{2}.
$$

(5.35)

Notice that the only difference between the EM and E2M algorithms lies in the E-step (Eqs. (3.9) and (5.23)). The general formulation for the M-step (Eqs. (3.10) and (5.35)) is the same. The update equations for the M-step of the E2M algorithm are thus

$$
r_{s}^{l+1} = \arg \max_{r_{s}} (\mathbf{c}_{s}^{l})^{H} \mathbf{G}_{s}^{H} \mathbf{c}_{s}^{l},
$$

(5.36)

and

$$
A_{s}^{l+1} = \frac{(\mathbf{G}_{s}^{l+1})^{H} \mathbf{c}_{s}^{l}}{(\mathbf{G}_{s}^{l+1})^{H} \mathbf{G}_{s}^{l+1}},
$$

(5.37)

where

$$
\hat{\mathbf{c}}_{s}^{l} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{c}_{st}^{l}.
$$

(5.38)

The strategy for estimating the sound sources locations and strengths using the E2M algorithm is summarized in Algorithm 2.

Algorithm 2 E2M Algorithm

For $l = 0$, pick starting values for the parameters $\mathbf{r}^{0}, \mathbf{A}^{0}$. For $l \geq 1$:

repeat

estimate the source contribution $\hat{\mathbf{c}}_{s}^{l}$ from Eqs. (5.38) and (5.23) (or (5.34) for Gaussian contour function case), $s = 1, \cdots, S$;

estimate the new source location estimates $\mathbf{r}_{s}^{l+1}, s = 1, \cdots, S$ by Eq. (5.36);

estimate the new source strength estimates $A_{s}^{l+1}, s = 1, \cdots, S$ by Eq. (5.37);

until the relative increase of the log-likelihood is less than a given threshold $\kappa$:

$$
\frac{\log L(\mathbf{r}^{l+1}, \mathbf{A}^{l+1}|p/l_{p}) - \log L(\mathbf{r}^{l}, \mathbf{A}^{l}|p/l_{p})}{\log L(\mathbf{r}^{l}, \mathbf{A}^{l}|p/l_{p})} < \kappa,
$$

(5.39)

where the log-likelihood $\log L(\mathbf{r}, \mathbf{A}|p/l_{p})$ is given by (D.4) in Appendix B.

It would be easy to conclude that the E2M algorithm is a generalization of the EM algorithm when addressing the problem of sound source localization. Indeed, when the data
5.4 Experiments

are certain, i.e., \( pl_t \) is a Dirac function taking unit impulse point at the measured pressure, the last term of Eq. (5.23) becomes \( \frac{1}{S} p_t \) (here for convenience \( p_t \) represents the measurement as well). Thus, Eqs. (3.9) and (5.23) are equivalent. In order to highlight the properties of the E2M algorithm and to show how it affects the results for the sound source localization problem, let us consider a particular case of Gaussian contour functions. Assume that the covariance matrix of \( pl_t(p_t) \) is diagonal: \( \Sigma^* = \text{diag}(\sigma_1^*, \ldots, \sigma_M^*) \). Then, the E-step consists in computing \( \hat{e}_m^* = (\hat{e}_1^*, \ldots, \hat{e}_M^*) \), in which

\[
\hat{e}_{mst} = [G^l A^l]_m + \frac{1}{S} \frac{\sigma^2 (\mu_{mt}^* - [G^l A^l]_m)}{\sigma^2 + \sigma_{mt}^2},
\]

where \([v]_m \) stands for the \( m \)-th entry of vector \( v \). Compared with the E-step of the EM algorithm (3.9), the E2M algorithm gives an extra weight \( \frac{\sigma^2}{\sigma^2 + \sigma_{mt}^2} \) to the measurement term. More particularly, the weight representing the degree of decrease \( \frac{\sigma^2}{\sigma^2 + \sigma_{mt}^2} \in (0, 1] \) is determined by the relative proportion of the variances of the noise and the uncertainty. Moreover, Eq. (5.40) illustrates that the E2M is a generalization of the EM in terms of level of belief: if the measurement is totally certain, i.e., \( \sigma_{mt}^* = 0 \), the last term of Eq. (5.40) (E-step) becomes \( \frac{1}{S} (\mu_{mt}^* - [G^l A^l]_m) \), and therefore the EM and E2M algorithms are identical.

5.4 Experiments

This section illustrates the proposed estimation strategy on both simulated and real data. As in Section 3.4, the experiments are based on the setup described in Figure 3.2. An array of 60 microphones is originally located on the plane \( z = 0 \) (the center of the array coincides with the origin). Two sound sources are placed in front of the microphone array. In this section, uncertainties in the microphone locations and wavenumber are introduced as in Example 3: the actual position of the center and the orientation of the array change due to motions of the array, and the wavenumber changes due to the variation of the temperature. The results are provided in Sections 5.4.1 for simulated data and in Section 5.4.2 for real experimental data respectively.

5.4.1 Simulated Data

In this simulation experiment, two sound sources are considered with strengths \( A_1 = A_2 = 0.8 \text{Pa} \) and locations \( r_1 = (-0.2m, 0.1m, 0.2m) \), \( r_2 = (0.2m, -0.1m, 0.2m) \); the sound frequency is \( f = 5000\text{Hz} \) for both sources. Then, uncertainties are introduced in the micro-
phone locations and in the wavenumber as follows. A wavenumber value is generated according to a Gaussian distribution with mean $\mu_k = \frac{2\pi f}{c}$ (with sound velocity $c = 340\text{m/s}$ and frequency $f = 5000\text{Hz}$) and predefined standard deviation $\sigma_k$. Similarly, a Gaussian distributed noise is introduced in the array center location $r_0$ with zero mean and predefined standard deviation $\sigma_{x_0}, \sigma_{y_0}, \sigma_{z_0}$ (for the x-, y-, z-coordinates, respectively), and in each of the rotation angles of the array $\theta_p$ ($p = 1, 2, 3$) with 0-mean and predefined standard deviation $\sigma_{\theta_p}$. We then generate pressures according to Eqs. (2.1) and (5.10), in which we set the measurement noise to be $\sigma = 0.05$.

For a given dataset, both the EM and E2M algorithms are run using these different starting values, retaining the solution with the highest log-likelihood. Remark that the covariance matrices $\Sigma^*$ of the contour functions used in E2M are computed using the parameters $r, A$ estimated via the EM algorithm. Since the data are randomly generated, the above procedure (from data generation to model estimation) is repeated 30 times, so that the mean square error (MSE) of the parameter values and 95% confidence intervals of the square error may be computed.

First, let $\sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.02\text{m}$, $\sigma_{\theta_p} = 15\frac{\pi}{180}$ for $p = 1, 2, 3$, $\sigma_k = 0.5m^{-1}$. Figure 5.4 shows the 30 estimation results for the x- and y-coordinates of the sources via the EM (blue circles) and the E2M (red crosses) algorithms, as well as the 95% confidence ellipses of the estimated source location. Similarly, Figure 5.5 exhibits a corresponding result for $\sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.04\text{m}$, $\sigma_{\theta_p} = 20\frac{\pi}{180}$ for $p = 1, 2, 3$, $\sigma_k = 0.5m^{-1}$. The estimates obtained via E2M clearly exhibit a smaller spread around the actual sources locations than the estimates obtained via EM, which illustrates its interest in terms of dealing with uncertain data.

Then, we investigate the estimation error of both algorithms. Let $\sigma_k = 0$ (the wavenumber is certain), and let us increase the degree of uncertainty on the microphone locations: $\sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.01l$ and $\sigma_{\theta_p} = 5l\frac{\pi}{180}$ ($p = 1, 2, 3$), for $l = 0, 1, 2, 3, 4$. The MSE of the estimated sound source locations and 95% confidence intervals of the square errors are displayed in Figure 5.6 (a). Alternatively, let $\sigma_k = \sigma_y = \sigma_z = \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 0$ (the microphone locations are precise) and let $\sigma_k$ increase from 0 to 1. Figure 5.6 (b) shows the corresponding estimation error trend. In both cases, when the level of uncertainty is low, the EM and E2M algorithms exhibit similar performances. As the amount of uncertainty increases, the accuracy of the estimates obtained using both algorithms decreases. However, the E2M proves to be much more robust to uncertainty than the EM, its estimation error staying under an acceptable level while the estimation error of EM estimate increases dramatically.
5.4 Experiments

5.4.2 Real Experiment

In this experiment, two sound sources with coordinates \( \mathbf{r}_1 = (-0.239m, -0.112m, 0.214m) \) and \( \mathbf{r}_2 = (0.172m, -0.012m, 0.214m) \) are considered; both radiate at a frequency \( f = 5025Hz \). The dataset is composed of 14 different signals: each has a duration of 30 seconds and was consequently divided into 30 snapshots in the frequency domain. Uncertainty was introduced in the microphone locations by applying a rotation (around 10° on average) and a translation (around 5cm on average) to the panel. The wavenumber was also considered as uncertain, since the temperature measured during the experiment ranged from 17°C to 23°C.

On each of the 14 datasets, ML estimates of the parameters were estimated using both the EM and E2M algorithms, each using 100 initial values obtained as follows. The starting values for each source location were generated using the same protocol as before, according to a three-dimensional uniform distribution on a 10cm × 10cm × 5cm cube. The initial values for the strengths of the sources were generated according to a complex-valued uniform distribution so that the sound pressure uncertainty was ±3 dB. Let \( \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = \frac{10}{180} \pi \), \( \sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.05m \), \( \sigma_k = 0.3m^{-1} \). For each of the fourteen signals, the EM and E2M
Fig. 5.5 Source location estimates obtained using the EM (blue circles) and E2M (red crosses) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. $\sigma_{x_0} = \sigma_{y_0} = \sigma_z = 0.04 m$, $\sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 20 \frac{\pi}{180}$, $\sigma_r = 0.5 m^{-1}$.

algorithms were run 100 times with different starting values, and the parameter estimates with maximum likelihood were kept.

The MSE of the 14 source locations estimates are 0.0435 using EM and 0.0153 using E2M, and the corresponding 95% confidence intervals of square error are $[0.0325, 0.0545]$ for EM and $[0.0129, 0.0177]$ for E2M. Furthermore, Figure 5.7 displays the x- and y-coordinate estimates of the source locations along with 95% confidence ellipses using EM and E2M. Figure 5.8 shows the source location estimates and 95% confidence ellipsoids in the 3-D space. As before, the E2M estimates exhibit a smaller spread over the actual source locations than the EM estimates. This shows the advantage of taking into account the uncertainty in the model.

5.5 Conclusion

In this chapter, we solved the multiple sound source localization problem for deterministic signals with model and data uncertainties. The uncertainties on the meta-parameters (micro-
phone locations and wavenumber) were considered. Instead of integrating the uncertainties on these meta-parameters directly in the model, we proposed to represent the uncertainties using contour functions. The level of uncertainty is directly reflected by the variance of the contour functions, estimated by first-order Taylor expansion. Then, model estimation is carried out using the E2M algorithm. Finally, we presented both simulation and real experiments to compare the results obtained via our approach with those obtained via the EM procedure described in Chapter 3, which does not take the uncertainties on the measurement into account. The results clearly showed the interest of the new approach in terms of accuracy of the model parameter estimation.
Fig. 5.7 Source location estimates obtained using EM (blue points) and E2M (red crosses), and corresponding 95% confidence ellipses on the x-y plane. Black crosses represent the actual source locations.

Fig. 5.8 Source location estimates obtained using EM (left) and E2M (right), and corresponding 95% confidence ellipsoids in 3D space. Black crosses represent the source locations.
Part III

Sound Source Localization for Random Signals
Chapter 6

Sound Source Estimation Under the Random Amplitude Assumption

6.1 Sound Source Model With Random Amplitude

In this chapter, we consider random signals [12, 33, 43]. Unlike the deterministic signals, the random signals cannot be exactly modeled by a formula, but a probabilistic term. Indeed, the randomness of a measurement quantifies not only the measurement noise but also its random source strength. In this case, the superimposed signal model is

\[
p_t = G(r)A_t + n_t = \sum_{s=1}^{S} G_s(r_s)A_{st} + n_t. \tag{6.1}
\]

Unlike in Eq. (3.2), the strengths \( A_t = (A_{1t}, \cdots, A_{St})^T \) are now considered to follow a \( S \)-dimensional complex-valued Gaussian distribution with mean 0 and covariance matrix

\[
\text{Cov}(A_t) = \mathbb{E}(A_t A_t^H) = \text{diag}(a_1^2, \cdots, a_S^2). \tag{6.2}
\]

As before, the noise \( n_t \) is represented by an \( M \)-dimensional Gaussian random vector with 0-mean and covariance matrix \( \Sigma = \sigma^2 I_M \). We further assume that \( A_{t_1} \) and \( n_{t_2} \) are independent for any \( t_1, t_2 \in \{1, \cdots, T\} \). The purpose is to estimate the sound source locations \( r \) and the amplitudes in terms of their covariances \( \mathbf{a}^2 = (a_1^2, \cdots, a_S^2)^T \).

Under the above assumptions, the sound pressure \( p_t \) measured by a microphone array
follows a complex-valued Gaussian distribution with 0-mean and covariance

\[
\Sigma_p(r, a^2) = \mathbb{E}(p_t p_t^H) = \mathbb{E}(GA_tA_t^H G^H) + \mathbb{E}(n_t n_t^H) = G \text{ diag}(a_1^2, \ldots, a_S^2) G^H + \Sigma. \tag{6.3}
\]

Then, the log-likelihood function of the measurements \( p = (p_1, \ldots, p_T) \) is

\[
\log L(r, a^2|p) = -\log (\det(\Sigma_p)) - \frac{1}{T} \sum_{t=1}^T p_t^H \Sigma_p^{-1} p_t
\]

\[
= -\log (\det(\Sigma_p)) - \frac{1}{T} \sum_{t=1}^T \text{tr} (\Sigma_p^{-1} p_t p_t^H)
\]

\[
= -\log (\det(\Sigma_p)) - \text{tr} (\Sigma_p^{-1} \hat{\Sigma}_p), \tag{6.4}
\]

where \( \hat{\Sigma}_p \) is the sample covariance matrix, defined by

\[
\hat{\Sigma}_p = \frac{1}{T} \sum_{t=1}^T p_t p_t^H. \tag{6.5}
\]

The sound source estimation problem can then be carried out by maximizing the log-likelihood function (6.4):

\[
\hat{r}, \hat{a}^2 = \arg\min_{p, a^2} (\log (\det(\Sigma_p)) + \text{tr} (\Sigma_p^{-1} \hat{\Sigma}_p)). \tag{6.6}
\]

As in the case of deterministic signals, this optimization problem is complicated. In the next section, we detail how the EM algorithm can be used to bypass this difficulty.

### 6.2 Model Estimation Via the EM Algorithm

As before, let the complete data \( c_t = (c_{1t}, \ldots, c_{6t}) \) represent the contributions of the sound sources to the pressures measured by the microphones:

\[
c_{st} = G_s A_{st} + n_{st}, s = 1, \ldots, S, \tag{6.7}
\]

where the noise component \( n_{st} \) is obtained, as before, by arbitrarily decomposing the total noise \( n_t \) into \( S \) components: \( \sum_{s=1}^S n_{st} = n_t \). Thus, as before, the incomplete data \( p_t \) is related
to the latent source contributions $c_{sf}$ by

$$ p_t = \sum_{s=1}^{S} c_{sf}. $$

(6.8)

Remark that $c_{sf}$ is Gaussian with 0-mean and covariance matrix

$$ \Sigma_{c_{sf}}(r_s, \alpha_s^2) = \mathbb{E}(c_{sf} c_{sf}^H) = \mathbb{E}(|A_{sf}|^2 G_s G_s^H) + \mathbb{E}(n_{sf} n_{sf}^H) $$

$$ = \alpha_s^2 G_s G_s^H + \frac{\sigma^2}{S} I_M. $$

(6.9)

Then, since the contributions are assumed to be independent from each other, the covariance matrix of the measurement $\Sigma_p(r, a^2)$ is the sum of the covariances of the contributions $\Sigma_{c_{sf}}(r_s, \alpha_s^2), s = 1, \cdots, S:

$$ \sum_{s=1}^{S} \Sigma_{c_{sf}}(r_s, \alpha_s^2) = \Sigma_p(r, a^2). $$

(6.10)

The log-likelihood function of the complete data $e = \{c_{sf}, s = 1, \cdots, S, t = 1, \cdots, T\}$ is

$$ L(r, a|e) = - \sum_{s=1}^{S} \log(\text{det}(\Sigma_{c_s})) - \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{S} c_{sf}^H \Sigma_{c_s}^{-1} c_{sf} $$

$$ = - \sum_{s=1}^{S} \log(\text{det}(\Sigma_{c_s})) - \sum_{s=1}^{S} \text{tr}(\Sigma_{c_s}^{-1} \hat{\Sigma}_{c_s}), $$

(6.11)

where we introduce the sample covariance matrix of $c_{sf}$

$$ \hat{\Sigma}_{c_s} = \frac{1}{T} \sum_{t=1}^{T} c_{sf} c_{sf}^H. $$

(6.12)

As previously, we propose to maximize the complete log-likelihood (6.11) via the EM algorithm. Given initial parameter estimates $r^0$ and $a^0$, the procedure then iterates the following E- and M-steps so as to update these estimates.

Given the estimates $r^l, a^l$ obtained in the $l$-th iteration, the E-step in $(l + 1)$-th iteration amounts to computing

$$ Q(\Phi|\Phi^l) = \mathbb{E} \left( L(r, a|e)|p_t, r^l, a^l \right) = - \sum_{s=1}^{S} \log(\text{det}(\Sigma_{c_s})) - \sum_{s=1}^{S} \text{tr} \left( \Sigma_{c_s}^{-1} e_s^l \right), $$

(6.13)

where $e_s^l = \mathbb{E}(\hat{\Sigma}_{c_s}|p_t, r^l, a^l)$ is the conditional expectation of the sample covariance matrix of the latent contribution $c_{sf}$. Below, we detail how this expectation may be computed.
Since the joint vector of pressures \((\mathbf{p}_t, \mathbf{c}_{st})^T\) is Gaussian with 0-mean and covariance matrix 
\[
\begin{pmatrix}
\Sigma_p & \Sigma_c \\
\Sigma_c & \Sigma_c
\end{pmatrix}
\]
, therefore by Theorem 1, the conditional distribution of \(\mathbf{c}_{st}\) given \(\mathbf{p}_t\) is Gaussian:
\[
\mathbf{c}_{st}|\mathbf{p}_t \sim \mathcal{N} \left( \mathbf{c}_{st}|\Sigma_{c} \Sigma_p^{-1} \mathbf{p}_t, \Sigma_c - \Sigma_c \Sigma_p^{-1} \Sigma_c \right). \tag{6.14}
\]

Therefore,
\[
\mathbf{e}_s^t = \mathbf{E}(\hat{\Sigma}_c|\mathbf{p}_t, \mathbf{r}^t, \mathbf{a}^t) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{E}(\mathbf{c}_{st}^t|\mathbf{p}_t, \mathbf{r}^t, \mathbf{a}^t)
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} \mathbf{E} \left( (\mathbf{c}_{st} - \Sigma_c \Sigma_p^{-1} \mathbf{p}_t)(\mathbf{c}_{st} - \Sigma_c \Sigma_p^{-1} \mathbf{p}_t)^H | \mathbf{p}_t, \mathbf{r}^t, \mathbf{a}^t \right) + \frac{1}{T} \sum_{t=1}^{T} \mathbf{E} \left( (\Sigma_c \Sigma_p^{-1} \mathbf{p}_t) \mathbf{c}_{st}^t | \mathbf{p}_t, \mathbf{r}^t, \mathbf{a}^t \right)
\]
\[
- \frac{1}{T} \sum_{t=1}^{T} \left( \Sigma_c \left( \Sigma_p^{-1} \mathbf{p}_t \right) \right) \left( \Sigma_c^{-1} \mathbf{p}_t \right) \right)^H.
\]
\[
= \Sigma_c - \Sigma_c \Sigma_p^{-1} \Sigma_c + \frac{1}{T} \sum_{t=1}^{T} \left( \Sigma_c \left( \Sigma_p^{-1} \mathbf{p}_t \right) \right) \left( \Sigma_c^{-1} \mathbf{p}_t \right) \right)^H,
\]
\[
= \Sigma_c - \Sigma_c \Sigma_p^{-1} \Sigma_c + \left( \Sigma_c \left( \Sigma_p^{-1} \mathbf{p}_t \right) \right)^H \mathbf{I}_M \tag{6.15}
\]
in which
\[
\Sigma_{c}^t = \Sigma_c \left( \mathbf{r}_s^t, (a_s^t)^2 \right) = (a_s^t)^2 \mathbf{G}_s (\mathbf{r}_s^t) \mathbf{G}_s (\mathbf{r}_s^t)^H + \frac{\sigma^2}{S} \mathbf{I}_M \tag{6.16}
\]
and
\[
\Sigma_{p}^t = \mathbf{G}(\mathbf{r}) \text{ diag}((a_1^t)^2, \cdots, (a_S^t)^2) \mathbf{G}(\mathbf{r})^H + \sigma^2 \mathbf{I}_M. \tag{6.17}
\]

The M-step consists in computing new estimates of the model parameters by solving:
\[
\mathbf{r}_s^{t+1}, a_s^{t+1} = \arg \min_{\mathbf{r}_s, a_s} \left( \log(\det(\Sigma_{c})) + \text{tr} \left( \Sigma_{c}^{-1} \mathbf{e}_s^t \right) \right), \tag{6.18}
\]
for \(s = 1, \cdots, S\). To simplify \(\log(\det(\Sigma_{c}))\), we compute the eigen decomposition of \(\Sigma_{c}\) by solving the equation \(\det(\lambda \mathbf{I}_M - \Sigma_{c}) = 0\). One eigenvalue of \(\Sigma_{c}\) is \(\left( \frac{\sigma^2}{S} + a_s^2 \mathbf{G}_s \right)^2 \) with eigenvector \(\mathbf{G}_s\). The remaining \(M - 1\) eigenvalues are all \(\frac{\sigma^2}{S}\), and the corresponding \(M - 1\) eigenvectors can be chosen as a basis for the orthogonal complement of the space \(\mathbf{G}_s\), by solving the
\((M - 1)\)-dimensional linear equation system \(\mathbf{G}_s \mathbf{x} = \mathbf{0}\). Therefore, there exists an \(M\)-by-\(M\) invertible matrix \(Q\) of eigenvectors such that

\[
\Sigma_{e_s} = Q \text{diag} \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2, \frac{\sigma^2}{S}, \ldots, \frac{\sigma^2}{S} \right) Q^{-1},
\]

which gives

\[
\det(\Sigma_{e_s}) = \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right)^M \frac{\sigma^2}{S}.
\]

Furthermore, by the Matrix Inversion Lemma, we have

\[
\Sigma^{-1}_{e_s} = \frac{S}{\sigma^2} \mathbf{I}_M - \frac{S}{\sigma^2} \frac{a_s \mathbf{G}_s \mathbf{G}_s^H}{\frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2}.
\]

Therefore, by Eqs. (6.20) and (6.21), the M-step (6.18) can be written as

\[
\begin{align*}
r_{s+1}^{J+1}, \epsilon_{s+1}^{J+1} &= \arg\min_{r_s, \epsilon_s} \left\{ \log \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right) \left( \frac{\sigma^2}{S} \right)^{M-1} \right\} + \text{tr} \left( \frac{S}{\sigma^2} \left( \mathbf{I}_M - \frac{a_s^2 \mathbf{G}_s \mathbf{G}_s^H}{\frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2} \right) \epsilon_s^J \right) \\
&= \arg\min_{r_s, \epsilon_s} \left\{ \log \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right) - \frac{S}{\sigma^2} \frac{a_s^2 \mathbf{G}_s \epsilon_s^J \mathbf{G}_s}{\frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2} \right\}.
\end{align*}
\]

To simplify further Eq. (6.22), let us define the following function of \(a_s^2\):

\[
g(a_s^2) = \log \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right) - \frac{S}{\sigma^2} \frac{a_s^2 \mathbf{G}_s \epsilon_s^J \mathbf{G}_s}{\frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2}.
\]

Its first and second derivatives can be computed as

\[
g'(a_s^2) = \frac{|\mathbf{G}_s|^2 \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right) - \mathbf{G}_s \epsilon_s^J \mathbf{G}_s}{\left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right)^2}
\]

and

\[
g''(a_s^2) = \frac{|\mathbf{G}_s|^4 \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right)^2 - 2 \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right) |\mathbf{G}_s|^2 \left[ |\mathbf{G}_s|^2 \left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right) - \mathbf{G}_s \epsilon_s^J \mathbf{G}_s \right]}{\left( \frac{\sigma^2}{S} + a_s^2 |\mathbf{G}_s|^2 \right)^4}.
\]
The unique stationary point of \( g(\alpha_s^2) \) is obtained by setting \( g'(\alpha_s^2) = 0 \):

\[
\alpha_s^2 = \frac{G_s^H e_s^i G_s}{|G_s|^4} - \frac{\sigma^2}{S} \frac{1}{|G_s|^2}.
\]

(6.26)

We may check that it corresponds to a minimum for \( g(\alpha_s^2) \):

\[
g''(\alpha_s^2) = \frac{|G_s|^4 \left( \frac{\sigma^2}{S} + \alpha_s^2 |G_s|^2 \right)^2}{\left( \frac{\sigma^2}{S} + \alpha_s^2 |G_s|^2 \right)^4} > 0,
\]

so that \( \alpha_s^2 \) minimizes \( g(\alpha_s^2) \).

Therefore, given any source location estimate \( r_s \), the source strength estimate is

\[
(\alpha_s^2)^{l+1} = \frac{G_s^H e_s^i G_s}{|G_s|^4} - \frac{\sigma^2}{S} \frac{1}{|G_s|^2}.
\]

(6.28)

Let us substitute Eq. (6.28) back into Eq. (6.22), the source location estimate is then obtained by

\[
r_s^{l+1} = \arg \min_{r_s} \left\{ \log \left( \frac{G_s^H e_s^i G_s}{|G_s|^2} \right) - \frac{S}{\sigma^2} \frac{G_s^H e_s^i G_s}{|G_s|^2} \right\}.
\]

(6.29)

The strategy for estimating the model parameters in the random amplitude case is summarized by Algorithm 3.

**Algorithm 3** EM algorithm for sound source localization in the case of random amplitude

For \( l = 0 \), pick starting values for the parameters \( r^0, a^0 \). For \( l \geq 1 \):

**repeat**

- estimate the covariance matrix of the latent source contributions \( e_s^l \) by Eq. (6.15), for \( s = 1, \ldots, S \);
- estimate the new source location \( r_s^{l+1} \), for \( s = 1, \ldots, S \), from Eq. (6.29);
- estimate the new variance of amplitude \( (\alpha_s^2)^{l+1} \), for \( s = 1, \ldots, S \) by substituting Eq. (6.29) back into Eq. (6.28);

**until** the relative increase of the observed data log-likelihood (6.4) is less than a given threshold \( \kappa \):

\[
\frac{L(r^{l+1}, a^{l+1} | p) - L(r^l, a^l | p)}{L(r^l, a^l | p)} < \kappa.
\]

(6.30)

Finally, let us consider the far-field case addressed in [12, 43], in which the optimization problem may be simplified. As is shown in Figure 6.1, since the source plane is far away from the microphone plane, the distance between each source and each microphone can be approximated as a constant \( R \). Therefore, the attenuation part of the Green function may
then be seen as a constant:

\[ G(r_s|r'_m) \approx \frac{e^{ik|r_s-r'_m|}}{4\pi R}, \]  

(6.31)

and thus

\[ ||G(r_s)|| = \frac{\sqrt{M}}{4\pi R} \approx \alpha. \]  

(6.32)

In this case, the M-step (6.22) becomes

\[ r^{l+1}_s, a^{l+1}_s = \arg \min_{r_s, a_s} \left\{ \log \left( \frac{\sigma^2}{S} + a_s^2 \alpha^2 \right) - \frac{S}{\sigma^2} \frac{a^2_s G_s^H \epsilon_s^l G_s}{\sigma^2 + a^2_s \alpha^2} \right\}. \]  

(6.33)

Therefore, the estimates of the sound source locations and of the variances of the strengths may be obtained separately:

\[ r^{l+1}_s = \arg \max_{r_s} G_s^H \epsilon_s^l G_s, \]  

(6.34)

and

\[ a^{l+1}_s = \arg \min_{a_s} \left\{ \log \left( \frac{\sigma^2}{S} + a_s^2 \alpha^2 \right) - \frac{S}{\sigma^2} \frac{a^2_s (G_s^{l+1})^H \epsilon_s^l G_s^{l+1}}{\sigma^2 + a^2_s \alpha^2} \right\} \]

\[ = \frac{(G_s^{l+1})^H \epsilon_s^l G_s^{l+1}}{\alpha^4} - \frac{\sigma^2}{S \alpha^2}. \]  

(6.35)

By comparing Eq. (6.34) with Eq. (2.8), it becomes clear that the source location estimation at each iteration of the EM algorithm amounts to perform a beamforming projection for each source.

Algorithm 4 summarizes the estimation procedure under the far-field assumption.

6.3 Experiments

This section illustrates the behavior of the proposed method via experiments realized on real data. The experimental setup and the distribution of the microphones are the same as those described in Section 3.4. Two sound sources are placed at \( r_1 = (-0.239m, -0.112m, 0.314m) \) and \( r_2 = (0.172m, -0.012m, 0.314m) \). In this experiment, random amplitude signals are played during 60 seconds. The signal is divided into 60 segments and then transformed in the frequency domain using a Discrete Fourier Transform, so that 60 snapshots are obtained in the frequency domain.
Fig. 6.1 Sound source estimation problem (2D case) in the far-field case.

\[ L(r^{l+1}, a^{l+1} | p) - L(r^l, a^l | p) \quad \frac{L(r^l, a^l | p)}{< \kappa.} \]
As explained in Section 3.4, a high number of sources may yield a more accurate estimate of the radiating source surfaces. Therefore, we compare three models, with \( S = 2 \), \( S = 6 \) and \( S = 10 \) point sources, respectively. The estimation procedure detailed in Algorithm 3 is run 100 times with initial parameter values selected at random as follows. For each source, its initial locations are generated according to a 2D uniform distribution having a 60 cm \( \times \) 60 cm support centered on the loudspeaker (half of the sources are related to the first loudspeaker, the remaining ones to the second). The initial standard deviations of the source amplitudes are taken as well at random according to an uniform distribution in a range of \( \pm 3 \text{dB} \). Next, we compare the EM results with those obtained via beamforming and SONAH with Tikhonov regularization. Figures 6.2, 6.3 and 6.4 display the estimation results at \( f = 400\text{Hz} \), \( f = 1200\text{Hz} \) and \( f = 2200\text{Hz} \), respectively. In each figure, (a) shows the 100 EM estimates obtained on the given \( z \)-plane (in all three methods, the \( z \)-coordinates of the sources are assumed to be known in advance) using blue points, and the retained estimate (corresponding to the highest likelihood value) using red cross, for \( S = 2 \). Subfigures (b)-(d) display the sound field reconstructed in the source plane via the EM estimates under the assumption of \( S = 2 \), \( S = 6 \) and \( S = 10 \) respectively. In the two latter cases, although the assumed number of sources is greater than the actual one, the sound fields obtained via the EM algorithm clearly indicate two sound sources. Moreover, Subfigures (e) and (f) show the corresponding sound pressure fields obtained via beamforming and SONAH. Beamforming cannot separate the two sources at low frequency (\( f = 400\text{Hz} \)) and generates several "ghost sources" at high frequency (\( f = 2200\text{Hz} \)). By contrast, SONAH can indicate two sources at high frequency (\( f = 1200, 2200\text{Hz} \)), but cannot clearly separate the two sources at low frequency (\( f = 400\text{Hz} \)). Furthermore, SONAH tends to underestimate the actual source pressure levels, which might be due to its intrinsic regularization.

Given the EM estimates \( \hat{r}_s \) and \( \hat{\alpha}_s^2 \) for \( s = 1, \ldots, S \), we can reconstruct the sound pressure at any position \( \mathbf{r} \) on the microphone plane by

\[
p(\mathbf{r}) = \sum_{s=1}^{S} G(\hat{r}_s, |\mathbf{r}|) \cdot |\hat{\alpha}_s|.
\]

(6.37)

Figures 6.6, 6.7 and 6.8 compare the measured sound field (Subfigure (a)) with the reconstructed sound field (Subfigures (b)-(d)) on the microphone plane obtained with the EM estimates at \( f = 400\text{Hz} \), \( f = 1200\text{Hz} \) and \( f = 2200\text{Hz} \), respectively. As the assumed number of sources increases, the quality of reconstruction is improved as well.
Fig. 6.2 Comparison between EM, beamforming and SONAH at f=400Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
6.3 Experiments

Fig. 6.3 Comparison between EM, beamforming and SONAH at $f=1200\text{Hz}$. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
Fig. 6.4 Comparison between EM, beamforming and SONAH at f=2200Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
Fig. 6.5 Comparison between EM, beamforming and SONAH at f=4300Hz. Black crosses represent the loudspeaker center. Subfigure (a): Estimated source position. Subfigures (b-f): Estimated pressure field on the source plane.
Fig. 6.6 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f=400$Hz.
Fig. 6.7 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f=1200$Hz.
Fig. 6.8 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f = 2200\text{Hz}$. 

Sound Source Estimation Under the Random Amplitude Assumption
6.3 Experiments

Fig. 6.9 Sound field on the microphone plane using (a) measurement (b) EM estimates, $S = 2$ (c) EM estimates, $S = 6$ (d) EM estimates, $S = 10$. $f = 4300Hz$. 

(a) Measurement

(b) EM estimates, $S = 2$

(c) EM estimates, $S = 6$

(d) EM estimates, $S = 10$
6.4 Conclusion

In this chapter, we presented how multiple sources can be estimated when their amplitudes are modeled by Gaussian random variables. Then, the EM algorithm may be used to estimate the locations and variances of the strengths of the sound sources. We provided closed forms for the parameter estimates for both the far-field and the near-field case when the random amplitudes are assumed to be Gaussian. Finally, experiments were conducted on real data, and showed the advantage of using our estimation procedure rather than beamforming or SONAH.
Chapter 7

Sound Source Localization in Presence of Uncertainty: Random Amplitude Case

7.1 Model

This chapter addresses the problem of random signal source localization problem in presence of model uncertainty. As in Chapter 6, the measured pressure vector is described by Eq. (6.1); our purpose is to estimate the locations \( r_s \) and variances \( \sigma_s^2 \) (\( s = 1, \cdots, S \)) of the sources. However, unlike in Chapter 6, we assume here that the microphone locations \( r'_m \) and wavenumber \( k \) in the propagation operator from the \( s \)-th source to the \( m \)-th microphone

\[
G(r_s|r'_m) = \frac{\exp(-j|\mathbf{r}_s - \mathbf{r}'_m|)}{4\pi |\mathbf{r}_s - \mathbf{r}'_m|}
\] 

(7.1)

are uncertain. As in Chapter 5, the uncertainties are described by the variances of the uncertainty sources. In Section 7.2, we use a first order Taylor method to quantify the uncertainty of the measurements according to the uncertainties on the microphone locations and the wavenumber. Thus, the data used in the parameter estimation process can be replaced by contour functions which represent the uncertain knowledge of the measured sound pressures. Finally, Section 7.3 describes how the locations and variances of strengths of sound sources can be estimated using the E2M algorithm.
7.2 Uncertainty Representation and Propagation

7.2.1 Uncertain Data

As in Section 5.2, we investigate here how meta-parameter uncertainties can be represented in the model. As before, the sound propagation model depends on the meta-parameters only through the measured pressure vector \( p_r \). Instead of integrating the uncertainty of these meta-parameters directly in the model, we propose to transpose it to the data at hand: we thus define contour functions \( pl_t(p_r) \), which specify the set of plausible values for the pressures measured at time \( t \).

Here, we detail the case of Gaussian contour functions: the imprecise knowledge of the measured pressure is described by a Gaussian PDF, the expectation of which is given by the measured pressure: \( \mu^*_t = p_r \). Its covariance matrix \( \Sigma^*_t \), which reflects the level of uncertainty to which the measurement is subject, is estimated from the meta-parameter uncertainties by

\[
\Sigma^*_t \approx \mathbb{E}\left( \nabla p_t(\mu_\Theta) \Sigma_\Theta \nabla p_t(\mu_\Theta)^H \right),
\]

(7.2)

where \( \Theta = (\theta_1, \ldots, \theta_J) \) denotes the set of uncertain meta-parameters as before, \( \nabla p_t \) is the gradient with respect to the meta-parameter \( \Theta \), \( \mu_\Theta \) and \( \Sigma_\Theta \) are the expectation and covariance matrix of \( \Theta \). The expectation in Eq. (7.2) is computed with respect to the random variable \( X_t \).

Let us assume that the array center and orientation are uncertain, due to motion of the array, as well as the wavenumber \( k \) due to a change in the temperature, as in Example 3 in Section 5.2. The actual coordinates of the microphones are given by Eqs. (5.10) - (5.13). Assume that partial information on the meta-parameter uncertainties is available in the form of variances \( \sigma^2_{\theta_1}, \sigma^2_{\theta_2}, \sigma^2_{\theta_3}, \sigma^2_{x}, \sigma^2_{y}, \sigma^2_{z} \) and \( \sigma^2_k \). These sources of uncertainty are again assumed as mutually cognitively independent, and having zero mean (except for the wavenumber which has mean \( k_0 \)). In this case, the covariance matrix of the contour function \( \Sigma^* \) can be estimated as follows:

\[
\Sigma^*_t \approx \mathbb{E}\left( \nabla p_t(\mu_\Theta) \Sigma_\Theta \nabla p_t(\mu_\Theta)^H \right) = \sum_{s=1}^{S} \frac{\sigma^2_s}{16\pi^2} Y_s \Sigma_\Theta Y_s^H,
\]

(7.3)

where \( Y_s \) and \( \Sigma_\Theta \) are given by Eqs. (E.9) and (E.18), respectively. The detailed computation of Eq. (7.3) can be found in Appendix E. Then, the meta-parameter uncertainty is quantified by a multivariate Gaussian contour function \( pl_t(p_r) \) with expectation \( \mu^*_t = p_r \) (here \( p_r \) denotes the measurement), and covariance matrix \( \Sigma^*_t \).
7.2.2 Likelihood Function of Uncertain Data

In this section, we detail the computation of the likelihood function of the model parameters given the uncertain data represented by multivariate contour functions. In this case, the counterpart to the certain data log-likelihood (6.4) for a given snapshot \( p_t \) is

\[
\log L(r, a|p_t) = \log \int f(p_t|r, a)p_l(p_t)dp_t.
\]

(7.4)

By Theorem 3,

\[
f(p_t|r, a)p_l(p_t) = \phi (p_t|0, \Sigma_p) \phi (p_t|\mu^*, \Sigma^*)
\]

\[
= \phi \left( p_t \left| \Sigma_p^{-1} + (\Sigma^*)^{-1} \right. \right)^{-1} \left( \Sigma^* \right)^{-1} \mu^*, \left( \Sigma_p^{-1} + (\Sigma^*)^{-1} \right)^{-1} \phi (\mu^*|0, \Sigma_p + \Sigma^*),
\]

(7.5)

where \( \Sigma_p \) is the covariance matrix of the sound pressure vector

\[
\Sigma_p = G \text{diag}(a_1^2, \cdots, a_M^2) G^H + \sigma^2 I_M.
\]

(7.6)

Finally, the likelihood function of the model parameters given the uncertain data corresponding to one snapshot is:

\[
\log L(r, a|p_t) = \log \phi (\mu^*|0, \Sigma_p + \Sigma^*)
\]

\[
= \log(\det(\Sigma_p + \Sigma^*)) - (\mu^*)^H(\Sigma_p + \Sigma^*)^{-1}\mu^*,
\]

(7.7)

and the likelihood of the model given a full record is

\[
\log L(r, a|p_{lp}) = \sum_{t=1}^{T} \log L(r, a|p_t).
\]

(7.8)

7.3 Model Estimation Via the Evidential EM Algorithm

The purpose is to estimate the model parameters by maximizing the likelihood function of the uncertain data \( L(r, a|p_{lp}) \). By introducing the latent variables corresponding to the contribution of the various sources: \( c_t = (c_{1t}, \cdots, c_{St}) \), in which

\[
c_{st} = G_s A_s + n_{st},
\]

(7.9)

the problem can be solved using the E2M algorithm introduced in Section 4.2.3. The E-step and M-step are detailed in the following.
E-step

The log-likelihood function of the complete data \( e = \{ e_{st}, s = 1, \ldots, S, t = 1, \ldots, T \} \) is given by Eq. (6.11). Given the parameter estimates at the \( l \)-th step \( \theta^l, \alpha^l \), the E-step consists in computing the conditional expectation of the complete likelihood. As in Section 6.2, this amounts to computing the expectation of the sample covariance of the latent contribution

\[
\mathbf{e}_t^l = \mathbb{E}(\hat{\Sigma}_e^l | p_l, \mathbf{r}_t^l, \mathbf{a}_t^l) ;
\]

(7.10)

note that this expectation is now computed with respect to the contour function \( p_l \), which we detail in the following.

For any measurable contour function \( p_l(\mathbf{p}_t) \), the probability of \( \mathbf{p}_t \) given its contour function is given by Eq. (5.19). Moreover, in the random amplitude case, the conditional PDF of the latent contribution \( e_{st} \) given the measurement \( \mathbf{p}_t \) can be obtained via Eq. (6.14)

\[
f(e_{st} | \mathbf{p}_t) = \phi \left( e_{st} | \Sigma_e, \Sigma_p^{-1} \mathbf{p}_t, \Sigma_e - \Sigma_e, \Sigma_p^{-1} \Sigma_e \right) .
\]

(7.11)

The probability of the source contribution \( e_{st} \) given the uncertain data \( p_l \) is thus

\[
f(e_{st} | p_l) = \int f(e_{st} | \mathbf{p}_t) f(\mathbf{p}_t | p_l(\mathbf{p}_t)) d\mathbf{p}_t ,
\]

(7.12)

and then

\[
\mathbb{E}(e_{st} e_{st}^H | p_l) = \int e_{st} e_{st}^H f(e_{st} | p_l(\mathbf{p}_t)) d\mathbf{p}_t = \int \int e_{st} e_{st}^H f(e_{st} | \mathbf{p}_t) d\mathbf{e}_{st} f(\mathbf{p}_t | p_l(\mathbf{p}_t)) d\mathbf{p}_t .
\]

(7.13)

Let us rewrite the term \( e_{st} e_{st}^H \):

\[
e_{st} e_{st}^H = (e_{st} - \Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)(e_{st} - \Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)^H + e_{st} (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)^H
\]

\[
+ (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t) e_{st}^H - (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)(\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)^H ,
\]

(7.14)

from which we obtain

\[
\int e_{st} e_{st}^H f(e_{st} | \mathbf{p}_t) d\mathbf{e}_{st} = \Sigma_e - \Sigma_e, \Sigma_p^{-1} \Sigma_e + (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)(\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)^H ,
\]

(7.15)

and therefore

\[
\mathbb{E}(e_{st} e_{st}^H | p_l) = \Sigma_e - \Sigma_e, \Sigma_p^{-1} \Sigma_e + (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t) \int p_t p_t^H f(p_t | p_l(\mathbf{p}_t)) d\mathbf{p}_t (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)^H
\]

\[
= \Sigma_e - \Sigma_e, \Sigma_p^{-1} \Sigma_e + (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t) \int p_l(\mathbf{p}_t) f(\mathbf{p}_t) d\mathbf{p}_t (\Sigma_e, \Sigma_p^{-1} \mathbf{p}_t)^H .
\]

(7.16)
Finally, we obtain

\[
e'_s = \Sigma_{c_s} - \Sigma_{c_s}(\Sigma_p^{-1})^{-1}\Sigma_{c_s} + \Sigma_{c_s}(\Sigma_p^{-1})^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} \int \frac{p_t p_t^H p_t (p_t) \phi(p_t; 0, \Sigma_p)}{\int p_t (p_t) \phi(p_t; 0, \Sigma_p) dp_t} (\Sigma_{c_s} (\Sigma_p^{-1})^{-1})^H \right),
\]

where

\[
\Sigma_{c_s}^l = (a_s^l)^2 \mathbf{G}_s^l (\mathbf{G}_s^l)^H + \frac{\sigma^2}{S} \mathbf{I}_M
\]

and

\[
\Sigma_p^l = \mathbf{G}^l \text{diag}((a_1^l)^2, \cdots, (a_S^l)^2)(\mathbf{G}^l)^H + \sigma^2 \mathbf{I}_M.
\]

Eq. (7.17) gives the computation of the E-step for any form of the contour function \( p_l \).

As before, we may obtain a closed form for \( e'_s \) in the particular case of Gaussian contour functions. Note that the sound pressure \( p_t \) is also Gaussian with PDF \( \phi(p_t; 0, \Sigma_p) \). Then, Corollary 5.3.1 can be used to obtain

\[
\frac{f(p_t) p_t^H}{\int f(p_t) p_t^H dp_t} = \phi(p_t| (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu_t^{*}, (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} \mu_t^{*}).
\]

By rewriting \( p_t p_t^H \) as

\[
p_t p_t^H = \left( p_t - (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu_t^{*} \right) \left( p_t - (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu_t^{*} \right)^H + \left( (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu_t^{*} \right) p_t^H + p_t \left( (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu_t^{*} \right)^H
\]

the last term in Eq. (7.17) can then be obtained by

\[
\frac{\int p_t p_t^H p_t (p_t) f(p_t) dp_t}{\int p_t (p_t) f(p_t) dp_t} = (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} + (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu_t^{*} (\mu_t^{*})^H (\Sigma^*)^{-1} (\Sigma_p^{-1} + (\Sigma^*)^{-1})^{-1}.
\]
By substituting (7.22) back to Eq. (7.17), we finally obtain

\[
e^l_s = \Sigma^{l+1}_c - \Sigma^{l+1}_c (\Sigma^l_p)^{-1} \Sigma^{l+1}_c + \Sigma^{l+1}_c (\Sigma^l_p)^{-1} (\Sigma^l_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^{l+1}_c (\Sigma^l_p)^{-1})^H \\
+ \Sigma^{l+1}_c (\Sigma^l_p)^{-1} \left[ \frac{1}{T} \sum_{t=1}^{T} (\Sigma^l_p^{-1} + (\Sigma^*)^{-1})^{-1} (\Sigma^*)^{-1} \mu^*_t (\mu^*_t)^H (\Sigma^*)^{-1} (\Sigma^l_p^{-1} + (\Sigma^*)^{-1})^{-1} \right] (\Sigma^{l+1}_c (\Sigma^l_p)^{-1})^H.
\]

(7.23)

M-step

The general formulation for the M-step is the same as the EM algorithm (Eq. (6.22)):

\[
r^{l+1}_s, \alpha^{l+1}_s = \arg \min_{r_s, \alpha_s} \left\{ \log \left( \frac{\sigma^2}{S} + \alpha^2_s |G_s|^2 \right) - \frac{S}{\sigma^2} \frac{\alpha^2_s G^H_s e^l_s G_s}{|G_s|^2} \right\}.
\]

(7.24)

Then, similar to the computation in Section 6.2, the estimates of the source locations and the strengths are given by

\[
r^{l+1}_s = \arg \min_{r_s} \left\{ \log \left( \frac{G^H_s e^l_s G_s}{|G_s|^2} \right) - \frac{S}{\sigma^2} \frac{G^H_s e^l_s G_s}{|G_s|^2} \right\}.
\]

(7.25)

and

\[
(\alpha^2_s)^{l+1} = \frac{(G^l_s e^l_s G^l_s)^H}{|G^l_s|^4} - \frac{\sigma^2}{S} \frac{1}{|G^l_s|^2}.
\]

(7.26)

The strategy for estimating the model parameters in the random amplitude case may be summarized by Algorithm 5.

Algorithm 5 E2M algorithm for sound source localization in the case of random amplitude

For \( l = 0 \), pick starting values for the parameters \( r^0, \alpha^0 \). For \( l \geq 1 \):

repeat

\begin{itemize}
  \item estimate the covariance matrix of the latent source contributions \( \epsilon^l_s \) from (7.23);
  \item estimate the new source location \( r^{l+1}_s \), for \( s = 1, \cdots, S \) from Eq. (6.29);
  \item estimate the new variance of amplitude \( (\alpha^2_s)^{l+1} \), for \( s = 1, \cdots, S \) from Eq. (7.26)
\end{itemize}

until the relative increase of the observed data log-likelihood (7.8) is less than a given threshold \( \kappa \):

\[
\frac{L(r^{l+1}, \alpha^{l+1}|p|_p) - L(r^l, \alpha^l|p|_p)}{L(r^l, \alpha^l|p|_p)} < \kappa.
\]

(7.27)

It would be easy to conclude that the E2M result in this section is a generalization of the EM algorithm: when the data are certain, i.e., \( p \) is a Dirac function taking unit impulse
point at the measured pressure, the last term of Eq. (7.17) becomes
\[
\sum_{c_i} \left( \sum_{p_l} \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} p_t P_t^H \left( \sum_{c_i} \left( \sum_{p_l} \right)^{-1} \right)^H,
\]
which means that Eqs. (6.15) and (7.17) are identical.

As in Section 6.2, we may consider the far-field case addressed in [12, 43], in which the attenuation part of the Green function may be ignored:
\[
G(r_s | r'_m) \approx \frac{e^{ik|r_s - r'_m|}}{4\pi R},
\]
where \( R \) is a constant. In this case, the M-step (7.24) becomes
\[
r_{s}^{l+1}, a_{s}^{l+1} = \arg \min_{r_s, a_s} \left\{ \log \left( \frac{\sigma^2}{S} + a_s^2 \alpha^2 \right) - \frac{S}{\sigma^2} \frac{\alpha^2 (G_s^{l+1})^H \epsilon_s^{l+1} G_s}{\alpha^2 + a_s^2 \alpha^2} \right\},
\]
where \( \alpha = \sqrt{\frac{\pi}{4\pi R}} \). Therefore, the estimates of the sound source location and the strength variance may be obtained separately:
\[
r_{s}^{l+1} = \arg \max_{r_s} G_s^H \epsilon_s^{l+1} G_s
\]
and
\[
a_{s}^{l+1} = \arg \min_{a_s} \left\{ \log \left( \frac{\sigma^2}{S} + a_s^2 \alpha^2 \right) - \frac{S}{\sigma^2} \frac{\alpha^2 (G_s^{l+1})^H \epsilon_s^{l+1} G_s}{\alpha^2 + a_s^2 \alpha^2} \right\}
= \frac{(G_s^{l+1})^H \epsilon_s^{l+1} G_s}{\alpha^4} - \frac{\sigma^2}{S \alpha^2}.
\]
Algorithm 6 summarizes the estimation procedure under the far-field assumption.

### 7.4 Experiments

In this section, we present the experiments realized on both simulated and real data to illustrate the advantage of the proposed estimation strategy. As before, the experimental setup and microphone distribution are described in Figure 3.2. The array of 60 microphones is originally located on the plane \( z = 0 \) and the center of the array coincides with the origin. Two sources playing random amplitude signals are placed in front of the microphone array. Uncertainties are introduced in the microphone locations and wavenumber. A motion of the microphone array and a temperature change are applied to the array. This leads to
Algorithm 6 E2M algorithm for far-field sound source localization for random amplitude signals

For $l = 0$, pick starting values for the parameters $r^0, a^0$. For $l \geq 1$: 
repeat
estimate the covariance matrix of the latent source contributions $c^l_s$ by (7.23);
estimate the new source location $r^{l+1}_s$ by (7.31), for $s = 1, \ldots, S$;
estimate the new variance of source amplitude $(a^l_s)^{l+1}$ by (7.32), for $s = 1, \ldots, S$;
until the relative increase of the observed data log-likelihood (7.8) is less than a given threshold $\kappa$:

$$\frac{L(r^{l+1}, a^{l+1}|p|_p) - L(r^l, a^l|p|_p)}{L(r^l, a^l|p|_p)} < \kappa. \quad (7.33)$$

variations, i.e., uncertainties, in the center of the array, in the orientation of the array, and in the wavenumber of the propagating waves. Sections 7.4.1 and 7.4.2 provide the results for simulated and real data respectively.

7.4.1 Simulated Data

In this simulated experiment, two sound sources are considered with amplitude standard deviation $a_1 = a_2 = 0.5$ Pa and locations $r_1 = (-0.2m, 0.1m, 0.2m), r_2 = (0.2m, -0.1m, 0.2m)$; the studied frequency is $f = 5$ kHz. Then, uncertainties in the microphone locations and in the wavenumber are introduced as in Section 5.4.1. More specifically, a wavenumber value is generated according to a Gaussian distribution with mean $\mu_k = \frac{2\pi f}{c}$ (sound velocity $c = 340$ m/s) and a predefined standard deviation $\sigma_k$. Similarly, a Gaussian distributed noise is introduced in the array center location $r_0$ with zero mean and predefined standard deviations $\sigma_x, \sigma_y, \sigma_z$ (for the $x, y, z$-coordinates of $r_0$, respectively), and in each rotation angles of the array $\theta_p (p = 1, 2, 3)$ with 0-mean and predefined standard deviation $\sigma_{\theta_p}$. We then generate pressures according to Eqs. (6.1) and (5.10), in which we set the measurement noise to be $\sigma = 0.01$.

For a given dataset, both the EM-based and the E2M-based procedures (defined by Algorithm 3 and Algorithm 5, respectively) are run using five different starting values, retaining the solution with the highest log-likelihood. The covariance matrix $\Sigma^*$ of the contour functions used in E2M is computed using the parameters $r, a$ estimated via the EM algorithm. Since the data are randomly generated, the above procedure (from data generation to model estimation) is repeated 30 times, so that the mean square error (MSE) of the parameter values and 95% confidence intervals of the square error may be computed.

First, let $\sigma_x = \sigma_y = \sigma_z = 0.02$m, $\sigma_{\theta_p} = \frac{10\pi}{180}$ for $p = 1, 2, 3$, $\sigma_k = 0.5$m$^{-1}$. Figure 7.1
shows the 30 estimation results for the x- and y-coordinates of the sources via the EM (blue circles) and the E2M (red crosses) algorithms, as well as the 95% confidence ellipses of the estimated source location. Similarly, Figure 7.2 displays the same result obtained with \( \sigma_{x_0} = \sigma_{y_0} = 0.03m, \sigma_{\theta_p} = 15 \frac{\pi}{180} \) for \( p = 1, 2, 3, \sigma_k = 0.5m^{-1} \). The estimates obtained via the E2M procedure clearly exhibit a smaller spread around the actual sources locations than the estimates obtained via EM, which illustrates that E2M can improve the estimation accuracy when the data are pervaded with uncertainties.

Then, we investigate the estimation error of both algorithms. Let \( \sigma_k = 0 \) (the wavenumber is certain), and let the degree of uncertainty on the microphone locations increase: \( \sigma_{x_0} = \sigma_{y_0} = 0.01l \) m and \( \sigma_{\theta_l} = 5l \frac{\pi}{180} (p = 1, 2, 3) \), for \( l = 0, 1, 2, 3, 4 \). The MSE of the estimated sound source locations and 95% confidence intervals of the square errors are displayed in Figure 7.3 (a). Alternatively, let \( \sigma_x = \sigma_y = \sigma_z = \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 0 \) (the microphone locations are precise) and let \( \sigma_k \) increase from 0 to 1. Figure 7.3 (b) shows the corresponding estimation error trend. In both cases, when the level of uncertainty is low, the EM and E2M algorithms exhibit similar performance. As the amount of uncertainty increases, the accuracy of the estimates obtained using both algorithms decreases. However, E2M proves to be much more robust to uncertainty than EM, its estimation error staying under an acceptable level while the estimation error of the EM estimate increases dramatically.

### 7.4.2 Real Experiment

In this experiment, two random amplitude sound sources \( r_1 = (-0.239m, -0.112m, 0.314m) \) and \( r_2 = (0.172m, -0.012m, 0.314m) \) are considered; the studied frequency is \( f = 2200Hz \). The microphone array is located on the plane \( z = 0 \) and the center of the array coincides with the origin. The distribution of the microphones is shown in Figure 3.2(b). However, in order to observe the how EM and E2M perform when the model is pervaded with uncertainties, we assume that the microphone locations are not precisely known. The microphone locations used in the estimation process are randomly generated by Eq. (5.10), in which \( \theta_1, \theta_2, \theta_3, \alpha_0, \beta_0, \gamma_0 \) are generated by Gaussian random numbers with 0-mean and variances \( \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 15\pi/180, \sigma_{\alpha_0} = \sigma_{\beta_0} = \sigma_{\gamma_0} = 0.03m \), respectively. Similarly, we assume the wavenumber is imprecisely known as well: the wavenumber used in the estimation process is generated by a Gaussian random number with mean \( \mu_k = 40.4m^{-1} \) and variance \( \sigma_k = 0.5m^{-1} \).

Then, we can estimate the source locations using EM (Algorithm 3) and E2M (Algo-
Fig. 7.1 Random amplitude signal source location estimates obtained using the EM (blue) and E2M (red) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. \( \sigma_{x_0} = \sigma_{y_0} = \sigma_{x_2} = \sigma_{y_2} = 0.02 \text{m} \), \( \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 10 \frac{\pi}{180} \), \( \sigma_k = 0.5m^{-1} \).

Algorithm 5) respectively, each using 100 initial values obtained as follows. The initial values for each source location are generated according to a three-dimensional uniform distribution on a 20cm \( \times \) 20cm \( \times \) 10cm cube. The initial values for the amplitude variances of the sources are generated according to a complex-valued uniform distribution so that the sound pressure uncertainty is \( \pm 3 \) dB. In order to observe the estimation error for both methods, we repeat the experiment 30 times, from generating the microphone locations and the wavenumber to the model parameter estimation. The MSE of the 30 source locations estimates are 0.0819 using EM and 0.0276 using E2M, and the corresponding 95% confidence intervals of square error are \([0.0625, 0.1012]\) for EM and \([0.0257, 0.0294]\) for E2M. Furthermore, Figure 7.4 displays the x- and y-coordinate estimates along with 95% confidence ellipses of the sources using EM and E2M. Figure 7.5 shows the 3D source location estimates and 95% confidence ellipsoids. Obviously, the E2M estimates exhibit a smaller spread over the actual source locations than the EM estimates. This shows the advantage of taking into account the uncertainty in the random signal model.
Fig. 7.2 Random amplitude signal source location estimates obtained using the EM (blue) and E2M (red) algorithms, and corresponding 95% confidence ellipses. Black crosses stand for the actual locations of the sources. $\sigma_{x_0} = \sigma_{y_0} = \sigma_{z_0} = 0.03m$, $\sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 15 \frac{\pi}{180}$, $\sigma_k = 0.5m^{-1}$.

7.5 Conclusion

In this chapter, we considered the multiple sound source estimation for random signals in the presence of model and data uncertainty. The uncertainties of the microphone locations and the wavenumber were considered and integrated to the measured data. We quantified the uncertainties using uncertain data in the form of contour functions. Then, the locations and variances of random strengths were estimated using the E2M algorithm. Finally, both the simulated and real experiments showed the interest of the E2M algorithm when the data at hand are pervaded with uncertainties.
Fig. 7.3 MSE and 95% confidence intervals for the (random amplitude) source location estimates. Case (a): $\sigma_k = 0$, $\sigma_{x_0} = \sigma_{y_0} = \sigma_{\theta_0} = 0.01l$ m, $\sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 5l/100$, $l = 0, 1, 2, 3, 4$. Case (b): $\sigma_{x_0} = \sigma_{y_0} = \sigma_{\theta_0} = \sigma_{\theta_1} = \sigma_{\theta_2} = \sigma_{\theta_3} = 0$, $\sigma_k = 0, 0.25, 0.5, 0.75, 1m^{-1}$.

Fig. 7.4 Random amplitude source location estimates obtained using EM (blue points) and E2M (red crosses), and corresponding 95% confidence ellipses on the x-y plane. Black crosses represent the actual source locations.
Fig. 7.5 Random amplitude source location estimates obtained using EM (left) and E2M (right), and corresponding 95% confidence ellipsoids in 3D space. Black crosses represent the source locations.
Chapter 8

Conclusion and Perspectives

8.1 Conclusion

This thesis considers the problem of multiple sound source localization using sound pressures measured by an array of microphones. Part I solves this problem for deterministic signals via maximum likelihood. The pressure measured by a microphone is interpreted as a mixture of latent signals emitted by the sources; then, both the sound source locations and the amplitudes are estimated using the EM algorithm. At each step of the procedure, the latent contributions of the sound sources can be estimated given current model parameter estimation; these new estimates are then used to update the model parameters. These two steps are repeated until a local maximum of the likelihood is attained.

In Part II, we consider the issue of uncertain measurements. While randomness is identified as the noise inherent to the measurement process, uncertainty is related to parameters of the model, generally assumed to be known, but that can be subject to errors. In particular, we consider two kinds of meta-parameters: the microphone locations and the wavenumber. In this work, the uncertainties on the meta-parameters are transposed to the data, via first-order approximations. The resulting uncertain pressures are then represented using contour functions. These mathematical objects, proposed within the framework of belief functions, provide a rich and flexible way of quantifying imprecise and uncertain knowledge of imperfectly observed variables. Eventually, model estimation can be carried out using the E2M algorithm, which was recently proposed to perform maximum likelihood estimation from uncertain data.

Part III considers the case of random signals, in which the parameters to estimate are the source locations and the variances of the amplitudes. As in Part I and Part II, we propose to estimate the parameters via the EM algorithm when the measurements are certain or via
the E2M algorithm when uncertainty pervades the model parameters. Explicit equations are provided for both algorithms.

In both the deterministic and random amplitude cases, experiments are realized on synthetic and real data. When the data are certain, the EM algorithm performs better than beamforming and SONAH that it clearly separates the sources and works well over a wide range of frequencies. On the other hand, when the data are pervaded with uncertainties, the E2M estimates display a smaller estimation error than the EM estimates.

8.2 Perspectives

Further work may be conducted in several directions.

First, in both EM and E2M algorithms, the number of sound sources must be pre-defined. As is shown in Sections 3.4 and 6.3, we may set a large source number before running the EM algorithm and then distinguish the number of sources by eyes. However, we may define a criterion to determine the number of sources. For example, we can remove those sources with low estimated strength using regularization methods.

Second, the meta-parameter uncertainties are modeled using Gaussian contour functions in Chapters 5 and 7, an assumption which may not always hold. Note that the proposed E2M method makes it possible to use any kind of contour function: simpler distributions, such as uniform or trapezoidal ones, may give better results when few is known about the data. However, for general contour functions, closed forms for the model parameter estimates may not be obtained, and numerical techniques (such as Monte-Carlo integrals) may have to be considered.
References


Appendix A

Complex Gaussian Random Variables

Let \( \mathbf{X} \) and \( \mathbf{Y} \) be Gaussian distributed random vectors in \( \mathbb{R}^k \). Then we say that the complex random vector

\[
\mathbf{Z} = \mathbf{X} + j\mathbf{Y}
\]

(A.1)

has a complex Gaussian distribution. This distribution can be described by 3 parameters: the expectation \( \mu \), the covariance matrix \( \Sigma \) and the relation matrix \( \mathbf{R} \):

\[
\mu = \mathbb{E}(\mathbf{Z}), \quad \Sigma = \mathbb{E}[(\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^H], \quad \mathbf{R} = \mathbb{E}[(\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^T].
\]

(A.2)

Here the location parameter \( \mu \) can be an arbitrary \( k \)-dimensional complex vector, the covariance matrix \( \Sigma \) must be Hermitian and non-negative definite, and the relation matrix \( \mathbf{R} \) should be symmetric. Moreover, the matrices \( \Sigma \) and \( \mathbf{R} \) are such that the matrix

\[
\mathbf{P} = \Sigma^{-c} - \mathbf{R}^H \Sigma^{-1} \mathbf{R}
\]

(A.3)

is also Hermitian and non-negative definite.

Let \( \mathbf{v} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \), which is a \( 2k \)-dimensional real-valued Gaussian with expectation and covariance matrix

\[
\mu_v = \begin{pmatrix} \mathbb{E}(\mathbf{X}) \\ \mathbb{E}(\mathbf{Y}) \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \Sigma_v = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix},
\]

(A.4)
in which

\[
\Sigma_{xx} = \mathbb{E} \left[ (X - \mu_x)(X - \mu_x)^T \right] = \frac{1}{2} \text{Re}(\Sigma + R); \\
\Sigma_{xy} = \mathbb{E} \left[ (X - \mu_x)(Y - \mu_y)^T \right] = \frac{1}{2} \text{Im}(\Sigma + R); \\
\Sigma_{yx} = \mathbb{E} \left[ (Y - \mu_y)(X - \mu_x)^T \right] = \frac{1}{2} \text{Im}(\Sigma + R); \\
\Sigma_{yy} = \mathbb{E} \left[ (Y - \mu_y)(Y - \mu_y)^T \right] = \frac{1}{2} \text{Re}(\Sigma - R).
\]

The PDF of the complex-valued Gaussian random vector \( \mathbf{Z} \) may be defined as the PDF of the real-valued random vector \( \mathbf{v} \)

\[
f_v(\mathbf{X}, \mathbf{Y}) = (2\pi)^{-k} \left| \text{det}(\Sigma_v) \right|^{-\frac{1}{2}} e^{-\frac{1}{2}((\mathbf{v} - \mu_v)^T \Sigma^{-1} \mathbf{v})}. \tag{A.9}
\]

As noted in [67], \( \mathbf{X} \) and \( \mathbf{Y} \) in (A.9) can be expressed in terms of \( \mathbf{Z} = \mathbf{X} + j\mathbf{Y} \) and \( \mathbf{Z}^c = \mathbf{X} - j\mathbf{Y} \), which introduces another PDF form. Let \( \mathbf{w} = \begin{pmatrix} \mathbf{Z} \\ \mathbf{Z}^c \end{pmatrix} \), which has expectation

\[
\mu_w = \begin{pmatrix} \mu \\ \mu^c \end{pmatrix}
\]

and covariance matrix \( \Sigma_w = \begin{pmatrix} \Sigma & R \\ R^H & \Gamma^c \end{pmatrix} \), the PDF of \( \mathbf{Z} \) can be expressed as

\[
f(\mathbf{Z}) = f_v(\mathbf{X}, \mathbf{Y}) = f_w(\mathbf{Z}, \mathbf{Z}^c)
\]

\[
= (2\pi)^{-k} \left| \text{det}(\Sigma_w) \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{w} - \mu_w)^T \Sigma_w^{-1} (\mathbf{w} - \mu_w)}
\]

\[
= \pi^{-k} |\text{det}(\Gamma)| \left| \text{det}(P) \right|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} ((\mathbf{Z} - \mu)^H, (\mathbf{Z} - \mu)^T) \begin{pmatrix} \Gamma & R \\ R^H & \Gamma^c \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z} - \mu \\ \mathbf{Z}^c - \mu^c \end{pmatrix} \right)
\]

\[
= \pi^{-k} |\text{det}(\Gamma)| \left| \text{det}(P) \right|^{-\frac{1}{2}} e^{-\frac{1}{2}[(\mathbf{Z} - \mu)^H \Gamma^{-1} (\mathbf{Z} - \mu) + \text{Re}[(\mathbf{Z} - \mu)^T (\Gamma^T)^{-1} \Gamma R^c (P^{-1})^c (\mathbf{Z} - \mu)]]}.
\]  \tag{A.10}

The derivation of Eq. (A.10) may be found in [53].

A case of particular importance: the Circular Symmetric complex Gaussian distribution, which appears when \( R = 0 \). In this case, the PDF of \( Z \) becomes

\[
f(\mathbf{Z}) = \frac{1}{\pi^k |\text{det}(\Gamma)|} e^{-\frac{1}{2}[(\mathbf{Z} - \mu)^H \Sigma^{-1} (\mathbf{Z} - \mu)].
\]  \tag{A.11}

Unless otherwise specified, the complex-valued Gaussian distribution is always assumed as Circular Symmetric.
Appendix B

Output Variance Approximated By Taylor Expansion

In this appendix, we remind the variance propagation method [13]. Assume that $X = (X_1, \ldots, X_r)$ consists of $r$ independent real-valued random variables with expectation $\mu_x = (\mu_1, \ldots, \mu_r)$ and covariance matrix $\Sigma_x$. Furthermore, we assume that

$$f(x_1, \cdots, x_r) = (f_1(x_1, \cdots, x_r), \cdots, f_M(x_1, \cdots, x_r))^T$$

is a twice differentiable $M$-dimensional complex-valued function. By 1-order Taylor expansion at $\mu$, we have

$$f(x_1, \cdots, x_r) \approx f(\mu_1, \cdots, \mu_r) + \nabla f(\mu_1, \cdots, \mu_r) \begin{pmatrix} x_1 - \mu_1 \\ \vdots \\ x_r - \mu_r \end{pmatrix},$$

(B.2)
where the gradient $\nabla f = \left( \begin{array}{c} \frac{\partial f_1}{\partial x_1} \\ \vdots \\ \frac{\partial f_M}{\partial x_1} \\ \frac{\partial f_1}{\partial x_r} \\ \vdots \\ \frac{\partial f_M}{\partial x_r} \end{array} \right)$. Then, the covariance matrix of $f(X_1, \cdots, X_r)$, denoted as $\Sigma_f$, may be obtained by

$$
\Sigma_f = E \left( [f(X_1, \cdots, X_r) - f(\mu_1, \cdots, \mu_r)][f(X_1, \cdots, X_r) - f(\mu_1, \cdots, \mu_r)]^H \right)
\approx E \left( \nabla f(\mu_1, \cdots, \mu_r) \begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_r - \mu_r \end{pmatrix} (X_1 - \mu_1, \cdots, X_r - \mu_r) \nabla f(\mu_1, \cdots, \mu_r)^H \right)
= \nabla f(\mu_x) \Sigma_x \nabla f(\mu_x)^H. \quad (B.3)
$$
Appendix C

Computation of the Gradient Matrix in Section 5.2

This appendix derives the gradient matrix $\nabla p_t(\mu_\Theta)$ in Eq. (5.14). First let the coordinate of the array center be $r_0 = (x_0, y_0, z_0)$, the (unknown) actual microphones coordinates be $r'_m = (x'_m, y'_m, z'_m)$, the assumed microphones coordinates be $r^*_m = (x^*_m, y^*_m, 0)$, the sources coordinates be $r_s = (x_s, y_s, z_s)$ and $r_{ms} = |r_s - r'_m|$. The partial derivatives of the measurement $p_{mt}$ with respect to the meta-parameter vector $\Theta = (\theta_1, \theta_2, \theta_3, x_0, y_0, z_0, k)$ are computed as

$$\frac{\partial p_{mt}}{\partial k} = \sum_{s=1}^{S} \frac{jA_s}{4\pi} e^{jkr_{ms}}, \quad (C.1)$$

and

$$\frac{\partial p_{mt}}{\partial \theta_1} = \frac{\partial p_{mt}}{\partial x'_m} \frac{\partial x'_m}{\partial \theta_1} + \frac{\partial p_{mt}}{\partial y'_m} \frac{\partial y'_m}{\partial \theta_1} + \frac{\partial p_{mt}}{\partial z'_m} \frac{\partial z'_m}{\partial \theta_1}, \quad (C.2)$$

$$\frac{\partial p_{mt}}{\partial \theta_2} = \frac{\partial p_{mt}}{\partial x'_m} \frac{\partial x'_m}{\partial \theta_2} + \frac{\partial p_{mt}}{\partial y'_m} \frac{\partial y'_m}{\partial \theta_2} + \frac{\partial p_{mt}}{\partial z'_m} \frac{\partial z'_m}{\partial \theta_2}, \quad (C.3)$$

$$\frac{\partial p_{mt}}{\partial \theta_3} = \frac{\partial p_{mt}}{\partial x'_m} \frac{\partial x'_m}{\partial \theta_3} + \frac{\partial p_{mt}}{\partial y'_m} \frac{\partial y'_m}{\partial \theta_3} + \frac{\partial p_{mt}}{\partial z'_m} \frac{\partial z'_m}{\partial \theta_3}, \quad (C.4)$$

$$\frac{\partial p_{mt}}{\partial x_0} = \frac{\partial p_{mt}}{\partial x'_m} \frac{\partial x'_m}{\partial x_0} + \frac{\partial p_{mt}}{\partial y'_m} \frac{\partial y'_m}{\partial x_0} + \frac{\partial p_{mt}}{\partial z'_m} \frac{\partial z'_m}{\partial x_0}, \quad (C.5)$$
\[
\frac{\partial p_{nt}}{\partial y_0} = \frac{\partial p_{nt}}{\partial x'_m} \frac{\partial x'_m}{\partial y_0} + \frac{\partial p_{nt}}{\partial y'_m} \frac{\partial y'_m}{\partial y_0} + \frac{\partial p_{nt}}{\partial z'_m} \frac{\partial z'_m}{\partial y_0},
\]
(C.6)

\[
\frac{\partial p_{nt}}{\partial z_0} = \frac{\partial p_{nt}}{\partial x'_m} \frac{\partial x'_m}{\partial z_0} + \frac{\partial p_{nt}}{\partial y'_m} \frac{\partial y'_m}{\partial z_0} + \frac{\partial p_{nt}}{\partial z'_m} \frac{\partial z'_m}{\partial z_0},
\]
(C.7)
in which

\[
\frac{\partial p_{nt}}{\partial x'_m} = \sum_{s=1}^{S} A_s \frac{jk}{4\pi} \left[ \frac{1}{r_{ms}^{2}} \left( e^{jk_{ms}} - e^{-jk_{ms}} \right) \right] \frac{x_s - x'_m}{r_{ms}},
\]
(C.8)

\[
\frac{\partial p_{nt}}{\partial y'_m} = \sum_{s=1}^{S} A_s \frac{jk}{4\pi} \left[ \frac{1}{r_{ms}^{2}} \left( e^{jk_{ms}} - e^{-jk_{ms}} \right) \right] \frac{y_s - y'_m}{r_{ms}},
\]
(C.9)

\[
\frac{\partial p_{nt}}{\partial z'_m} = \sum_{s=1}^{S} A_s \frac{jk}{4\pi} \left[ \frac{1}{r_{ms}^{2}} \left( e^{jk_{ms}} - e^{-jk_{ms}} \right) \right] \frac{z_s - z'_m}{r_{ms}},
\]
(C.10)
and

\[
\frac{\partial x'_m}{\partial \theta_1} = 0,
\]
(C.11)

\[
\frac{\partial y'_m}{\partial \theta_1} = \left( -\cos \theta_1 \cos \theta_2 \cos \theta_3 - \sin \theta_1 \sin \theta_3 \right) (x'_m - x_0)
+ (\cos \theta_1 \sin \theta_2 \sin \theta_3 - \sin \theta_1 \cos \theta_3) (y'_m - y_0) + \cos \theta_1 \cos \theta_2 z_0,
\]
(C.12)

\[
\frac{\partial z'_m}{\partial \theta_1} = \left( -\sin \theta_1 \sin \theta_2 \cos \theta_3 + \cos \theta_2 \sin \theta_3 \right) (x'_m - x_0)
+ (\sin \theta_1 \sin \theta_2 \sin \theta_3 + \cos \theta_1 \cos \theta_3) (y'_m - y_0) + \sin \theta_1 \cos \theta_2 z_0,
\]
(C.13)

\[
\frac{\partial x'_m}{\partial \theta_2} = -\sin \theta_2 \cos \theta_3 (\alpha'_m - \alpha_0) + \sin \theta_2 \sin \theta_3 (\beta'_m - \beta_0) + \cos \theta_2 \gamma_0,
\]
(C.14)

\[
\frac{\partial y'_m}{\partial \theta_2} = -\sin \theta_1 \cos \theta_2 \cos \theta_3 (x'_m - x_0)
+ \sin \theta_1 \cos \theta_2 \sin \theta_3 (y'_m - y_0) - \sin \theta_1 \sin \theta_2 z_0,
\]
(C.15)
\[
\frac{\partial z'_m}{\partial \theta_2} = \cos \theta_1 \cos \theta_2 \cos \theta_3 (x'_m - x_0) \\
- \cos \theta_1 \cos \theta_2 \sin \theta_3 (y'_m - y_0) + \cos \theta_1 \sin \theta_2 z_0,
\]

(C.16)

\[
\frac{\partial x'_m}{\partial \theta_3} = -\cos \theta_2 \sin \theta_3 (x'_m - x_0) - \cos \theta_2 \cos \theta_3 (y'_m - y_0),
\]

(C.17)

\[
\frac{\partial y'_m}{\partial \theta_3} = (\sin \theta_1 \sin \theta_2 \sin \theta_3 + \cos \theta_1 \cos \theta_3) (x'_m - x_0) \\
+ (\sin \theta_1 \sin \theta_2 \cos \theta_3 - \cos \theta_1 \sin \theta_2) (y'_m - y_0),
\]

(C.18)

\[
\frac{\partial z'_m}{\partial \theta_3} = (- \cos \theta_1 \sin \theta_2 \sin \theta_3 + \sin \theta_1 \cos \theta_3) (x'_m - x_0) \\
+ ( - \cos \theta_1 \sin \theta_2 \cos \theta_3 - \sin \theta_1 \sin \theta_3) (y'_m - y_0),
\]

(C.19)

\[
\frac{\partial x'_m}{\partial x_0} = -\cos \theta_2 \cos \theta_3,
\]

(C.20)

\[
\frac{\partial y'_m}{\partial x_0} = \sin \theta_1 \sin \theta_2 \cos \theta_3 - \cos \theta_1 \sin \theta_3,
\]

(C.21)

\[
\frac{\partial z'_m}{\partial x_0} = - (\cos \theta_1 \sin \theta_2 \cos \theta_3 + \sin \theta_1 \sin \theta_3),
\]

(C.22)

\[
\frac{\partial x'_m}{\partial y_0} = \cos \theta_2 \sin \theta_3,
\]

(C.23)

\[
\frac{\partial y'_m}{\partial y_0} = - (\sin \theta_1 \sin \theta_2 \sin \theta_3 + \cos \theta_1 \cos \theta_3),
\]

(C.24)

\[
\frac{\partial z'_m}{\partial y_0} = \cos \theta_1 \sin \theta_2 \sin \theta_3 - \sin \theta_1 \cos \theta_3,
\]

(C.25)

\[
\frac{\partial x'_m}{\partial z_0} = \sin \theta_2,
\]

(C.26)
\[
\frac{\partial y'_m}{\partial z_0} = \sin \theta_1 \cos \theta_2, \tag{C.27}
\]

\[
\frac{\partial z'_m}{\partial z_0} = -\cos \theta_1 \cos \theta_2. \tag{C.28}
\]

Then, we obtain

\[
\frac{\partial x'_m}{\partial \theta_1} (\mu_{\Theta}) = 0, \quad \frac{\partial y'_m}{\partial \theta_1} (\mu_{\Theta}) = 0, \quad \frac{\partial z'_m}{\partial \theta_1} (\mu_{\Theta}) = y'_m, \tag{C.29}
\]

\[
\frac{\partial x'_m}{\partial \theta_2} (\mu_{\Theta}) = 0, \quad \frac{\partial y'_m}{\partial \theta_2} (\mu_{\Theta}) = 0, \quad \frac{\partial z'_m}{\partial \theta_2} (\mu_{\Theta}) = x'_m, \tag{C.30}
\]

\[
\frac{\partial x'_m}{\partial \theta_3} (\mu_{\Theta}) = -y'_m, \quad \frac{\partial y'_m}{\partial \theta_3} (\mu_{\Theta}) = x'_m, \quad \frac{\partial z'_m}{\partial \theta_3} (\mu_{\Theta}) = 0, \tag{C.31}
\]

\[
\frac{\partial x'_m}{\partial x_0} (\mu_{\Theta}) = -1, \quad \frac{\partial y'_m}{\partial x_0} (\mu_{\Theta}) = 0, \quad \frac{\partial z'_m}{\partial x_0} (\mu_{\Theta}) = 0, \tag{C.32}
\]

\[
\frac{\partial x'_m}{\partial y_0} (\mu_{\Theta}) = 0, \quad \frac{\partial y'_m}{\partial y_0} (\mu_{\Theta}) = -1, \quad \frac{\partial z'_m}{\partial y_0} (\mu_{\Theta}) = 0, \tag{C.33}
\]

\[
\frac{\partial x'_m}{\partial z_0} (\mu_{\Theta}) = 0, \quad \frac{\partial y'_m}{\partial z_0} (\mu_{\Theta}) = 0, \quad \frac{\partial z'_m}{\partial z_0} (\mu_{\Theta}) = -1. \tag{C.34}
\]

Therefore, let \( r'^{s}_{ms} = \sqrt{(x_s - x'_m)^2 + (y_s - y'_m)^2 + z_s^2} \), we have

\[
\frac{\partial p_{mt}}{\partial \theta_1} (\mu_{\Theta}) = y'_m \frac{\partial p_{mt}}{\partial z'_m} (\mu_{\Theta})
= y'_m \sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r'^{s}_{ms}} e^{j\mu_k r'^{s}_{ms}} - \frac{1}{(r'^{s}_{ms})^2} e^{-j\mu_k r'^{s}_{ms}} \right] \frac{z_s}{r'^{s}_{ms}}, \tag{C.35}
\]

\[
\frac{\partial p_{mt}}{\partial \theta_2} (\mu_{\Theta}) = x'_m \frac{\partial p_{mt}}{\partial z'_m} (\mu_{\Theta})
= x'_m \sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r'^{s}_{ms}} e^{j\mu_k r'^{s}_{ms}} - \frac{1}{(r'^{s}_{ms})^2} e^{-j\mu_k r'^{s}_{ms}} \right] \frac{z_s}{r'^{s}_{ms}}. \tag{C.36}
\]
\[
\begin{align*}
\frac{\partial p_{nl}}{\partial \theta}(\mu_\Theta) &= -y_m^* \frac{\partial p_{nl}}{\partial x_m^*}(\mu_\Theta) + x_m^* \frac{\partial p_{nl}}{\partial y_m^*}(\mu_\Theta) \\
&= -y_m^* \sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^*} \right] e^{\frac{1}{(r_{ms}^*)^2}}e^{-j\mu_k r_{ms}^*} \left[ \frac{x_s - x_m^*}{r_{ms}^*} \right] \\
&+ x_m^* \sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^*} \right] e^{\frac{1}{(r_{ms}^*)^2}}e^{-j\mu_k r_{ms}^*} \left[ \frac{y_s - y_m^*}{r_{ms}^*} \right], \\
(C.37)
\end{align*}
\]

\[
\begin{align*}
\frac{\partial p_{nl}}{\partial x_0}(\mu_\Theta) &= -\frac{\partial p_{nl}}{\partial x_m^*}(\mu_\Theta) \\
&= -\sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^*} \right] e^{\frac{1}{(r_{ms}^*)^2}}e^{-j\mu_k r_{ms}^*} \left[ \frac{x_s - x_m^*}{r_{ms}^*} \right], \\
(C.38)
\end{align*}
\]

\[
\begin{align*}
\frac{\partial p_{nl}}{\partial y_0}(\mu_\Theta) &= -\frac{\partial p_{nl}}{\partial y_m^*}(\mu_\Theta) \\
&= -\sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^*} \right] e^{\frac{1}{(r_{ms}^*)^2}}e^{-j\mu_k r_{ms}^*} \left[ \frac{y_s - y_m^*}{r_{ms}^*} \right], \\
(C.39)
\end{align*}
\]

\[
\begin{align*}
\frac{\partial p_{nl}}{\partial z_0}(\mu_\Theta) &= -\frac{\partial p_{nl}}{\partial z_m^*}(\mu_\Theta) \\
&= -\sum_{s=1}^{S} \frac{A_s}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^*} \right] e^{\frac{1}{(r_{ms}^*)^2}}e^{-j\mu_k r_{ms}^*} \left[ \frac{z_s}{r_{ms}^*} \right], \\
(C.40)
\end{align*}
\]

and

\[
\frac{\partial p_{nl}}{\partial k}(\mu_\Theta) = \sum_{s=1}^{S} \frac{jA_s}{4\pi} e^{\mu_k r_{ms}^*}, \\
(C.41)
\]

in which \(r_{ms}^* = \sqrt{(x_s - x_m^*)^2 + (y_s - y_m^*)^2 + z_s^2} \). Denote \(\frac{\partial p}{\partial \mu_\Theta} = \left( \frac{\partial p_1}{\partial \mu_\Theta}, \ldots, \frac{\partial p_M}{\partial \mu_\Theta} \right)^T\), the gradient matrix is obtained

\[
\nabla p_i(\mu_\Theta) = \left( \frac{\partial p_1}{\partial \theta_1}(\mu_\Theta), \frac{\partial p_i}{\partial \theta_2}(\mu_\Theta), \frac{\partial p_i}{\partial \theta_3}(\mu_\Theta), \frac{\partial p_i}{\partial x_0}(\mu_\Theta), \frac{\partial p_i}{\partial y_0}(\mu_\Theta), \frac{\partial p_i}{\partial z_0}(\mu_\Theta), \frac{\partial p_i}{\partial \mu_\Theta} \right). \quad (C.42)
\]
Appendix D

Computation of the Log-Likelihood of the Uncertain Data

The likelihood function of observed data in the form of contour function is

\[
L(r, A|pl_p) = \prod_{t=1}^{T} L(r, A|pl_t) = \prod_{t=1}^{T} \int f(p_t)pl_t(p_t)dp_t. \tag{D.1}
\]

By Theorem (3),

\[
f(p_t)pl_t(p_t) = \phi(p_t|GA, \sigma^2I_M) \phi(p_t|\mu_t^*, \Sigma^*)
= \phi \left( p_t \left| \left( \frac{I_M}{\sigma^2} + (\Sigma^*)^{-1} \right)^{-1} \left( \frac{I_M}{\sigma^2}GA + (\Sigma^*)^{-1}\mu_t^* \right), \left( \frac{I_M}{\sigma^2} + (\Sigma^*)^{-1} \right)^{-1} \right) \right)
\phi \left( \mu_t^*|GA, \sigma^2I_M + \Sigma^* \right), \tag{D.2}
\]

so we have

\[
\log L(r, A|pl_t) = \phi(\mu_t^*|GA, \sigma^2I_M + \Sigma^*)
= (\mu_t^* - GA)^H(\sigma^2I_M + \Sigma^*)^{-1}(\mu_t^* - GA). \tag{D.3}
\]

Finally we obtain

\[
\log L(r, A|pl_p) = - \sum_{t=1}^{T} (\mu_t^* - GA)^H(\sigma^2I_M + \Sigma^*)^{-1}(\mu_t^* - GA). \tag{D.4}
\]
Appendix E

Computation of the Gradient Matrix in Section 7.2

Similar to Eqs. (C.35)-(C.41), we obtain the following partial derivatives

\[
\frac{\partial p_{mt}}{\partial \theta_1}(\mu_\Theta) = y^m \sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{z_s}{r_{ms}^e}, \tag{E.1}
\]

\[
\frac{\partial p_{mt}}{\partial \theta_2}(\mu_\Theta) = x^m \sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{z_s}{r_{ms}^e}, \tag{E.2}
\]

\[
\frac{\partial p_{mt}}{\partial \theta_3}(\mu_\Theta) = -y^m \sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{x_s - x^m}{r_{ms}^e} + x^m \sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{y_s - y^m}{r_{ms}^e}. \tag{E.3}
\]

\[
\frac{\partial p_{mt}}{\partial x_0}(\mu_\Theta) = -\sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{x_s - x^m}{r_{ms}^e}, \tag{E.4}
\]

\[
\frac{\partial p_{mt}}{\partial y_0}(\mu_\Theta) = -\sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{y_s - y^m}{r_{ms}^e}, \tag{E.5}
\]

\[
\frac{\partial p_{mt}}{\partial z_0}(\mu_\Theta) = -\sum_{s=1}^S \frac{A_{st}}{4\pi} \left[ \frac{j\mu_k}{r_{ms}^e} e^{j\mu_k r_{ms}^e} - \frac{1}{(r_{ms}^e)^2} e^{-j\mu_k r_{ms}^e} \right] \frac{z_s}{r_{ms}^e}. \tag{E.6}
\]
and
\[
\frac{\partial p_{nt}}{\partial k}(\mu_\Theta) = \sum_{s=1}^{S} \frac{jA_{st}}{4\pi} e^{j\mu_k r_{ms}}^*,
\]
(E.7)
in which \( r_{ms}^* = \sqrt{(x_s - x_m)^2 + (y_s - y_m)^2 + z_s^2} \) represents the distance between the assumed location of the \( m \)-th microphone and the \( s \)-th source. Thus, the gradient matrix is
\[
\nabla p_t(\mu_\Theta) = \sum_{s=1}^{S} \frac{A_{st}}{4\pi} Y_s,
\]
(E.8)
in which
\[
Y_s = (Y_s(\theta_1), Y_s(\theta_2), Y_s(\theta_3), Y_s(x_0), Y_s(y_0), Y_s(z_0), Y_s(k)),
\]
(E.9)
and
\[
Y_s(\theta_1) = \left( y_s^* \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{z_s}{r_{ms}^*} \right)^M_{m=1},
\]
(E.10)
\[
Y_s(\theta_2) = \left( x_s^* \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{z_s}{r_{ms}^*} \right)^M_{m=1},
\]
(E.11)
\[
Y_s(\theta_3) = \left( -y_s^* \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{z_s}{r_{ms}^*} \right) + x_s^* \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{y_s}{r_{ms}^*} \right)^M_{m=1},
\]
(E.12)
\[
Y_s(x_0) = -\left( \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{z_s}{r_{ms}^*} \right)^M_{m=1},
\]
(E.13)
\[
Y_s(y_0) = -\left( \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{z_s}{r_{ms}^*} \right)^M_{m=1},
\]
(E.14)
\[
Y_s(z_0) = -\left( \left[ \frac{j\mu_k}{r_{ms}} e^{j\mu_k r_{ms}^*} - \frac{1}{(r_{ms}^*)^2} e^{-j\mu_k r_{ms}^*} \right] \frac{z_s}{r_{ms}^*} \right)^M_{m=1},
\]
(E.15)
\[ Y_s(k) = \left( j e^{j \mu_n^m} \right)^M_{m=1}. \]  

(E.16)

The covariance matrix of the contour function (7.2) can be computed by

\[
\Sigma' \approx \mathbb{E} \left( \nabla p_I(\mu_\Theta) \Sigma_\Theta \nabla p_I(\mu_\Theta)^H \right) = \sum_{s=1}^{S} \frac{a_s^2}{16\pi^2} Y_s Y_s^H, \]

(E.17)

where

\[
\Sigma_\Theta = \text{diag} \left( \sigma_{\theta_1}^2, \sigma_{\theta_2}^2, \sigma_{\theta_1}^2, \sigma_{x_0}^2, \sigma_{y_0}^2, \sigma_{z_0}^2, \sigma_{\kappa}^2 \right). \]

(E.18)