Optimization and estimation techniques for passive acoustic source localization

Jan Neering

ED n°84 : Science et technologies de l'information et de la communication

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|__|

T H E S E

pour obtenir le grade de

DOCTEUR DE L'ECOLE NATIONALE SUPERIEURE DES MINES DE PARIS

Spécialité "Informatique temps réel, robotique et automatique"

présentée et soutenue publiquement par
Jan NEERING

15.04.2009

Optimization and Estimation Techniques for Passive Acoustic Source Localization

Directeur de thèse : Nadia Maïzi

Jury

M. Jean-Pierre Le Cadre	Rapporteur
M. Philippe Vanheeghe	Rapporteur
M. Jean Demartini	Examinateur
M. Nadia Maïzi	Examinateur
M. Rémi Drai	Examinateur
M. Marc Bordier	Examinateur

Résumé

La performance d'un système de localisation de source passive acoustique au moyen d'un ensemble de capteurs ne dépend pas uniquement du choix des algorithmes d'estimation mais aussi de la géométrie du réseau de capteurs et de la position de la source. Cette thèse s'intéresse à de nouveaux estimateurs de position via un positionnement optimal de capteurs.

Pour une source "quasi-statique", trois mesures de performance sont successivement traitées, comparées et évaluées : la borne inférieure de Cramer-Rao (CRLB), la dilution de précision géométrique (GDOP) et le nombre de conditionnement. Les deux premières décrivent l'influence du bruit de mesure décrit par une fonction de probabilité connue, tandis que la dernière est une mesure non-statistique. En considérant le bruit Gaussien et l'estimateur d'un modèle linéarisé, il est montré que la même configuration optimale des capteurs est obtenue par ces trois mesures.

Ces estimateurs sont étendus à une source mobile au moyen de deux approches. Dans la première, la zone de surveillance est décrite par plusieurs points représentatifs et on minimise la moyenne des mesures de performance de tous ces points. La deuxième est une approche dynamique qui modélise le mouvement de la source par des équations d'état. Des estimateurs récursifs Bayesiens, comme par exemple le filtre de Kalman dans le cas de systèmes linéaires, sont ensuite appliqués afin de prédire la position future de la source. On sélectionne alors parmi tous les microphones, un sous-ensemble qui minimise la mesure de performance pour la position prédite. Ce sous-ensemble est alors utilisé pour l'estimation.

Abstract

The performance of acoustic passive source localization based on a multiple sensor system does not only depend on the chosen estimation algorithms, but is also strongly correlated to the geometry of the sensor network and the position of the source. This thesis approaches the optimization of the estimation procedure, by utilizing an optimal microphone setup.

In order to carry out this optimization procedure for a "quasi-static" source, three performance measures, the Cramer-Rao Lower Bound (CRLB), the Geometric Dilution of Precision (GDOP) and the condition number, are addressed, compared and evaluated. While the two former describe the influence of measurement noise with known probability density function, the latter is a non-statistical measure. Considering zero-mean Gaussian noise and a linearized model estimator, it is shown that all three approaches lead to the same configuration.

The performance measures are extended for a moving source proposing two approaches. The first one is to represent the surveillance area by multiple representative points. In order to assure a good coverage of the zone the average performance measure of all these points is minimized. The second, dynamic approach models the movement of the source using a state-space representation. Recursive Bayesian estimators, such as the Kalman filter for linear systems, predict the most likely upcoming position of the source. Utilizing an adaptive microphone network, only those microphones, which minimize the cost function for this predicted position, are then selected to carry out the estimation procedure.

Acknowledgements

First of all, I would like to thank my advisor, Professor Nadia Maïzi, director of the Center for Applied Mathematics (CMA), for offering me the opportunity to carry out this work, for supporting me over my five wonderful years at the CMA, and for giving me so much freedom to explore and discover new areas. In addition I would like to thank the committee members, Professor Jean Demartini, Professor Philippe Vanheeghe, Professor Jean-Pierre Le Cadre, Rémi Drai, and Marc Bordier, for their interest in my work, their constructive comments, and advices. I am especially grateful to Marc Bordier, in whom I did not only find a mentor, but also a friend.

I would like to thank Professor Jean-Paul Mamorat and Professor Yves Rouchaleau for supporting me over the years.

I would like to thank my colleague and great friend Christian Fischer, who provided plenty of valuable comments and ideas.

My special thanks go to Dominique Micollier, who greatly helped me with all administration issues, travels, the planning of my defense, and many other concerns.

I would like to thank Gilles Guerassimoff and Jean-Charles Bonin who provided valuable computer support. Further, I would like to thank Valérie Roy for offering the opportunity to get some teaching experiences, and for overworking French texts of mine. I would also like to thank Mathilde Drouineau, Lionel Daniel, and Johann Thomas for advising and correcting me on the French language.

I would like to thank my many other friends and colleagues at the CMA with whom I have had the pleasure of working over the years.

Finally, above all, I am grateful for my family and Sarah, their confidence, and loving support.

Contents

List of Figures

List of Tables

Symbolism and Abbreviations

In this work scalars are indicated by thin lower case letters. Bold lowercase letters, including Greek letters, represent vectors. Matrices are denoted by bold uppercase letters, including Greek letters.

Symbolism

$\boldsymbol{A}^T$	transpose of the matrix $\boldsymbol{A}$
$\boldsymbol{A}^\dagger$	Moore-Penrose pseudo-inverse of matrix $\boldsymbol{A}$
$\boldsymbol{A}^*$	conjugate complex of $\boldsymbol{A}$
$\hat{\boldsymbol{\theta}}$	Estimate of vector $\boldsymbol{\theta}$
$\boldsymbol{a} \circ \boldsymbol{b}$	Hadamard product (element-wise product) of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$
$\boldsymbol{a} \times \boldsymbol{b}$	cross product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$
$x * y$	convolution of signals x and y
$\lceil a \rceil$	rounding operator to the next higher integer of a

Abbreviations

ADC	Analog to Digital Conversion/Converter
AED	Adaptive Eigenvalue Decomposition
CRLB	Cramer Rao Lower Bound
DAC	Digital to Analog Conversion/Converter
DOA	Direction Of Arrival
DSP	Digital Signal Processor
EKF	Extended Particle Filter
FFT	Fast Fourier Transform
GCC	Generalized Cross Correlation
GDOP	Geometric Dilution of Precision
GNSS	Global Navigation Satellite System
GPS	Global Positioning System

IFFT	Inverse Fourier Transform
i.i.d.	independent and identically-distributed
LCLS	Linear Correction Least-Squares (estimator)
LI	Linear Intersection (estimator)
LLS	Linear Least-Squares (estimator)
LMS	Linear Mean Square (estimator)
LSE	Least-Squares Estimator
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
MVU	Minimum Variance Unbiased (estimator)
NLLS	Nonlinear Least-Squares (estimator)
pdf	probability density function
PHAT	PHAse Transform (GCC)
PF	Particle Filter
RD	Range Difference
RMSE	Root Mean Square Error
SI	Spherical Interpolation (estimator)
SIR	Sequential Importance Resampling
SIS	Sequential Importance Sampling
SNR	Signal to Noise Ratio
RADAR	RAdio Detection And Ranging
SONAR	SOund Navigation And Ranging
SVD	Singular Value Decomposition
TDE	Time Delay Estimate
TDOA	Time Difference of Arrival
UART	Universal Asynchronous Receiver/Transmitter
UKF	Unscented Particle Filter
UT	Unscented Transformation
WLLS	Weighted Linear Least-Squares (estimator)

Chapter 0

Introduction Française

L'estimation de la position d'un objet revêt une importance de tout premier plan dans un très grand nombre de domaines d'applications. Citons entre autres:

- La robotique: les robots autonomes doivent être capables d'estimer leur position courante de manière à suivre un parcours déterminé tout en évitant les obstacles;
- La navigation: les systèmes de positionnement par satellite (Global Navigation Satellite Systems ou GNSS) qui trouvent leurs utilisations dans les applications suivantes:
  - Systèmes de guidage de véhicules, avions, bateaux, missiles, satellites, etc.;
  - Systèmes anticollision, comme par exemple le *système d'identification automatique* (*Automatic Identification System* ou AIS) utilisé en navigation maritime pour éviter les collisions entre navires;
- La sismique: par exemple la détection des mines en sous sol au moyen d'ondes sismiques;
- La surveillance civile ou militaire: estimation de la position de troupes ennemies, ou détection d'intrusion;
- La localisation d'orateur: l'estimation de la position d'un orateur dans une salle de conférence permettent d'améliorer:
  - les téléconférences: en pointant automatiquement une camera vers un conférencier pendant son discours;

- l'amélioration de la qualité de la voix (*speech enhancement*) par filtres spatiaux: les signaux sonores provenant de la direction de l'orateur sont amplifiés tandis que les bruits issus des autres directions sont supprimés;
- les applications d'immersion sonore virtuelle d'une personne dans une réalité virtuelle;
- les appareils d'aides auditives (*hearing aids*).

Ces applications peuvent recourir à une grande variété de systèmes de positionnements, comme par exemple le *Global Positioning System* (ou GPS), éventuellement assistés de capteurs inertiels, de systèmes de traitement de signaux vidéo dans l'infrarouge ou le visible, de radars *RAdio Detection And Ranging* ou encore de sonars *SOund Navigation And Ranging* (ou SONAR). Les deux derniers types de capteurs se subdivisent en deux classes: les capteurs actifs et les capteurs passifs. Les systèmes actifs émettent un signal, écoutent son écho sur l'objet cible au moyen d'un ou plusieurs récepteurs et en déduisent une estimation de la position de la cible. Les systèmes passifs n'émettent pas des signaux, ils écoutent directement les signaux provenant de l'objet cible au moyen des capteurs et utilisent les informations ainsi recueillies pour inférer la position de l'objet.

La systèmes de positionnements passifs se subdivisent en deux groupes: ceux utilisant des méthodes indirectes et ceux utilisant des méthodes directes. Les méthodes indirectes résolvent le problème de l'estimation de la position en deux étapes: i) la première étape consiste à estimer les retards de temps (*time differences of arrival* ou TDOA) entre les paires de capteurs; ii) la seconde étape utilise ces délais pour estimer la position de la source. Les méthodes directes, quant à elles, estiment généralement la position de la source en une seule étape, en transférant les signaux provenant des microphones dans le domaine fréquentiel. Les méthodes les plus souvent appliquées de la classe des estimateurs directs, sont le MUltiple SIgnal Classification (MUSIC) un algorithme proposé par Schmidt[Sch86] ainsi qu'une variété de techniques de "beamforming", comme par exemple le *steered response power beamformer*[DBS01] ou encore le *steered filter-and-sum beamformer*[SMR99].

Le choix de la sous-classe, la mieux adaptée à un scénario d'estimation, dépend fortement des caractéristiques physiques et des particularités de la source ainsi que des propriétés de propagation des milieux où se déplace les signaux sonores entre la source et le réseau de capteurs. Chen et al.[CYH02] en identifient les caractéristiques fondamentales et les propriétés de la manière suivante:

- *Bande étroite* versus *large bande*. Les signaux se classent dans ces deux

catégories de bandes suivant le ratio de fréquence: fréquence haute versus basse. Ainsi, le son provenant de véhicules roulant ou tractés contient des fréquences allant de 20Hz à 2kHz. Leur ratio s'évalue à environ 100, ce qui les classe comme des signaux à larges bandes. Par contre, des signaux de transmissions radio sont généralement des signaux à bande étroite puisque leur ratio est typiquement proche de l'unité.

- *Champ proche* versus *champ éloigné*. Le front d'onde d'un signal émis est courbe et sa courbure, à une position donnée, dépend de la distance du signal à la source. Quand cette distance est petite on parle de champ proche. Quand cette distance grandit, le front de l'onde s'aplatit et on parle alors de champ éloigné. Dans le cas d'un champ éloigné, seule la direction d'arrivée (*direction of arrival* ou DOA) de la source peut être estimée, dans le cas d'un champ proche on peut aussi estimer sa position.

- *Propagation en espace dégagé* versus *propagation en espace réverbérant*. Dans un espace dégagé, le signal émis n'est reflété par aucun obstacle, comme des murs, et il parvient au capteur en parcourant une trajectoire directe (une ligne droite entre la source et le capteur). La plupart des applications qui se situent en intérieur sont assez réverbérantes, aussi le signal reçu par un capteur est la somme de la trajectoire directe ainsi que des signaux reflétés par des murs et autres obstacles présents. Les environnements fortement réverbérants rendent particulièrement complexe l'estimation de la position d'une source puisqu'un obstacle reflétant un signal peut être vu comme une source indépendante.

- *Une source* versus *multiples sources*. Le choix d'un estimateur dépend naturellement du nombre de sources présentes.

Les méthodes directes furent tout d'abord dérivées pour l'estimation des directions d'arrivées de signaux en bandes étroites et en champ éloigné. Aujourd'hui, des extensions existent pour des signaux en larges bandes en champs proches ou éloignés. Généralement, les extensions concernant des signaux en large bande, subdivisent tout d'abord la bande des signaux, appliquent ensuite les algorithmes standards pour bandes étroites à chaque sous-ensemble de fréquence considéré, et finalement, utilisent des méthodes de fusion pour trouver une unique estimation de la position. Évidemment ces approches sont très calculatoires et nécessitent des ordinateurs performants.

Cette thèse étudie la classe des estimateurs de position basée sur les retards de temps d'arrivées (TDOA). Dans le chapitre 3, nous donnons une

vue globale des estimateurs existants et discutons de leurs avantages et de leurs faiblesses. Ces estimateurs sont quasiment indépendants des largeurs de bandes des signaux et peuvent être aussi bien appliqués en champs proches qu'en champs éloignés. Alors que la littérature dispose d'un grand nombre d'articles sur la meilleure manière de mettre en oeuvre ces estimateurs dans toutes sortes de situations, très peu d'informations existent sur la meilleure manière de repartir les capteurs dans l'espace. Intuitivement, nous pouvons considérer que les signaux reçus par les capteurs se ressemblent fortement lorsque ces capteurs se trouvent dans un voisinage immédiat et en conclure que l'estimation de la position sera de qualité médiocre. Au contraire, si les capteurs sont répartis dans l'espace, l'information obtenue par un seul capteur devient significative vis à vis des autres et la qualité de l'estimation en sera améliorée. Ainsi, la deuxième question importante à laquelle nous nous intéressons dans chapitre 3 concerne la manière de répartir les capteurs afin d'obtenir une performance optimale dans l'estimation de la position. Le chapitre 2 présente des méthodes permettant de mesurer la performance d'un estimateur et décrit la manière d'utiliser ces mesures de qualité pour optimiser la répartition spatiale des capteurs.

Dans le chapitre 3, nous montrons que les mesures de performance ne dépendent pas uniquement des positions de capteurs, mais aussi de la position de la source. Pour une source en mouvement, l'estimation de la configuration optimale des capteurs se complexifie drastiquement. Une manière de résoudre ce problème consiste à prédire la future position de la source et d'adapter la configuration des capteurs à cette prédiction. Le chapitre 4 traite de cette prédiction en utilisant pour cela des filtres Bayesiens récursifs.

Le chapitre 5 décrit la réalisation d'un système de localisation passif mettant en œuvre un ordinateur et plusieurs processeurs de traitement numérique de signal (*digital signal processors* ou DSPs). Ces processeurs sont utilisés de manière intensive dans la téléphonie et bénéficient ainsi d'une distribution de masse qui les rend peu coûteux. Les signaux, reçus par les microphones, sont convertis par les convertisseurs analogiques/numériques des processeurs et sont ensuite transférés vers l'ordinateur qui exécute le programme d'estimation de position.

Chapter: Estimator Optimization

Ce chapitre introduit les fondements mathématiques des idées proposées dans cette thèse. Il présente les aspects fondamentaux de la théorie de l'estimation statistique. Une variété d'estimateurs classiques sont présentés ainsi que deux mesures pour valider leur justesse: la *borne inférieure de Cramer Rao*

(*Cramer Rao Lower Bound* ou CRLB) et la *dilution géométrique de précision* (*Geometric Dilution of Precision* ou GDOP). Ces estimateurs optimaux et sous-optimaux utilisent des informations statistiques sur les mesures bruitées comme leur vecteur moyen et leur matrice de variance. Dans le cas où de telles informations ne sont pas disponibles la théorie de l'estimation statistique ne s'applique pas. Elle est remplacée par une approche fondée sur la minimisation de l'erreur entre modèle et mesures (moindres carrés ou *least-squares estimators*). Un critère de performance appelé le *nombre de conditionnement* (*condition number*) est introduit pour ce dernier estimateur. Nous discuterons de la meilleure manière d'exploiter les mesures en choisissant les paramètres internes de l'estimateur de manière optimale. Les paramètres internes les plus courants sont : la position des capteurs (par exemple pour l'estimateur de position), la dimension ou la taille des capteurs (par exemple les jauges de déformation pour l'estimation de forces), la fréquence de coupure des filtres intégrés, etc. Pour les estimateurs utilisant les TDOA, les paramètres internes sont les positions des microphones.

Chapter: TDOA-Based Passive Source Localization

Le sonar passif consiste en un ensemble de microphones ou d'hydrophones répartis spatialement et utilise cette information spatiale pour estimer la position de l'objet. Le son se propage avec une vitesse finie variable suivant les milieux. Dans l'air à $20°C$, la vitesse du son est d'environ $343m/s$. Dans l'eau, cette vitesse dépend principalement de la pression, de la température et de la salinité du milieu, et se situe typiquement autour de $1500m/s$[Pie81] et [NE92]. Due à cette vitesse de propagation limitée, le son produit par la source arrivera aux microphones à des instants différents, ce sont ces retards qui seront utilisés pour estimer la position de l'objet.

La classe des estimateurs passifs de position basés sur des TDOAs estime la position en deux étapes. Dans la première étape, on estime les retards de temps (les TDOAs) de chaque paire de microphones distribués dans l'espace. Ensuite ces TDOAs sont utilisés dans une seconde étape pour estimer la position courante de l'objet. A cause des bruits additionnels et des environnements réverbérants, les TDOAs estimés peuvent être vus comme des variables aléatoires composées à partir des valeurs réelles et d'un terme additionnel de bruits. Afin de diminuer l'influence des termes concernant les bruits, le nombre de microphones est habituellement choisi plus grand que le nombre minimal de capteurs nécessaires (qui est la dimension de l'espace aug-

mentée de deux), et la théorie des estimateurs statistiques est appliquée afin de trouver la position de la source. En outre, la performance de l'estimation est fortement corrélée à la distribution de capteurs.

La méthode de localisation passive de sources par TDOA se décompose alors en trois parties: i) le choix d'un estimateur des retards de manière à obtenir le vecteur de données; ii) le choix d'un estimateur de position basé sur le modèle d'estimation des équations; iii) le choix du nombre de microphones N et l'optimisation de leur configuration.

Bien que tous ces points soient traités dans le chapitre 3, ce chapitre se focalise essentiellement sur le développement des point deux et trois.

Par soucis de complétude, trois estimateurs de TDOA seront présentés dans la section 3.2:

1. l'estimateur utilisant la *cross-corrélation généralisée* (*Generalized Cross-Correlation estimator* ou GCC);

2. l'estimateur utilisant les *moindres carrés linéaires* (*Linear Mean Squares estimator* ou LMS);

3. l'estimateur utilisant la *décomposition adaptative en valeurs propres* (*Adaptive Eigenvalue Decomposition estimator* ou AED).

Si on se base sur les résultats de simulations menées en environnement réverbérant avec différents ratios de rapports signal sur bruit, les TDOAs estimés peuvent être considérés comme la somme des TDOAs réels et d'un vecteur aléatoire Gaussien avec une valeur moyenne de zéro. L'estimateur itératif de vraisemblance maximale utilisant ce modèle de TDOA sera présenté dans la section 3.4.1. Cet estimateur est habituellement le meilleur choix d'estimateur dans la théorie de l'estimation statistique puisqu'il atteint asymptotiquement la borne inférieure de Cramer-Rao lorsque le nombre de mesures tend vers l'infini. En réalité, le nombre de microphones est loin d'être infini. D'autre part, il est nécessaire de connaître les informations statistiques sur le bruit (valeur moyenne et variance) afin de pouvoir appliquer l'estimateur de vraisemblance maximale, alors que, dans les applications acoustiques, ces informations sont le plus souvent difficiles à obtenir ou en constante variation. Pour toutes ces raisons, nous présenterons dans la section 3.4.2, un estimateur itératif linéaire qui ne prend pas en compte d'informations statistiques sur le bruit.

Dans la section 3.5, nous dérivons plusieurs estimateurs analytiques qui peuvent, soit être appliqués directement pour trouver une estimation de position, soit être utilisés comme initialisation des estimateurs linéaires ou de l'estimation du maximum de vraisemblance. Ces estimateurs analytiques

considèrent uniquement les retards de temps par rapport à un microphone de référence. Leur biais et leurs matrices de covariances sont utilisés dans le calcul de la dilution géométrique de précision (GDOP) et sont également importants pour le suivi de trajectoires basé sur les TDOAs. Ces idées sont présentées dans la chapitre 4.

Des configurations optimales de microphones pour les différents estimateurs seront données en section 3.7. Alors que les configurations des capteurs basées sur le CRBL sont indépendantes de l'estimateur choisi, les configurations trouvées soit par minimisation de la dilution géométrique de précision soit par le nombre de conditionnement dépendent de l'estimateur sélectionné.

Nous considérerons ensuite plusieurs configurations optimales. La première est la configuration optimale par rapport à la GDOP pour l'estimateur linéarisé, considérant le bruit comme gaussien et de valeur moyenne nulle. La seconde est la configuration optimale par rapport à la borne inférieure de Cramer-Rao. Elle est décrite par les sommets de solides platoniques (tétraèdre, hexaèdre ou cube, octaèdre, dodécaèdre, icosaèdre, fig. 3.9) ayant pour centre la position de la source. Nous montrerons que ces configurations sont identiques et qu'elles restent optimales quand le nombre de conditionnent devient la fonction de coût. Nous montrerons aussi que la performance des estimateurs itératifs augmente fortement quand ceux-ci sont initialisés à l'aide d'estimateurs analytiques. En conséquence, il serait intéressant d'avoir une configuration optimale non seulement pour les estimateurs itératifs, mais aussi pour les estimateurs analytiques. Dans la section 3.7.3 nous décrivons les configurations optimales par rapport au nombre de conditionnement. La section 3.7.4 présentera une configuration optimale pour une combinaison des deux estimateurs.

A la fin du chapitre, en section 3.8, nous décrirons un estimateur analytique basé sur les configurations optimales par rapport au nombre de conditionnement.

Chapter: Source Tracking

Le chapitre 4 "*Source Tracking*" traite le problème d'estimation récursive des positions successives d'un source mobile. Ces positions ne sont pas indépendantes les unes des autres et leur dynamique est représentée au moyen d'un modèle d'états dynamiques et d'une équation de mesure. Le vecteur d'état contient toutes les données pertinentes nécessaires à la description de la dynamique du système. Dans les applications, il est en général composé de la position et de la vitesse du système, ainsi que des informations dynamiques additionnelles comme l'angle de braquage d'un véhicule.

L'estimation Bayésienne récursive fournit une estimation du vecteur d'état à un instant k en se basant sur toutes les observations antérieures à k. Cette méthode cherche à calculer la fonction de densité de probabilité à posteriori du vecteur d'état. Une fois cette fonction de probabilité obtenue, l'estimation de l'état optimal du système se calcule au moyen d'un critère statistique quelconque comme la moyenne ou la médiane. Dans le cas d'un système linéaire sous l'hypothèse de bruits Gaussien, l'estimateur récursif optimal est équivalent à un *filtre de Kalman*. En cas d'équations non-linéaires ou de bruit non-Gaussien, le filtre optimal généralement n'est plus calculable et on doit utiliser des filtres sous-optimaux. Le plus courant est le *filtre de Kalman étendu* (*Extended Kalman Filter* ou EKF) qui linéarise les équations et considère les bruits comme étant Gaussien. Récemment, une intéressante alternative au EKF, le *filtre de Kalman non-parfumé* (*Unscented Kalman Filter* ou UKF) a suscité un intérêt grandissant dans la communauté du traitement du signal. Plutôt que de linéariser les équations, le UKF approxime les fonctions de densité de probabilité par des densités Gaussiennes, grâce à un ensemble minimal et bien choisi de points d'échantillonnage déterministes (appelés les *sigma points*). Ces points d'échantillonnage expriment les vraies valeurs de la moyenne et de la covariance d'une variable aléatoire Gaussienne et, quand ils sont propagés à travers un système non-linéaire, les points d'échantillonnage permettent de calculer les valeurs à posteriori de la moyenne et de la covariance au second ordre près.

Le troisième type de filtre étudié dans le cadre de cette thèse est le *filtre particulaire* (*particle filter* ou PF) qui est plus approprié que les deux précédents dans les cas où les fonctions sont fortement non-linéaires ou lorsque les bruits sont clairement non-Gaussiens. Son principe est de représenter la fonction de densité de probabilité à posteriori par un ensemble de points d'échantillonnage pondérés appelés *les particules* et, de calculer l'estimation souhaitée de l'état du système à partir de ces particules. Contrairement à l'UKF, ces particules sont arbitrairement extraites de la fonction de densité de probabilité. D'un point de vue calculatoire, les particules de ce filtre étant beaucoup plus nombreuses que les points d'échantillonnage d'UKF, l'utilisation de ce filtre entraîne un plus grand nombre de calculs.

Ces filtres sont utilisés et comparés dans le cas du suivi de sources avec la méthode TDOA, où le système d'équations est considéré comme linéaire et l'équation de mesures, donnant la relation entre les TDOAs, la position de la source et les positions des capteurs, non linéaire. De plus, on considère que les bruits additionnels utilisés suivent des distributions Gaussiennes. Dans un tel cas, on montre que les filtres présentés ont tous des performances équivalentes, et que par conséquent on peut en déduire que soit les filtres EKF soit les filtres UKF sont plus attractifs, compte tenu de leur avantage

calculatoire.

Par ailleurs, nous montrons comment une configuration optimale des capteurs, exhibée dans les chapitres précédents, peut modifier favorablement la sortie du filtrage. Enfin, dans une discussion conclusive, nous présentons comment utiliser le pas de prédiction de ces filtres récursifs pour mieux estimer les TDOA.

Chapter: DSP-based Passive Source Localization

Le chapitre 5 présente un système de localisation passif de sources acoustiques par traitement numérique du signal. L'objectif est de proposer une solution meilleure marché que celle existant dans la littérature. Le point central de ces systèmes est la conversion analogique/numérique (*analog to digital conversion* ou AC) des signaux provenant des microphones. Ces convertisseurs sont en général issus de l'industrie musicale. Ils sont connectés à un PC dédié à l'estimation de la position. Le prix minimum de ces systèmes est d'environ 1000€ et n'a pas de limite supérieure. Chen et Li [CL04] et Bian et al. [BRA04] proposent deux architectures pour cette approche. Dans J. C. Chen et al.[CYW+03] et M. Chen et al. [CLH+07] les systèmes de localisation sans fil utilisent respectivement des pocket PC et des ordinateurs portables. De tels systèmes peuvent être intéressants pour localiser un orateur dans une salle de conférence. En effet, la plupart des participants à une conférence sont aujourd'hui munis soit de pocket PC soit d'ordinateurs portables dont les microphones embarqués peuvent alors servir à localiser le conférencier. Tandis que le problème de synchronisation est traité dans [CYW+03], M. Chen et al. l'évitent en utilisant un algorithme de localisation basé sur une fonction de l'énergie, plutôt que sur les TDOA.

Le besoin d'un système à bas coût est guidé par ses applications potentielles. Un système de pointage automatique d'une caméra sur une cible n'est économiquement intéressant que dans la mesure où le prix du système est bas. En effet, plus on réalise d'économies sur le système de localisation/suivi, plus la part d'argent mise dans la caméra peut être élevée. Des applications alternatives de la localisation passive de sources sont les kits mains-libres pour la téléphonie. La direction de l'interlocuteur est estimée à l'aide des algorithmes présentés et, un filtrage spatial (aussi connu sous le nom de *beamforming*) est utilisé pour d'une part amplifier le son provenant de la direction de l'interlocuteur et d'autre part atténuer les sons provenant des autres directions. Un tel système peut trouver des application dans le

domaine de l'industrie automobile puisque la plupart des pays restreignent l'usage du téléphone portable par le conducteur aux systèmes mains-libres. Cependant, les bruits parasites ambiants dans l'habitacle du véhicule i. e. le moteur, les roues, les autres véhicules, l'aérodynamisme, ... sont de manière générale plutôt importants et pénalisent l'utilisation des systèmes mains-libres. L'utilisation d'algorithmes de localisation de sources peut permettre de supprimer ces bruits et d'amplifier la voix de l'interlocuteur. Enfin, de tels systèmes peuvent être aussi particulièrement intéressants pour des applications de réalité virtuelle ou des jeux vidéos. Plusieurs microphones pourraient être montés sur ou intégrés dans l'écran de l'ordinateur et la voix du joueur pourrait être utilisée pour améliorer son immersion virtuelle dans le jeu. Tous ces systèmes deviennent économiquement attractifs, si l'estimation de la position peut être faite à un coût raisonnable.

Tandis que le coût des systèmes présentés précédemment reste supérieur à 1000 Euros, l'objectif de ce chapitre est de présenter un système réalisable pour un coût bien inférieur.

D'autre part, le système présenté dans cette section, est utilisé pour évaluer la précision des algorithmes présentés dans la première partie de cette thèse.

Chapter 1

Introduction

The estimation of an object's position is of utmost importance for a large number of application domains. Amongst others the following fields of application strongly rely on an accurate position estimate:

- Robotics: autonomous robots need to be capable to estimate their position in order to follow their intended paths and to avoid obstacles
- Navigation: navigation systems such as the global navigation satellite systems (GNSS) find their application in fields such as
  - Guidance systems for vehicles, aircrafts, ships, missiles, satellites, etc.
  - Anti-collision systems such as the *autonomous identification system* (AIS) for vessel collision avoidance
- Seismic subsurface object detection, e.g. buried landmines
- Military and Surveillance: position estimation of own and enemies position, intrusion detection
- Speaker Localization: estimation of a speaker's position in a room for e.g.
  - Hearing Aids
  - Teleconference: automatically pointing a camera toward the speaker during a conference
  - Speech Enhancement by spatial filtering: signals coming from the speaker's direction are amplified, while sounds from the other directions are suppressed

  - Acoustic virtual immersion of a person into a virtual reality

- etc.

The presented systems can access a large number of existing position estimators, such as the *global positioning system* (GPS), possibly supported by inertial sensors, infrared and video processing procedures, *radio detection and ranging* (RADAR) and *sound navigation and ranging* (SONAR) systems. The two latter can be divided into the classes of passive and active sensors. Active sensors transmit a signal, which is reflected by the target and then detected by one or multiple receivers of the sensor. The captured reflected signals are then used to estimate the object's position. Passive sensors do not emit a signal, but detect the signals produced by the object itself by multiple sensors and use the collected data of this *sensor array* to infer the source's position.

1.1 Passive Source Localization using Acoustic Sensor Arrays

The class of passive source localization estimators based on the measurements of an acoustic sensor array, can be roughly subcategorized into direct and indirect methods. The indirect methods first estimate the time differences of arrival between multiple sensor pairs and then use these delays to estimate the source position in a second step. Direct methods, on the other hand, generally carry out the position estimation procedure in one step, by transferring the received microphone signals into the frequency domain. Out of the class of direct methods probably the *MUltiple SIgnal Classification* (MUSIC) algorithm due to Schmidt [Sch86], and a variety of beamforming techniques, such as the steered response power beamforming approach [DBS01] and the steered filter-and-sum beamformer [SMR99] are currently the most widely applied.

The choice of which subclass is best suited for a given estimation scenario is strongly dependent on the physical characteristics and features of the source(s) and the propagation properties between the source(s) and the sensor array, Chen et al. [CYH02] characterize the basic features and properties as follows

- Narrow-band versus wide-band signal: signals can be divided into the class of narrow-band or wide-band sources by the ratio of their highest to their lowest frequency component. The sound produced from

wheeled and tracked vehicles may range from 20Hz-2kHz, resulting in a ratio of about 100, and consequently are referred to as wide-band signals. On the contrary, transmitted radio signals are usually narrow-band, since their ratio is typically close to unity.

- Far-field versus near-field source: the wavefront of an emitted signal is curved and the curvature depends on the distance to the source. If the sensors are assumed to be close to the source, the signal is said to be in the near-field. If the distance becomes large, the wavefront becomes planar, and the source is said to be in the far-field. For far-field sources, only the direction of arrival (DOA) can be estimated, meaning that the direction towards the source can be estimated. For near-field sources its position can be estimated.

- Free-space versus reverberant space propagation: in free-space the emitted signal is not reflected by obstacles, such as walls, and the emitted signal arrives at the microphones only by the direct path (straight line from source to sensor). Most indoor applications are fairly reverberant, meaning that the received signal is the sum of the direct path signal and signals reflected from the walls and other obstacles. Strongly reverberant environments make the position estimation rather difficult, since the reflected signals might be interpreted as being an independent source.

- Single versus multiple source: the choice of estimator is made upon to the number of present sources.

The direct techniques were originally derived for narrow-band sources in the far-field, estimating the direction of arrival (DOA) only. A variety of extensions for broad-band signals in the near-, as well as in the far-field are proposed in literature. These extensions do generally divide the entire frequency range of the broad-band signal into frequency bins and then carry out the estimation procedure for each frequency bin, using the standard algorithms for narrow-band signals. This procedure results in position estimates for each frequency bin, which must then be combined to obtain a single source position estimate. Obviously, this results in a considerable computational burden.

On the other hand, the computational burden of the TDOA-based position estimators is quasi independent of the bandwidth of the emitted source signal, and can easily be utilized in near-field as well as far-field scenarios.

1.2 Aim and Objectives of this Thesis

This thesis investigates the class of time difference of arrival (TDOA) based position estimators and tries to give a comprehensive overview of the existing TDOA-based estimators and discusses their strengths and weaknesses.

While a large number of articles has been published on how to best implement the TDOA-based estimators in all kind of situations, relatively little information can be found on how to best place the spatially separated microphones. Intuitively it can be argued that if the microphones are all placed in the near neighborhood of each other, that the received signals will be very similar. Consequently, only little information can be obtained from the individual microphones and the estimation procedure is likely to be of poor quality. Contrary, if the microphones are spread, the information obtained from an individual microphone becomes more important and the estimation quality will possibly increase. The question of how to best place the sensors in order to obtain an optimal performance builds the second major part of this thesis. It will be discussed how the quality of an estimator can be measured, and how this quality measure can then be used to optimize the sensor configuration.

It will be seen that a sensor configuration being optimal for one source position might be insufficient for other locations of the object. This can lead to severe limitations of the estimation procedure, if a moving source is considered. One way of addressing this problem is to predict the future source position and then to adapt the sensor network. This procedure is studied using recursive Bayesian estimation techniques.

Finally, a low-cost system is presented, which uses a low-cost digital signal processor (DSP) to analog-digital convert the data of multiple microphones and then transmit these to a standard PC, which carries out the actual estimation procedure.

1.3 State of the Art

1.3.1 TDOA-based Position Estimation

Time difference of arrival (TDOA) based estimators are two step procedures and can be applied to near-field as well as the far-field situations considering narrow-band or wide-band signals in free- or reverberant-space. In a first step the relative time differences between microphone pairs are estimated, which are then, in the second step, used to estimate the source position. Assuming a spatially separated microphone pair i and j and a single source,

the sound produced by the source will arrive at the sensors as delayed and filtered version of the origin signal. A variety of time delay estimators exist, which are capable of estimating the relative time delays of microphone arrays. Knowing the sensor positions and the estimated TDOA of a sensor pair, the position of the source lies on one half of an hyperboloid. Estimating the TDOAs of multiple independent sensor pairs, the source's position can be estimated by the intersection of these hyperboloids. Consequently, in 3-D space, three such hyperboloids, and consequently at least four microphones are needed to get a unique position estimate. Fang [Fan90] derived an exact solution of this intersection problem for situations where the number of TDOA estimates is equal to the number of spatial dimensions. However, this solution cannot be extended to extra TDOA measurements. But, additional TDOA measurements can greatly increase the estimation accuracy, if disturbance noise is present. In such a case the multiple hyperboloids will usually not intersect in a single point, and some compromise between the resulting intersection points needs to be made.

The problem of the multiple intersections can be written as a nonlinear maximum likelihood estimation problem, which needs to be solved via numerical optimization. The maximum likelihood estimator is approximately optimal if the number of observations, or in this case the number of TDOA estimates, becomes large. However, a large number of TDOA estimates, corresponds to a large number of microphones, and consequently results in a large computational and financial burden. Further, the noise level in acoustic scenarios most often strongly varies over time, which makes the application of the maximum likelihood estimator even more troublesome. In order to avoid any inaccuracy due to changing, or unknown statistical properties of the noise, the TDOA-based position estimation problem is quite often described as a nonlinear least-squares estimation problem. Iterative methods which start with an initial guess and successively approximate the optimal solution via a linearized least-squares estimate at each step in the procedure exist [Foy76], [Tor84]. Obviously, such estimates suffer of numerical optimization problems: they are not guaranteed to converge in many instances, and tend to be sensitive to the choice of the initial guess.

A large number of sub-optimal, closed-form TDOA-based estimators exist, which approximate the solution of the nonlinear problem [SA87], [BAS97], [CH94], [HBE00], [HBEM01], [SR87], etc. These estimators are computationally undemanding and in [NFBM08] we showed that they suffer little performance decrease compared to the optimal estimators, if an appropriate sensor configuration is selected. Further, their outcome can be used as the initial guess for the iterative estimators, which due to the accurate initialization become quite robust [NBM07].

The advantages of these TDOA-based position estimators over the direct methods is their flexibility of application (near-field vs. far-field, narrow-band vs. wide-band, free-space vs. reverberant-space), and their significantly smaller computation time. Their main inconvenience is their difficulty handling multi-source situations. However, recently a number of multi-source time delay estimators have been proposed, which seem capable of resolving this problem.

Multiple Source TDOA-based Position Estimation

Recently a number of TDOA estimation procedures for the multi-source scenarios have been proposed. The majority of these time delay estimators are based on the theory of blind source separation. Blind source separation deals with separating multiple sources from a set of multiple mixed signals, without any or only little knowledge on the source signals or the mixing process.

Assuming multiple acoustic sources and multiple microphones, the signal of an individual microphone can be seen as the sum of the filtered versions of the emitted signals. While the main goal of blind source separation would be to reconstruct the individual emitted signals, for time delay estimates the focus lies on identifying the channel impulse response, responsible for the filtering process of the signals. If these channel impulse responses can be identified, time delay estimates for each sensor pair and each source can be obtained from these filters.

Independent component analysis (ICA) is currently the most popular procedure for carrying out this separation procedure, if at least as many microphones as sources are present.

The TRINICON algorithm due to Buchner, Aichner and Kellermann [BAK07], as well as the frequency domain ICA approach due to Sawada et al. [SMAM05] both use the convolutive version of the independent component analysis to estimate the corresponding time delays of multiple sources.

In case of fewer microphones than sources Emile, Comon and Le Roux [ECR98] have proposed a procedure for estimating the time delays based on higher oder statistics.

While the three above presented methods can all be seen as blind source separation techniques, Scheuing and Yang [SY06], [SY07] have proposed a correlation-based technique for identifying the time delays of multiple sources.

1.3.2 Optimal Sensor Configuration

The performance of a position estimator based on a, not necessarily acoustic, sensor array is strongly dependent on the geometry of the microphone network and the relative position of the source w.r.t. this array. The problem of maximizing the estimator's accuracy by using an optimal sensor configuration has been addressed in literature for a variety of different position estimator, minimizing a variety of cost functions, describing the estimator's performance.

One of the most often utilized cost functions is the trace of the Cramer Rao lower bound (CRLB). This performance measure was used by e.g. Abel [Abe90], Aranda et al. [AMB05], and Yang and Scheuing [YS05], who all considered the problem of passive acoustic source localization. While Abel [Abe90] derived an analytic solution, constraining the sensors to lie on a line segment, Yang and Scheuing derived an analytic solution without any constraints on the sensor positions: the optimal configuration is obtained if the sensors are evenly spread around the source position. Consequently, the configuration will usually be optimal for a single source position. Aranda et al. found the same configuration and relaxed the problem of a single source position, by assuming a moving sensor network.

An extension of the Cramer Rao lower bound is used by Jourdan and Roy [JR06] for ultra-wideband ranging sensors. Instead of using a moving sensor network for achieving an optimal performance over an entire area or trajectory, they minimize the average cost function over the considered area or path.

Zhang [Zha95] and Levanon [Lev00] find the same sensor configurations as Yang and Scheuing for two-dimensional estimation problems, using alternative cost functions. While Levanon minimizes the trace of the geometric dilution of precision (GDOP) of a TDOA-based position estimator, Zhang considers the estimation covariance matrix of an estimator composed of multiple acoustic transceivers.

Usually the geometric dilution of precision is utilized as a performance measure for the global positioning system GPS. However, McKay and Pachter [MP97] argued that the condition number is the better suited as a cost function for their inverted pseudolite positioning system. The condition number was also used by Hegazy and Vachtsevanos [HV03] for optimizing the performance of an energy-based acoustic position estimator.

In his PhD thesis Rabinkin [Rab98] focused on the signal to noise ratio as a cost function for passive acoustic position estimation.

1.3.3 Passive Acoustic Source Tracking

The goal of passive acoustic source tracking is to estimate the object's trajectory using noise-corrupted position or bearing measurements, from a passive source localization estimator. While the object's movement is usually described as a dynamic state model, a measurement equation links the position or bearing estimates to the state of this dynamic model.

If the two equations are linear and it is assumed that additive Gaussian noise is disturbing both equations, then it is well known that the classical Kalman filter is the optimal filter for the tracking problem [Kal60].

In tracking applications it is quite common to model the target motion as a linear state model, such as the *white noise acceleration model* [BC07] [1]. For human source tracking applications Vermaak and Blake [VB01] proposed a linear state model describing a Langevin process. This model has shown to work well in practice [LJN07] and was adapted by e.g. Ward et al. [WLW03], Lehmann and Johansson [LJ07] and Fallon and Godsill [FG07].

On the other hand, the measurement model is typically nonlinear, which means that the classical Kalman filter is not applicable. In [WLW03], [LJN07], [LJ07], [FG07] the steered beamformer approach is utilized for carrying out the position estimation, which results in a nonlinear measurement equation. The tracking problem becomes even more difficult if not consecutive position estimates, but only bearing estimates are available. In such a case the measurement model is nonlinear and the state vector becomes unobservable (no range estimate is available), or at best only poorly observable [Cad98]. Bearing-only tracking systems find their application in e.g. marine applications, when the sensor geometry only allows a bearing estimate (e.g. sensors installed on a cord dragged by a vessel).

For nonlinear state space models a variety of alternatives to the classical Kalman filter are available. They range from the extended Kalman filter over the unscented Kalman filter originally proposed by Julier and Uhlmann [JU97], grid based filters [Ber99], to recursive particle filters, such as the bootstrap filter [GSS93], or sequential importance sampling filters. An exhaustive overview of these particle filters can be found in technical report of Doucet et al. [DGA00], in the tutorial of Arulampalam et al. [AMGC02], and the PhD thesis of Bergman [Ber99], which also presents the ideas of the grid based approaches.

The unobservability of the source's range is compensated by the bearing-only tracking approach by the movement of the sensor array. A large number of recursive filters have been proposed for carrying out the actual tracking

[1] The survey article of Li and Jilkov [LJ03] gives a thorough overview of dynamic state models for tracking applications

procedure. Arulampalam and Ristic [AR00] have demonstrated the superiority of particle filter approaches over the modified extended Kalman filter approaches. However, Fearnhead [Fea98] demonstrated that the use of the standard particle filter can nevertheless lead to a divergence of the position estimation. This was also observed by Bréhard and Le Cadre [BC07] , who propose to use the particle filter in combination with a coordinate transformation to resolve these divergence problems. Le Cadre and Trémois proposed to solve the tracking problem by a grid based method using hidden Markov models [CT98].

The majority of the beamformer based position trackers utilize the particle filter approach [LJN07], [LJ07], [FG07]. Vermaak and Blake utilize the generalized cross-correlation based approach [VB01], and Ward et al. [WLW03] do a comparative study of these two approaches.

1.4 Structure of the Thesis

Chapter 2 builds the framework for acoustic passive source localization and its optimal sensor configuration. The basic ideas of estimation theory are presented, and the *Cramer Rao Lower Bound* (CRLB) and the Geometric Dilution of Precision (GDOP) are introduced as possible measures for statistical estimators, which use the available information about the present disturbance noises. If no such information is available, the statistical estimators are usually replaced by the least-squares estimator, which can be seen as a "fitting the data" estimator. It minimizes the squares of the residuals, which are the difference of an overdetermined model outcome and the observed data. If the model is linear, the estimate is obtained by the multiplication of the pseudo-inverse of a matrix and the observation vector. The calculation of a pseudo-inverse can numerically be delicate, if the concerned matrix is ill-conditioned. A measure for the condition of a matrix is the *condition number*, which in the next chapter will be introduced as an alternative performance measure to the CRLB and the GDOP. Further, it will be discussed how these three measures can be used to optimize the performance of a given estimator.

In chapter 3 the theory presented in chapter 2 is applied to the problem of acoustic TDOA-based passive source localization. A detailed overview of existing estimators is presented, and their advantages and inconveniences are discussed. Depending on the problem at hand either one of the presented algorithms might be favorable over the others. Due to their different structure, a sensor configuration being optimal for one estimator might not be optimal for the others. Consequently, the sensor configuration also depends on the

chosen estimator. Procedures for optimizing the sensor configuration will be presented for all estimators, considering all three performance measures. It will be shown that under the assumption of independent and identically distributed (i.i.d.) Gaussian noise all three measures lead to the same sensor configuration for the class of statistical estimators.

In chapter 4 the ideas of recursive Bayesian estimation are presented and the Kalman filter, the extended Kalman filter, the unscented Kalman filter and Bootstrap filter in particular introduced. In combination with the presented TDOA-based passive source localization estimators the estimation accuracy can greatly be increased. Further, these recursive estimators have the ability to predict the future position as a random variable with known mean vector and covariance matrix. This statistical information will be used to support the time delay estimation and the adaptive sensor configuration procedures.

In chapter 5 a low-cost TDOA-based source localization system is presented, which is based on a digital signal processor (DSP) and a standard PC.

Finally a conclusion and an outlook are given in chapter 6.

Chapter 2

Estimator Optimization

This chapter builds the mathematical framework for the ideas presented in this thesis. The main aspects of statistical estimation theory will be presented. A variety of common estimators are introduced, and two measures for validating their accuracy are described. The presented optimal, or suboptimal estimators are based on available statistical information on the noisy measures, such as their means and covariance matrices. In case that no such information is available, this theory cannot be applied, and might be replaced by an approach, which minimizes the sum of square differences between a defined model and the observed measures. This process is commonly known as the least-squares approach. A third performance measure, the *condition number*, is introduced for this non statistical estimator.

It will be discussed how the three presented measures can be exploited for increasing the performance of the presented estimators by choosing optimal values for the internal parameters of the estimator at hand. Those parameters are variables which can usually be arbitrarily chosen by the operator. Common internal parameters are the position of the sensors (e.g. of localization estimators), the dimension or size of the sensors (e.g. of resistance strain gauges for force estimates), the cutoff frequencies of integrated filters, and so forth.

2.1 Statistical Estimation Theory

The general idea of estimation theory is to infer values of unknown parameters

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T \tag{2.1}$$

based on a set of discrete measurement data

$$\boldsymbol{x} = [x[0], \ldots, x[m-1]]. \tag{2.2}$$

The discrete measurement data is usually found by the sampling process of a computer or microcontroller of one or multiple continuous waveforms, which are dependent on the unknown parameters collected in the vector $\boldsymbol{\theta}$. This unknown parameter vector can be of statistical nature, but is considered to be deterministic in this and the next chapter. In determining a good estimator the first step is to mathematically model this dependency:

$$\tilde{\boldsymbol{x}} = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i), \tag{2.3}$$

where $\tilde{\boldsymbol{x}}$ is the model output, which approximates the true observation $\boldsymbol{x}$, and $\boldsymbol{x}_i$ denotes the internal parameters of the estimator in vector notation. Since the measurement data is usually subject to disturbing noises $\boldsymbol{n}$, which are modeled as random variables, the measurement data are random as well. Throughout this thesis it is considered that the disturbing noise is of additive nature. Hence, the ideal model of eq. (2.3) transforms to

$$\boldsymbol{x} = \tilde{\boldsymbol{x}} + \boldsymbol{n} = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i) + \boldsymbol{n}. \tag{2.4}$$

Since it is desired to determine $\boldsymbol{\theta}$ based on the data, the goal of estimation theory is to derive an estimator $\boldsymbol{g}$, such that

$$\hat{\boldsymbol{\theta}} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{x}_i, \boldsymbol{n}), \tag{2.5}$$

is an estimate of $\boldsymbol{\theta}$ fulfilling some optimality criterion. Usually such an estimate is based on the first two moments, the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{C}_n$ of the additive noise $\boldsymbol{n}$.

2.1.1 Unbiased and Minimum Variance Estimators

The interest is to find an estimator which fulfills some optimality criterion. A natural one is the mean square error (MSE), defined as

$$\begin{aligned} mse(\hat{\boldsymbol{\theta}}) &= E((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})) \\ &= E([(\boldsymbol{\theta} - E(\hat{\boldsymbol{\theta}})) + (E(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\theta}})]^T[(\boldsymbol{\theta} - E(\hat{\boldsymbol{\theta}})) + (E(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\theta}})]) \\ &= E(|\boldsymbol{\theta} - E(\hat{\boldsymbol{\theta}})|_2^2) + E(|E(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\theta}}|_2^2) \\ &= |\boldsymbol{\theta} - E(\hat{\boldsymbol{\theta}})|_2^2 + E(|E(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\theta}}|_2^2) \\ &= |E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}|_2^2 + \mathrm{trace}(\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}), \end{aligned} \tag{2.6}$$

where $E()$ denotes the expected value, $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ represents the covariance matrix of the estimate and $|.|_2$ denotes the Euclidean vector norm. The bias $\boldsymbol{b}$ is defined as

$$\boldsymbol{b} = E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}. \tag{2.7}$$

Definition 1. *An estimator is said to be unbiased if* $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$*, otherwise it is said to be biased.*

In order to evaluate the bias $\boldsymbol{b}$ of an estimator, the true value of the unknown parameter needs to be known. Therefore, the bias is a rather theoretical measure, and is generally unavailable in practice. Consequently, since the MSE is the sum of the trace of the estimator's covariance and the bias, an estimator minimizing the MSE is generally not realizable. An alternative is to minimize the variance $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ of the estimate, since this is only a function on $\hat{\boldsymbol{\theta}}$ and not on the true unknown parameters $\boldsymbol{\theta}$. The estimators, out of the set of all unbiased estimators, with minimum variance are called *minimum variance unbiased* MVU estimators. In order to find a MVU estimator, or to test whether a given estimator is a MVU estimator, the Cramer Rao Lower Bound usually needs to be derived.

2.1.2 Cramer Rao Lower Bound

The Cramer Rao Lower Bound (CRLB) defines a lower limit on the variance of any unbiased estimator. This lower bound is of great importance for the development of estimators. At best, the minimum variance unbiased (MVU) estimator can directly be derived from it. At worst, it provides a benchmark against which the performance of any unbiased estimator can be compared. Furthermore, the theory of the CRLB offers a possibility to determine whether an estimator, which actually attains this bound, exists [Kay93].

In the following it is assumed that the measured data is given in vector notation, that is $\boldsymbol{x} \in \mathbb{R}^m$.

Knowing the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ the likelihood function for observing $\boldsymbol{x}$, is denoted by the probability density function (pdf) $p(\boldsymbol{x}; \boldsymbol{\theta})$. The ";" in this pdf represents its dependency on the deterministic unknown parameter $\boldsymbol{\theta}$. Consequently, the expectation of the data $\boldsymbol{x}$ taken with respect to $p(\boldsymbol{x}; \boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ as well, and is denoted by $\boldsymbol{\mu}(\boldsymbol{\theta})$.

The vector version of the CRLB is given by the following theorem.

Theorem 1 (Cramer-Rao Lower Bound - Vector Parameter)**.** *It is assumed that the probability density function* $p(\boldsymbol{x}; \boldsymbol{\theta})$ *satisfies the "regularity" conditions*

$$E\left[\frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0} \text{ for all } \boldsymbol{\theta}, \tag{2.8}$$

where the expectation is taken with respect to $p(\boldsymbol{x}; \boldsymbol{\theta})$*, and*

$$\frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1}, \quad \ldots \quad \frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_p} \right). \tag{2.9}$$

Then, the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ *of any unbiased estimator* $\hat{\boldsymbol{\theta}}$ *satisfies*

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \boldsymbol{F}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0} \tag{2.10}$$

where ≥ 0 *indicates a positive semi-definite matrix. The Fisher information matrix (FIM)* $\boldsymbol{F}(\boldsymbol{\theta})$ *is given by*

$$[\boldsymbol{F}(\boldsymbol{\theta})]_{ij} = -E\left[\frac{\partial^2 \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\theta^i \partial\theta^j}\right], \tag{2.11}$$

where the derivatives are evaluated at the true value of $\boldsymbol{\theta}$ *and the expectation is taken with respect to* $p(\boldsymbol{x};\boldsymbol{\theta})$*. Furthermore, an unbiased estimator may be found that attains the bound in that* $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{F}^{-1}(\boldsymbol{\theta})$ *if and only if*

$$\left(\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T = \boldsymbol{F}(\boldsymbol{\theta})\left(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{\theta}\right) \tag{2.12}$$

for some p-dimensional function $\boldsymbol{g}$ *and some* $p \times p$ *matrix* $\boldsymbol{F}$*. That estimator, which is the minimum variance unbiased (MVU) estimator, is* $\hat{\boldsymbol{\theta}} = \boldsymbol{g}(\boldsymbol{x})$*, and its covariance matrix is* $\boldsymbol{F}^{-1}(\boldsymbol{\theta})$*.*

The proof of this theorem is omitted, but can be found in estimation theory literature, such as in [Kay93].

In the case of a Gaussian pdf $p(\boldsymbol{x};\boldsymbol{\theta})$ with mean vector $\boldsymbol{\mu}$ and covariance $\boldsymbol{C}$ the Fisher information matrix results in [Kay93]

$$\begin{aligned}[\boldsymbol{F}(\boldsymbol{\theta})]_{ij} &= \left[\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right]^T \boldsymbol{C}^{-1}\left[\frac{\partial\boldsymbol{\mu}}{\partial\theta^j}\right] \\ &\quad + \frac{1}{2}\mathrm{trace}\left[\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\theta^i}\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\theta^j}\right],\end{aligned} \tag{2.13}$$

with

$$\frac{\partial\boldsymbol{\mu}}{\partial\theta^i} = \left[\frac{[\partial\boldsymbol{\mu}]_1}{\partial\theta^i}, \ldots, \frac{[\partial\boldsymbol{\mu}]_N}{\partial\theta^i}\right]^T. \tag{2.14}$$

Assuming that the measures $\boldsymbol{x}$ are equal to the nonlinear model of equation (2.4) perturbed by additive additive zero-mean Gaussian noise with constant covariance $\boldsymbol{C}_n$, $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_n)$,

$$\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i) + \boldsymbol{n}, \tag{2.15}$$

the likelihood function $p(\boldsymbol{x};\boldsymbol{\theta})$ is Gaussian, too. Rewriting eq. (2.15) as

$$\boldsymbol{n} = \boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i), \tag{2.16}$$

its pdf follows to

$$\begin{aligned} p(\boldsymbol{n}) &= \frac{1}{(2\pi)^{p/2}\det(\boldsymbol{C}_n)^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{n}^T\boldsymbol{C}_n^{-1}\boldsymbol{n}\right) \\ &= \frac{1}{(2\pi)^{p/2}\det(\boldsymbol{C}_n)^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i))^T\boldsymbol{C}_n^{-1}(\boldsymbol{x}-\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i))\right) \\ &= p(\boldsymbol{x};\boldsymbol{\theta}). \end{aligned} \tag{2.17}$$

One sees that the likelihood function $p(\boldsymbol{x};\boldsymbol{\theta})$ is completely described by the mean

$$\boldsymbol{\mu} = \boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i) \tag{2.18}$$

and the covariance $\boldsymbol{C}_n$. Applying eq. (2.13), the Fisher information matrix is found to be

$$\boldsymbol{F}(\boldsymbol{\theta},\boldsymbol{x}_i) = \boldsymbol{\mathcal{J}}^T(\boldsymbol{\theta},\boldsymbol{x}_i)\boldsymbol{C}_n^{-1}\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i), \tag{2.19}$$

where $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i)$ denotes the Jacobian matrix given by

$$\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)}{\partial \boldsymbol{\theta}}. \tag{2.20}$$

If further it is assumed that the noise is independent identically distributed, such that $\boldsymbol{C}_n = \sigma^2\boldsymbol{I}$, then the bound on the unbiased estimators covariances is given by

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} \geq \sigma^2 \cdot (\boldsymbol{\mathcal{J}}^T(\boldsymbol{\theta},\boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i))^{-1}, \tag{2.21}$$

for which the inverse of $\boldsymbol{\mathcal{J}}^T(\boldsymbol{\theta},\boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i)$ needs to exist.

Definition 2. *An estimator, which is unbiased and attains the CRLB, is said to be efficient in that it efficiently uses the data.*

2.1.3 Linear Model Estimators

As discussed above, the determination of the MVU is in general a difficult task. However, for linear estimation models of the form

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{n}, \tag{2.22}$$

with zero mean Gaussian noise $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{C}_n)$, the MVU can directly be derived using eq. (2.12). Replacing the nonlinear data model $\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)$ by the linear term $\boldsymbol{H}\boldsymbol{\theta}$ the pdf $p(\boldsymbol{x};\boldsymbol{\theta})$ follows to

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{p/2}\det(\boldsymbol{C}_n)^{1/2}}\exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\theta})^T\boldsymbol{C}_n^{-1}(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\theta})\right). \tag{2.23}$$

Defining

$$\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \theta_1}, \quad \ldots, \quad \frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \theta_p} \right) \tag{2.24}$$

it follows that

$$\begin{aligned} \left(\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T &= \boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{x} - \boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H} \boldsymbol{\theta} \\ &= (\boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H}) \left((\boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{x} - \boldsymbol{\theta} \right). \end{aligned} \tag{2.25}$$

Comparing eq. (2.25) with eq. (2.12) one sees that the MVU for the linear model is

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{x}, \tag{2.26}$$

and its covariance follows to

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = (\boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H})^{-1}. \tag{2.27}$$

2.1.4 Maximum Likelihood Estimator

It is more complicated, and quite often even impossible to derive the minimum variance unbiased estimator for nonlinear systems. In such cases the most popular alternative is the maximum likelihood (ML) estimator, which is probably due to its approximate efficiency, for large data sets, $\boldsymbol{x} = [\boldsymbol{x}[0], \ldots, \boldsymbol{x}[m-1]]$, with $m \to \infty$, and its ease of implementation [Kay93]. It is defined as the value that maximizes the probability density function $p(\boldsymbol{x};\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{x};\boldsymbol{\theta}). \tag{2.28}$$

Because many density functions contain exponential functions, it is often more convenient to find the maximum of the log likelihood function $\ln(p(\boldsymbol{x};\boldsymbol{\theta}))$. Since the log function is a strictly monotonically increasing function the estimate that maximizes $p(\boldsymbol{x};\boldsymbol{\theta})$ clearly also maximizes $\ln(p(\boldsymbol{x};\boldsymbol{\theta}))$. The optimization problem then results to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln(p(\boldsymbol{x};\boldsymbol{\theta})). \tag{2.29}$$

For zero-mean Gaussian noise, with covariance matrix $\boldsymbol{C}_n$, the probability density function $p(\boldsymbol{x};\boldsymbol{\theta})$ results to (see eq. (2.17))

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{p/2} \det(\boldsymbol{C}_n)^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i))^T \boldsymbol{C}_n^{-1} (\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)) \right). \tag{2.30}$$

If the covariance matrix is assumed to be independent of $\boldsymbol{\theta}$, the corresponding optimization problem of eq. (2.29) can be simplified:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \ln(p(\boldsymbol{x};\boldsymbol{\theta})) \\ &= \arg\min_{\boldsymbol{\theta}} (\boldsymbol{x}-\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i))^T \boldsymbol{C}_n^{-1} (\boldsymbol{x}-\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)). \end{aligned} \tag{2.31}$$

A necessary condition for the maximum of eq. (2.30) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln(p(\boldsymbol{x};\boldsymbol{\theta})) = \mathbf{0}. \tag{2.32}$$

Since $\boldsymbol{C}_n$ is a symmetric matrix,

$$\left(\frac{\partial}{\partial \boldsymbol{y}} (\boldsymbol{y}^T \boldsymbol{C}_n^{-1} \boldsymbol{y})\right)^T = 2\boldsymbol{C}_n^{-1}\boldsymbol{y}, \tag{2.33}$$

and hence eq. (2.32) follows to

$$\left(\frac{\partial \ln p(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T = \boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i)^T \boldsymbol{C}_n^{-1} (\boldsymbol{x}-\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)), \tag{2.34}$$

with $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i)$ denoting the Jacobian of $\boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)$ w.r.t. $\boldsymbol{\theta}$:

$$\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)}{\partial \boldsymbol{\theta}}. \tag{2.35}$$

The matrix product of $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i)^T \boldsymbol{C}_n^{-1}$ is of dimension $\mathbb{R}^{p,m}$, with p being the dimension of the unknown estimation vector $\boldsymbol{\theta} \in \mathbb{R}^p$, and m being the length of the observation vector $\boldsymbol{x} \in \mathbb{R}^m$. Usually $m > p$ and hence the solution of eq. (2.34) is $\boldsymbol{\theta}$, such that $\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i)$ belongs to the kernel of $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta},\boldsymbol{x}_i)^T \boldsymbol{C}_n^{-1}$. Of course, if there exists a $\boldsymbol{\theta}$ such that $\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta},\boldsymbol{x}_i) = 0$, this $\boldsymbol{\theta}$ will be the ML estimate. However, most often, the solution of this maximum likelihood problem cannot be found in close-form and numerical optimization must be applied for finding the maximum likelihood estimate.

Numerical Solution If the maximum cannot be found in closed-form, a variety of numerical algorithms exist, which are suitable for carrying out the search for the maximum.

The safest way of finding this maximum numerically is to perform a grid search over the allowable region of the unknown parameter $\boldsymbol{\theta}$. If, however, the unknown parameter $\boldsymbol{\theta}$ is not constraint to a finite region, the grid search

approach is likely to become computationally infeasible. In such a case iterative maximization procedures, such as the expectation-maximization (EM) algorithm [DLR77], the Quasi-Newton, the simplex downhill or the Newton-Raphson method [Fle00] can replace the grid search. All these methods are of iterative nature and their performance is strongly correlated to their initialization. Meta-heuristic optimization techniques, such as the particle swarm optimization [KE95], exist, which avoid the problem of needing a good initial guess. Nevertheless, they do not guarantee to converge to the global maximum, either, and are computationally quite demanding.

In this thesis the Newton-Raphson algorithm is applied for such nonlinear estimation problems. When additive Gaussian noise is considered, this algorithm exploits the pseudo-inverse of the estimation model

$$\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}, \boldsymbol{x}_i) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)}{\partial \boldsymbol{\theta}}, \tag{2.36}$$

which will play a fundamental role for the optimization of the estimator's performance in the following sections and chapters.

Approximative Efficiency In numerical analysis the Newton-Raphson algorithm is probably the best known method for finding the roots of a function. The idea is as follows: one starts with an initial guess as close as possible to the true root. The function is then approximated by its tangent at this point, and the root of the resulting linear term is calculated. This value is then used as the new approximation of the root, and the procedure is repeated. For multidimensional problems $\boldsymbol{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ the iterative step is given by

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{\mathcal{J}}(\boldsymbol{x}_n)^\dagger \boldsymbol{f}(\boldsymbol{x}_n), \tag{2.37}$$

with

$$\boldsymbol{\mathcal{J}}(\boldsymbol{x}_n) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}|_{\boldsymbol{x}=\boldsymbol{x}_n}. \tag{2.38}$$

An inconvenience of the ML estimator is that it only approximates the minimum variance estimator for infinite data sets. For a finite number of observations, however, the ML estimator might be inefficient, or even biased. However, strictly speaking, if the maximum likelihood estimator is biased its performance cannot be evaluated by comparing it to the Cramer Rao Lower Bound and alternative measures need to be considered. One alternative performance measure, which is also applicable to biased estimators is the *dilution of precision*.

2.1.5 Dilution of Precision

In estimation theory it is often of great importance to have a measure for the current accuracy of an estimator. A well-known application, which uses this information, is a navigation system, which estimates the position of an object by the sensor fusion of data obtained from a global navigation satellite system (GNSS), such as GPS or Galileo, and some alternative sensors, such as inertial sensors measuring the translational and radial accelerations of the object. Most often the information obtained from the inertial sensors is used to estimate the position of the object, when no GNSS position is available, or if this measure can not be trusted. In GPS applications the measure on how much a position estimate can be trusted is the *dilution of precision* (DOP). It describes the geometric strength of satellite configuration on the GPS accuracy. Roughly speaking, when visible satellites are close together it is said to be weak and the DOP value is high. When the visible satellites are far apart, the geometry is said to be strong and the DOP value is low. The dilution of precision can be expressed as a number of separate measures in GPS applications. One normally distinguishes between the horizontal (HDOP), the vertical (VDOP), the position (3D) (PDOP) and the time dilution of precision (TDOP). The combination of all these measures is called the geometric dilution of precision, which is defined as the ratio of the root mean square error (RMSE) of the position estimate $\boldsymbol{x}_s$ and the RMSE of the range error $\boldsymbol{e}$

$$\mathrm{GDOP}(\boldsymbol{x}_s) = \frac{\mathrm{RMSE}(\boldsymbol{x}_s)}{\mathrm{RMSE}(\boldsymbol{e})}, \tag{2.39}$$

with the RMSE being defined as the square-root of the MSE of eq. (2.6)

$$\mathrm{RMSE}(\boldsymbol{y}) = \sqrt{E(\boldsymbol{y} - \hat{\boldsymbol{y}})^T(\boldsymbol{y} - \hat{\boldsymbol{y}})}. \tag{2.40}$$

Appendix B presents more information on the GDOP for GPS applications.

For other classes of estimation problems the dilution of precision can be defined as the ratio of the root mean square error (RMSE) of the observations and the RMSE of the noise

$$\mathrm{GDOP}(\hat{\boldsymbol{\theta}}) = \frac{\mathrm{RMSE}(\hat{\boldsymbol{\theta}})}{\mathrm{RMSE}(\boldsymbol{n})}. \tag{2.41}$$

In eq. (2.6) it was shown that the MSE, and therefore also the RMSE is a combination of the estimator's bias and its covariance. Therefore, the dilution of precision is an appropriate measure for all kind of statistical estimators, for which these values can be determined.

This section presented some main ideas of statistical estimation theory. Sometimes however, it is difficult to obtain the necessary statistical information for deriving the proposed estimators, or for calculating the CRLB or the GDOP. An alternative, non-statistical approach to estimation theory, is the theory of least-squares, presented in the following section.

2.2 Least-Squares Estimation

The least-squares estimator (LSE) is widely used in practice due to its easy implementation, and its availability to offer estimates even when no statistical information about the noisy measures is available. It dates back to 1795 when Gauss used the method to study planetary motion. It is defined as the estimate $\hat{\boldsymbol{\theta}}$ which minimizes the sum of the squared difference between the true observations $\boldsymbol{x}$ and the outcome $\tilde{\boldsymbol{x}}$ of an observation model

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{x} - \tilde{\boldsymbol{x}}|_2 = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)|_2, \tag{2.42}$$

with $\tilde{\boldsymbol{x}} = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)$, as defined in eq. 2.3.

2.2.1 Linear Least-Squares Estimator

If the mathematical model is linear $\tilde{\boldsymbol{x}} = \boldsymbol{H}\boldsymbol{\theta}$, the least-squares estimator can be written as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}|_2, \tag{2.43}$$

where $\boldsymbol{H}$ is a known upright matrix with full column rank, referred to as the observation matrix. Thus, the LSE is found by minimizing

$$J(\boldsymbol{\theta}) = (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta})^T(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}). \tag{2.44}$$

The gradient of $J(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$ is easily found to be

$$\left(\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T = -2\boldsymbol{H}^T\boldsymbol{x} + 2\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\theta}. \tag{2.45}$$

Since the matrix $\boldsymbol{H}$ is assumed to be of full column rank, the square matrix $\boldsymbol{H}^T\boldsymbol{H}$ is of full rank, too. Consequently, the linear least-squares estimator results in

$$\hat{\boldsymbol{\theta}} = \boldsymbol{H}^{\dagger}\boldsymbol{x}, \tag{2.46}$$

with $\boldsymbol{H}^{\dagger} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T$ denoting the pseudo-inverse of $\boldsymbol{H}$.

An extension of the LLSE is the weighted linear least-squares (WLLS) estimator. Instead of minimizing (2.44), a positive definite matrix $\boldsymbol{W}$ is included, such that

$$J(\boldsymbol{\theta}) = (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta})^T \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\theta}) \tag{2.47}$$

will be minimized. The idea of using such a weighting matrix is that measures, which are assumed to be more reliable, will have a greater influence on the estimate. The weighted least-squares estimate then follows to

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{H}^T \boldsymbol{W} \boldsymbol{H})^{-1} \boldsymbol{H} \boldsymbol{W} \boldsymbol{x}. \tag{2.48}$$

Note that if the weighting matrix $\boldsymbol{W}$ is set equal to the inverse of the noise's covariance matrix of eq.(2.26), it can be seen that the weighted least-squares estimator is equal to the statistical linear model estimator of section 2.1.3.

In the general nonlinear case, the LSE of equation (2.42) can generally not be derived in closed-form, and numerical analysis must be applied to find the minimizing parameters.

2.2.2 Nonlinear Least-Squares Estimator

It is quite easily seen that the general nonlinear case results in the same equations as the ones for the maximum likelihood estimator, replacing the covariance matrix $\boldsymbol{C}_n$ by the identity matrix. The extension to the weighted nonlinear least squares estimator, which is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} (\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i))^T \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)), \tag{2.49}$$

also results in the same equations. This time, however, the inverse of the covariance matrix is replaced by the weighting matrix $\boldsymbol{W}$.

2.2.3 Linearized Estimator

Ideally an estimate $\hat{\boldsymbol{\theta}}_n$ will be found which results in no residual of the minimization problem of eq. (2.42), or equivalently in

$$\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i) = \boldsymbol{0}. \tag{2.50}$$

Such an estimate can generally be found by the iterative Newton-Raphson algorithm, which is based on the linearized model of $\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)$ in the neighborhood of the estimate $\hat{\boldsymbol{\theta}}_n$:

$$\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i) \approx \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i) + \underbrace{\frac{\partial \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i)}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}}_{\boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n). \tag{2.51}$$

The iterative estimator is initialized by an initial guess of the estimator's outcome $\boldsymbol{\theta}_0$ and then updated by

$$\hat{\boldsymbol{\theta}}_{n+1} = \hat{\boldsymbol{\theta}}_n + \boldsymbol{\mathcal{J}}^{\dagger}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)(\boldsymbol{x} - \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)). \tag{2.52}$$

For shorter notation the bracket of the Jacobian will from now on be omitted: $\boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)$.

Even though this estimator does not guarantee to find the optimal estimate, nor is it guaranteed to converge, it does usually perform quite well, if the initial guess $\hat{\boldsymbol{\theta}}_0$ is in a close neighborhood of the true unknown vector $\boldsymbol{\theta}$. This makes the linearized estimator a standard estimator in estimation theory, and particularly for position estimation systems, such as the global positioning system [Zog06] and passive source localization estimators [HBC06].

Note that this linearized estimator does not take possible information about the disturbance noise into consideration. Still it is obviously of interest how the estimator performs if statistical information about the noise is available. Again, additive noise is considered:

$$\begin{aligned} \boldsymbol{x} &= \tilde{\boldsymbol{x}} + \boldsymbol{n} \\ &= \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i) + \boldsymbol{n}. \end{aligned} \tag{2.53}$$

Replacing $\boldsymbol{x}$ in eq. (2.52) by eq. (2.53) leads to

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{n+1} &= \hat{\boldsymbol{\theta}}_n + \boldsymbol{\mathcal{J}}^{\dagger} \cdot (\boldsymbol{x} - \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)) \\ &= \hat{\boldsymbol{\theta}}_n + \boldsymbol{\mathcal{J}}^{\dagger} \cdot \left(\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{x}_i) + \boldsymbol{n} - \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i) \right). \end{aligned} \tag{2.54}$$

Further, if $\boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)$ is replaced by the approximation of eq. (2.51), the estimate $\hat{\boldsymbol{\theta}}_{n+1}$ follows to

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{n+1} &\approx \hat{\boldsymbol{\theta}}_n + \boldsymbol{\mathcal{J}}^{\dagger} \cdot \left(\boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i) + \boldsymbol{\mathcal{J}} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) + \boldsymbol{n} - \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i) \right) \\ &= \hat{\boldsymbol{\theta}}_n + \boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{\mathcal{J}} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) + \boldsymbol{n}) \\ &= \boldsymbol{\theta} + \boldsymbol{\mathcal{J}}^{\dagger} \cdot \boldsymbol{n}. \end{aligned} \tag{2.55}$$

The mean and the covariance of the estimate of the linearized estimator then follow to

$$E(\hat{\boldsymbol{\theta}}_{n+1}) \approx \boldsymbol{\theta} + \boldsymbol{\mathcal{J}}^{\dagger}\boldsymbol{\mu} \tag{2.56}$$

and

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}_{n+1}) \approx \boldsymbol{\mathcal{J}}^{\dagger}\boldsymbol{C}_n(\boldsymbol{\mathcal{J}}^{\dagger})^T. \tag{2.57}$$

Thus, for zero-mean additive noise $\boldsymbol{\mu} = 0$ the linearized estimator is unbiased and its geometric dilution of precision results to

$$\text{GDOP} = \frac{\sqrt{\text{trace}(\text{Cov}(\hat{\boldsymbol{\theta}}_n))}}{\sqrt{\text{trace}(\boldsymbol{C}_n)}}. \tag{2.58}$$

If the individual terms of the noise vector $\boldsymbol{n}$ are assumed to be uncorrelated and identically distributed, its covariance matrix is of the form

$$\boldsymbol{C}_n = \sigma^2 \boldsymbol{I}. \tag{2.59}$$

In such a case the covariance matrix of the linearized estimator becomes

$$\text{Cov}(\hat{\boldsymbol{\theta}}_{n+1}) \approx \sigma^2 (\boldsymbol{\mathcal{J}}^T \boldsymbol{\mathcal{J}})^{-1}. \tag{2.60}$$

Note that the covariance matrix of the linearized model and the CRLB are then of the same form. The difference of these two are the different parameters for which the Jacobians are evaluated. While the Jacobians of the linearized estimator are evaluated for the estimate found at time instant n, the Jacobians of the Fisher information matrix are evaluated for the true parameter. Consequently, the initialization of the linearized estimator is of utmost importance. If ideally it would be initialized by the true parameter, its covariance would be equal to the Fisher information matrix and hence, the linearized estimator would be efficient.

In case that no statistical information about the disturbance noise is available, it would still be interesting and important to have a quality measure of the proposed estimators. In this thesis the condition number is studied as a possible performance measure for the non statistical LLS estimator and the linearized estimator.

2.3 Condition Number

The condition number is primarily used in numerical computation as a measure for the amenability of a problem to digital computation. It describes the influence of disturbances on the input data on the solutions of a problem. A problem with low condition number is said to be well-posed, or well-conditioned, while a problem with high condition number is said to be ill-conditioned, or ill-posed.

The condition number is most widely used for linear problems given in matrix, vector notation, such as

$$\boldsymbol{H} \cdot \boldsymbol{x} = \boldsymbol{b}. \tag{2.61}$$

Matrix $\boldsymbol{H}$ and vector $\boldsymbol{b}$ are known, but are subject to uncertainties such as round-off errors or noise, and the unknown $\boldsymbol{x}$ is to be calculated numerically. Assuming a full column matrix $\boldsymbol{H}$ the solution of eq. (2.61) is found by multiplying both sides by the pseudo-inverse of $\boldsymbol{H}$. However, most often numerical algorithms are forced to use rounded values of the exact entries of $\boldsymbol{H}$ and $\boldsymbol{b}$. The question now arises of how the uncertainties in the matrix $\boldsymbol{H}$ and the vector $\boldsymbol{b}$ influence the solution $\boldsymbol{x}$. As an example for an ill-conditioned system consider the matrix

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 \\ 1 & 1+e \end{bmatrix}, \tag{2.62}$$

and the vector

$$\boldsymbol{b} = \begin{pmatrix} 2 \\ 2+e \end{pmatrix}. \tag{2.63}$$

The inverse of the matrix $\boldsymbol{H}$ amounts to

$$\boldsymbol{H}^{-1} = \begin{bmatrix} 1+1/e & -1/e \\ -1/e & 1/e \end{bmatrix}, \tag{2.64}$$

and the solution results to

$$\boldsymbol{x} = \boldsymbol{H}^{-1}\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{2.65}$$

In a first case consider that the vector $\boldsymbol{b}$ is somehow perturbed, such that

$$\tilde{\boldsymbol{b}} = \boldsymbol{b} + \begin{pmatrix} \Delta b_1 \\ \Delta b_2 \end{pmatrix}. \tag{2.66}$$

The solution of the perturbed system then results to be

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{H}^{-1} \begin{pmatrix} \Delta b_1 \\ \Delta b_2 \end{pmatrix} = \boldsymbol{x} + \begin{pmatrix} \Delta b_1 + (\Delta b_1 - \Delta b_2)/e \\ (\Delta b_2 - \Delta b_1)/e \end{pmatrix}. \tag{2.67}$$

For example, if $e = 10^{-3}$, $\Delta b_1 = 0.01$ and $\Delta b_2 = -0.01$, then the perturbed solution becomes

$$\tilde{\boldsymbol{x}} = \begin{pmatrix} 21.01 \\ -19 \end{pmatrix}, \tag{2.68}$$

which is obviously far from the true solution $\boldsymbol{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. Thus, uncertainties in $\boldsymbol{b}$ of less than 1% can result in a complete miss-calcuation of $\boldsymbol{x}$.

In a second case it is now assumed, that the vector $\boldsymbol{b}$ is perfectly known and that now some kind of uncertainty occurs in matrix $\boldsymbol{H}$:

$$\tilde{\boldsymbol{H}} = \boldsymbol{H} + \begin{bmatrix} 0 & 0 \\ 0 & \Delta h \end{bmatrix}. \tag{2.69}$$

The inverse of this matrix calculates to

$$\tilde{\boldsymbol{H}}^{-1} = \begin{bmatrix} 1+(e+\Delta h)^{-1} & -(e+\Delta h)^{-1} \\ -(e+\Delta h)^{-1} & (e+\Delta h)^{-1} \end{bmatrix}, \tag{2.70}$$

and the solution of the perturbed system becomes

$$\tilde{\boldsymbol{x}} = \begin{pmatrix} 2 - \dfrac{e}{e+\Delta h} \\ \dfrac{e}{e+\Delta h} \end{pmatrix}. \tag{2.71}$$

If $e = 10^{-3}$ and $\Delta h = -e+10^{-5}$ the solution of the perturbed system becomes

$$\tilde{\boldsymbol{x}} = \begin{pmatrix} -98 \\ 100 \end{pmatrix}. \tag{2.72}$$

Again, the solution of the perturbed system is not even related to the true outcome.

Indeed, one can say that the above given example is ill-conditioned: a small error in either $\boldsymbol{H}$ or $\boldsymbol{b}$ results in huge solution errors.

If the matrix $\boldsymbol{H} \in \mathbb{R}^{m,n}$ with $m \geq n$ has full column rank, the inverse is replaced by the pseudo-inverse $\boldsymbol{H}^{\dagger}$, which minimizes the linear least-squares problem. The perturbed solution

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \Delta\boldsymbol{x} \tag{2.73}$$

to the perturbed system

$$\tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}, \tag{2.74}$$

with $\tilde{\boldsymbol{H}} = \boldsymbol{H} + \Delta\boldsymbol{H}$ and $\tilde{\boldsymbol{b}} = \boldsymbol{b} + \Delta\boldsymbol{b}$, is the sum of the true solution $\boldsymbol{x}$ and an absolute error $\Delta\boldsymbol{x}$. In the following lines an upper bound for the norm of this error will be presented and the condition number is introduced. In the following only induced (or operator) matrix norms are considered, which are induced from vector norms:

$$\|\boldsymbol{H}\| = \sup_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{|\boldsymbol{H}\boldsymbol{x}|}{|\boldsymbol{x}|}. \tag{2.75}$$

These norms have the additional property that they are sub-multiplicative, that is $\|\boldsymbol{H}\boldsymbol{B}\| \leq \|\boldsymbol{H}\|\|\boldsymbol{B}\|$. Using this property the following theorem can be derived.

Theorem 2. *Suppose $\boldsymbol{H} \in \mathbb{R}^{m,n}$, with $m \geq n$, is of full rank. Then, provided that $\|\boldsymbol{H}^{\dagger}\|\|\Delta\boldsymbol{H}\| < 1$, with $\Delta\boldsymbol{H}$ describing an uncertainty about matrix $\boldsymbol{H}$, the relative error of the solution vector $\boldsymbol{x}$ is bounded by*

$$\frac{|\Delta\boldsymbol{x}|}{|\boldsymbol{x}|} \leq \frac{\kappa(\boldsymbol{H})}{1 - \kappa(\boldsymbol{H})\dfrac{\|\Delta\boldsymbol{H}\|}{\|\boldsymbol{H}\|}} \left(\frac{|\Delta\boldsymbol{b}|}{|\boldsymbol{b}|} + \frac{\|\Delta\boldsymbol{H}\|}{\|\boldsymbol{H}\|} \right), \tag{2.76}$$

with

$$\kappa(\boldsymbol{H}) = \|\boldsymbol{H}\|\|\boldsymbol{H}^{\dagger}\| \tag{2.77}$$

being the condition number of $\boldsymbol{H}$.

The proof can be found in the appendix.

Using the sub-multiplicative property of the induced matrix norms it can easily be shown that

$$\kappa(\boldsymbol{H}) \geq 1. \tag{2.78}$$

From now on only the spectral (or Euclidean) norm is considered

$$\|\boldsymbol{H}\|_2 = \sup_{|\boldsymbol{x}|_2=1} |\boldsymbol{H}\boldsymbol{x}|_2. \tag{2.79}$$

In the appendix it will be shown that the spectral norm of $\boldsymbol{H}$ is just its largest singular value σ:

$$\|\boldsymbol{H}\|_2 = \max \sigma_i(\boldsymbol{H}) = \sigma_{\max}. \tag{2.80}$$

Further, for a full rank matrix $\boldsymbol{H}$ the spectral norm of its pseudo-inverse corresponds to the inverse of the smallest singular value of $\boldsymbol{H}$

$$\|\boldsymbol{H}^{\dagger}\|_2 = \frac{1}{\min \sigma_i(\boldsymbol{H})} = \frac{1}{\sigma_{\min}}. \tag{2.81}$$

The corresponding condition number $\kappa_2(\boldsymbol{H})$ then follows to

$$\kappa_2(\boldsymbol{H}) = \frac{\sigma_{\max}}{\sigma_{\min}}. \tag{2.82}$$

Hence, the minimum condition number is obtained if the singular values of $\boldsymbol{H}$ are all identical $\sigma = \sigma_1 = \cdots = \sigma_n$.

Theorem 3. *Assume that $\boldsymbol{H} \in \mathbb{R}^{m,n}$ with $m \geq n$. The singular values of $\boldsymbol{H}$ are identical, $\sigma_1 = \cdots = \sigma_n$ if and only if*

$$\boldsymbol{H}^T\boldsymbol{H} = \sigma^2\boldsymbol{I} \tag{2.83}$$

Proof. Since $\boldsymbol{H}^T\boldsymbol{H}$ is positive definite, its eigenvalues are real values greater than zero and the corresponding eigenvectors orthogonal. Further, it is diagonalizable, such that

$$\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{V}^T\Delta\boldsymbol{V}, \tag{2.84}$$

with Δ being a diagonal matrix, having the squares of the singular values σ_i of $\boldsymbol{H}$ as its entries. $\boldsymbol{V}$ is a matrix consisting of the corresponding eigenvectors of $\boldsymbol{H}^T\boldsymbol{H}$. Assume that all singular values are identical $\sigma = \sigma_1 = \cdots = \sigma_n$, then

$$\boldsymbol{H}^T\boldsymbol{H} = \sigma^2\boldsymbol{V}^T\boldsymbol{V}, \tag{2.85}$$

Since the eigenvectors of $\boldsymbol{H}^T\boldsymbol{H}$ are orthogonal $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$. Thus, $\boldsymbol{H}^T\boldsymbol{H} = \sigma^2\boldsymbol{I}$. □

As a result, searching for a matrix $\boldsymbol{H}$ with $\kappa_2(\boldsymbol{H}) = 1$ is equivalent to searching for a matrix, which fulfills eq. (2.83).

The condition number can directly be applied to the LLSE and the linearized estimator. The smaller the condition number of the observation matrix $\boldsymbol{H}$ of the LLSE, or the Jacobian $\boldsymbol{\mathcal{J}}$ of the linearized estimator, the smaller will be the relative error of the estimate $\hat{\boldsymbol{\theta}}$.

2.4 Optimal Internal Parameter Selection for Localization Estimators

In the preceding section two performance measures, the CRLB and the GDOP, for the statistical estimation problem and the condition number for non statistical estimation problems were introduced. In this section we will discuss how these measures can be used, to optimize the performance of the used estimators. While the three performance measures might be applied to any kind of estimation problem, the problem of source localization is from now on considered. The parameter to be estimated $\boldsymbol{\theta}$ denotes the unknown position of an object. The known internal parameters $\boldsymbol{x}_i$ are the positions of the sensors and the measure $\boldsymbol{x}$ is found by some kind of localization or bearing estimator.

As discussed in the theory of the CRLB, the variance of unbiased estimators is lower bounded by the Fisher information matrix. It was shown that for linear models, the linear model estimator attains that bound, and that for nonlinear models the linearized estimator approximates it, if it is initialized by a position in the neighborhood of the true source location. Evidently, if the trace of the Fisher information matrix is maximized, the trace

of the covariance matrix of these estimators is minimized. An optimal sensor configuration w.r.t. the CRLB can then be stated as

$$\arg\max_{\boldsymbol{x}_i} \left(\text{trace}(\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{x}_i))\right) = \arg\min_{\boldsymbol{x}_i} \left(\text{trace}(\boldsymbol{F}^{-1}(\boldsymbol{\theta}, \boldsymbol{x}_i))\right), , \tag{2.86}$$

with $\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{x}_i)$ being the Fisher information matrix given in eq. (2.19) and $\boldsymbol{x}_i$ being a vector consisting of the sensor positions. Note, that the CRLB is independent of the chosen estimator and that the Fisher information matrix is positive definite, if the covariance matrix of the noise is positive definite $\boldsymbol{C}_n > 0$, which is usually fulfilled, and if the Jacobian $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}, \boldsymbol{x}_i)$ is of full rank. Otherwise, it will be positive semi-definite.

The two other methods can directly be implemented by either minimizing the GDOP of eq. (2.41) for statistical estimators or the condition number of equation (2.77) for the LLSE or the linearized estimator. In optimization theory the function to be minimized is often referred to as the cost function. In the following lines this term will be used to represent either one of the three proposed measures. Since all measures are functions of the source position $\boldsymbol{\theta}$ and the sensor positions $\boldsymbol{x}_i$, it will be denoted by $J(\boldsymbol{x}_i, \boldsymbol{x}_s)$, which is not to be confounded with the Jacobian $\boldsymbol{\mathcal{J}}$.

2.4.1 Comparison of Cost Functions

Which cost function to use, is strongly dependent on the situation, the chosen estimator, and the knowledge about the noise. If an unbiased estimator is known, which attains the CRLB, and further sufficient information about the ambient noise is available, then the minimization of the trace of the Fisher information matrix would probably be the favorable cost function. In this case, the covariance matrix of the estimator is equal to the inverse of the Fisher information matrix. No other unbiased estimator then exists, which provides a better estimate w.r.t. to its covariance.

On the contrary, for biased, or non-efficient estimators the CRLB is probably not the favorable cost function. It offers no measure for biased and gives only a lower bound for unbiased estimators. Assume that an unbiased estimator is used, which does not attain the CRLB (non-efficient). In such a case, it is known that the covariance of the estimator can not be smaller than the Fisher information matrix. However, no upper bound for it is available. Therefore, minimizing the trace of the Fisher information matrix, does not assure a better performance of non-efficient estimators.

For such estimators, the GDOP or the condition number might be the favorable cost function, depending on the information on the ambient noise. If the statistical information about this noise is available, in order to calculate

the GDOP, the GDOP is a powerful measure to carry out the optimization of the sensor geometry.

The condition number has its advantages if little, or no information about the noise is known, so that the CRLB and the GDOP cannot be calculated. Since it offers an upper bound on the relative estimation error, it is well suited as a cost function. However, it is limited to linear or linearized problems, such as the LLSE and the linearized estimator.

Due to the dependency of the proposed cost functions on the source position, a sensor configuration, optimal for one source location, might be inadequate for others.

2.4.2 Quasi-static Source

In this thesis an object, which is only slightly changing its position, is referred to as a quasi-static source. One possible application considering such a quasi-static source might be an automatic camera-steering algorithm for a tele-conference. Assume that the talk of a speaker in a conference is supposed to be transmitted to a distant conference room. In such a case the voice of the speaker as well as a live stream of its presentation would be transmitted over the internet. The speaker is standing most of its time, e.g. behind the lectern, or in front of the blackboard and a camera is supposed to be automatically directed toward his face. For such a scenario an acoustic source localization algorithm could estimate the position of the speaker's mouth and the orientation of the camera could be adjusted accordingly. The optimization procedure for the sensor geometry then becomes straightforward: a representative position, e.g. at the height of the speaker's mouth right behind the lectern, and the estimator, e.g. the linearized estimator, are chosen. Further, the number of sensors is assigned and one of the three presented cost functions is minimized for the selected reference source position.

Optimal Sensor Configuration for the Linearized Estimator

Assume that some kind of position estimation is to be implemented for a quasi-static source. An additive zero-mean, i.i.d. Gaussian noise is considered, with covariance $\boldsymbol{C}_n = \sigma^2 \boldsymbol{I}$ and the linearized estimator of eq. (2.52) is chosen to carry out this procedure. We assume that its initial value $\hat{\boldsymbol{\theta}}_0$ is found by some kind of closed-form estimator, which offers an unbiased estimate of the source position

$$E(\hat{\boldsymbol{\theta}}_0) = \boldsymbol{\theta}. \tag{2.87}$$

Since, statistical information about the additive noise is available, it is reasonable to either choose the Cramer-Rao Lower Bound, or the geometric dilution of precision as the cost function for optimizing the sensor configuration. The Cramer-Rao Lower bound for this unbiased estimator was found in section 2.1.2, eq. (2.21), to be

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} \geq \sigma^2 \cdot (\boldsymbol{\mathcal{J}}^T(\boldsymbol{\theta}, \boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}, \boldsymbol{x}_i))^{-1} \tag{2.88}$$

Further, in section 2.2.3, the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ of the estimate at iteration step $n+1$ of the linearized estimator was found to be (eq. (2.57))

$$\begin{aligned} \boldsymbol{C}_{\hat{\boldsymbol{\theta}}} &\approx \boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)^{\dagger}\boldsymbol{C}_n(\boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)^{\dagger})^T \\ &= \sigma^2 \cdot (\boldsymbol{\mathcal{J}}^T(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i))^{-1}. \end{aligned} \tag{2.89}$$

Note that since the Jacobian is a function of the estimate at iteration step n, this covariance matrix will be a random matrix. Considering the first iteration $\hat{\boldsymbol{\theta}}_1$ it results to

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} \approx \sigma^2 \cdot (\boldsymbol{\mathcal{J}}^T(\hat{\boldsymbol{\theta}}_0, \boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_0, \boldsymbol{x}_i))^{-1}. \tag{2.90}$$

If now the random vector $\hat{\boldsymbol{\theta}}_0$ is replaced by its mean, which above was assumed to be the true vector, equality in eq. (2.88) holds. Consequently, minimizing the CRLB is equal to minimizing the covariance matrix of this estimator. The minimization of a matrix is usually carried out by minimizing its trace. The performance of the estimator is minimized by choosing the optimal internal values collected in $\boldsymbol{x}_i$:

$$\boldsymbol{x}_i = \arg\min_{\boldsymbol{x}_i} \left(\text{trace}(\boldsymbol{\mathcal{J}}^T(\boldsymbol{\theta}, \boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}, \boldsymbol{x}_i))^{-1}\right). \tag{2.91}$$

Just as well the geometric dilution of precision can be used for the presented problem. In section 2.1.5 it was defined by eq. (2.41) to be

$$\text{GDOP}(\hat{\boldsymbol{\theta}}) = \frac{\text{RMSE}(\hat{\boldsymbol{\theta}})}{\text{RMSE}(\boldsymbol{n})}. \tag{2.92}$$

Since, the linearized estimator is unbiased, the root mean square error of the estimator is simply the trace of its covariance matrix

$$\text{RMSE}(\hat{\boldsymbol{\theta}}) = \text{trace}(\sigma^2 \cdot (\boldsymbol{\mathcal{J}}^T(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i))^{-1}), \tag{2.93}$$

and hence the GDOP follows to

$$\text{GDOP}(\hat{\boldsymbol{\theta}}) = \sqrt{\text{trace}(\boldsymbol{\mathcal{J}}^T(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i)\boldsymbol{\mathcal{J}}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{x}_i))^{-1}/p}, \tag{2.94}$$

with p being dimension of the noise vector $\boldsymbol{n}$. Due to its dependency on the random variable $\hat{\boldsymbol{\theta}}_n$, the GDOP is random, too. Arguing as before, the random vector $\hat{\boldsymbol{\theta}}_n$ is replaced by the true unknown parameter $\boldsymbol{\theta}$. Now minimizing the GDOP w.r.t. the internal parameters $\boldsymbol{x}_i$, leads to the same minimization problem as the one using the Cramer-Rao lower bound given by eq. (2.91).

Hence, in the case of additive zero-mean, i.i.d. Gaussian noise, the GDOP of the linearized estimator and the CRLB supply the same cost function for an optimal sensor configuration.

Since,

$$\text{trace}(\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}}) = \sum_{i=1}^{M}\sigma_i^2 \quad \text{and} \quad \text{trace}(\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}})^{-1} = \sum_{i=1}^{M} 1/\sigma_i^2, \tag{2.95}$$

with M being the number of columns of $\boldsymbol{\mathcal{J}}$, it might be of interest to find a sensor configuration, which assures that all the singular values of $\boldsymbol{\mathcal{J}}$ are equal. This can easily be shown using the Cauchy-Schwarz inequality

$$\begin{aligned} M &= [\sigma_1, \ldots, \sigma_M][1/\sigma_1 \ldots 1/\sigma_M]^T \\ &\leq \sqrt{\text{trace}(\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}})} \cdot \sqrt{\text{trace}(\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}})^{-1}} \end{aligned} \tag{2.96}$$

Consequently,

$$\text{trace}(\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}})^{-1} \geq \frac{M^2}{\text{trace}(\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}})}. \tag{2.97}$$

Since, equality holds if and only if $\sigma_1 = \cdots = \sigma_M$, it might be a good choice to choose a configuration, which fulfills this condition. In section 2.3 it was shown that if all singular values of $\boldsymbol{\mathcal{J}}$ are equal, then

$$\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}} = \sigma^2\boldsymbol{I}, \tag{2.98}$$

and consequently, that the condition number of $\boldsymbol{\mathcal{J}}$ is equal to one, which optimizes the sensor configuration w.r.t. the condition number as well.

Note that the condition $\sigma_1 = \cdots = \sigma_M$ is sufficient and necessary for the condition number approach. However, for the CRLB and the GDOP this might generally not be true, even though inequality (2.97), makes it very likely.

Hence, the CRLB and the GDOP deliver the same sensor configurations for the linearized estimator, if additive zero-mean i.i.d. Gaussian noise is considered. Further, it is very likely that the found configuration will also be optimal w.r.t. the condition number of the estimator. In chapter 3 it will be

shown that this is indeed true in case of TDOA-based passive source localization. The optimization of the sensor configuration thus becomes equivalent to finding a setup for the static position $\boldsymbol{\theta}$ such that the singular values of $\boldsymbol{\mathcal{J}}$ become all equal.

Things become more difficult for moving sources, because a trade off between multiple source locations must be found.

2.4.3 Moving Source

While the quasi-static source assumptions is sufficient for a variety of localization problems, most often a moving source must be considered. In robotics a robot must estimate its position, in order to autonomously navigate through a defined or undefined region, without hitting any obstacles. In surveillance applications, the intrusion of an threatening object, or person into an industrial, military, or strongly populated area must be detected and its trajectory must be monitored. In such cases, the selection of a representative position, for which the sensor geometry is optimized, might be difficult, or might even be impossible, because an entire area must usually be covered.

Weighted Average Cost Function

In case of a coverage of multiple positions, trajectories, such as entry paths of airplanes arriving at the airports, or areas, one solution for optimizing the performance of the localization estimator is to discretized the coverage zone. An equi-spaced grid might be placed on it, and the average cost function over the corresponding points could be minimized:

$$\arg\min_{\boldsymbol{x}_i} \sum_{j=1}^{P} w_j \cdot J(\boldsymbol{x}_i, \boldsymbol{x}_{s_j}), \tag{2.99}$$

with P being the number of source positions. Defining different weighting terms w_j, the importance of the individual positions can be taken into account. Positions, which are more important to be well covered, obtain a larger weight than others.

An alternative approach could be based on moving sensors. An optimal sensor geometry could be attained for each source position, by constantly rearranging the sensor in order to achieve an optimal performance. However, such a network would be quite expensive and complex to implement. Each sensor would have to be installed on a robot and the observations would need to be transmitted to a base station. In this thesis, an adaptive sensor network is considered.

Adaptive Sensor Network

The basic idea of an adaptive sensor network is the use of a large number of sensors, from which only a subset, assuring an optimal performance, are active. In case of acoustic passive source localization the sensors are microphones, which are cheap in production. A large number of these can easily be integrated in an estimation system, without causing a cost explosion of the overall system.

If the source is considered to be slowly moving and the estimation rate is considered to be high, then it can be assumed, that the position of the object is not drastically changing from one time instant to another. In this case the position estimate from time instant k is used to reconfigure the sensor network for time instant $k+1$.

Let us consider the problem of the tele-conference, again. This time however, it is not assumed, that the speaker is standing at one position, but that he is constantly moving around in the room. Beforehand, the room is equipped with e.g. 20 microphones, which are installed at convenient positions, such as the walls, the roof, at the edges of tables, etc. Again, a grid over the accessible region of the speaker is placed, and a number of microphones (e.g. 5) for each grid point are chosen, which minimize one of the proposed cost functions. The positions together with their optimal configurations are then stored in a database, which is accessed each time a new position estimate is found. The position in the database closest to the estimated position is determined and its corresponding sensor configuration is selected for carrying out the next position estimation.

This approach can be extended by using the tracking algorithms, which will be presented in chapter 4. A linear or nonlinear state-space model of the source's movement is established. The observation sequence up to time instant k is estimated and the position of time instant $k+1$ is predicted, by means of recursive Bayesian estimation. The Kalman filter is optimal for this prediction, if the source model is linear and the additive Gaussian noise is assumed. In other cases alternative recursive filters, such as the extended Kalman filter, the unscented transformation filter, or particle filters might achieve better results, depending on the modelization of the source's movement and the chosen position estimator. Once, the predicted source position is found, the sensor configuration of the database that is best adapted for the predicted source position, is selected to carry out the upcoming estimation procedure.

2.5 Chapter Summary

This section presented the general ideas of estimating unknown parameters on a series of measures. The statistical approach to estimation theory, based on known statistical information about the measures was presented, and two performance measures, the Cramer Rao lower bound and the geometric dilution of precision were derived. Further, alternative estimators, mainly based on the theory of least-squares were introduced, which can replace the statistical ones in case that no, or only little statistical information about the measures is available. For those non statistical estimators, the condition number was proposed as an adequate performance measures.

All these three measures are functions of the unknown estimation parameters, the observed data, and internal selectable parameters of the estimator. The problem of position estimation was presented, and it was discussed, how the performance of the estimators can be increased, by optimizing one of the performance measures w.r.t. the internal parameters, which in the presented problem are mainly the positions of the sensors.

In the next chapter this general framework will be applied to the problem of passive source localization, based on time difference of arrival measures.

Chapter 3

TDOA-Based Passive Source Localization

The problem of passive source localization is addressed in this chapter. Active sensors, such as radar or active sonar, emit a signal and estimate the position or the bearing of an object by the reflected version of the transmitted signal. Passive sonar does not transmit, but only uses the sound produced by the object captured by spatially separated multiple microphones (in air), or hydrophones (in water).

Sound travels with a finite velocity, which varies for different mediums. In air at a temperature of $20°C$, the speed of sound is around $c \approx 343m/s$. In water, the sound speed mainly depends on the pressure, the ambient temperature, and the salinity. It is typically around $1500m/s$ [Pie81], [NE92].

Due to this limited propagation speed, the sound produced by the object will arrive at the spatially separated microphones at different time instances. These different arrival times can be used to estimate the position of the source.

3.1 Problem Statement

As already discussed in the introductory chapter, the TDOA-based passive position estimator can be seen as a two step procedure. In the first step the time differences of arrival of multiple spatially separated microphone pairs are estimated, which are then used in the second step to estimate the actual position of the source. Due to disturbance noises and reverberant environments, the estimated time differences can be seen as random variables consisting of the true value and an additive noise term. Denoting the source position by $\boldsymbol{x}_s$ and the sensor positions of microphones i and j by $\boldsymbol{x}_i$ and

$\boldsymbol{x}_j$, respectively, the time delay estimate τ_{ij} can be represented as the sum of range difference of the microphone pair

$$d_{ij}(\boldsymbol{x}_s, \boldsymbol{x}_i, \boldsymbol{x}_j) = |\boldsymbol{x}_i - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_j - \boldsymbol{x}_s|_2 \tag{3.1}$$

and an additive noise term n_{ij}:

$$\tau_{ij} = d_{ij}(\boldsymbol{x}_s, \boldsymbol{x}_i, \boldsymbol{x}_j)/c + n_{ij}. \tag{3.2}$$

Note that assuming a three dimensional position estimation, eq. (3.1) describes one half of a hyperboloid with focal points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Consequently, having a single time delay estimate τ_{ij} does not specify a unique source location $\boldsymbol{x}_s$, but only restricts the potential location to lie on the hyperboloid defined by the microphone positions and the time delay estimate. Having multiple microphone pairs, the source location is restricted to lie on all resulting hyperboloids. The estimation of the source's position can thus be seen as the intersection of multiple hyperboloids. In order to obtain a unique position estimate at least three hyperboloids are needed, which means that at least four microphones must be used for a unique three dimensional position estimate. Since the true range differences d_{ij} are unavailable, they are estimated by the time delay estimates τ_{ij} times the propagation speed. However, the additive noise terms influences the shape of the hyperboloids and consequently the estimate of the source position. In order to decrease the influence of these noise terms, the number of microphone pairs is usually chosen greater than three.

Collecting multiple time delay estimates of N microphones in a row vector $\boldsymbol{\tau}$, eq. (3.2) can be expanded to

$$\boldsymbol{\tau} = \boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)/c + \boldsymbol{n}. \tag{3.3}$$

Consequently, with $\boldsymbol{x}_s = \boldsymbol{\theta}$ and $\boldsymbol{\tau} = \boldsymbol{x}$, the TDOA-based position estimation problem is written as the general estimation problem presented in the previous chapter (eq. (2.4)), with

$$\boldsymbol{h}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N). \tag{3.4}$$

Note that the internal parameters of the previous chapter are now the sensor coordinates $\boldsymbol{x}_i$. Since in section 2 it was shown that for the given problem the presented performance measures (the Cramer Rao lower bound, the geometric dilution of precision and the condition number) are all functions of the estimation vector and the internal parameters, the performance of the presented estimators can be increased by choosing optimal internal parameters.

The problem of optimal TDOA-based passive source localization is thus composed of three parts:

1. choose an appropriate time delay estimator in order to obtain the data vector $\boldsymbol{\tau}$

2. choose an appropriate position estimator based on the estimation model of eq. (3.3)

3. choose an appropriate number of sensors N and optimize their configuration

These three points will be addressed in this section. However the focus is directed on points 2 and 3. For completeness three time delay estimators, the generalized cross correlator, the linear mean square error estimator, and the adaptive eigenvalue decomposition estimator, will be presented in section 3.2. Based on a number of simulations, considering varying reverberant environments and varying signal to noise ratios, it will be argued that the found time delay estimates can indeed be represented by eq. (3.3), and that it is reasonable to assume that $\boldsymbol{n}$ is a zero-mean Gaussian random vector. This in mind, the most often utilized position estimators for step two, based on the noisy model of eq. (3.3), are presented in section 3.3. The Cramer Rao lower bound for this problem is derived and it will be shown that for the TDOA-based position estimation problem no unbiased estimator exists, which attains this bound.

The maximum likelihood would then usually be the favorable choice of estimator, since it asymptotically attains the CRLB, when $N \to \infty$. However, the number of used microphones is typically far from infinity. Further, in order to apply the maximum likelihood estimator, the statistical information about the additive noise term $\boldsymbol{n}$ needs to be known. However, for a large number of applications this information is hard to obtain, or will constantly change. Therefore, the linearized estimator is usually the favorable choice of estimator for TDOA-based passive source localization. Both, the maximum likelihood estimator and the linearized estimator, are for the TDOA-based position estimation of iterative nature.

A number of closed-form estimators exist, which can either be used by their own, or can be used to initialize the linearized estimator or the maximum likelihood estimator. Further, if the statistical information about the additive noise is available, these closed-form estimators can be implemented in a way to approximate the maximum likelihood estimator.

Following some estimators' derivations, their bias and covariance matrix will be analytically derived. These terms are not only needed for the evaluation of the geometric dilution of precision, but are also of great interest for the TDOA-based source tracking procedure presented in chapter 4.

In section 3.7 optimal sensor configurations for the individual estimators are derived. While the sensor configurations based on the CRLB are independent of the chosen estimator, the configurations found by either minimizing the geometric dilution of precision or the condition number will depend on the estimator selection.

Finally in section 3.8 an analytic estimator is proposed, which is derived from the optimal sensor configuration of the linear approximation estimators.

3.2 Time Delay Estimation

A wavefront, emanating from a radiating sound source, arrives at spatially separated microphones at different time instances. The difference of the arrival time offer important information on the position of the source. Therefore, it is of fundamental importance to reliably estimate these.

Sound radiating from one point in a closed environment usually arrives at the sensors through multiple paths. The main path is the direct path and is usually the least attenuated. However, due to reflections at e.g. walls, additional versions of the produced sound also arrive at the sensors, having an additional time delay. Since, only a part of the emitted signal hitting a reflector is reflected, with the other part being absorbed, a reflected version arriving at the microphone will usually be more attenuated than the direct path signal. If the sound produced by the object is denoted by $s(t)$, the received signals of two microphones $x_i(t)$ and $x_j(t)$ can be modeled as the convolution between the source signal and the channel impulse responses $h_i(t)$ and $h_j(t)$ from the source to the receivers

$$\begin{aligned} x_i(t) &= h_i(t) * s(t) + n_i(t) \\ x_j(t) &= h_j(t) * s(t) + n_j(t), \end{aligned} \tag{3.5}$$

with $n_i(t)$ and $n_j(t)$ denoting additive ambient noise, and $*$ being the convolution operator.

However, for the derivation of the generalized cross-correlator and the least mean square adaptive (LMS) estimator it is convenient to simplify the multi-path model to a single-path model. This ideal case, assumes that each microphone output is only a delayed and attenuated copy of the source signal corrupted by additive noise. Equation (3.5) then becomes

$$\begin{aligned} x_i(t) &= \alpha_i s(t - T_i) + n_i(t) \\ x_j(t) &= \alpha_j s(t - T_j) + n_j(t), \end{aligned} \tag{3.6}$$

where $\alpha_i, \alpha_j \in [0, 1]$ are the attenuation factors and T_i, T_j correspond to the unknown propagation times from the source to the microphone i and j,

respectively:

$$T_i = |\boldsymbol{x}_i - \boldsymbol{x}_s|_2/c. \tag{3.7}$$

In the following it is assumed that the additive noises are uncorrelated with each other, as well as with the source signal.

3.2.1 Generalized Cross-Correlation

Assuming the idealized propagation model of eq. (3.6) the cross-correlation of the two microphone outputs is given by

$$\begin{aligned} r_{x_i,x_j}(\tau) &= \int_{-\infty}^{\infty} x_i(t)x_j(t+\tau)dt \\ &= \alpha_i\alpha_j \int_{-\infty}^{\infty} s(t-T_i)s(t-T_j+\tau)dt \\ &= \alpha_i \cdot \alpha_j \cdot r_{s,s}(\tau - T_i + T_j), \end{aligned} \tag{3.8}$$

with $r_{s,s}(\tau)$ denoting the autocorrelation of the source signal. Note that use has been made of the assumption that the source signal and the additive noise terms are uncorrelated.

Using the Cauchy-Schwarz inequality it can be shown that the autocorrelation of any signal $s(t)$ has the property that [GG95]

$$r_{ss}(\tau) \leq r_{ss}(0) \qquad \forall\tau. \tag{3.9}$$

Assuming a non-periodic source signal $s(t)$ equality in (3.9) cannot be obtained and "$\leq$" can be replaced by "$<$". Consequently, $r_{x_i,x_j}(\tau)$ will attain its maximum value at

$$\tau = T_i - T_j, \tag{3.10}$$

which is equal to the time difference of arrival of the microphone pair i and j. Hence, the TDOA τ_{ij} of a microphone pair i and j can be calculated by

$$\tau_{ij} = \arg\max_{\tau} r_{x_i,x_j}(\tau). \tag{3.11}$$

Depending on the signal $s(t)$ the correlation function might have multiple local maxima and will be flat or steep in the neighborhood of τ_{ij}. A flat function, or one with multiple local maxima will make the search for the maximum value more difficult, and might easily result in a wrong estimation outcome. Consequently, the resulting correlation function would ideally be a time-shifted Dirac function.

The cross spectral density function R_{x_i,x_j} is defined as the Fourier transform of the cross-correlation function of $x_i(t)$ and $x_j(t)$

$$\begin{aligned} R_{x_i,x_j}(\omega) &= \int_{-\infty}^{\infty} r_{x_i,x_j}(\tau) e^{-j\omega\tau} d\tau \\ &= X_i(\omega) X_j^*(\omega) \\ &= S(\omega) S^*(\omega) e^{-j\omega(T_i - T_j)} \end{aligned} \tag{3.12}$$

with $X_i(\omega)$ being the Fourier transform of $x_i(t)$, $X_j^*(\omega)$ being the conjugate complex Fourier transform of $x_j(t)$, and $S(\omega)$ being the Fourier transform of the source signal $s(t)$. In order to obtain a sharp cross-correlation function, a variety of filters have been proposed, which are based on this cross spectral density function. Multiplying the cross spectral density function in eq. (3.12) by the Fourier transform $\Xi(\omega)$ of a filter $\xi(t)$ and applying the inverse Fourier transform to this product, leads to the so-called generalized cross-correlation function

$$g_{x_i x_j}(\tau) = \int_{-\infty}^{\infty} \Xi(\omega) S(\omega) S^*(\omega) e^{-j\omega(T_i - T_j)} e^{j\omega\tau} d\omega. \tag{3.13}$$

If ideally $\Xi(\omega)$ would be chosen to

$$\Xi(\omega) = \frac{1}{|S(\omega) S^*(\omega)|} \tag{3.14}$$

the generalized cross-correlation function would result in the desired time shifted Dirac function

$$g_{x_i x_j}(\tau) = \int_{-\infty}^{\infty} e^{-j\omega(T_i - T_j)} e^{j\omega\tau} d\omega = \delta(t - \tau_{ij}). \tag{3.15}$$

However, the cross spectral density function of the source signal is usually unavailable, and a filter must be designed based on the available signals $x_i(t)$ and $x_j(t)$. Commonly used filters are the Roth processor [Rot71], the smoothed coherence transform (SCOT) [CNC73], the Eckart filter, the phase transform (PHAT), and the maximum likelihood (ML) processor [KC76]. Knapp and Carter [KC76] give an overview of these filters.

In literature the maximum likelihood and the PHAT filter seem to be the favorable choices. Champagne et al. [BCS94, CBS96] tested the performance of the ML estimator in reverberant environments. They found that for small reverberation times, this estimator shows an efficient performance, meaning that it attains the Cramer Rao Lower Bound. However, once the reverberation time increases, its performance drastically decreases and no reliable TDOA measurements can be carried out. On the contrary Gustafsson et al. [GRT02], [GRT03] showed that the PHAT filter is optimal among the class of

cross-correlation based time delay-estimators for reverberant environments. It is given by the transfer function

$$\Xi(\omega) = \frac{1}{|R_{x_i,x_j}(\omega)|}. \tag{3.16}$$

While in theory this filter perfectly whitens the spectrum, and hence results in a Dirac function in the time domain, in practice only an approximation of $R_{x_ix_j}(\omega)$ can be calculated. Obviously, in practical applications the time signals $x_i(t)$ and $x_j(t)$ are not available from $t = -\infty$ to $t = \infty$, but only a finite version, obtained by windowing the signals. Further, such systems are usually implemented using digital computations and hence, the Fourier transform is usually approximated by the fast Fourier transform (FFT) and its inverse (IFFT).

Consequently the true cross spectral density $R_{x_ix_j}(\omega)$ can only be approximated by the discrete finite time series, whose sampling frequency and window size are of crucial importance. If the sampling frequency is chosen too small, aliasing effects can ruin the estimation procedure. Further, since the time delay is estimated by the maximum of the now discrete time series, its resolution is the inverse of the sampling frequency. Contrary, if the sampling frequency is chosen to be large, a large number of data must be processed, which makes the estimation procedure time consuming. A large window time will have the same influence on the computation time. Further, if a moving source is considered, Doppler shifts can influence the estimation accuracy. For smaller window sizes, the computational time will decrease and the moving source might be considered to be stationary.

3.2.2 Least Mean Square Adaptive Estimator

The least mean square (LMS) estimator proposed by Reed et al. [RFB81] also considers the ideal signal model in discrete form

$$\begin{aligned} x_i[k] &= \alpha_i s[k - T_i] + n_i[k] \\ x_j[k] &= \alpha_j s[k - T_j] + n_j[k], \end{aligned} \tag{3.17}$$

where k is the discrete time instant. It adaptively estimates a channel impulse response $\hat{\boldsymbol{h}}$, between the microphone signals

$$x_i[k] = \hat{\boldsymbol{h}}^T \boldsymbol{x}_j[k], \tag{3.18}$$

and finds the time delay of this sensor pair as the largest component of this filter. Considering the ideal discrete model of eq. (3.17) the microphone

output $x_i[k]$ can be written as a attenuated delayed version of the microphone signal $x_j[k]$

$$x_i[k] = \alpha x_j[k - \tau_{ij}] + n[k], \tag{3.19}$$

with $\alpha = \alpha_i/\alpha_j$, and $n[k] = n_i[k] - n_j[k - \tau_{ij}]/\alpha_j$. Assuming a windowed signal

$$\boldsymbol{x}_j[k] = (x_j[k - L], \ldots, x_j[k + L])^T, \tag{3.20}$$

eq. (3.19) can be written as

$$x_i[k] = \boldsymbol{h}^T \boldsymbol{x}_j[k] + n[k], \tag{3.21}$$

with

$$\boldsymbol{h}^T = (0, \ldots, 0, \alpha, 0, \ldots, 0) \tag{3.22}$$

being a finite impulse response filter of length $2L+1$, with entry α at position $\tau_{ij} + L$.

The goal of the LMS estimator is now estimate $\hat{\boldsymbol{h}}$, which will then lead to the following time delay estimate

$$\tau_{ij} = \arg\max |\hat{\boldsymbol{h}}| - L. \tag{3.23}$$

In order to find this estimate an error signal can be formulated as

$$e[k] = x_i[k] - \boldsymbol{h}^T[k] \boldsymbol{x}_j[k]. \tag{3.24}$$

The linear mean-square estimate of $\boldsymbol{h}$ is achieved by minimizing $E(e^2[k])$, using either a batch or an adaptive algorithm. In [CBH07] the adaptive algorithm

$$\hat{\boldsymbol{h}}[k + 1] = \hat{\boldsymbol{h}}[k] + \mu e[k] \boldsymbol{x}_j[k], \tag{3.25}$$

with μ being a small positive adaptation step size, is proposed for carrying out this estimation procedure.

The problem of the two so far presented time delay estimators (TDE) is that they both assume an ideal transmission model, and do not consider possible reverberations.

3.2.3 Adaptive Eigenvalue Decomposition

Benesty [Ben00] proposed a method that considers the more realistic sound propagation model of eq. (3.5) in discrete form

$$\begin{aligned} x_i[k] &= \boldsymbol{h}_i^T \boldsymbol{s}[k] + n_i[k] \\ x_j[k] &= \boldsymbol{h}_j^T \boldsymbol{s}[k] + n_j[k], \end{aligned} \tag{3.26}$$

with $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$ being two finite impulse responses of length L and

$$\boldsymbol{s}[k] = (s[k], s[k-1], \ldots, s[k-L+1])\,. \tag{3.27}$$

As opposed to the LMS estimator, the channel impulse responses $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$ are now assumed to have nonzero entries, due to the considered reverberations. Usually the entries of $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$ corresponding to the direct path of the signal to the microphones, will have the maximum value, because the direct path signals will usually be less attenuated than the signals resulting from reflections.

The goal of the adaptive eigenvalue decomposition (AED) estimator is thus to blindly estimate the channel impulse responses $\boldsymbol{h}_i(t)$ and $\boldsymbol{h}_j(t)$ and then infer the time delay estimate by

$$\tau_{ij} = \arg\max_t |\hat{\boldsymbol{h}}_i| - \arg\max_t |\hat{\boldsymbol{h}}_j|, \tag{3.28}$$

where $\hat{\boldsymbol{h}}_n$ denotes the estimate of the nth channel impulse response.

Writing the received microphone signals of length L in vector notation leads to

$$\boldsymbol{x}_i[k] = \boldsymbol{S}[k] \cdot \boldsymbol{h}_i + \boldsymbol{n}_i[k] \tag{3.29}$$

$$\boldsymbol{x}_i[k] = \boldsymbol{S}[k] \cdot \boldsymbol{h}_i + \boldsymbol{n}_i[k], \tag{3.30}$$

with

$$\boldsymbol{x}_i[k] = \begin{pmatrix} x_i[k], & x_i[k-1], & \ldots, & x_i[k-L+1] \end{pmatrix}^T \tag{3.31}$$

$$\boldsymbol{x}_j[k] = \begin{pmatrix} x_j[k], & x_j[k-1], & \ldots, & x_j[k-L+1] \end{pmatrix}^T \tag{3.32}$$

$$\boldsymbol{S}[k] = \begin{bmatrix} s[k] & s[k-1] & \ldots, & s[k-L+1] \\ s[k-1] & s[k-2] & \ldots, & s[k-L] \\ \vdots & \vdots & \vdots & \vdots \\ s[k-L+1] & s[k-L] & \ldots & s[k-2L+2] \end{bmatrix}, \tag{3.33}$$

$$\boldsymbol{n}_i[k] = (n_i[k], \ldots, n_i[k-L+1])^T, \tag{3.34}$$

$$\boldsymbol{n}_j[k] = (n_j[k], \ldots, n_j[k-L+1])^T. \tag{3.35}$$

Since $\boldsymbol{S}$ is a symmetric matrix,

$$\boldsymbol{h}_i^T \boldsymbol{S} \boldsymbol{h}_j = \boldsymbol{h}_j^T \boldsymbol{S} \boldsymbol{h}_i, \tag{3.36}$$

and consequently

$$\boldsymbol{x}^T[k]\boldsymbol{u} = (\boldsymbol{n}_j^T[k], \boldsymbol{n}_i^T[k])\boldsymbol{u}, \tag{3.37}$$

with

$$\boldsymbol{x}[k] = \begin{pmatrix} \boldsymbol{x}_i[k] \\ \boldsymbol{x}_j[k] \end{pmatrix}, \tag{3.38}$$

and

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{h}_j^T \\ -\boldsymbol{h}_i \end{pmatrix}. \tag{3.39}$$

Assuming that the noise terms $\boldsymbol{n}_{i,j}[k]$ and the microphone vectors $\boldsymbol{x}_i^T[k]$ are uncorrelated, left multiplying eq. (3.37) by $\boldsymbol{x}[k]$ and taking the expectation yields

$$\boldsymbol{R}_x\boldsymbol{u} = \sigma^2\boldsymbol{u}, \tag{3.40}$$

where $\boldsymbol{R}_x = E(\boldsymbol{x}[k]\boldsymbol{x}^T[k])$ is the covariance matrix of $\boldsymbol{x}[k]$, and σ^2 the variance of the individual noise terms

$$\sigma^2 = E(n_n[k]^2). \tag{3.41}$$

The collection $\boldsymbol{u}$ of two channel impulse responses can thus be seen as the eigenvector of $\boldsymbol{R}_x$ corresponding to the eigenvalue $\lambda = \sigma^2$. In [HBC06] the authors argue that this eigenvalue is the smallest eigenvalue of $\boldsymbol{R}_x$ and offer the following adaptive algorithm for finding the corresponding eigenvector and consequently the desired channel impulse responses collected in $\boldsymbol{u}$:

$$\hat{\boldsymbol{u}}[k+1] = \frac{\hat{\boldsymbol{u}}[k] - \mu e[k]\boldsymbol{x}[k]}{\|\hat{\boldsymbol{u}}[k] - \mu e[k]\boldsymbol{x}[k]\|_2}, \tag{3.42}$$

with the constraint that $\|\hat{\boldsymbol{u}}[k]\|_2 = 1$,

$$e[k] = \hat{\boldsymbol{u}}^T[k]\boldsymbol{x}[k] \tag{3.43}$$

is an error signal and μ a constant, positive adaptation step. With the estimated impulse responses collected in $\hat{\boldsymbol{u}}$, the time delay estimate can be found by eq. (3.28).

3.2.4 Time Delay Estimator Evaluation

The three presented TDE are compared by a series of computer simulations, considering different signal to noise ratios (SNRs) and different reverberation times. The reverberation time RT_{60} is a measure for the magnitude of reverberation and is defined as the required time for the sound pressure level in a room to decay to a value one millionth of its original intensity, or to drop $60dB$. The simulation setup is taken from [CBH07]. A rectangular room, with plane surfaces, is considered. All the walls, as well as the ceiling and the floor have identical reflection coefficients, which can vary between 0 and

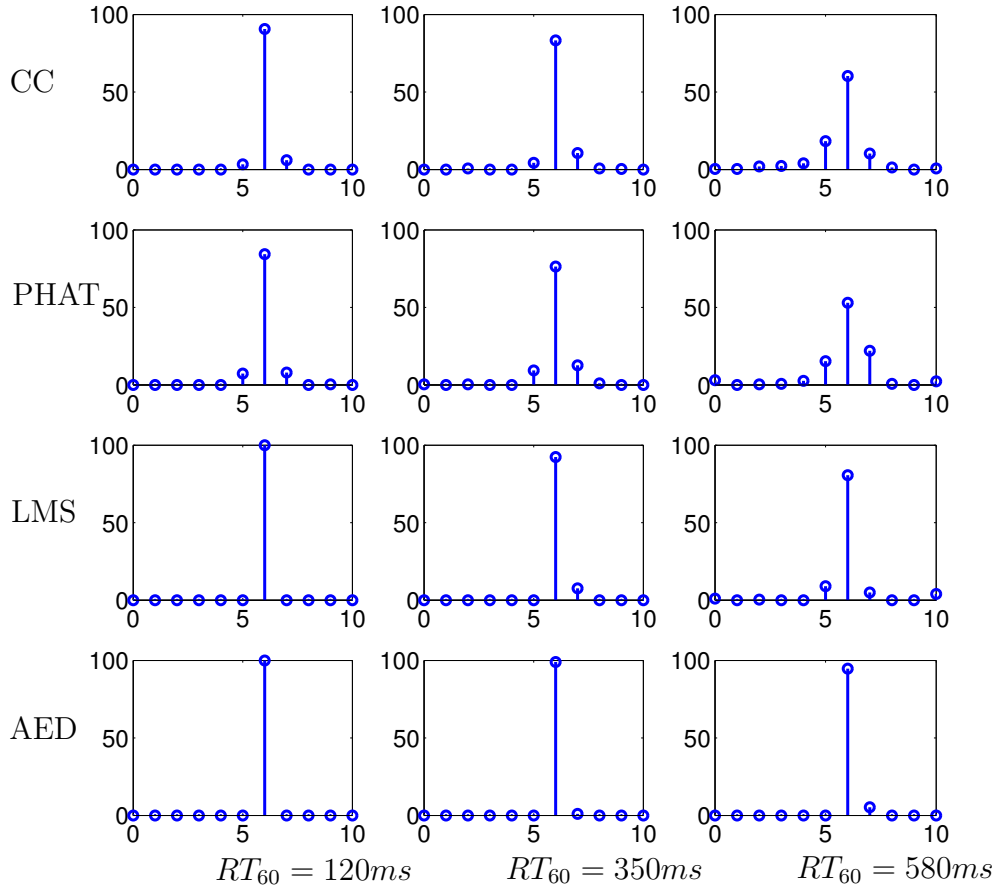

Figure 3.1: time delay estimator comparison for varying reverberation times

1. The dimensions of the room are given by $3.05m \times 4.57m \times 3.81m$. While the source position is set to be $2.54m \times 2.54m \times 1.02m$, two microphone positions are set to $0.51m \times 0.254m \times 1.02m$ and $0.71m \times 0.254m \times 1.02m$, respectively. The room impulse responses from the source to the two sensors are calculated by the image method, due to Allen and Berkley [AB79], for 3 different reflection coefficients. The resulting room impulse responses have reverberation times of $RT_{60} = 120ms$, $RT_{60} = 350ms$, and $RT_{60} = 580ms$.

John F. Kennedy's speech "Ich bin ein Berliner" was samples with 16 bits and a frequency of $f_s = 16kHz$ and the microphone inputs were calculated by the convolution of this signal and the individual room impulse responses. Zero-mean, white Gaussian noise was added to the signal, resulting in SNRs

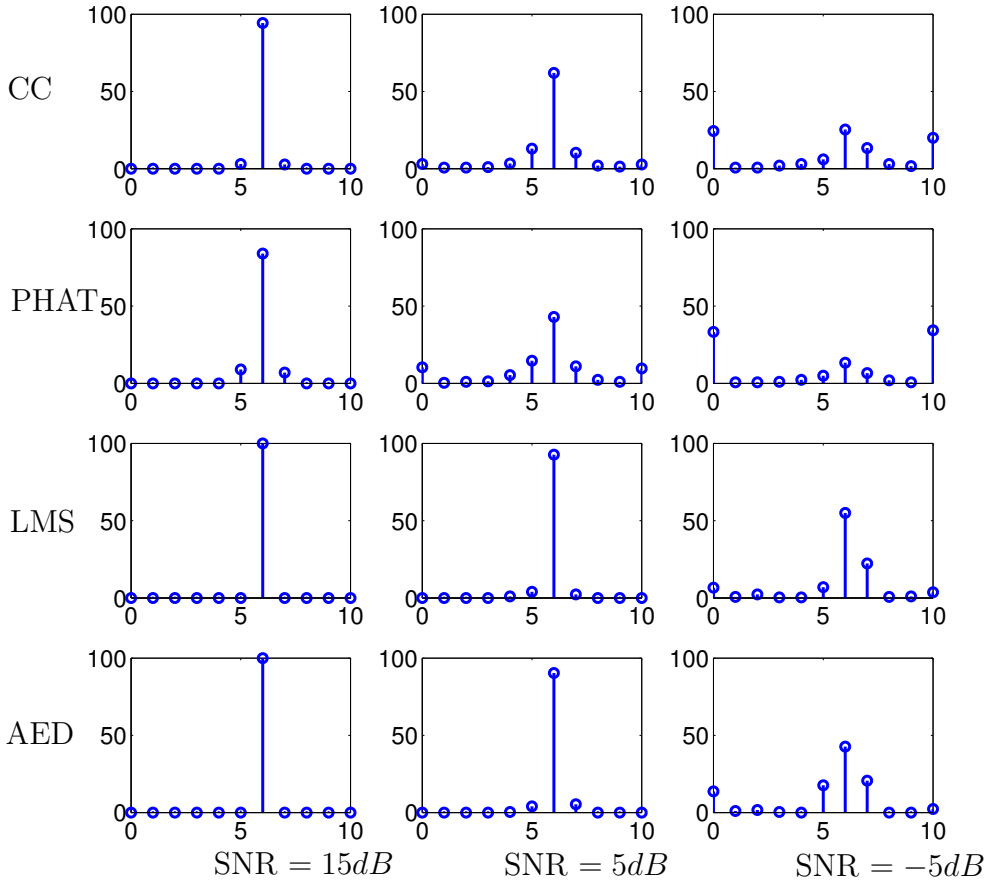

Figure 3.2: time delay estimator comparison for varying signal to noise ratios

of $15dB$, $5dB$, and $-5dB$.

The filter lengths $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ were set to 1024 for the AED, as well as for the LMS estimator. The step sizes for the adaptive updates were chosen for the AED to be $\mu = 0.01$ and for the LMS to be $\mu = 0.001$. The correlator-based estimators were carried out using a uniform window of length 1024.

Figure 3.1 displays the first series of simulations, comparing the performance of the four TDEs w.r.t. the reverberation time, using 300 trials each. Note that the true TDOA calculates to

$$\tau_{21} = d_{21}/c \approx 38ms, \tag{3.44}$$

which corresponds to sample $k = \tau_{21} \cdot f_s \approx 6$. Since the AED is the only TDE

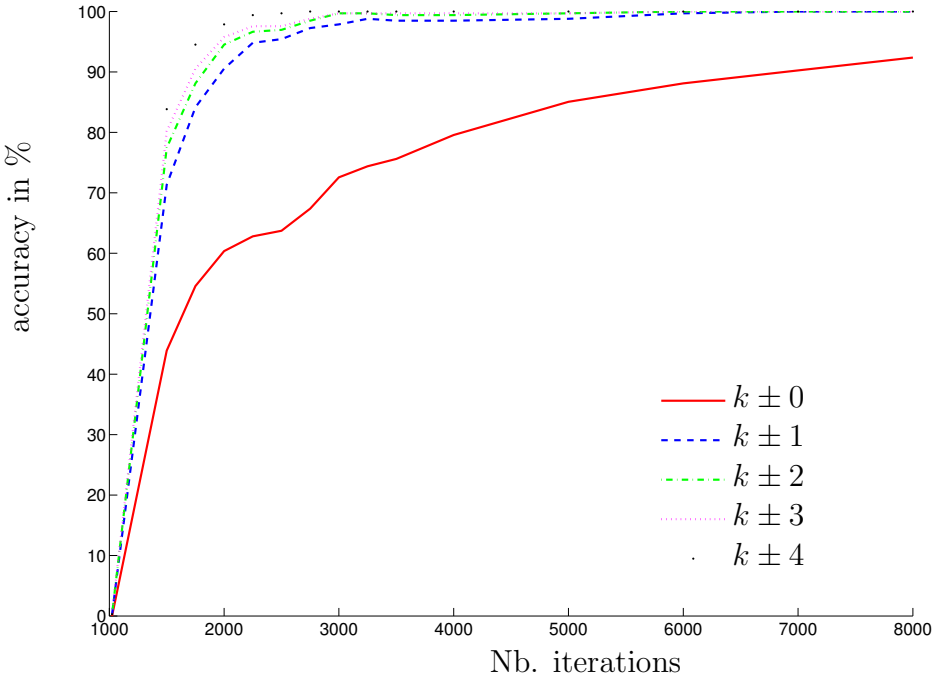

Figure 3.3: performance evaluation of the AED estimator for varying iterations, $k \pm i$: TDOA estimate is accurate within $\pm i$ samples

which considers the room reverberations, it outperforms the three others, as expected. Also the LMS estimator seems to be quite robust against multipath problems.

The second series of simulations studies the effects of the additive noise. Its results are shown in fig 3.2. None of the presented estimators have problems of estimating the TDOAs for moderate sensor noise, resulting in a SNR of $15dB$. However, as the SNR decreases, so does the performance of all the presented estimators. Again, the AED and the LMS estimator seem to outperform the correlation-based estimators. However, due to their iterative nature, they are computationally quite expensive. As can be seen in figures 3.3 and 3.4, a large number of iterations is needed in order to obtain the right TDOA estimates. At a sampling frequency of $f_s = 16000kHz$ a TDOA which is wrong estimated by one sample corresponds to an range difference error of $1 \cdot c/f_s \approx 2cm$. Such an error might still be acceptable for carrying out a position estimation. Therefore, figures 3.3 and 3.4 also present the estimation accuracy if the TDOAs are estimated within a sample precision. For

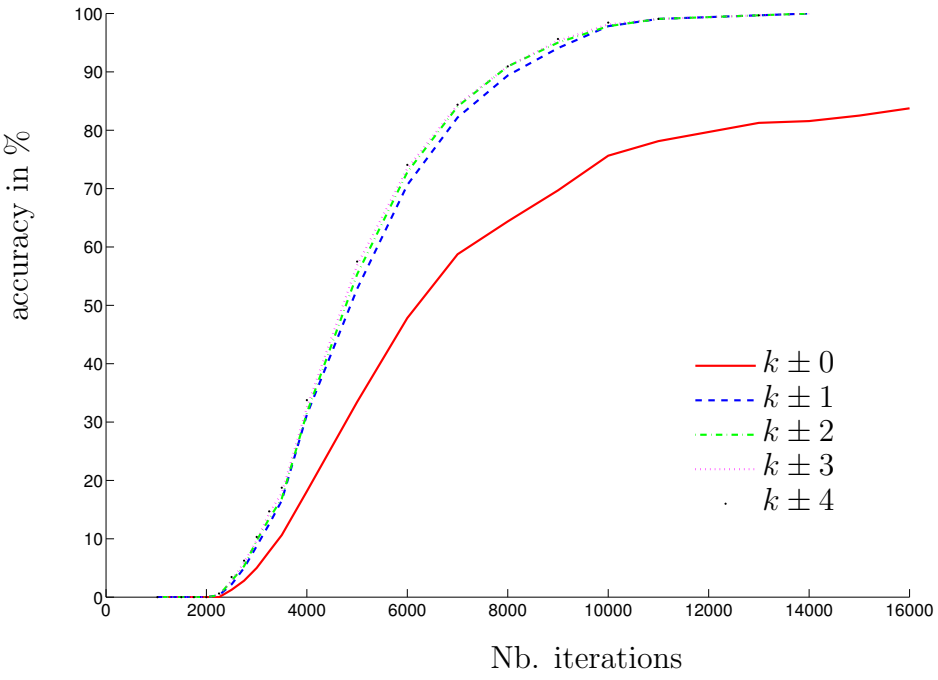

Figure 3.4: performance evaluation of the LMS estimator for varying iterations, $k \pm i$: TDOA estimate is accurate within $\pm i$ samples

real-time implementations, the number of iterations and the size of filters for the AED and the LMS estimator must be kept as small as possible. In such cases, the performance of the LMS and AED estimators might drastically decrease, and the correlation-based estimators might be favorable.

3.2.5 Time Delay Estimation Summary

In this section the most widely used time delay estimators were presented and compared to each other. It is generally not possible to conclude that one method outperforms the others. Rather, the selection of the best adapted estimator, depends on the room geometry, on the selection of used components, and possible restrictions on the calculation time.

Looking at the shape of figures 3.1 and 3.2 it seems reasonable to represent the found time delay estimates as a sum of the true outcome and a random variable, as it was assumed in eq. (3.2):

$$\tau_{ij} = d_{ij}/c + n_{ij}. \tag{3.45}$$

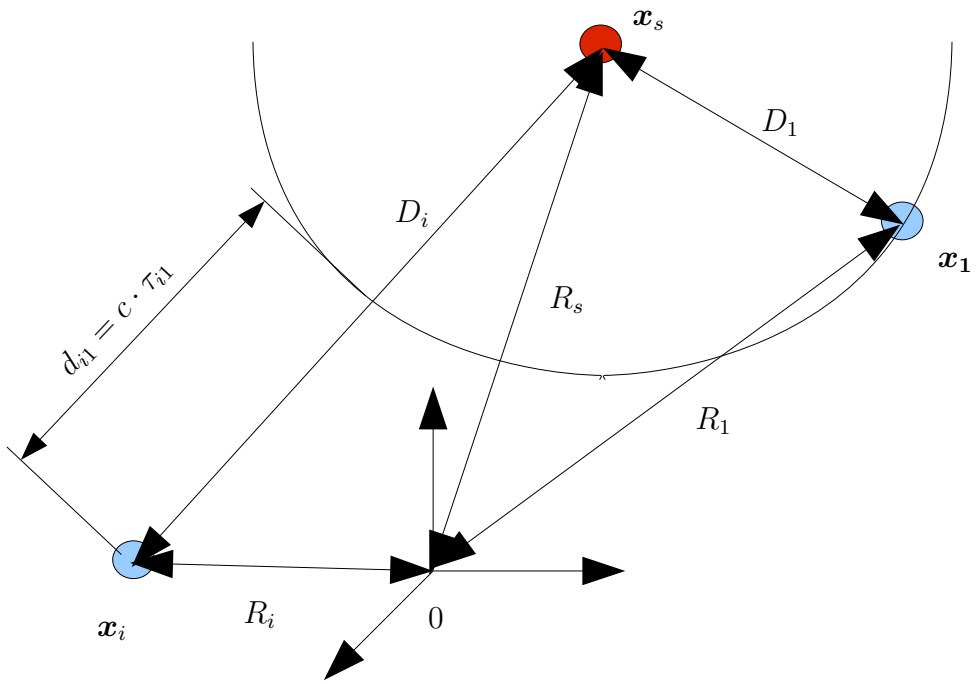

Figure 3.5: Passive Source Localization Setup

Further, by inspection, the random variable might be seen as being zero-mean Gaussian distributed. Consequently, from now on it is assumed that the time delay estimation vector $\boldsymbol{\tau}$ has already been estimated and is treated as a random vector with its mean being equal to the true vector.

3.3 TDOA-based Measurement Model

Based on the argumentation of the previous section that the time delay estimates can indeed be seen as the sum of the corresponding range difference d_{ij} divided by the propagation speed and a zero-mean noise term n_{ij}, the measurement model for TDOA-based position estimation is now derived, on which the presented estimators of chapter 2 are then applied. Above it was argued that for a three dimensional position estimate at least four microphones are necessary. However, in order to decrease the influence of the additive noise terms more microphones are usually utilized.

Assuming N spatially distributed microphones, $N(N-1)/2$ distinct TDOAs

can be calculated, knowing that $\tau_{ii} = 0, \forall i$ and noting that $\tau_{ij} = -\tau_{ji}$. However, if the additional noise in eq. (3.45) is neglectable, $(N-1)$ independent TDOA measures are sufficient to define all the others. Denoting the distance between the microphone i and the source by

$$D_i = |\boldsymbol{x}_i - \boldsymbol{x}_s|_2, \tag{3.46}$$

eq. (3.45) can be written as

$$\begin{aligned} \tau_{ij} &= (D_i - D_j)/c \\ &= (D_i - D_k - (D_j - D_k))/c \\ &= \tau_{ik} - \tau_{jk}. \end{aligned} \tag{3.47}$$

Thus, selecting a reference sensor, the TDOA measures w.r.t. to this microphone define a minimum spanning tree. In case of small ambient noise it might thus be sufficient to calculate the $(N-1)$ TDOAs w.r.t. a reference sensor. Combining these in vector notation

$$\boldsymbol{\tau}_1 = (\tau_{21}, \ldots, \tau_{N1})^T, \tag{3.48}$$

and defining the range difference (RD) vector

$$\boldsymbol{d}_1(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = (d_{21}, \ldots, d_{N1})^T, \tag{3.49}$$

the reference sensor based measurement model becomes

$$\boldsymbol{\tau}_1 = \boldsymbol{d}_1(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)/c + \boldsymbol{n}_1. \tag{3.50}$$

In case of larger ambient noise it might be favorable to utilize all the $N(N-1)/2$ TDOA measures. In fact Yang and Scheuing [YS05] proved that using all $N(N-1)/2$ time delay estimates result in a more accurate position estimate. In this case the measurement model of eq. (3.50) is modified to

$$\boldsymbol{\tau} = \boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)/c + \boldsymbol{n}. \tag{3.51}$$

with

$$\boldsymbol{\tau} = \boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)/c + \boldsymbol{n}. \tag{3.52}$$

and

$$\boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = (d_{21}, d_{31}, d_{32}, \ldots, d_{N,N-1})^T. \tag{3.53}$$

For shorter notations the dependency of $\boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ on the source and microphone positions will from now on either be written as

$$\boldsymbol{d}(\boldsymbol{x}_s, \boldsymbol{X}),$$

with

$$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N], \tag{3.54}$$

or will completly be omitted.

3.3.1 Cramer Rao Lower Bound

In the previous section two possible measurement models were derived, one with respect to a reference sensor (eq. (3.50)) and one considering all $N(N-1)/2$ available time delay estimates (eq. (3.51)). Due to the greater estimation accuracy, the model of eq. (3.51) is now considered. Further it is assumed that $\boldsymbol{n}$ is drawn from a zero-mean Gaussian distribution with covariance $\boldsymbol{C}_n$, as argued in section 3.2.

It was shown in section 2.1.2 that the Fisher information matrix of the CRLB for such a problem is given by

$$\boldsymbol{F}(\boldsymbol{x}_s, \boldsymbol{X}) = \boldsymbol{\mathcal{J}}^T(\boldsymbol{x}_s, \boldsymbol{X}) \boldsymbol{C}_n^{-1} \boldsymbol{\mathcal{J}}(\boldsymbol{x}_s, \boldsymbol{X}), \tag{3.55}$$

with $\boldsymbol{\mathcal{J}}(\boldsymbol{x}_s, \boldsymbol{X})$ being the Jacobian of $(\boldsymbol{d}/c)$ w.r.t. the source position $\boldsymbol{x}_s$. The individual entries of $\boldsymbol{d}$ are the range differences $d_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_j - \boldsymbol{x}_s|_2$, and its partial derivative w.r.t. $\boldsymbol{x}_s$ follow to

$$\boldsymbol{g}_{ij} = \left(\frac{\partial d_{ij}}{\partial \boldsymbol{x}_s} \right)^T = \left(\frac{\boldsymbol{x}_i - \boldsymbol{x}_s}{|\boldsymbol{x}_i - \boldsymbol{x}_s|_2} - \frac{\boldsymbol{x}_j - \boldsymbol{x}_s}{|\boldsymbol{x}_j - \boldsymbol{x}_s|_2} \right). \tag{3.56}$$

The Jacobian of $\boldsymbol{d}$ then results in

$$\boldsymbol{\mathcal{J}}(\boldsymbol{x}_s, \boldsymbol{X}) = \frac{1}{c} \left[\boldsymbol{g}_{21}, \boldsymbol{g}_{31}, \boldsymbol{g}_{32}, \dots \boldsymbol{g}_{N,N-1} \right]^T. \tag{3.57}$$

Consequently, the covariance matrix $\mathrm{Cov}(\boldsymbol{x}_s)$ of any unbiased TDOA-based position estimator is lower bounded by

$$\mathrm{Cov}(\boldsymbol{x}_s) - (\boldsymbol{\mathcal{J}}^T(\boldsymbol{x}_s, \boldsymbol{X}) \boldsymbol{C}_n^{-1} \boldsymbol{\mathcal{J}}(\boldsymbol{x}_s, \boldsymbol{X}))^{-1} \geq \boldsymbol{0}. \tag{3.58}$$

Further theorem 1 states that equality holds if and only if the partial derivative of $\ln p(\boldsymbol{\tau}; \boldsymbol{x}_s)$

$$\frac{\partial \ln p(\boldsymbol{\tau}; \boldsymbol{x}_s)}{\partial \boldsymbol{x}_s}$$

can be broad in the following form

$$\left(\frac{\partial \ln p(\boldsymbol{\tau}; \boldsymbol{x}_s)}{\partial \boldsymbol{x}_s} \right)^T = \boldsymbol{F}(\boldsymbol{x}_s) \left(\boldsymbol{g}(\boldsymbol{\tau}) - \boldsymbol{x}_s \right) \tag{3.59}$$

for some p-dimensional function $\boldsymbol{g}$ and some $p \times p$ matrix $\boldsymbol{F}$. That estimator, which would then be the minimum variance unbiased (MVU) estimator, would be given by $\boldsymbol{x}_s = \boldsymbol{g}(\boldsymbol{\tau})$, and its covariance matrix by $\boldsymbol{F}^{-1}(\boldsymbol{\theta})$.

Theorem 4. *Assuming a Gaussian error model of the form*

$$\boldsymbol{\tau} = \boldsymbol{d}/c + \boldsymbol{n}, \tag{3.60}$$

with the covariance matrix of $\boldsymbol{n}$ *being denoted by* $\boldsymbol{C}_n$*, no unbiased estimator attains the Cramer Rao lower bound.*

Proof. Using eq. (2.34) the partial derivative of $\ln p(\boldsymbol{\tau}; \boldsymbol{x}_s)$ w.r.t. $\boldsymbol{x}_s$ is found to be

$$\left(\frac{\partial \ln p(\boldsymbol{\tau}; \boldsymbol{x}_s)}{\partial \boldsymbol{x}_s}\right)^T = \boldsymbol{\mathcal{J}}^T(\boldsymbol{x}_s, \boldsymbol{X})\boldsymbol{C}_n^{-1}(\boldsymbol{\tau} - \boldsymbol{d}/c). \tag{3.61}$$

The unknown source position $\boldsymbol{x}_s$ appears in eq. (3.61) as a nonlinear term in $\boldsymbol{d}$, and consequently eq. (3.61) is not of the same form as eq. (3.59), which proofs the theorem. □

3.4 Iterative TDOA-based Estimators

In the previous section it was shown that no unbiased estimator can attain the Cramer Rao lower bound. In such a case the most often utilized estimator is the maximum likelihood estimator, since it approximates this bound, if the number of measurements tends to infinity. However, in case of TDOA-based passive source localization that would mean, that the TDOA vector $\boldsymbol{\tau}$ would need to tend to infinity and consequently the number of utilized microphones must tend to infinity. This obviously contradicts with the demand of a low cost system and the possible need for a realtime implementation. Further, the covariance matrix of the additive noise needs to be known. However, in acoustic environments this matrix is usually constantly changing. In outdoor applications possible influences on this covariance matrix are changing wind speeds and the number of present vehicles. In indoor applications such as teleconferences, the number of present spectators, possibly utilizing notebooks equipped with noisy fans, and the state of the windows, closed or open, result in varying signal to noise ratios.

Least-squares estimators enjoy a large attention in TDOA-based position estimation, since they do not need any information about the ambient noise.

3.4.1 Maximum Likelihood Estimator

Assuming an additive Gaussian noise vector $\boldsymbol{n}$, with covariance matrix $\boldsymbol{C}_n$, the maximum likelihood estimator introduced in section 2.1.4 can readily be applied to the problem of TDOA-based source localization, using either one of the measurement models of eqs. (3.50) and (3.51). In the following

the complete range difference vector $\boldsymbol{d}$, considering all $N(N-1)/2$ different TDOA estimates is selected due to its higher precision. Assuming that the covariance matrix $\boldsymbol{C}_n$ is independent of the source position $\boldsymbol{x}_s$, the maximum likelihood estimator follows from eq. (2.31):

$$\hat{\boldsymbol{x}}_s = \arg\min_{\boldsymbol{x}_s} (\boldsymbol{\tau} - \boldsymbol{d}/c)^T \boldsymbol{C}_n^{-1} (\boldsymbol{\tau} - \boldsymbol{d}/c). \tag{3.62}$$

3.4.2 Linearized Estimator

The linearized estimator was introduced in chapter 2.2.3 as an alternative to the maximum likelihood estimator, for estimation problems, in which the noise's covariance matrix is unknown, or constantly changing. This iterative method can directly be derived from eq. 2.52 for the TDOA-based localization approach:

$$\boldsymbol{x}_{s_{n+1}} = \boldsymbol{x}_{s_n} + \boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{x}_{s_n}, \boldsymbol{X})(\boldsymbol{\tau} - \boldsymbol{d}(\boldsymbol{x}_{s_n})/c). \tag{3.63}$$

Bias and Covariance Assuming an additive Gaussian noise and using eqs. (2.56) and (2.57) the TDOA-based linearized estimator is seen to be unbiased and its covariance matrix follows to

$$\mathrm{Cov}(\boldsymbol{x}_{s_{n+1}}) \approx \boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{x}_{s_n}, \boldsymbol{X}) \boldsymbol{C}_n (\boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{x}_{s_n}, \boldsymbol{X}))^T. \tag{3.64}$$

In section 2.2.3 it was pointed out that the covariance matrix of the linearized estimator and the inverse of Fisher information matrix of the problem are practically identical, if the noise's covariance matrix is assumed to be of the form

$$\boldsymbol{C}_n = \sigma^2 \boldsymbol{I}. \tag{3.65}$$

The only difference is that the Jacobian in eq. (2.57) is evaluated at the position estimate found at iteration step n, while the Jacobians in the Fisher information matrix are evaluated at the true source position. Consequently, if the linearized estimator would be initialized by the true source's position it would be efficient, in that its covariance matrix of the first iteration step would be identical to the Fisher information matrix. As a direct consequence it can be seen that the initialization of the linearized estimator is of crucial importance.

3.4.3 Iterative Estimator Evaluation

In order to evaluate the performance of the linearized estimator, it is compared to other iterative non-derivative optimization techniques applied to the

maximum likelihood estimator. These are the "non-derivative quasi-Newton method" (QN) using the Broyden-Fletcher-Goldfarb-Shanno algorithm, the "Levenberg-Marquardt algorithm" (LM), and the "Nelder and Mead Simplex Downhill Method" (SD). All three are readily implemented in Matlab's optimization toolbox [Mat06].

An independent identically distributed additive noise vector $\boldsymbol{n}$ is assumed, such that its covariance matrix is of the from

$$\boldsymbol{C}_n = \sigma^2 \boldsymbol{I}. \tag{3.66}$$

Following eq. (2.31) it can be seen that the maximum likelihood estimator can then be written as

$$\hat{\boldsymbol{x}}_s = \arg\min_{\boldsymbol{x}_s} |\boldsymbol{\tau} - \boldsymbol{d}/c|_2, \tag{3.67}$$

and is identical to the nonlinear least-squares estimator.

The evaluation is carried out by computer simulations, for a localization procedure in a room with dimensions $10m, 10m, 10m$. In the first test it is assumed that no additive noise is disturbing the TDOA measures and the linearized estimator as well as the three versions (QN, LM, and SD) of the maximum likelihood estimator are initialized by placing the estimated source at the origin of the coordinate frame. 1000 Monte Carlo trials are carried out, by randomly placing 7 microphones within the room dimensions and by randomly selecting a source position. Hence, good as well as bad configurations are equally likely. Table 3.1 compares the outcomes of the four iterative estimators. The mean error and standard deviation of the estimation processes are denoted by $\mu(error)$ and $\sigma(error)$, respectively, the mean number of iterations and its standard deviation are denoted by $\mu(iter)$ and $\sigma(iter)$. Further, the mean computation time, using Matlab on a Pentium4 PC with 3GHz clock rate and 1GByte RAM, are indicated in the row "$\mu(time)$". In order to neglect the outliers of the estimation procedure, estimation outcomes with positions 100m or further from the origin were omitted. The number of outliers for each of the estimators is presented in the last row of table 3.1.

Neglecting the number of wrong estimates, one sees that all four methods achieve to estimate the true source position, when no ambient noise is present. However, if the number of wrong estimates is considered, it becomes obvious that the linearized estimator, as well as the maximum likelihood estimator utilizing eiter the Levenberg-Marquardt, or the simplex downhill method have severe problems of finding the source location, if they are initialized by some fixed position. Only, the quasi-Netwon implementation of the maximum likelihood estimator seems to be quite robust against the choice of initialization, if no ambient noise is present.

	Levenberg-Maquardt	Quasi Newton	Simplex Downhill	Linearized Model
μ(error)	0.00	0.00	0.20	0.00
σ(error)	0.00	0.00	1.76	0.00
μ(iter)	8.36	31.20	209.18	7.08
σ(iter)	2.40	2.84	35.32	1.95
μ(time)	0.0210	0.0761	0.0791	0.0044
wrong est.	230	34	335	747

Table 3.1: comparison of iterative estimators using the origin as the initial guess, no noise is added, μ and σ stand for the mean and the standard deviation, respectively, error is the distance between the true and the estimated position, "iter" denotes the number of iterations, "time" denotes the computation time,"wrong estimates" is the number of times error$\geq 100m$ (these estimates are not considered in the calculation of the other values)

	Levenberg-Maquardt	Quasi Newton	Simplex Downhill	Linearized Model
μ(error)	0.13	0.13	0.13	0.13
σ(error)	0.11	0.11	0.11	0.11
μ(iter)	4.64	10.06	53.69	4.21
σ(iter)	0.92	2.08	9.41	4.82
μ(time)	0.0130	0.0176	0.0218	0.0027
wrong est.	0	0	0	0

Table 3.2: comparison of iterative estimators using the true source position as the initial guess, with zero mean Gaussian noise ($\sigma = 0.1m$), μ and σ stand for the mean and the standard deviation, respectively, error is the distance between the true and the estimated position, "iter" denotes the number of iterations, "time" denotes the computation time,"wrong estimates" is the number of times error$\geq 100m$ (these estimates are not considered in the calculation of the other values)

	Levenberg-Maquardt	Quasi Newton	Simplex Downhill	Linearized Model
μ(error)	1.33	1.33	1.34	1.36
σ(error)	1.71	1.71	1.71	1.76
μ(iter)	9.27	11.46	69.00	17.56
σ(iter)	2.08	2.73	14.97	28.66
μ(time)	0.0238	0.0179	0.0268	0.0106
wrong est.	1	1	1	1

Table 3.3: comparison of iterative estimators using the true source position as the initial guess, with zero mean Gaussian noise ($\sigma = 1.0m$), μ and σ stand for the mean and the standard deviation, respectively, error is the distance between the true and the estimated position, "iter" denotes the number of iterations, "time" denotes the computation time,"wrong estimates" is the number of times error$\geq 100m$ (these estimates are not considered in the calculation of the other values)

Looking at table 3.2 it can be seen that the initialization is of utmost importance. This time the estimators were initialized using the true source position. Even though such an initialization is quite unrealistic (if the true position would be known, it would not be necessary to estimate it), it is very helpful to point out the importance of the initialization procedure. The values were again found by 1000 Monte Carlo trials with arbitrarily sensor and source positions using 7 microphones. All estimators show the same performance and none of them has any divergence problems, even though a small additive noise, with standard deviation equivalent to $0.1m$ was added to the TDOA measures:

$$\sigma = 0.1m/c. = \frac{0.1}{343}s. \tag{3.68}$$

Tables 3.3 and 3.4 show the outcomes of the same trials with larger ambient noise. The results in 3.4 are found for additive noise with standard deviation equivalent to $1m$, while the results in 3.4 represent the outcomes with standard deviation equivalent to $2m$. Looking at the number of wrong estimates it becomes obvious that the linearized estimator is the least robust w.r.t. ambient noise.

In chapter 2 it was shown that the robustness of the linearized estimator is directly related to the condition number of the Jacobian. Figures 3.6 and 3.7 are found by using an optimal sensor configuration for the reference point $5m, 5m, 5m$. 6 microphones are installed with a radius of $5m$ from this

	Levenberg-Maquardt	Quasi Newton	Simplex Downhill	Linearized Model
μ(error)	3.05	3.05	3.07	3.15
σ(error)	4.73	4.73	4.76	5.04
μ(iter)	10.87	13.10	76.46	23.20
σ(iter)	2.40	2.84	35.32	1.95
μ(time)	0.0283	0.0196	0.0298	0.0140
wrong est.	12	12	11	55

Table 3.4: comparison of iterative estimators using the true source position as the initial guess, with zero mean Gaussian noise ($\sigma = 2.0m$), μ and σ stand for the mean and the standard deviation, respectively, error is the distance between the true and the estimated position, "iter" denotes the number of iterations, "time" denotes the computation time,"wrong estimates" is the number of times error$\geq 100m$ (these estimates are not considered in the calculation of the other values)

reference point, building the edges of an octahedron (see fig. 3.9). In section 3.7 it will be shown that this microphone configuration indeed is optimal for the linearized estimator, as well as the maximum likelihood estimator.

The covariance matrix of the disturbance noise is chosen to be $\sigma^2\boldsymbol{I}$, with standard deviation equivalent to $\sigma = 2m$. The x-axis of the figures represent the uncertainties of the initialization. Gaussian distributed noise with covariance matrix $\sigma^2_{\boldsymbol{x}_{s_0}}\boldsymbol{I}$ is added to the true sensor position. The Quasi-Newton and the linearized estimator are then initialized by this random position. Looking at fig. 3.6 one sees that the two estimators perform nearly the same and that their accuracy decreases for initializations with larger $\sigma_{\boldsymbol{x}_{s_0}}$. Again, estimates with a norm of larger than $100m$ are neglected, and the number of those wrong estimates is plotted in fig. 3.7. For good initializations no wrong estimates appear, and only as the initialization becomes worse the problem of wrong estimates appears for the two estimators. Comparing these results with the ones of table 3.4, one sees that the use of an optimal sensor configuration does not only increase the accuracy of the estimators, but also helps to solve the problem of the number of wrong estimates.

3.5 Linear Approximation Estimators

In the previous section it was shown that the accuracy of an iterative TDOA-based position estimator strongly relies on an accurate initialization. This

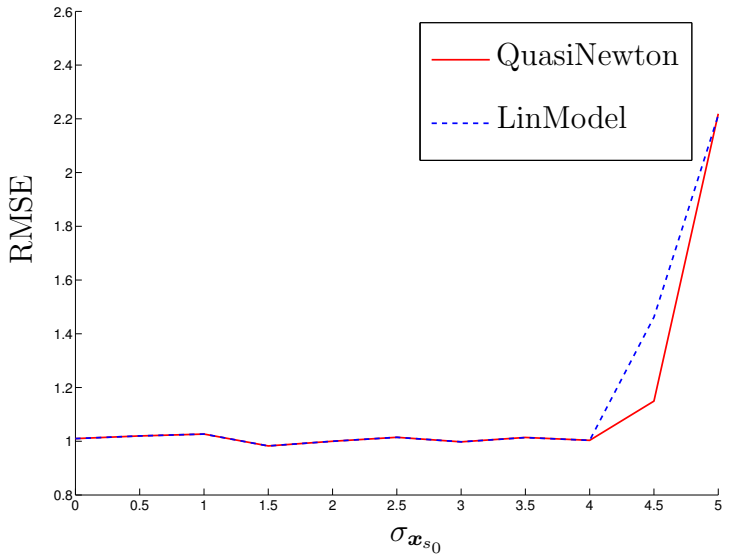

Figure 3.6: Uncertain linearized estimator initialization

section presents a number of linear approximation estimators, which can either estimate the source position in closed-form, or are self-initializing. All these estimators only consider the time differences of arrival w.r.t. to a reference sensor. This reference sensor-based approach offers the possibility of writing the estimation problem as a constrained linear least-squares problem. In the following the first microphone, at position $\boldsymbol{x}_1$, is selected as the reference sensor, which leads to the TDOA and range difference vectors as defined in eqs.(3.48) and (3.49)

$$\boldsymbol{\tau}_1 = (\tau_{21}, \ldots, \tau_{N1})^T,$$

and

$$\boldsymbol{d}_1 = (d_{21}, \ldots, d_{N1})^T,$$

respectively. While microphone one is chosen to be the reference sensor, the remaining microphones are often referred to as the "slave" sensors.

Squaring the entries of the range difference vector results in

$$d_{i1}^2 = D_i^2 + D_1^2 - 2D_iD_1. \tag{3.69}$$

With $D_i = d_{i1} + D_1$, the squared range difference can be written as

$$d_{i1}^2 = D_i^2 - D_1^2 - 2d_{i1}D_1. \tag{3.70}$$

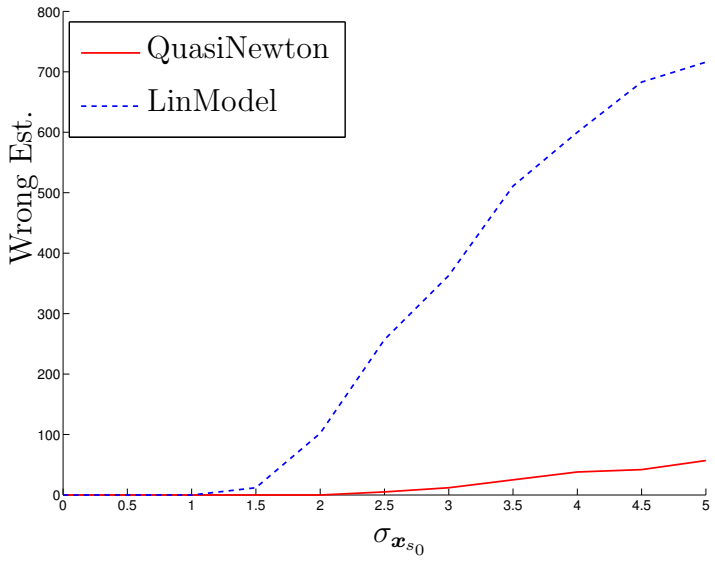

Figure 3.7: Uncertain linearized estimator initialization, wrong estimates

Utilizing the relation

$$D_i^2 = R_i^2 - 2\boldsymbol{x}_i^T\boldsymbol{x}_s + R_s^2 \tag{3.71}$$

this leads to

$$d_{i1}^2 = R_i^2 - R_1^2 - 2(\boldsymbol{x}_i - \boldsymbol{x}_1)^T\boldsymbol{x}_s - 2D_1 d_{i1}, \tag{3.72}$$

with R_i begin the distance form the origin to sensor i at position $\boldsymbol{x}_i$ (see fig. 3.5). Replacing d_{i1} by

$$d_{i1} = c \cdot \tau_{i1} + c \cdot n_{i1}, \tag{3.73}$$

eq. (3.72) becomes

$$R_i^2 - R_1^2 - c\tau_{i1}^2 - 2(\boldsymbol{x}_i - \boldsymbol{x}_1)^T\boldsymbol{x}_s - 2D_1 c\tau_{i1} = \epsilon_{i1} \tag{3.74}$$

with

$$\begin{aligned} \epsilon_{i1} &= 2c^2\tau_{i1}n_{i1} - c^2n_{i1}^2 + 2D_1 cn_{i1} \\ &= 2c(c\tau_{i1} + D_1)n_{i1} - c^2n_{i1}^2. \end{aligned} \tag{3.75}$$

being an error term due to the additive noise n_{i1}. Considering all $N-1$ time differences the following equation holds

$$\boldsymbol{b} - 2\boldsymbol{S}\boldsymbol{x}_s - 2D_1 c\boldsymbol{\tau}_1 = \boldsymbol{\epsilon}, \tag{3.76}$$

with

$$\boldsymbol{b} = \begin{pmatrix} R_2^2 - R_1^2 - (c \cdot \tau_{21})^2 \\ R_3^2 - R_1^2 - (c \cdot \tau_{31})^2 \\ \vdots \\ R_N^2 - R_1^2 - (c \cdot \tau_{N1})^2 \end{pmatrix}, \tag{3.77}$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_2^T - \boldsymbol{x}_1^T \\ \boldsymbol{x}_3^T - \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_N^T - \boldsymbol{x}_1^T \end{bmatrix}, \tag{3.78}$$

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{21}, \epsilon_{31}, \vdots, \epsilon_{N1} \end{pmatrix}^T. \tag{3.79}$$

The goal of TDOA-based source localization can then be seen as finding the source position $\boldsymbol{x}_s$ that minimizes the error term $\boldsymbol{\epsilon}$. Schau and Robinson [SR87] proposed to set the error term equal to zero and then to solve the resulting quadratic equation. We refer to this solution as the spherical intersection estimator.

3.5.1 Spherical Intersection Estimator

Schau and Robinson [SR87] proposed to estimate the source position by simply setting the error term $\boldsymbol{\epsilon}$ in eq. (3.76) equal to zero and then to solve the resulting quadratic equation. With $\boldsymbol{\epsilon} = \boldsymbol{0}$, solving eq. (3.76) for $(\boldsymbol{x}_s - \boldsymbol{x}_1)$ leads to

$$\boldsymbol{x}_s - \boldsymbol{x}_1 = \frac{1}{2}\boldsymbol{S}^\dagger(\boldsymbol{b} - 2D_1 c\boldsymbol{\tau}_1 - 2\boldsymbol{S}\boldsymbol{x}_1). \tag{3.80}$$

With

$$D_1^2 = (\boldsymbol{x}_s - \boldsymbol{x}_1)^T(\boldsymbol{x}_s - \boldsymbol{x}_1), \tag{3.81}$$

one obtains a quadratic equation in D_1:

$$aD_1^2 + bD_1 + c = 0, \tag{3.82}$$

with

$$a = 1 - \boldsymbol{d}_1^T(\boldsymbol{S}^\dagger)^T\boldsymbol{S}^\dagger\boldsymbol{d}_1, \tag{3.83}$$

$$b = (\boldsymbol{b} - 2\boldsymbol{S}\boldsymbol{x}_1)^T(\boldsymbol{S}^\dagger)^T\boldsymbol{S}^\dagger\boldsymbol{d}_1, \tag{3.84}$$

$$c = (\boldsymbol{b} - 2\boldsymbol{S}\boldsymbol{x}_1)^T(\boldsymbol{S}^\dagger)^T\boldsymbol{S}^\dagger(\boldsymbol{b} - 2\boldsymbol{S}\boldsymbol{x}_1). \tag{3.85}$$

Since $D_1 = |\boldsymbol{x}_s - \boldsymbol{x}_1|_2 \geq 0$, only the positive root(s) of eq. (3.82) is (are) possible estimates of D_1. Assuming a single positive root, the position estimate can be found by substitution of this root into eq. (3.80).

However, if two positive roots are found, the position estimate of this spherical intersection approach is not unique. Further if no positive root exists, no position estimate can be obtained. Both situations are quite likely to happen, if the sensor geometry is not well chosen and if the additive noise becomes large. More robust position estimators are based on the theory of weighted least-squares estimation.

3.5.2 Weighted Least-Squares Estimator

The weighted least-squares estimator for the reference sensor, TDOA-based position estimation problem is defined as the estimator that minimizes the weighted cost function $J(\boldsymbol{x}_s, \boldsymbol{X})$ on the error term $\boldsymbol{\epsilon}$ of eq. (3.76)

$$J(\boldsymbol{x}_s, \boldsymbol{X}) = \boldsymbol{\epsilon}^T \boldsymbol{W} \boldsymbol{\epsilon}, \tag{3.86}$$

with $\boldsymbol{W}$ denoting a weighting matrix:

$$\hat{\boldsymbol{x}}_s = \arg\min_{\boldsymbol{x}_s} J(\boldsymbol{x}_s, \boldsymbol{X}) = \arg\min_{\boldsymbol{x}_s} \boldsymbol{\epsilon}^T \boldsymbol{W} \boldsymbol{\epsilon}. \tag{3.87}$$

If no statistical information about the additive noise is available, the weighting matrix is usually set to be the identity matrix. However, for a zero-mean Gaussian noise vector $\boldsymbol{n}_1$, a weighting matrix can be constructed such that the weighted least-squares estimator approximates the maximum likelihood estimator.

Gaussian Noise Weighting Matrix

Neglecting the second order noise terms in eq. (3.75) the error term ϵ_{i1} can be approximated by

$$\begin{aligned} \epsilon_{i1} &\approx 2c(c\tau_{i1} + D_1)n_{i1} \\ &= 2(c\tau_{i1} + D_1)(c\tau_{i1} - D_1) \\ &\approx 2(d_{i1} + D_1)(c\tau_{i1} - D_1) \\ &= 2D_i(c\tau_{i1} - D_1). \end{aligned} \tag{3.88}$$

The error vector $\boldsymbol{\epsilon}$ then becomes

$$2\mathrm{Diag}(D_2, \ldots, D_N)(c \cdot \boldsymbol{\tau}_1 - \boldsymbol{d}_1), \tag{3.89}$$

and the cost function (3.86) results to

$$J(\boldsymbol{x}_s, \boldsymbol{X}) = (c \cdot \boldsymbol{\tau}_1 - \boldsymbol{d}_1)^T \mathrm{Diag}(D_2, \ldots, D_N) \boldsymbol{W} \mathrm{Diag}(D_2, \ldots, D_N)(c \cdot \boldsymbol{\tau}_1 - \boldsymbol{d}_1). \tag{3.90}$$

Note that the factor 2 is omitted, since it does not influence the optimization outcome of the cost function.

Assuming zero-mean additive Gaussian noise, with constant covariance matrix $\boldsymbol{C}_n$, and choosing the weighting matrix to be

$$\boldsymbol{W} = \mathrm{Diag}(D_2, \ldots, D_N)^{-1} \boldsymbol{C}_n^{-1} \mathrm{Diag}(D_2, \ldots, D_N)^{-1}, \tag{3.91}$$

the cost function results to

$$J(\boldsymbol{x}_s, \boldsymbol{X}) = (c \cdot \boldsymbol{\tau}_1 - \boldsymbol{d}_1)^T \boldsymbol{C}_n^{-1} (c \cdot \boldsymbol{\tau}_1 - \boldsymbol{d}_1). \tag{3.92}$$

Note that minimizing this cost function results in the maximum likelihood estimator of eq. (3.62), if only the TDOAs w.r.t. the reference sensor are considered. However, the matrix $\mathrm{Diag}(D_2, \ldots, D_N)$ is usually unavailable, since its entries D_i are dependent on the unknown sensor position $\boldsymbol{x}_s$. As a solution, Smith and Abel [SA87] proposed an iterative approach: choose the initial entries of $\boldsymbol{W}$ to be the identity matrix $\boldsymbol{W} = \boldsymbol{I}$, and iteratively update those by calculating D_i with the previous position estimate. They argue that one or two iterations are sufficient for an appropriate position estimate.

So far it has been shown that estimators minimizing the cost function defined in eq. (3.86) can approximate the maximum likelihood estimator, if Gaussian noise is considered. In the following it will be shown how this cost-function can be minimized in closed-form.

3.5.3 Weighted Linear Least-Squares Estimator

Collecting the source position $\boldsymbol{x}_s$ and the distance between the reference sensor and the source D_1 in a vector $\boldsymbol{\theta} = [\boldsymbol{x}_s^T, D_1]^T$ and introducing matrix

$$\boldsymbol{A} = 2[\boldsymbol{S}, c \cdot \boldsymbol{\tau}_1] \tag{3.93}$$

eq. (3.76) can be written as a constrained linear equation

$$\begin{aligned} \boldsymbol{\epsilon} &= \boldsymbol{b} - \boldsymbol{A}\boldsymbol{\theta} \\ &\text{subject to} \quad , \\ D_1 &= |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2 \end{aligned} \tag{3.94}$$

which results in the constrained, linear estimation problem

$$\begin{aligned} \hat{\boldsymbol{x}}_s = \arg\min_{\boldsymbol{x}_s} &(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{\theta})^T \boldsymbol{W} (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{\theta}) \\ &\text{subject to} \\ D_1 &= |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2. \end{aligned} \tag{3.95}$$

Huang et al. [HBE00] proposed to just neglect the constraint of this optimization problem, which then simply results in the weighted linear-least squares problem

$$\begin{aligned} \boldsymbol{J} &= \boldsymbol{\epsilon}^T \boldsymbol{W} \boldsymbol{\epsilon} \\ &= (\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{b})^T \boldsymbol{W} (\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{b}). \end{aligned} \tag{3.96}$$

Using the results of chapter 2.2.1, and more precisely eq. (2.48) the minimizing vector $\hat{\boldsymbol{\theta}}^{WLLS}$ results to

$$\begin{aligned} \hat{\boldsymbol{\theta}}^{WLLS} = \begin{pmatrix} \hat{\boldsymbol{x}}_s^{WLLS} \\ \hat{D}_1 \end{pmatrix} &= (\boldsymbol{A}^T \boldsymbol{W} \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{W} \boldsymbol{b} \\ &= \boldsymbol{A}_w \boldsymbol{b}, \end{aligned} \tag{3.97}$$

Note that since the constraint was neglected that the relation between $\hat{D}_1$ and $\hat{\boldsymbol{x}}_s^{WLLS}$

$$\hat{D}_1 = |\boldsymbol{x}_1 - \hat{\boldsymbol{x}}_s^{WLLS}|_2 \tag{3.98}$$

does usually not hold. However, if the additive noise vector is considered to be small, it can be argued that this estimator still offers an unbiased estimate of the position.

Bias and Covariance In the approximation of eq. (3.88) error vector $\boldsymbol{\epsilon}$ is approximated to

$$\boldsymbol{\epsilon} \approx \mathrm{Diag}(2c(c\tau_{21} + D_1), \ldots, 2c(c\tau_{N1} + D_1))\boldsymbol{n}. \tag{3.99}$$

Assuming zero-mean noise with covariance matrix $\boldsymbol{C}_n$, this error vector can then be seen as a zero-mean random vector with covariance matrix

$$\boldsymbol{C}_\epsilon \approx \mathrm{Diag}(2c(c\tau_{21}+D_1), \ldots, 2c(c\tau_{N1}+D_1))\boldsymbol{C}_n \mathrm{Diag}(2c(c\tau_{21}+D_1), \ldots, 2c(c\tau_{N1}+D_1)). \tag{3.100}$$

Consequently, since $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, the mean and covariance of the WLLS estimator can be calculated to

$$\boldsymbol{\mu}_{WLLS} = E(\hat{\boldsymbol{\theta}}^{WLLS}) = \boldsymbol{A}_w \boldsymbol{b} - \boldsymbol{A}_w \boldsymbol{\mu}_\epsilon \approx \boldsymbol{A}_w \boldsymbol{b} \tag{3.101}$$

and

$$\boldsymbol{C}_{WLLS} \approx \boldsymbol{A}_w \boldsymbol{C}_\epsilon \boldsymbol{A}_w^T. \tag{3.102}$$

Hence, the weighted linear least-squares estimator of eq. (3.97) is approximately unbiased, even though the constrained optimization problem was converted into an unconstrained optimization, by simply neglecting the constraint.

Linear Least-Squares Estimator

If the weighting matrix is set to be the identity matrix $\boldsymbol{W} = \boldsymbol{I}$, the position estimation simply becomes the pseudo-inverse of $\boldsymbol{A}$ times the vector $\boldsymbol{b}$ and the linear least-squares (LLS) estimator results:

$$\hat{\boldsymbol{\theta}}^{LLS} = ((\hat{\boldsymbol{x}}_s^{LLS})^T, \hat{D}_1)^T = \boldsymbol{A}^\dagger \boldsymbol{b}. \tag{3.103}$$

Bias and Covariance The mean and the covariance matrix of this estimator then result in

$$\boldsymbol{\mu}_{LLS} = E(\hat{\boldsymbol{\theta}}^{LLS}) = \boldsymbol{A}^\dagger \boldsymbol{b} - \boldsymbol{A}^\dagger \boldsymbol{\mu}_\epsilon \approx \boldsymbol{A}^\dagger \boldsymbol{b} \tag{3.104}$$

and

$$\boldsymbol{C}_{LLS} \approx \boldsymbol{A}^\dagger \boldsymbol{C}_\epsilon (\boldsymbol{A}^\dagger)^T. \tag{3.105}$$

Spherical Interpolation

Abel [SA87] proposed to solve eq. (3.76) for $\boldsymbol{x}_s$ by first premultiplying it by a projection matrix for either $\boldsymbol{\tau}_1$ or $\boldsymbol{x}_s$, which will eliminate the corresponding term. The resulting linear estimation problem is then being solved by the LLS approach. This procedure is now presented with a projection matrix for $\boldsymbol{\tau}_1$, which is given by

$$\boldsymbol{P}_\tau = \boldsymbol{I} - \boldsymbol{\tau}_1 (\boldsymbol{\tau}_1^T \boldsymbol{\tau}_1)^{-1} \boldsymbol{\tau}_1^T \tag{3.106}$$

Eq. (3.86) then turns out to be

$$\boldsymbol{\epsilon}^T \boldsymbol{P}_\tau \boldsymbol{W} \boldsymbol{P}_\tau \boldsymbol{\epsilon} = (\boldsymbol{P}_\tau \boldsymbol{b} - 2\boldsymbol{P}_\tau \boldsymbol{S} \boldsymbol{x}_s)^T \boldsymbol{W} (\boldsymbol{P}_\tau \boldsymbol{b} - 2\boldsymbol{P}_\tau \boldsymbol{S} \boldsymbol{x}_s), \tag{3.107}$$

which is maximized for

$$\hat{\boldsymbol{x}}_s^{LI} = 1/2 (\boldsymbol{S}^T \boldsymbol{P}_\tau \boldsymbol{W} \boldsymbol{P}_\tau \boldsymbol{S})^{-1} \boldsymbol{S}^T \boldsymbol{P}_\tau \boldsymbol{W} \boldsymbol{P}_\tau \boldsymbol{b}. \tag{3.108}$$

When the weighting matrix $\boldsymbol{W}$ is set to be the identity matrix $\boldsymbol{W} = \boldsymbol{I}$ it can be shown that eqs. (3.108) and (3.103) become the same and hence, the methods based on the nonweigted linear least-squares approach of eq. (3.103) and the non-weighted linear interpolation of eq. (3.108) are identical. Consequently the mean and the covariance of the non-weighted spherical interpolation estimator are identical to the ones of the LLS estimator (eqs. (3.104) and (3.105)).

3.5.4 Linear Correction Least-Squares Estimator

The obvious problem of the weighted linear least-squares approach is that it neglects the constraint of eq. (3.94). Huang et. al [HBEM01] proposed an iterative procedure, which takes this dependency into account based on the theory of Lagrange multipliers. If the coordinate frame is chosen such that the reference sensor $\boldsymbol{x}_1$ lies on the origin, the constraint of the optimization problem defined in eq. (3.95) becomes

$$R_s = |\boldsymbol{x}_s|_2, \tag{3.109}$$

and $\boldsymbol{\theta}$ becomes

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{x}_s \\ \hat{R}_s \end{pmatrix}. \tag{3.110}$$

Consequently, the constraint can be written as

$$\boldsymbol{\theta}^T \boldsymbol{\Sigma} \boldsymbol{\theta} = 0, \tag{3.111}$$

with

$$\boldsymbol{\Sigma} = \mathrm{Diag}(1, 1, 1, -1). \tag{3.112}$$

If further the weighting matrix $\boldsymbol{W}$ is set to be the identity matrix, the problem of passive source localization becomes to find a $\boldsymbol{\theta}$ which minimizes the following linear, constrained optimization problem:

$$\begin{gathered} \min_{\boldsymbol{\theta}} (\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{b})^T (\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{b}) \\ \text{subject to} \\ \boldsymbol{\theta}^T \boldsymbol{\Sigma} \boldsymbol{\theta} = 0. \end{gathered} \tag{3.113}$$

The Lagrangian of this constrained problem follows to

$$\mathcal{L}(\boldsymbol{\theta}, \lambda) = (\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{b})^T (\boldsymbol{A}\boldsymbol{\theta} - \boldsymbol{b}) + \lambda \boldsymbol{\theta}^T \boldsymbol{\Sigma} \boldsymbol{\theta}, \tag{3.114}$$

where λ is the Lagrange multiplier. Necessary conditions for minimizing eq. (3.114) can be obtained by taking the gradient of $\mathcal{L}(\boldsymbol{\theta}, \lambda)$ with respect to $\boldsymbol{\theta}$ and finding $\boldsymbol{\theta}$ such that this gradient becomes zero. Solving for $\boldsymbol{\theta}$ yields the constrained least-squares estimate

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^T \boldsymbol{A} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{A}^T \boldsymbol{b}, \tag{3.115}$$

where λ is yet to be determined such that the constraint of (3.114) is fulfilled. Substituting (3.115) into the constraint of (3.113) leads to the following equation

$$\boldsymbol{b}^T \boldsymbol{A} (\boldsymbol{A}^T \boldsymbol{A} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma} (\boldsymbol{A}^T \boldsymbol{A} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{A}^T \boldsymbol{b} = 0. \tag{3.116}$$

With $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}$, $(\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{\Sigma})^{-1}$ follows to

$$(\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{\Sigma})^{-1} = \boldsymbol{\Sigma}(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma} + \lambda\boldsymbol{I})^{-1}. \tag{3.117}$$

Consequently, eq. (3.116) can be written as

$$\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma} + \lambda\boldsymbol{I})^{-1}(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} = 0. \tag{3.118}$$

This nonlinear equation must now be solved for λ. Assuming that

$$\boldsymbol{A}^T\boldsymbol{A} > \boldsymbol{0}, \tag{3.119}$$

the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}\Sigma$ will be real [1] and consequently the proposed eigenvalue factorization in [HBEM01] holds:

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}, \tag{3.120}$$

with $\boldsymbol{\Lambda} = \mathrm{Diag}(\gamma_1, \ldots, \gamma_4)$ and γ_i denoting the ith eigenvalue of the matrix $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma}$. The matrix $\boldsymbol{U}$ is composed of the corresponding eigenvectors, and consequently, the constraint of eq. (3.118) becomes

$$\boldsymbol{p}^T(\boldsymbol{\Lambda} + \lambda\boldsymbol{I})^{-2}\boldsymbol{q} = \sum_{i=1}^{4} \frac{p_i q_i}{(\lambda + \gamma_i)^2} = 0, \tag{3.121}$$

with $\boldsymbol{p} = \boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{A}^T\boldsymbol{b}$ and $\boldsymbol{q} = \boldsymbol{U}^{-1}\boldsymbol{A}^T\boldsymbol{b}$. Since the solution to (3.121) is not unique, it is proposed to use the secant method [PFTV02] to determine its roots. If the ambient noise is small, the solution will be close to the linear least-squares estimate of eq. (3.103). Hence, the Lagrange multiplier will be small. Therefore Huang et al. [HBEM01] propose that the initial points of the secant method should be chosen either as $\lambda_0 = 0$ or as $\lambda_0 = \beta$, where β is a small number dependent on the array geometry. They state that five iterations should be sufficient to give an accurate approximation to the root.

In summary the two steps of the linear-correction least-squares (LCLS) estimation are as follows: the Lagrangian multiplier λ is determined by finding the root of (3.121) around zero. Using this λ the position estimate is found by eq. (3.115).

Even though this estimator uses an iterative procedure to calculate the Lagrangian multiplier from the nonlinear eq. (3.118), it is integrated in the section of the linear approximation estimator, since it is self-initializing.

[1] eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma}$ are real valued, proof: since $\boldsymbol{A}^T\boldsymbol{A} > \boldsymbol{0}$, $\boldsymbol{A}^T\boldsymbol{A}\Sigma$ can be written as $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Sigma} = \sqrt{\boldsymbol{A}^T\boldsymbol{A}}(\sqrt{\boldsymbol{A}^T\boldsymbol{A}})^T\boldsymbol{\Sigma}\sqrt{\boldsymbol{A}^T\boldsymbol{A}}(\sqrt{\boldsymbol{A}^T\boldsymbol{A}})^{-1}$, with $\sqrt{\boldsymbol{A}^T\boldsymbol{A}}$ being a lower triangular matrix obtained by the Cholesky factorization . Consequently the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}\Sigma$ are equal to the eigenvalues of $(\sqrt{\boldsymbol{A}^T\boldsymbol{A}})^T\boldsymbol{\Sigma}\sqrt{\boldsymbol{A}^T\boldsymbol{A}}$, which are real valued, since, with $\boldsymbol{\Sigma}$ being symmetric, $(\sqrt{\boldsymbol{A}^T\boldsymbol{A}})^T\boldsymbol{\Sigma}\sqrt{\boldsymbol{A}^T\boldsymbol{A}}$ is symmetric, leading to real eigenvalues

3.5.5 Hyperbolic Interpolation Estimator

Chan and Ho [CH94] extend the weighted LLS estimator of eq. (3.97) to a two step procedure. In a first step a first position estimate is found by the weighted linear least-squares estimator of eq. (3.97), using the presented iterative procedure for finding the appropriate weighting matrix:

$$\hat{\boldsymbol{\theta}}^{WLLS} = \boldsymbol{A}_w \boldsymbol{b}. \tag{3.122}$$

It was shown that under the assumption of zero-mean Gaussian noise, this estimate is approximately unbiased and Gaussian with covariance matrix $\boldsymbol{C}_{WLLS}$ given by eq. (3.102). Consequently, this estimate can be written as the sum of the true position/range vector $\boldsymbol{\theta}$ and an zero-mean Gaussian distributed error vector $\Delta\boldsymbol{\theta}$, with covariance matrix $\boldsymbol{C}_{WLLS}$. Assuming the general three dimensional estimation problem with

$$\hat{\boldsymbol{\theta}}^{WLLS} = (\hat{\theta}_x, \hat{\theta}_y, \hat{\theta}_z, \hat{\theta}_D)^T, \tag{3.123}$$

$$\Delta\boldsymbol{\theta} = (\Delta\theta_x, \Delta\theta_y, \Delta\theta_z, \Delta\theta_D)^T, \tag{3.124}$$

$$\boldsymbol{x}_s = (x_s, y_x, z_s)^T, \tag{3.125}$$

$$\boldsymbol{x}_1 = (x_1, y_1, z_1), \tag{3.126}$$

the following relation holds

$$\begin{pmatrix} (\hat{\theta}_x - x_1)^2 \\ (\hat{\theta}_y - y_1)^2 \\ (\hat{\theta}_z - z_1)^2 \\ \hat{\theta}_{D_1}^2 \end{pmatrix} = \begin{pmatrix} (x_s + \Delta\theta_x - x_1)^2 \\ (y_s + \Delta\theta_y - y_1)^2 \\ (z_s + \Delta\theta_z - z_1)^2 \\ (D_1 + \Delta\theta_D)^2 \end{pmatrix}. \tag{3.127}$$

Neglecting the quadratic error terms this equation can be approximated by

$$\underbrace{\begin{pmatrix} (\hat{\theta}_x - x_1)^2 \\ (\hat{\theta}_y - y_1)^2 \\ (\hat{\theta}_z - z_1)^2 \\ \hat{\theta}_{D_1}^2 \end{pmatrix}}_{\boldsymbol{h}} \approx \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}}_{\boldsymbol{\Sigma}_{hl}} \underbrace{\begin{pmatrix} (x_s - x_1)^2 \\ (y_s - y_1)^2 \\ (z_s - z_1)^2 \end{pmatrix}}_{\boldsymbol{\theta}_{hl}} + \underbrace{\begin{pmatrix} 2(x_s - x_1)\Delta\theta_x \\ 2(y_s - y_1)\Delta\theta_y \\ 2(z_s - z_1)\Delta\theta_z \\ 2D_1\Delta\theta_D \end{pmatrix}}_{\boldsymbol{\zeta}}, \tag{3.128}$$

which once more can be seen as a weighted linear-least squares problem.

With $\boldsymbol{C}_{WLLS}$ being the covariance matrix of $\Delta\boldsymbol{\theta}$, the covariance matrix of the zero-mean random vector $\boldsymbol{\zeta}$ follows to

$$\boldsymbol{C}_{hl} \approx \boldsymbol{B}_{hl} \boldsymbol{C}_{WLLS} \boldsymbol{B}_{hl}, \tag{3.129}$$

with $\boldsymbol{B}_{hl} = \text{Diag}(x_s - x_1, y_s - y_1, z_s - z_1, D_1)$. The weighted linear least-squares solution of $\boldsymbol{\theta}_{hl}$ is then found to be

$$\hat{\boldsymbol{\theta}}_{hl} = (\boldsymbol{\Sigma}_{hl}^T \boldsymbol{C}_{hl}^{-1} \boldsymbol{\Sigma}_{hl})^{-1} \boldsymbol{\Sigma}_{hl}^T \boldsymbol{C}_{hl}^{-1} \boldsymbol{h}. \tag{3.130}$$

Again, the true values of $\boldsymbol{C}_{hl}$ are not known, since they contain the true values of the source localization. An iterative approach just as before, has to be applied, with the starting points chosen as the values found by eq. (3.97). The final position is then obtained from $\boldsymbol{\theta}_{ml}$

$$\hat{\boldsymbol{x}}_s^{hl} = \pm \sqrt{\begin{pmatrix} \theta_{x_{hl}} \\ \theta_{y_{hl}} \\ \theta_{z_{hl}} \end{pmatrix}} + \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \tag{3.131}$$

Since eq. (3.131) results in eight possible locations, Chan and Ho [CH94] propose to choose the outcome which lies in a predefined search region. However, if multiple of these outcomes lie within this region, all these estimates must be considered.

Example 1. *The reference sensor is placed at the origin and four other sensors are arbitrarily placed in a cube of dimension* $[-5m, 5m]$*. The source location is chosen to be* $[1, -2, 3]$ *and a zero-mean Gaussian noise vector with standard deviation of* $0.01m$ *is added to the corresponding RD vector* $\boldsymbol{d} = c \cdot \boldsymbol{\tau}$*. If the hyperbolic algorithm as described in [CH94] is applied, the outcomes of eq. (3.131) are*

$$\hat{\boldsymbol{x}}_{s_1}^{hl} \approx (\pm 1.0058, \pm 2.0157, \pm 3.0036)^T. \tag{3.132}$$

In order to avoid this ambiguity, we propose an alternative decision procedure. The initial guess for the iterative calculation of eq. (3.130) is found by eq. (3.97). Therefore, we propose to calculate the non-squared entries of vector $\boldsymbol{h}$,

$$\begin{pmatrix} \hat{\theta}_x - x_1 \\ \hat{\theta}_y - y_1 \\ \hat{\theta}_z - z_1 \end{pmatrix} = \hat{\boldsymbol{x}}_s^{WLLS} - \boldsymbol{x}_1,$$

and save their sign. Correspondingly to these saved signs, the signs of the individual entries of the square root of eq. (3.131) are chosen. In the example above the initial guess was found to be $(1.0057, -2.0153, 3.0027)$ and the signs of $(\hat{\boldsymbol{x}}_s^{WLLS} - \boldsymbol{x}_1)$ should be memorized $(+, -, +)$. The position estimate then results to

$$\hat{\boldsymbol{x}}_s^{HL} = \begin{pmatrix} + \\ - \\ + \end{pmatrix} \sqrt{\begin{pmatrix} \theta_{x_{hl}} \\ \theta_{y_{hl}} \\ \theta_{z_{hl}} \end{pmatrix}} + \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \tag{3.133}$$

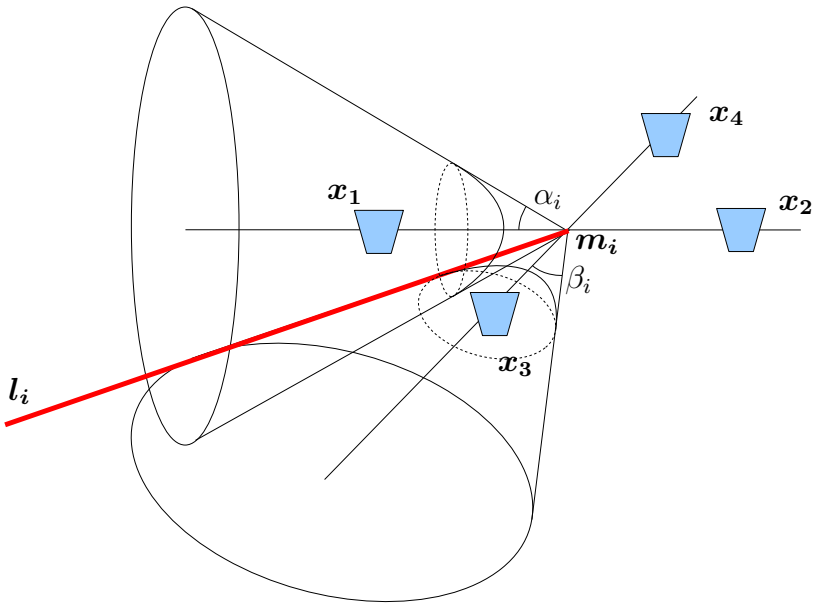

Figure 3.8: Approximated hyperbolic intersections by cones

and the right position estimate results

$$\hat{\boldsymbol{x}}_s^{HL} \approx (1.0058, -2.0157, 3.0036)^T. \tag{3.134}$$

The modified hyperbolic location estimation algorithm is thus as follows: use the WLLS estimator of eq. (3.97) with weighting matrix (3.91) for an initial guess of the source position. Use this vector for the calculation of eq. (3.130). The source position is then estimated to be

$$\hat{\boldsymbol{x}}_s^{HL} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \text{sign}(\hat{\boldsymbol{x}}_s^{WLLS} - \boldsymbol{x}_1) \circ \sqrt{\begin{pmatrix} \theta_{x_{hl}} \\ \theta_{y_{hl}} \\ \theta_{z_{hl}} \end{pmatrix}}, \tag{3.135}$$

where $\circ$ denotes the Hadamard product (element-wise product) of two vectors and sign(.) denotes the element-wise sign function of a vector.

3.6 Linear Intersection Estimator

Brandstein [BAS97] proposed a completely different closed-form approach to the TDOA-based localization problem. It is based on the fact that, if a time difference of arrival τ_i of a microphone pair with positions $\boldsymbol{x}_{i,1}$, $\boldsymbol{x}_{i,2}$ is

known, the possible positions of a source are described by a hyperboloid. If the source position is assumed to be far from the sensors, this hyperboloid can be approximated by a cone with opening angle

$$\alpha_i = \cos^{-1}(c\tau_i/|\boldsymbol{x}_{i,1} - \boldsymbol{x}_{i,2}|_2). \tag{3.136}$$

and apex

$$\boldsymbol{m}_i = (\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})/2. \tag{3.137}$$

Now, two sensor pairs $\{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\}$ and $\{\boldsymbol{x}_{i3}, \boldsymbol{x}_{i4}\}$ are considered with their measured TDOAs $\tau_{i,12}$ and $\tau_{i,34}$, respectively. The two lines connecting the sensor pairs are constrained to be orthogonal and to intersect in their midpoints (fig. 3.8):

$$\boldsymbol{m}_i = (\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})/2 = (\boldsymbol{x}_{i3} - \boldsymbol{x}_{i4})/2. \tag{3.138}$$

A local coordinate frame, with origin $\boldsymbol{m}_i$ and the unit vectors chosen as

$$\boldsymbol{x} = \frac{\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}}{|\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}|_2}, \tag{3.139}$$

$$\boldsymbol{y} = \frac{\boldsymbol{x}_{i3} - \boldsymbol{x}_{i4}}{|\boldsymbol{x}_{i3} - \boldsymbol{x}_{i4}|_2}, \tag{3.140}$$

and

$$\boldsymbol{z} = \boldsymbol{x} \times \boldsymbol{y}, \tag{3.141}$$

with $\times$ denoting the crossproduct of two three dimensional vectors, is established. The range difference of the first sensor pair approximates a cone with constant direction angle α_i relative to the $\boldsymbol{x}$-axis. The second sensor pair defines a constant direction angle β_i relative to the $\boldsymbol{y}$-axis. Since both cones represent the possible source locations, they will share at least one common point. In fact, the intersection of those two cones are two lines (one with $z < 0$ and one with $z > 0$) on which the source position has to lie. If the sensor pairs are installed such that $z < 0$ is physically not possible [2], only one bearing line $\boldsymbol{l}_i$ results

$$\boldsymbol{l}_i = r_i \boldsymbol{a}_i + \boldsymbol{m}_i, \tag{3.142}$$

with

$$\boldsymbol{a} = \begin{pmatrix} \cos(\alpha_i) \\ \cos(\beta_i) \\ \cos(\gamma_i) \end{pmatrix} \tag{3.143}$$

[2] example: installing the sensor quadruples on the walls or the ceiling of a room, leads to a z-coordinate pointing into the room. Consequently, one of the two resulting bearing lines of a single quadruple points into the room, while the other points outwards. The outward pointing line can then be neglected

being the direction vector of the line (or bearing line) and r_i being the distance to a point on the line. The remaining direction angle γ_i of the direction vector $\boldsymbol{a}_i$ may be calculated from the identity

$$\cos^2(\alpha_i) + \cos^2(\beta_i) + \cos^2(\gamma_i) = 1 \tag{3.144}$$

with $0 \leq \gamma_i \leq \frac{\pi}{2}$.

Given two or more of those sensor quadruples, each offering a bearing line, the source location can be calculated in the noiseless case as the intersection point of those bearings. However, if noise is present those bearing lines will usually not intersect. The approach proposed by Brandstein is to calculate a number of potential source locations from the points of closest intersection for all pairs of bearing lines and use a weighted average of these locations to generate a final estimate. Given two bearing lines

$$\boldsymbol{l}_i = r_i \boldsymbol{a}_i + \boldsymbol{m}_i \tag{3.145}$$

$$\boldsymbol{l}_j = r_j \boldsymbol{a}_j + \boldsymbol{m}_j \tag{3.146}$$

the shortest distance of those is measured as the length of the two points on those lines intersecting with their common normal and is given by

$$\delta_{ij} = \frac{|(\boldsymbol{a}_i \times \boldsymbol{a}_j) \cdot (\boldsymbol{m}_i - \boldsymbol{m}_j)|_2}{|\boldsymbol{a}_i \times \boldsymbol{a}_j|_2}. \tag{3.147}$$

Two possible source locations are then found for each bearing line pair: the intersection of line $\boldsymbol{l}_i$ and the normal, resulting in an estimate $\hat{\boldsymbol{x}}_{s_{ij}}$, and the intersection of line $\boldsymbol{l}_j$ and the normal, resulting in $\hat{\boldsymbol{x}}_{s_{ji}}$. Those points are found by first solving the overconstrainted equation

$$r_i \boldsymbol{a}_i - r_j \boldsymbol{a}_j = \boldsymbol{m}_j - \boldsymbol{m}_i - \delta_{ij}(\boldsymbol{a}_i \times \boldsymbol{a}_j) \tag{3.148}$$

for r_i and r_j and then setting those results into eqs. (3.145) and (3.146), respectively. Assuming M sensor quadruples, and consequently M bearing lines, leads to a total of $M(M-1)$ possible source positions $\hat{\boldsymbol{x}}_{ij}$. The overall source position is then calculated as the weighted sum over all these estimates:

$$\hat{\boldsymbol{x}}_s^{LI} = \sum_{j=1}^{M} \sum_{k=1, k\neq j}^{M} w_{jk} \hat{\boldsymbol{x}}_{jk}, \tag{3.149}$$

with w_{ik} denoting the weight corresponding to the estimate found by the intersection of line i and its common normal with line k. If no statistical information on the additive noise terms is available, the weights are set to unity. If this information is at hand, the reader is referred to [BAS97], [Bra95] on how to best choose the weighing terms.

3.7 Optimal Sensor Configuration

In the last three sections a number of TDOA-based position estimators were presented. While some are of iterative nature and others can be written in closed-form, they all have in common that their accuracy depends on the geometry of the sensor network and the source position. In chapter 2 it was argued that the accuracy of an estimator could be increased by choosing optimal internal parameters, which in the case of passive source localization were identified as the sensor positions. Further, since the performance does also depend on the actual source position, it was proposed to carry out this optimization procedure in a first step w.r.t. a reference source position. In this section optimal configurations for the previously presented TDOA-based passive source location estimators are derived.

Assuming that the iterative maximum likelihood and linearized estimators approximately attain the Cramer Rao lower bound, the first configuration is derived using this bound. Considering the geometric dilution of precision, as well as the condition number, it will be shown that all three performance measures result in the same configuration for the linearized estimator.

The found configuration is not optimal for the linear approximation estimators of section 3.5. This is mainly due to the fact that these linear approximation estimators do not consider all $N(N-1)/2$ TDOA estimates. For this class of estimators, it is proposed to find the optimal configuration w.r.t. the condition number. Again, a close dependency, between the condition number and the geometric dilution of precision based configuration can be seen. Since these estimators are ideally suited as initialization procedures for the iterative linearized estimator, it is proposed to minimize the condition numbers of both estimators at the same time. A minimum number of microphone setup will be presented, which achieves to render both condition numbers equal to one.

The linear intersection estimator of section 3.6, strongly differs from the other estimators, in that it assumes that the source is located in the far-field of the individual sensor quadruples. Consequently, this estimator has to be treated separately. Procedures for optimizing its performance are discussed, again considering the condition number.

Yang and Scheuing [YS05] used the CRLB to derive an optimal sensor configuration for TDOA-based passive source localization. The following subsection is basically a summary of their work.

3.7.1 CRLB Optimal Sensor Configuration

Theorem 5. *If all $N(N-1)/2$ available TDOA measures of N microphones are available, and their covariance matrix is given by $\boldsymbol{C}_{\boldsymbol{\tau}} = \sigma^2 \boldsymbol{I}$, then the inverse of the Fisher information matrix is lower bounded by*

$$\operatorname{trace}(\boldsymbol{F}^{-1}) \geq (c \cdot \sigma)^2 \frac{p^2}{N^2}. \tag{3.150}$$

Equality holds if and only if

1. $\sum_{i=1}^{N} \boldsymbol{g}_i = 0$, *with*
$$\boldsymbol{g}_i = \frac{\boldsymbol{x}_i - \boldsymbol{x}_s}{|\boldsymbol{x}_i - \boldsymbol{x}_s|_2}. \tag{3.151}$$
2. *the matrix $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_N]$, satisfies $\boldsymbol{G}\boldsymbol{G}^T = (N/p)\boldsymbol{I}$, i.e. $\boldsymbol{G}$ has orthogonal row vectors with equal row norm.*

The reader is referred to [YS05] for a proof of this theorem, which gives necessary and sufficient conditions for an optimal sensor configuration with respect to the CRLB. The question now arises, of how to find a microphone setup, which fulfills them. Note that the bound does not only depend on the sensor positions, but also on the source position. Consequently, a sensor configuration which is optimal for one prospective source position, is very likely to not be optimal for others. In the following a sensor configuration is presented, which fulfills the conditions of theorem 5 for one representative position. Without loss of generality, it is assumed that this position is the origin of the coordinate frame $\boldsymbol{x}_s = \boldsymbol{0}$. Again, the reader is referred to [YS05], which proposes optimal configurations in the two and three dimensional case.

Since the $\boldsymbol{g}_i$'s are normalized vectors, Yang and Scheuing propose solutions which lie on the unit circle in $\mathbb{R}^2$ or the unit sphere in $\mathbb{R}^3$. In $\mathbb{R}^2$ it is easily verified that a uniform angular array defined by the sensor positions

$$\boldsymbol{g}_i = (\cos\alpha_i, \sin\alpha_i)^T, \tag{3.152}$$

with

$$\alpha_i = \alpha_1 + \frac{2\pi}{N}(i-1), \tag{3.153}$$

satisfies both conditions of theorem 5.

In $\mathbb{R}^3$ the vectors $\boldsymbol{g}_i$ which are equally distributed on the unit sphere, fulfill the conditions of theorem 5 . They can be represented by a combination of the extremities of the five platonic solids (or symmetrical polyhedra) [BSMM99]: tetrahedron, hexahedron (cube), octahedron, dodecahedron, and icosahedron, see fig. 3.9.

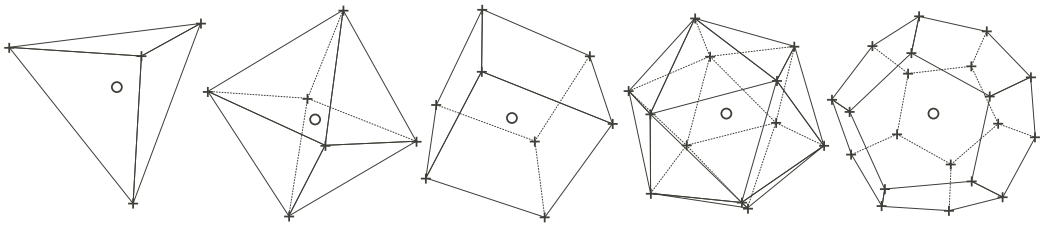

Figure 3.9: Tetrahedron, Octahedron, Hexahedron (cube), Dodecahedron, and Icosahedron

Yang and Scheuing [YS05] also showed that no unbiased estimator for TDOA-based position estimation can exist, which attains the CRLB. Therefore, minimizing the CRLB, does not necessarily optimize the performance of the chosen estimator. As a consequence the geometric dilution of precision and the condition number are now studied as possible cost functions for increasing the performance of the linearized estimator.

3.7.2 Optimal Sensor Configuration for the Linearized Estimator

The GDOP of the general linearized estimator was derived in chapter 2. It can readily be applied to the linearized estimator of the TDOA-based position estimation problem. If the TDOA measures are again assumed to be independent measures, with additive Gaussian noise with covariance $\boldsymbol{C}_{\boldsymbol{\tau}} = \sigma^2 \boldsymbol{I}$, the GDOP follows to

$$\text{GDOP} = \sqrt{\text{trace}(\boldsymbol{\mathcal{J}}^T(\boldsymbol{x}_{s_n}, \boldsymbol{X})\boldsymbol{\mathcal{J}}(\boldsymbol{x}_{s_n}, \boldsymbol{X}))^{-1}/p}, \tag{3.154}$$

with p being the dimension of the estimation problem (2D or 3D). Neglecting this dimension term and the square root does not change the global minimum of this equation. Further, as before, it is assumed that the estimator is initialized by an unbiased closed-from estimator, such as the LLS estimator. Then it is reasonable to optimize this GDOP for the initial value $\boldsymbol{x}_{s_n} = \boldsymbol{x}_s$, and optimizing the GDOP becomes identical to optimizing the CRLB. Consequently, the same sensor configuration as for the CRLB approach results.

A necessary condition for minimizing these measures is that $\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}} = \gamma^2\boldsymbol{I}$. However, theorem 3 states that this is equivalent to a minimum condition number of the Jacobian

$$\kappa_2(\boldsymbol{\mathcal{J}}) = 1. \tag{3.155}$$

Consequently, the sensor configurations that optimize the CRLB and the GDOP of the linearized estimator, are also optimal w.r.t. the condition number of the linearized estimator.

3.7.3 Optimal Sensor Configuration for the LLS Estimator

The presented linear approximation estimators of section 3.5 are all somehow related to the calculation of the pseudo-inverse of $\boldsymbol{A}$ defined by eq. (3.93). While the LLS estimator is directly calculated from this pseudo-inverse, the WLLS, the hyperbolic interpolation, and the LCLS estimator are initialized by the LLS estimator's outcome. The non-weighted SI estimator was shown to be equivalent to the LLS estimator. Consequently, we propose to derive an optimal sensor configuration w.r.t. the condition number of $\boldsymbol{A}$, such that

$$\boldsymbol{A}^T\boldsymbol{A} = \sigma^2\boldsymbol{I}, \tag{3.156}$$

with matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = 2[\boldsymbol{S}, \boldsymbol{d}_1] = 2\begin{bmatrix} \boldsymbol{x}_2^T - \boldsymbol{x}_1^T, & d_{21} \\ \boldsymbol{x}_3^T - \boldsymbol{x}_1^T, & d_{31} \\ \vdots & \\ \boldsymbol{x}_N^T - \boldsymbol{x}_1^T, & d_{N1} \end{bmatrix}. \tag{3.157}$$

Since the true RDs d_{21} are unknown, they are replaced by the measured TDOAs times the propagation speed $d_{i1} = c \cdot \tau_{i1}$. However, since these measures are subject to noise, $\boldsymbol{n} = c\boldsymbol{\tau} - \boldsymbol{d}$, matrix $\boldsymbol{A}$ should rather be replaced by $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \Delta\boldsymbol{A}$, where

$$\Delta\boldsymbol{A} = [\boldsymbol{0}, \boldsymbol{n}] \tag{3.158}$$

is a matrix that takes the noise vector $\boldsymbol{n}$ into account, and $\boldsymbol{0}$ denotes a matrix of appropriate size with all its entries being equal to 0:

$$\tilde{\boldsymbol{A}} = 2[\boldsymbol{S}, c \cdot \boldsymbol{\tau}_1] = 2\begin{bmatrix} \boldsymbol{x}_2^T - \boldsymbol{x}_1^T, & c \cdot \tau_{21} \\ \boldsymbol{x}_3^T - \boldsymbol{x}_1^T, & c \cdot \tau_{31} \\ \vdots & \\ \boldsymbol{x}_N^T - \boldsymbol{x}_1^T, & c \cdot \tau_{N1} \end{bmatrix}. \tag{3.159}$$

The LLS estimator is then given by

$$\hat{\boldsymbol{\theta}}^{LLS} = ((\hat{\boldsymbol{x}}_s^{LLS})^T, \hat{D}_1)^T = \tilde{\boldsymbol{A}}^{\dagger}\tilde{\boldsymbol{b}}, \tag{3.160}$$

where $\tilde{\boldsymbol{b}} = \boldsymbol{b} + \Delta\boldsymbol{b}$ denotes the perturbed version of $\boldsymbol{b}$, with

$$\Delta\boldsymbol{b} = \boldsymbol{n} \circ \boldsymbol{n}, \tag{3.161}$$

where $\circ$ denotes the Hadamard product of two vectors. Hence, the LLS estimator is perfectly described by theorem 2. This theorem clearly states that the upper bound for the estimation error decreases for smaller condition numbers. Therefore, it is desirable to minimize the condition number of the LLS estimator.

Fig. 3.10 shows the dependency of the relative estimation error of the LLS estimator as a function of the condition number $\kappa_2(\boldsymbol{A})$: 30 sensor configurations using 4 sensors were found for each $\kappa_2(\boldsymbol{A}) = [2, 3, \ldots, 50]$, when the source was set to the position $[1, 1]$. For each $\kappa_2(\boldsymbol{A})$ 1000 Monte-Carlo trials were carried out with arbitrary sensor configurations. The corresponding range differences were calculated and then perturbed by an additive zero-mean Gaussian noise with standard deviation $\sigma = 0.1m$. Note that the relative error of the LLS estimator seems to linearly increase with increasing $\kappa_2(\boldsymbol{A})$.

The attention is now drawn on how to minimize the condition number of the LLS estimator. Without loss of generality, it is assumed that the reference sensor is placed at the origin $\boldsymbol{x}_1 = \boldsymbol{0}$. Matrix $\boldsymbol{A}$ can then be written as

$$\boldsymbol{A} = 2 \begin{bmatrix} \boldsymbol{p}_2^T \\ \boldsymbol{p}_3^T \\ \vdots \\ \boldsymbol{p}_N^T \end{bmatrix}, \tag{3.162}$$

with

$$\boldsymbol{p}_i = \begin{pmatrix} \boldsymbol{x}_i \\ d_{i1} \end{pmatrix} \in \mathbb{R}^{p+1}. \tag{3.163}$$

Eq. (2.83) then gives an idea of how the sensors need to be installed in order to obtain $\kappa_2(\boldsymbol{A}) = \sigma^2\boldsymbol{I}$. The column vectors of $\boldsymbol{A}$ need to be orthogonal and of equal length in $\mathbb{R}^{N-1}$ and the row vectors $\boldsymbol{p}_i$ need to fulfill the range difference equations

$$d_{i1} = |\boldsymbol{x}_i - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_s|_2, \tag{3.164}$$

which describe a cone in the extended space $\mathbb{R}^{p+1}$. E.g. in the two dimensional case, the x-coordinate and the y-coordinate describe the position of the slave sensors, while the z-coordinate indicates the RD. The cone has its apex at $(\boldsymbol{x}_s^T, -|\boldsymbol{x}_s|_2)^T$. The cones axis of revolution passes through the source position $\boldsymbol{x}_s$ and at $d = 0$ the cone includes the position of the reference sensor, which was set to zero (see fig. 3.11).

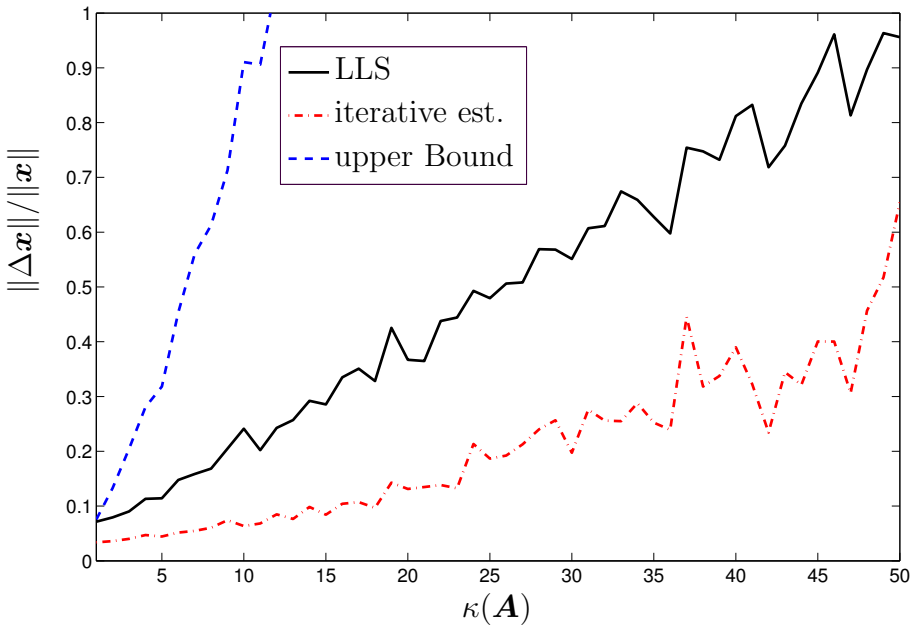

Figure 3.10: Relative position error of the LLS estimator as a function of the condition number $\kappa(\boldsymbol{A})$, with the number of sensors $N = 4$, upper bound given by (2.76), reference sensor at $[0, 0]$, source position at $[1, 1]$

The condition that $\kappa_2(\boldsymbol{A}) = 1$, or equivalently $\boldsymbol{A}^T\boldsymbol{A} = \sigma^2\boldsymbol{I}$, can be interpreted as that the row vectors $\boldsymbol{p}_i$ are random vectors with correlation matrix

$$\boldsymbol{R} = \frac{\sigma^2}{4(N-1)}\boldsymbol{I}. \tag{3.165}$$

This can be easily seen by

$$\boldsymbol{A}^T\boldsymbol{A} = 4 \cdot \sum_{i=2}^{N} \boldsymbol{p}_i\boldsymbol{p}_i^T = \sigma^2\boldsymbol{I}. \tag{3.166}$$

The correlation matrix $\boldsymbol{R}$ is calculated by

$$\boldsymbol{R} = \frac{1}{N-1}\sum_{i=2}^{N} \boldsymbol{p}_i\boldsymbol{p}_i^T. \tag{3.167}$$

A more detailed geometrical interpretation of these conditions is rather complicated to obtain, when $N > p + 2$. In the following subsection analytical

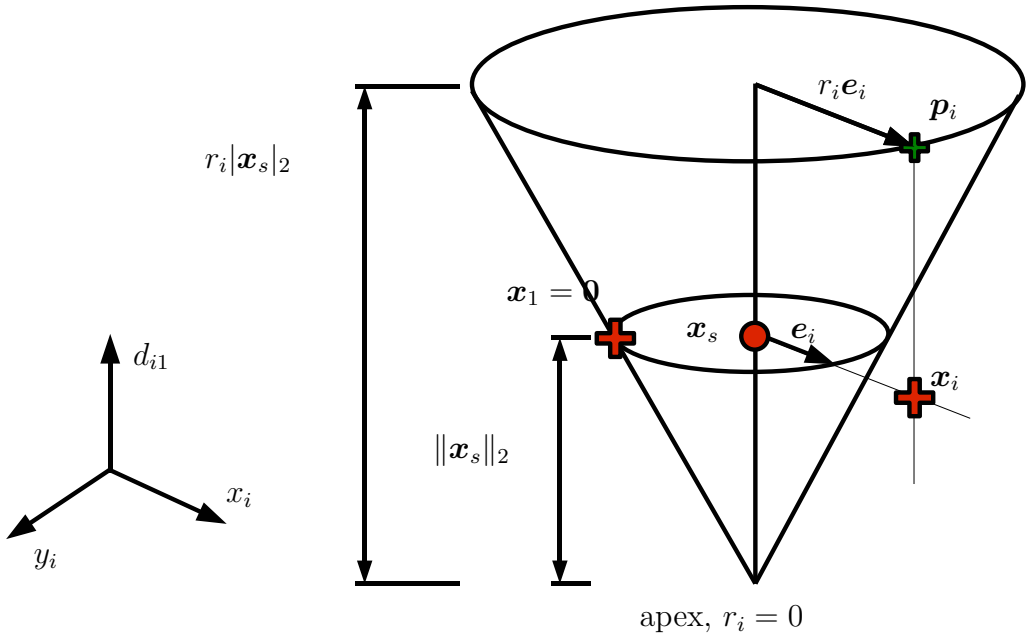

Figure 3.11: Cone representation of the RD equation

equations are derived, which offer an optimal sensor configuration with respect to the condition number, when a minimum number of sensors for the LLS approach $N = p + 2$ is considered.

Closed-form Sensor Configuration

When the minimum number of sensors is considered, the matrix $\boldsymbol{A}$ becomes square and therefore

$$\boldsymbol{A}^T\boldsymbol{A} = \sigma^2\boldsymbol{I} = \boldsymbol{A}\boldsymbol{A}^T. \tag{3.168}$$

Hence, three conditions for an optimal sensor configuration with respect to $\kappa_2(\boldsymbol{A}) = 1$ need to be fulfilled. The extended sensor positions $\boldsymbol{p}_i$ must all

$$\begin{aligned} &1) \quad \text{lie on the cone defined by (3.164)} \\ &2) \quad \text{lie on the same sphere } 4\boldsymbol{p}_i^T\boldsymbol{p}_i = \sigma^2, \forall i \in \{2, \ldots, N\} \\ &3) \quad \text{be orthogonal } \boldsymbol{p}_i^T\boldsymbol{p}_j = 0, \forall j \neq i \ . \end{aligned} \tag{3.169}$$

Conditions 2) and 3) immediately result from (3.168). Since the vectors need to lie on the cone of condition 1), as well as on the sphere defined by condition 2), these two conditions can be combined: the extended vectors $\boldsymbol{p}_i$ need to lie on the intersection of the cone and the sphere. Further, condition 3) rests to be fulfilled.

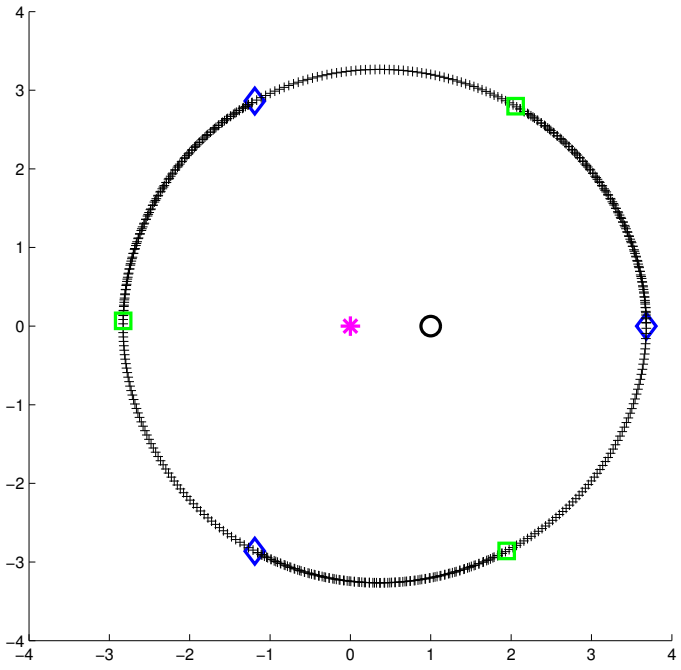

Figure 3.12: 2D minimal number of sensors configuration with $\kappa_2(\boldsymbol{A}) = 1$ and $\phi_i \in [0°, 180°]$, $\circ$: Source, $*$: reference sensor, $+$: slave sensors, $\Diamond$: slave sensors with $\boldsymbol{x}_2$ at $\phi_2 = 0°$, $\square$: slave sensors with $\boldsymbol{x}_2$ at $\phi_2 = 180°$

In order to simplify these equations, a special parameterization of $\boldsymbol{p}_i$ is used. Since all $\boldsymbol{p}_i$ must lie on the cone defined by eq. (3.164), the cone is parameterized as follows:

$$\boldsymbol{p}_i = \begin{pmatrix} \boldsymbol{x}_i \\ d_{i1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_s + r_i \boldsymbol{e}_i \\ (r_i - 1)|\boldsymbol{x}_s|_2 \end{pmatrix}, \tag{3.170}$$

with $|\boldsymbol{e}_i|_2 = |\boldsymbol{x}_s|_2$, which is illustrated for the two dimensional case in fig. 3.11. Further, since

$$\begin{aligned} d_{i1} &= |\boldsymbol{x}_i - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2 \\ &= |\boldsymbol{x}_i - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_s|_2 \\ &\geq -|\boldsymbol{x}_s|_2, \end{aligned} \tag{3.171}$$

the height parameter r_i is non negative.

Using the cone representation of (3.170) condition 2) leads to:

$$r_i^2 + r_i(\cos\phi_i - 1) + 1 = k,\ k = \frac{\sigma^2}{8|\boldsymbol{x}_s|_2^2}, \tag{3.172}$$

with

$$\boldsymbol{x}_s^T \boldsymbol{e}_i = \boldsymbol{x}_s^T \boldsymbol{x}_s \cos\phi_i, \tag{3.173}$$

and $\phi_i = \angle(\boldsymbol{x}_s, \boldsymbol{e}_i)$ being the angle between $\boldsymbol{x}_s$ and $\boldsymbol{e}_i$.

The third condition $\boldsymbol{p}_i^T \boldsymbol{p}_j = 0$ for all $i \neq j$ then results in

$$r_i^2 + r_j^2 - r_i r_j(\cos\phi_{ij} + 1) = 2k, \tag{3.174}$$

with $\phi_{ij} = \angle(\boldsymbol{e}_i, \boldsymbol{e}_j)$ being the angle between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$.

Two dimensional analytic solution

In order to obtain further insight, the two dimensional case ($p = 2$) is now investigated. The cosine of the relative sensor angles ϕ_{ij} may then be expressed in terms of the cosines of the absolute sensor angles ϕ_i, ϕ_j:

$$\cos\phi_{ij} = \cos(\phi_i - \phi_j). \tag{3.175}$$

The equations can be simplified by placing one of the slave sensors on the $\{\boldsymbol{x}_1, \boldsymbol{x}_s\}$ axis:

$$\phi_2 = \pi \Rightarrow \cos\phi_2 = -1. \tag{3.176}$$

Eq. (3.172) then becomes

$$r_2^2 - 2r_2 + 1 = k. \tag{3.177}$$

Thus, the parameter r_2 results to

$$r_2 = 1 \pm \sqrt{k}. \tag{3.178}$$

As a consequence eq. (3.174) follows to

$$r_i^2 + r_2^2 - r_i r_2(\cos\phi_{i2} + 1) = 2k;\ i \in \{3, 4\}. \tag{3.179}$$

With eq. (3.175) and eq. (3.176) this leads to

$$r_i^2 + r_2^2 + r_i r_2(\cos\phi_i - 1) = 2k. \tag{3.180}$$

Further, using eq. (3.178) this results to

$$\underbrace{r_i^2 + r_i(\cos\phi_i - 1) + 1 - k}_{=0,\ \text{eq. (3.172)}} \pm \sqrt{k}(r_i(\cos\phi_i - 1) + 2) = 0. \tag{3.181}$$

Since, $\sigma > 0 \Rightarrow k > 0$ and $r_i > 0$

$$\begin{array}{ll} r_i & = \sqrt{1+k} \\ \cos\phi_i & = 1 - \frac{2}{\sqrt{1+k}}, \end{array} \tag{3.182}$$

with $i \in \{3, 4\}$. Hence, this sensor setup is symmetric w.r.t. the $\{\boldsymbol{x}_1, \boldsymbol{x}_s\}$ axis. This symmetry leads to

$$\cos\phi_{34} = \cos(\phi_3 - \phi_4) = \cos(2\phi_i) = 2\cos^2\phi_i - 1;\ i \in \{3, 4\} \tag{3.183}$$

This and $r_3 = r_4 = r_i$ together with (3.174) for sensors 3 and 4 gives

$$r_i^2(1 - \cos^2\phi_i) = k;\ i \in \{3, 4\} \tag{3.184}$$

and with (3.182) this leads to a unique solution

$$k = 8 \tag{3.185}$$

since $\sigma > 0 \Rightarrow k > 0$.

With this value of k, the slave sensor positions may be computed using (3.176), (3.178), (3.182) and (3.170). The sensor configuration obtained this way is illustrated in fig. 3.12 (represented by the $\square$) and the values in the case $\boldsymbol{x}_1 = [0, 0]^T$ and $\boldsymbol{x}_s = [1, 0]^T$ are given below:

- Source: $\boldsymbol{x}_s = [1, 0]^T$
- Sensor 1 (reference): $\boldsymbol{x}_1 = [0, 0]^T$
- Sensor 2: $r_2 = 1 + \sqrt{8}, \phi_2 = \pi$ hence $\boldsymbol{x}_2 = [-\sqrt{8}, 0]^T$
- Sensor 3: $r_3 = 3, \phi_3 = \arccos\frac{1}{3}$ hence $\boldsymbol{x}_3 = [2, \sqrt{8}]^T$
- Sensor 4: $r_4 = 3, \phi_4 = -\phi_3$ hence $\boldsymbol{x}_4 = [2, -\sqrt{8}]^T$

In the general two dimensional case, where ϕ_2 is arbitrary, numerical analysis showed that the parameter k varies between 8 and 8.192 (e.g. when $\phi_2 = 0$). Fig. 3.13 shows the evaluations of the sensor parameters k, r_i and ϕ_i as a function of ϕ_2, when the above described two dimensional case is considered.

Analog results can be found in a three dimensional setup. In this case the slave sensors are placed on the surface of revolution defined by (3.172). The resulting surface is the shape of fig. 3.12 rotated around the $\boldsymbol{x}_1 - \boldsymbol{x}_s$ axis.

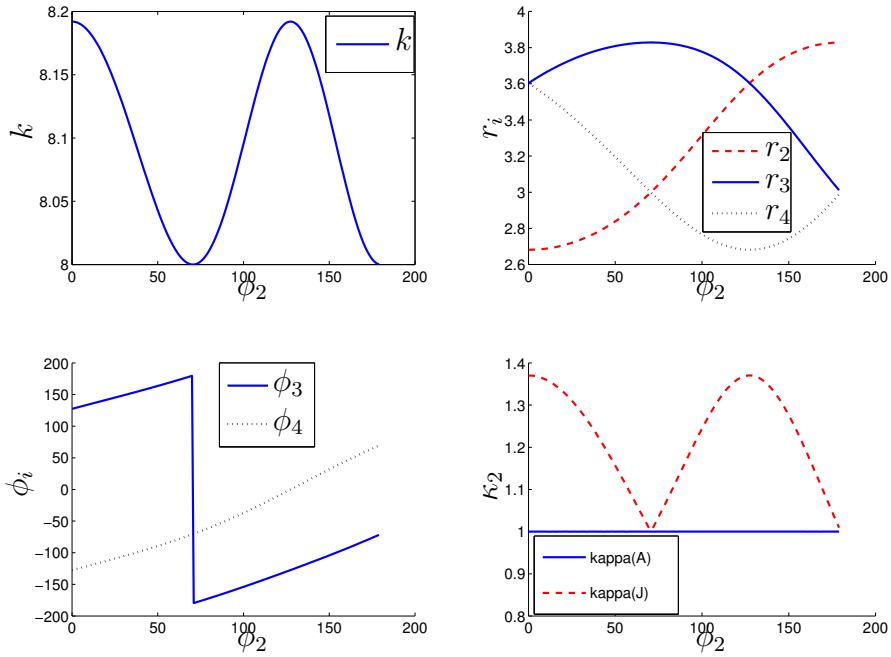

Figure 3.13: 2D minimal number of sensors configuration with $\kappa_2(\boldsymbol{A}) = 1$ and $\phi_i \in [0°, 180°]$, parameter evolution

Performance Evaluation of the Analytic Estimators

Fig. 3.14 shows the performance evaluation of the above mentioned estimators. The LLS, the WLLS, as well as the SI estimator perform identically, and are clearly outperformed by the hyperbolic location and the linear correction least-squares estimators. The difference of these two estimators compared to the others is that they take the constraint $D_1 = \|\boldsymbol{x}_s - \boldsymbol{x}_1\|_2$ into account. All of the presented estimators show a dependency on the sensor configuration in that their performances all increase for small condition numbers. Thus, if either one of these estimators is chosen to carry out the localization procedure, it might be desirable to use an optimal sensor configuration w.r.t. the condition number $\kappa_2(\boldsymbol{A})$.

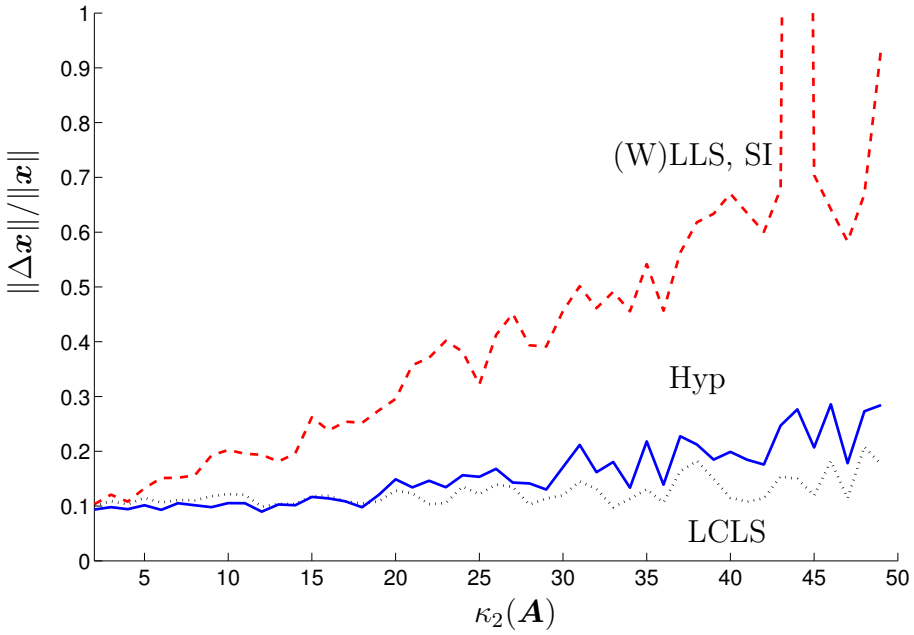

Figure 3.14: Comparison of the LLS, WLLS, SI, Hyp, and LCLS estimator for varying condition number $\kappa_2(\boldsymbol{A})$, the LLS, the WLLS, and the SI estimator show an identical outcome and are denoted only by LLS

3.7.4 Optimal Sensor Configuration for the Combined LLS, Linearized Estimator

If one of the linear approximation estimators of section 3.5 is used for finding an initial guess for the linearized estimator, it might be interesting to minimize the condition number of the Jacobian $\kappa_2(\boldsymbol{\mathcal{J}})$ as well as the condition number of the LLS estimator $\kappa_2(\boldsymbol{A})$. The problem can then be written as finding the roots of the two-component function

$$\boldsymbol{J}(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N) = \begin{bmatrix} \kappa_2(\boldsymbol{\mathcal{J}}) - 1 \\ \kappa_2(\boldsymbol{A}) - 1 \end{bmatrix} = \boldsymbol{0}. \tag{3.186}$$

In section 3.7.3 analytic sensor configurations were found for the 2 dimensional case. All the configurations found by applying the parameters represented in fig. 3.13 assured that the condition number of the LLS estimator results to one. This figure also shows the development of the condition number of the linearized estimator $\kappa_2(\boldsymbol{\mathcal{J}})$. It shows that, when the second sensor is placed at the position with $\phi_2 = 180°$ that not only the condition num-

	Source	Sensor 1	Sensor 2	Sensor 3	Sensor 4
$r_i/\|\boldsymbol{x}_s\|_2$	-	-	3.828	3	3
ϕ_i	-	-	$-180°$	$70.52°$	$-70.52°$
$x_i/\|\boldsymbol{x}_s\|_2$	1	0	-2.83	2	2
$y_i/\|\boldsymbol{x}_s\|_2$	0	0	0	2.83	-2.83

Table 3.5: Sensor positions fulfilling $\kappa_2(\boldsymbol{A}) = \kappa_2(\boldsymbol{\mathcal{J}}) = 1$. Reference sensor $\boldsymbol{x}_1$ and $\boldsymbol{x}_1 - \boldsymbol{x}_s$ axis define the origin and the x-axis of the cartesian coordinate frame, respectively

ber of the LLS estimator, but also the condition number of the linearized estimator is equal to one

$$\kappa_2(\boldsymbol{A}) = \kappa_2(\boldsymbol{\mathcal{J}}) = 1. \tag{3.187}$$

Hence, the resulting sensor configuration fulfills the condition of eq. (3.186) and the individual sensor positions as well as the source position are given in table 3.5.

3.7.5 Optimal Sensor Configuration of the LI estimator

The problem of an optimal sensor configuration for the LI estimator is twofold. The first part is the installation of the individual sensor quadruples w.r.t. the source position. The bearing line of one sensor quadruple is calculated by the intersection of two cones, which approximate the TDOA hyperboloids of the two sensor pairs. For disadvantageous source positions and ambient noise, theses two cones might not intersect, and imaginary position estimates results. Assume that the source lies in the plane of the sensor quadruple. If no ambient noise is present, the two resulting cones will tangent each other. If now ambient noise is present it becomes very likely that the two do not touch and a position estimation becomes impossible. Contrary, if the source lies on the normal of the sensor quadruple, the two cones become planes, which intersect at an angle of 90°. Even if large disturbance noise is present, it is very unlikely that the two resulting cones do not intersect. Consequently, it is favorable to install the quadruples in such a way that their normal points towards the source position. An alternative argue for such an installation is the approximation of the hyperboloid by the cone. It is argued that, if the source is far from the sensors, the resulting hyperboloids representing the TDOAs of the sensor pairs can be approximated by cones.

However, if the source lies on the normal of the sensor quadruple, the hyperboloid as well as the approximated cone become the same plan. In this case no approximation error results.

The second part of finding an optimal sensor configuration for the LI estimator is concerned with the relative sensor position of two or more sensor quadruples. Equation (3.148) is easiest solved for r_i and r_j by rearranging it to

$$[\boldsymbol{a}_i, \boldsymbol{a}_j] \begin{pmatrix} r_i \\ -r_j \end{pmatrix} = \boldsymbol{m}_j - \boldsymbol{m}_i - d_{ij}(\boldsymbol{a}_i \times \boldsymbol{a}_j). \tag{3.188}$$

Left multiplying eq. (3.188) by the Moore-Penrose pseudo-inverse of $[\boldsymbol{a}_i, \boldsymbol{a}_j]$ gives the result. In the preceding chapter it was shown that the calculation of the pseudo-inverse of a matrix is most robust, if the condition number of the matrix is equal to 1. For the LI estimator this can be translated to the problem of finding a geometry which fulfills

$$[\boldsymbol{a}_i, \boldsymbol{a}_j]^T [\boldsymbol{a}_i, \boldsymbol{a}_j] = \sigma^2 \boldsymbol{I}. \tag{3.189}$$

Equation (3.189) is fulfilled if the bearing lines $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ of the sensor quadruples i and j are orthogonal to each other. Consequently, if three sensor quadruples are considered, they should be installed such that their bearing lines are all orthogonal to each other for a specified reference point.

3.8 Analytic Linear Correction Least-Squares Estimator

A disadvantage of the procedure of the linear correction least-squares estimator of section 3.5.4 is the need of an iterative procedure for finding the Lagrangian multiplier λ. The result of the estimator will then strongly depend on the initial value of λ, and so might the number of iterations. An alternative, analytic approach is proposed in this section. It is based on the assumption that the condition number with respect to the 2-norm of matrix $\boldsymbol{A}$ is approximately one. Let us thus suppose that

$$\boldsymbol{A}^T \boldsymbol{A} = \sigma^2 \boldsymbol{I}. \tag{3.190}$$

The position estimate resulting from the linear correction least-squares estimator of eq. (3.115) then becomes

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\boldsymbol{A}^T \boldsymbol{A} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{A}^T \boldsymbol{b} \\ &= (\sigma^2 \boldsymbol{I} + \lambda \boldsymbol{\Sigma})^{-1} \begin{pmatrix} \boldsymbol{c} \\ d \end{pmatrix}, \end{aligned} \tag{3.191}$$

with

$$\begin{pmatrix} \boldsymbol{c} \\ d \end{pmatrix} = \boldsymbol{A}^T \boldsymbol{b}. \tag{3.192}$$

Hence, the constraint of eq. (3.113) becomes

$$\begin{aligned} \hat{\boldsymbol{\theta}}^T \boldsymbol{\Sigma} \hat{\boldsymbol{\theta}} &= (\boldsymbol{c}^T, d)^T (\sigma^2 \boldsymbol{I} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma} (\sigma^2 \boldsymbol{I} + \lambda \boldsymbol{\Sigma})^{-1} \begin{pmatrix} \boldsymbol{c} \\ d \end{pmatrix} \\ &= \frac{\boldsymbol{c}^T \boldsymbol{c}}{(\sigma^2 + \lambda)^2} - \frac{d^2}{(\sigma^2 - \lambda)^2} = 0. \end{aligned} \tag{3.193}$$

This quadratic equation offers two possible solutions for the Lagrangian multiplier

$$\lambda_1 = \sigma^2 \frac{\boldsymbol{c}^T \boldsymbol{c} - d}{\boldsymbol{c}^T \boldsymbol{c} + d} \quad \text{and} \quad \lambda_2 = \sigma^2 \frac{\boldsymbol{c}^T \boldsymbol{c} + d}{\boldsymbol{c}^T \boldsymbol{c} - d}. \tag{3.194}$$

A necessary and sufficient condition for that one of the λ's is a local minimum is that the Hessian matrix of the Lagrangian with respect to the estimation vector $\boldsymbol{\theta}$ is positive definite:

$$\mathcal{H}(\boldsymbol{\theta}, \lambda) = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \mathcal{L}(\boldsymbol{\theta}, \lambda) > 0. \tag{3.195}$$

It is easily calculated to

$$\mathcal{H}(\boldsymbol{\theta}, \lambda) = \boldsymbol{A}^T \boldsymbol{A} + \lambda \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I} + \mathrm{Diag}(\lambda, \lambda, \lambda, -\lambda). \tag{3.196}$$

In order to be positive its eigenvalues $\alpha_{1,2,3} = \sigma^2 + \lambda$ and $\alpha_4 = \sigma^2 - \lambda$ need to be positive. Assuming a small additive noise, this is only fulfilled by the Lagrangian multiplier λ_1. Supposing that no noise is present

$$\begin{pmatrix} \boldsymbol{c} \\ d \end{pmatrix} = \boldsymbol{A}^T \boldsymbol{b} = \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{\theta} = \sigma^2 \begin{pmatrix} \boldsymbol{x}_s \\ \|\boldsymbol{x}_s\|_2 \end{pmatrix}, \tag{3.197}$$

which would result in the multipliers $\lambda_1 = 0$ and $\lambda_2 = \infty$. Thus, λ_1 results in a positive definite Hessian and gives the solution to the constrained linear least-squares estimator. Now assuming a small additive noise vector $\boldsymbol{n}$, resulting in a small error vector $\boldsymbol{\epsilon}$,

$$\begin{pmatrix} \boldsymbol{c} \\ d \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_s \\ \|\boldsymbol{x}_s\|_2 \end{pmatrix} + \boldsymbol{A}\boldsymbol{\epsilon}, \tag{3.198}$$

Therefore, $\boldsymbol{c}$ will still be close to $\sigma^2 \boldsymbol{x}_s$ and d will still be in the neighborhood of $|\boldsymbol{x}_s|_2$, which will lead to a small multiplier λ_1 and to a large multiplier λ_2.

In the case when additive noise is considered, matrix $\boldsymbol{A}$ will be perturbed and so will its condition number. As a result the matrix product $\boldsymbol{A}^T \boldsymbol{A}$ will

no longer be diagonal with identical elements. However, if the condition number of the non-perturbed system is equal to one, the diagonal elements of $\boldsymbol{A}^T\boldsymbol{A}$ will still be close to the singular value of the non-perturbed $\boldsymbol{A}$ and will still be big compared to the off-diagonal elements. It is then reasonable to replace the ideal singular value of (3.116) by

$$\sigma^2 = \text{trace}(\boldsymbol{A}^T\boldsymbol{A})/4. \tag{3.199}$$

The Lagrangian multiplier then results to

$$\lambda = \frac{1}{4}\text{trace}(\boldsymbol{A}^T\boldsymbol{A})\frac{\boldsymbol{c}^T\boldsymbol{c} - d}{\boldsymbol{c}^T\boldsymbol{c} + d}. \tag{3.200}$$

Once this multiplier is calculated, two options for calculating the position estimate are left. It can either be calculated by eq. (3.115) using the term $\sigma^2\boldsymbol{I}$

$$\hat{\boldsymbol{\theta}}_1 = (\sigma^2\boldsymbol{I} + \lambda\boldsymbol{\Sigma})^{-1}\boldsymbol{A}^T\boldsymbol{b}, \tag{3.201}$$

or the term $\boldsymbol{A}^T\boldsymbol{A}$ is being used

$$\hat{\boldsymbol{\theta}}_2 = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{\Sigma})^{-1}\boldsymbol{A}^T\boldsymbol{b}. \tag{3.202}$$

3.8.1 Evaluation

Fig. 3.15 compares the performances of the proposed constrained linear least-squares estimator and the unconstrained linear least-squares estimator. 50 sensor configurations, using 7 microphones were found for a source position of $(6m, 3m, 2m)$, assuring that the condition number $\kappa(\boldsymbol{A})$ is equal to one. Zero mean Gaussian noise with varying standard deviation ($\sigma \in (0, 2)$), was added to the calculated range difference vector $\boldsymbol{d} = c \cdot \boldsymbol{\tau}$, and 1000 Monte Carlo trials were carried out for each σ, while arbitrarily choosing one of the 50 sensor configurations. The root mean square error (RMSE) of the position estimate as well as of the Lagrange multipliers are plotted over the standard deviation of the added noise. It clearly shows, how the estimators based on the Lagrangian outperform the linear least-squares estimator, which neglects the constraint. Further, the analytic estimator based on eqs. (3.201) and eq. (3.200) shows a clear advantage over the others.

Unfortunately, the matrix $\boldsymbol{A}$ is dependent on the ambient noise. If additive Gaussian noise is assumed on the TDOA vector, $\boldsymbol{d}_1 = c \cdot \boldsymbol{\tau}_1 + \boldsymbol{n}_1$, the real matrix $\tilde{\boldsymbol{A}}$ will be a perturbed version of the ideal matrix $\boldsymbol{A}$:

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \Delta\boldsymbol{A}, \tag{3.203}$$

with $\Delta\boldsymbol{A} = [0, \boldsymbol{n}_1]$. Thus, the condition $\kappa(\tilde{\boldsymbol{A}}) = 1$ will usually not hold.

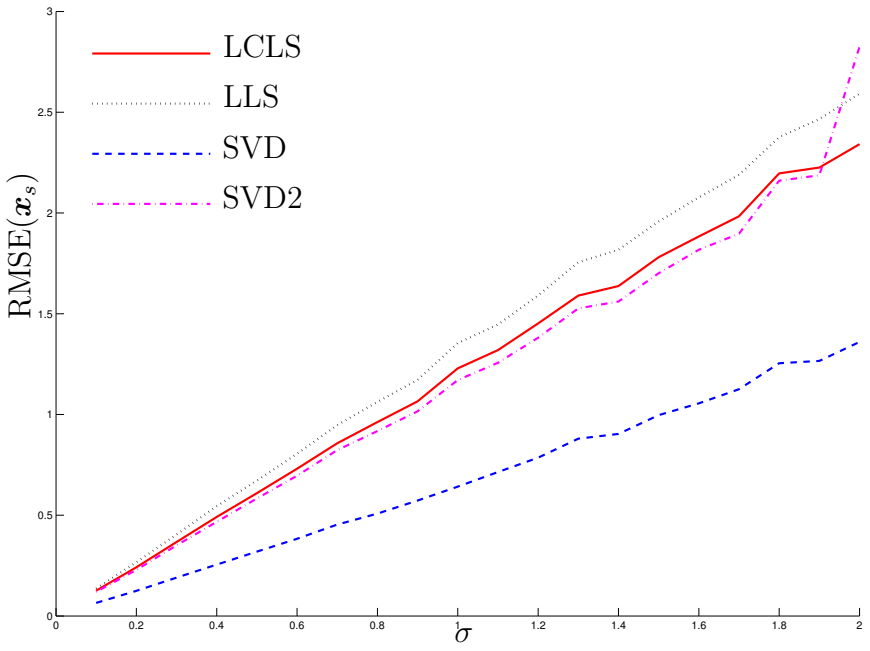

Figure 3.15: Position estimation with $\kappa(\boldsymbol{A}) \approx 1$ and Gaussian noise with standard deviation between 0 and 2, LCLS: linear correction least-squares estimator (eqs. (3.121) and 3.115), LLS: linear least-squares estimator (eq.(3.103)), SVD: analytic constrained linear least-squares estimator (eq. (3.201)) and (eq. (3.200)), SVD2: analytic constrained linear least-squares estimator (eq. (3.202)) and (eq. (3.200))

Fig. 3.16 studies the influence of a condition number $\kappa(\boldsymbol{A})$ not equal to one. Arbitrary sensor configurations for $\kappa(\boldsymbol{A} \in (1, 2)$ were found and a series of Monte Carlo trials were carried out. This time the standard deviation of the zero mean Gaussian noise was kept constant at $\sigma = 1m$. The LCLS procedure of section 3.5.4 was tested with two different starting points of the iterative search of λ. λ_0 was either set to zero (LCLS) or was calculated by eq. (3.200) (LCLS2). These two starting points result in identical outcomes up to a condition number of around 1.5. At higher values the performance of the LCLS2 estimator plunges. An even more drastic decrease appears for the analytic constrained linear least-squares estimators (SVD and SVD2). While they outperform the others for condition numbers smaller than 1.2, their results become unacceptable for higher condition numbers. Consequently,

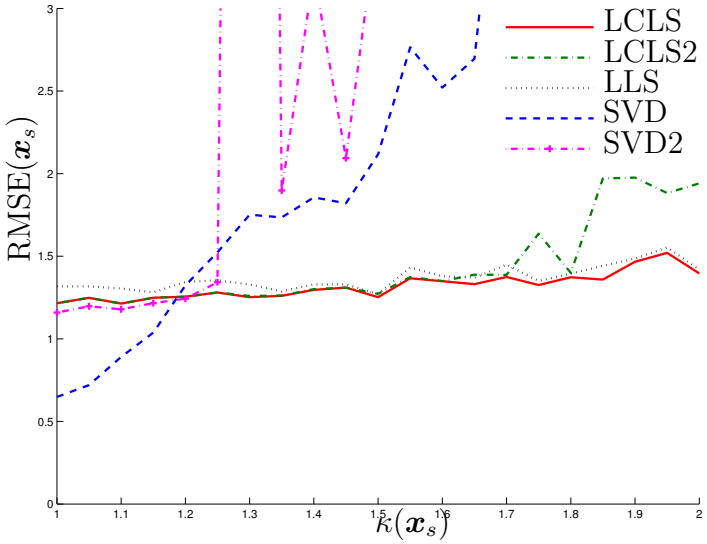

Figure 3.16: Position estimation with $\kappa(\boldsymbol{A})$ between 1 and 2 and Gaussian noise with standard deviation equal to 1. LCLS: linear correction least-squares estimator (eqs. (3.121) and 3.115) with $\lambda_0 = 0$, LCLS2: same as LCLS, only λ_0 is calculated using eq. (3.200), LLS: linear least-squares estimator (eq.(3.103)), SVD: analytic constrained linear least-squares estimator (eq. (3.201)) and (eq. (3.200)), SVD2: analytic constrained linear least-squares estimator (eq. (3.202)) and (eq. (3.200))

in order to apply the analytic linear least-squares estimator it is of utmost importance to assure that $\kappa_2(\boldsymbol{A})$ is kept very close to unity. If this cannot be guaranteed, the standard linear correction least-squares estimator is the better choice initialized by $\lambda_0 = 0$.

3.9 Chapter Summary

This chapter summarized the TDOA-based passive source localization algorithms, and concentrated on how to optimize the performance of those, by using optimal sensor configurations. Three different time delay estimators were presented and evaluated. The right choice of such an estimator is dependent on multiple factors, such as the signal to noise ratio, the reverberation time, and the computational capacities.

The most widely used TDOA-based position estimators were introduced,

and extensions for the hyperbolic interpolation and the linear correction least-squares estimator are proposed, to further increase their accuracy.

The main contribution of this chapter was the derivation of optimal sensor configurations for the linearized, the linear interpolation, and all the reference sensor based estimators. Further, an optimal, minimum number of sensor configuration was presented for the combined linear least-squares/ linearized estimator.

All these optimal sensor configurations were found for a reference source position. However, if moving sources are considered, such a reference source position might become difficult to define. The problem of moving sources is addressed in the next chapter.

Chapter 4

Source Tracking

The previous two chapters addressed the problem of optimizing TDOA-based acoustic passive source localization by choosing optimal positions for the microphones. It was assumed that a reference source position could be chosen for which the configuration was to be optimized. The true position of the source was assumed to be a deterministic value and no dependency between two consecutive position estimates were assumed.

This chapter addresses the problem of recursively estimating the source position, when consecutive source positions are no longer independent of each other, but can be modeled by a discrete dynamic state-space approach. The central part of such a state-space model are the discrete system equation, which models the evolution of the system with time, and the measurement equation, which describes the dependency of the measures on the state vector of the system. The state vector contains all relevant information to describe the modeled system. In tracking applications this state vector usually is composed of the position, the velocity, and additional dynamic information about the system, such as the steering angle of a vehicle.

Recursive Bayesian estimation deals with estimating the state vector at time instant k, based on the previous state vector at $k-1$, and all the observations up to k. It attempts to calculate the posterior probability density function of the state vector, which contains all statistical information about it. Once this pdf is obtained the optimal state estimate can be obtained w.r.t. any statistical criterion, such as the mean or the median. In case of a linear system, a linear measurement model, and additive Gaussian noise, the recursive Bayesian estimator w.r.t. any statistical criterion results in the Kalman filter.

In the following section the general framework of recursive Bayesian estimation is presented, and a variety of filters are derived for varying system and measurement models, considering Gaussian and non-Gaussian noise. This

framework is then applied to TDOA-based source tracking and an evaluation of the different filters is carried out for this problem, focusing on the sensor geometry.

4.1 Recursive Bayesian Estimation

Assume that the state vector $\boldsymbol{x}_{k+1}$ at time instant $k+1$ can be modeled by the system equation

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{w}_k), \tag{4.1}$$

where $\boldsymbol{f}$ is a possibly nonlinear function of the state vector $\boldsymbol{x}_k$, a system input $\boldsymbol{u}_k$, and a process noise $\boldsymbol{w}_k$. The objective of recursive Bayesian estimation is then to recursively estimate $\boldsymbol{x}_k$ from the measures

$$\boldsymbol{z}_k = \boldsymbol{g}(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{v}_k, \tag{4.2}$$

where $\boldsymbol{g}$ is a possibly nonlinear function of the true state vector and the input vector pertured by a measurement noise vector $\boldsymbol{v}_k$. More precisely, the estimation is not only based on the measure at time instant k, but on the set of all available measurements up to time k denoted by

$$\boldsymbol{Z}_{1:k} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}. \tag{4.3}$$

Assume that at time instant 0 no measures are available, and that the initial posterior pdf $p(\boldsymbol{x}_0|\boldsymbol{z}_0)$ is equal to the available prior $p(\boldsymbol{x}_0)$. Then, $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ can recursively be obtained by a prediction and an update stage. The prediction stage is based on the observation that the system equation (eq. (4.1)) describes a Markov process of order one, such that

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Z}_{1:k-1}) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}). \tag{4.4}$$

Consequently the state prediction pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1})$ follows to

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1}) &= \int_{-\infty}^{\infty} p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{Z}_{1:k-1}) d\boldsymbol{x}_{k-1} \\ &= \int_{-\infty}^{\infty} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Z}_{1:k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Z}_{1:k-1}) d\boldsymbol{x}_{k-1} \\ &= \int_{-\infty}^{\infty} p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Z}_{1:k-1}) d\boldsymbol{x}_{k-1}, \end{aligned} \tag{4.5}$$

which is known as the Chapman-Kolmogorov equation [AMGC02].

The update stage uses the measurement model of eq. (4.2) and the known statistics of $\boldsymbol{v}_k$ to calculate

$$p(\boldsymbol{z}_k|\boldsymbol{Z}_{1:k-1}) = \int_{-\infty}^{\infty} p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1})d\boldsymbol{x}_k. \tag{4.6}$$

Using this equation the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ calculates to

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1})}{p(\boldsymbol{z}_k|\boldsymbol{Z}_{1:k-1})}. \tag{4.7}$$

Recursively applying eqs. (4.5) and (4.7) results in the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$, which can then be used to estimate $\boldsymbol{x}_k$ w.r.t. to any statistical optimality criterion. The most common criteria are:

- Maximize the probability

$$\hat{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}_k} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) \tag{4.8}$$

  Using Bayes' theorem

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) = \frac{p(\boldsymbol{Z}_{1:k}|\boldsymbol{x}_k)p(\boldsymbol{x}_k)}{p(\boldsymbol{Z}_{1:k})}, \tag{4.9}$$

  it is easily seen that if $p(\boldsymbol{x}_k)$ is uniformly distributed, this estimate is equal to the classical maximum likelihood estimate:

$$\begin{aligned} \hat{\boldsymbol{x}}_k &= \arg\max_{\boldsymbol{x}_k} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) \\ &= \arg\max_{\boldsymbol{x}_k} \frac{p(\boldsymbol{Z}_{1:k}|\boldsymbol{x}_k)p(\boldsymbol{x}_k)}{p(\boldsymbol{Z}_{1:k})} \\ &= \arg\max_{\boldsymbol{x}_k} \frac{p(\boldsymbol{Z}_{1:k}|\boldsymbol{x}_k)}{p(\boldsymbol{Z}_{1:k})} \\ &= \arg\max_{\boldsymbol{x}_k} p(\boldsymbol{Z}_{1:k}|\boldsymbol{x}_k). \end{aligned} \tag{4.10}$$

- Minimum Variance Estimate

$$\hat{\boldsymbol{x}}_k = \arg\min_{\boldsymbol{y}} E(|\boldsymbol{y} - \boldsymbol{x}_k|_2^2|\boldsymbol{Z}_{1:k}), \tag{4.11}$$

  with $E(|\boldsymbol{y} - \boldsymbol{x}_k|_2^2|\boldsymbol{Z}_{1:k})$ being the expectation of $|\boldsymbol{y} - \boldsymbol{x}_k|_2^2$ taken over $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$. It can be shown (e.g. [AM79]) that $\hat{\boldsymbol{x}}_k$ results in:

$$\hat{\boldsymbol{x}}_k = E(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}). \tag{4.12}$$

- Median Estimator

$$\hat{\boldsymbol{x}}_k = \min \max |\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k|_2 \tag{4.13}$$

Note that if the conditional pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ is symmetric, the three presented optimality criteria result in the same estimate. Among others this is true for the Gaussian density function.

The approach of optimal recursive Bayesian estimation is of theoretical nature, in that in general the posterior density cannot be determined analytically. However, in case of linear system and measurement equations with additive Gaussian noise vectors $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$, respectively, the solution can be written in closed-form.

4.1.1 Linear System, Gaussian Noise: Kalman Filter

Under the assumption of white Gaussian observation and measurement noise $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k)$ and $\boldsymbol{v}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_k)$, it can be shown that the posterior density function of $\boldsymbol{x}_{k+1}$ is Gaussian, too, if the system and measurement equations are linear, such that

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k\boldsymbol{x}_k + \boldsymbol{B}_k\boldsymbol{u}_k + \boldsymbol{G}_k\boldsymbol{w}_k, \tag{4.14}$$

and

$$\boldsymbol{z}_k = \boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{v}_k. \tag{4.15}$$

In the following it is assumed that $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are independent of each other, such that

$$E(\boldsymbol{w}_k\boldsymbol{v}_k^T) = \boldsymbol{0}. \tag{4.16}$$

In [HL64] the following corresponding probability density functions are derived

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{Z}_{1:k-1}) \sim \mathcal{N}(\boldsymbol{x}_{k-1|k-1}, \boldsymbol{\Sigma}_{k-1|k-1}), \tag{4.17}$$

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1}) \sim \mathcal{N}(\boldsymbol{x}_{k|k-1}, \boldsymbol{\Sigma}_{k|k-1}), \tag{4.18}$$

and

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) \sim \mathcal{N}(\boldsymbol{x}_{k|k}, \boldsymbol{\Sigma}_{k|k}), \tag{4.19}$$

where

$$\boldsymbol{x}_{k|k-1} = \boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1|k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}, \tag{4.20}$$

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{C}_k\boldsymbol{x}_{k|k-1}), \tag{4.21}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{A}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}^T + \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^T, \tag{4.22}$$

$$\boldsymbol{\Sigma}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\boldsymbol{\Sigma}_{k|k-1}, \tag{4.23}$$

and

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T(\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T + \boldsymbol{R}_k)^{-1} \tag{4.24}$$

denotes the Kalman gain. No matter which optimality criterion is applied on $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$, the estimate results in the mean $\boldsymbol{x}_{k|k}$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k|k}. \tag{4.25}$$

The Kalman filter is usually initialized by the mean and covariance matrix of the prior density function $p(\boldsymbol{x}_0) \sim \mathcal{N}(\bar{\boldsymbol{x}}_0, \boldsymbol{P}_0)$:

$$\boldsymbol{x}_{0|-1} = \bar{\boldsymbol{x}}_0, \tag{4.26}$$

and

$$\boldsymbol{\Sigma}_{0|-1} = \boldsymbol{P}_0. \tag{4.27}$$

The update stage of the Kalman filter is then composed of calculating the Kalman gain, using eq. (4.24), which is used for calculating the state estimate by eq. (4.21), once a new measurement $\boldsymbol{z}_k$ is observed. The last part of the update stage then consists in calculating $\boldsymbol{\Sigma}_{k+1|k}$ using eq. (4.22).

The prediction stage of the Kalman filter is composed of the calculation of the state prediction $\boldsymbol{x}_{k+1|k}$ using eq. (4.20) and its covariance matrix $\boldsymbol{\Sigma}_{k+1|k}$ using eq. (4.22).

In order to understand the functioning of the Kalman filter, assume that the matrix $\boldsymbol{C}_k$ is a square nonsingular matrix and that the prediction covariance matrix $\boldsymbol{\Sigma}_{k|k-1}$ is positive definite. If the measurement noise $\boldsymbol{v}_k$ is assumed to be small, the covariance matrix $\boldsymbol{R}_k$ will be small, too. Assuming that, when compared to $\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T$, it can be neglected, then the Kalman gain follows to

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T(\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T)^{-1} = \boldsymbol{C}_k. \tag{4.28}$$

Consequently, the state estimate follows to

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{C}_k^{-1}(\boldsymbol{z}_k - \boldsymbol{C}_k\boldsymbol{x}_{k|k-1}) = \boldsymbol{C}_k^{-1}\boldsymbol{z}_k, \tag{4.29}$$

which is independent on the state prediction vector and only uses the observation for estimating the state vector.

Contrary, if the system noise $\boldsymbol{w}_k$ is considered to be small, then the covariance matrix of the predicted state vector $\boldsymbol{\Sigma}_{k|k-1}$ will be small. Consequently the Kalman gain will be small, too, and the state estimate is basically the predicted state vector. Hence, if the system noise is large compared to the measurement noise, the Kalman filter will trust the true observation, more than the state prediction. On the contrary, if the system noise is smaller than the measurement noise, more weight will be put on the prediction than on the observation.

4.1.2 Nonlinear System, Gaussian Noise

In the general nonlinear case the probability density functions of eqs. (4.17), (4.18), (4.19) can usually not be calculated analytically and approximations must be made. Due to its wide spread in control theory, the standard approximation is the extended Kalman filter. In classical control theory nonlinear systems are linearized using the first oder Taylor approximation

$$\boldsymbol{f}(\boldsymbol{x}) \approx \boldsymbol{f}(\boldsymbol{x}_0) + \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}|_{\boldsymbol{x}=\boldsymbol{x}_0} (\boldsymbol{x} - \boldsymbol{x}_0)\,, \tag{4.30}$$

with $\boldsymbol{x}_0$ denoting the operation point around which $\boldsymbol{f}(\boldsymbol{x})$ is linearized. Once the linear approximations are obtained, the classical linear control theory can be applied. One of the most often used controllers is the state feedback estimator, which feedbacks the measured states to the input of system. However, often these states cannot completely be measured, because no adequate sensors exist. Also, as the number of states becomes large, the measurement of all of them might become quite expensive. In such cases the extended Kalman filter is used, which uses the measures $\boldsymbol{z}_k$ to infer the states of the system. Those can then be fed back to the input of the control loop.

4.1.3 Extended Kalman Filter

If the classical Kalman filter is applied to a linearized system, one speaks of the extended Kalman filter. Its prediction and update stages are only slightly modified versions of the classical Kalman filter.

Assume that the system equation (4.1) is first order linearized, such that

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{w}_k) \\ &\approx \boldsymbol{A}_k(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}) + \boldsymbol{G}_k(\boldsymbol{w}_k - \boldsymbol{0}), \end{aligned} \tag{4.31}$$

with

$$\boldsymbol{A}_k = \frac{\partial \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{w}_k)}{\partial \boldsymbol{x}_k}|_{\boldsymbol{x}_k=\boldsymbol{x}_{k|k}, \boldsymbol{w}_k=\boldsymbol{0}}, \tag{4.32}$$

and

$$\boldsymbol{G}_k = \frac{\partial \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{w}_k)}{\partial \boldsymbol{w}_k}|_{\boldsymbol{x}_k=\boldsymbol{x}_{k|k}, \boldsymbol{w}_k=\boldsymbol{0}}. \tag{4.33}$$

Similarly, the measurement equation (4.2) is first order linearized, such that

$$\begin{aligned} \boldsymbol{z}_k &= \boldsymbol{g}(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{v}_k \\ &\approx \boldsymbol{C}_k(\boldsymbol{x}_k - \boldsymbol{x}_{k|k}) + \boldsymbol{v}_k, \end{aligned} \tag{4.34}$$

with

$$\boldsymbol{C}_k = \frac{\partial \boldsymbol{g}(\boldsymbol{x}_k, \boldsymbol{u}_k)}{\partial \boldsymbol{x}_k}|_{\boldsymbol{x}_k=\boldsymbol{x}_{k|k}}. \tag{4.35}$$

The extended Kalman filter is initialized such as the standard Kalman filter by eqs. (4.26) and (4.27). Its update and prediction stages are then trivial variations of the standard Kalman filter [AM79].

Update Stage :

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T(\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T + \boldsymbol{R}_k)^{-1} \tag{4.36}$$

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{g}(\boldsymbol{x}_{k|k-1}, \boldsymbol{u}_k)) \tag{4.37}$$

$$\boldsymbol{\Sigma}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\boldsymbol{\Sigma}_{k|k-1} \tag{4.38}$$

Prediction Stage :

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{0}) \tag{4.39}$$

$$\boldsymbol{\Sigma}_{k+1|k} = \boldsymbol{A}_k\boldsymbol{\Sigma}_{k|k}\boldsymbol{A}_k^T + \boldsymbol{G}_k\boldsymbol{Q}_k\boldsymbol{G}_k^T \tag{4.40}$$

4.1.4 Unscented Kalman Filter

Julier and Uhlmann [JU97] proposed an alternative approach to the extended Kalman filter for nonlinear, Gaussian systems. Rather than linearizing the nonlinear equations, they propose to approximate the resulting probability density functions as Gaussian densities. This approximation is carried out by the unscented transformation, which is a method for calculating the statistics of a random variable $\boldsymbol{x}$ which undergoes a nonlinear transformation

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}). \tag{4.41}$$

A set of so-called *Sigma points* $\mathcal{X}_i$ are chosen such that their sample mean $\bar{\boldsymbol{x}}$ and their sample covariance $\boldsymbol{P}_{\boldsymbol{xx}}$ are equal to the mean and covariance of the random variable $\boldsymbol{x}$. The nonlinear function $\boldsymbol{f}(\boldsymbol{x})$ is applied on each sigma point $\mathcal{X}_i$, which results in the corresponding points

$$\mathcal{Y}_i = \boldsymbol{f}(\mathcal{X}_i). \tag{4.42}$$

These $\mathcal{Y}_i$ are then used to approximate the mean and the covariance of $\boldsymbol{y}$. As opposed to Monte Carlo-type methods, the sigma points are not selected at random, but by a specific, deterministic algorithm. Julier proposed an extension to the classical unscented transformation [JU97], called the scaled unscented transformation, in [Jul02]. For an n-dimensional random variable

$\boldsymbol{x}$ with mean $\bar{\boldsymbol{x}}$ and covariance $\boldsymbol{P_{xx}}$ the $2n+1$ scaled sigma points are given by

$$\begin{array}{lcllcl} \mathcal{X}_0 & = & \bar{\boldsymbol{x}}, & w_0 & = & \lambda/(n+\lambda), \\ \mathcal{X}_i & = & \bar{\boldsymbol{x}} + (\sqrt{(n+\lambda)\boldsymbol{P_{xx}}})_i, & w_i & = & \lambda/(n+\lambda), \\ \mathcal{X}_{i+n} & = & \bar{\boldsymbol{x}} - (\sqrt{(n+\lambda)\boldsymbol{P_{xx}}})_i, & w_i & = & \lambda/(n+\lambda), \end{array} \tag{4.43}$$

with $\lambda = \alpha^2(n+\kappa) - n$, $\kappa \in \mathbb{R}$, and $0 \leq \alpha \leq 1$ being scaling parameters. $(\sqrt{(n+\kappa)\boldsymbol{P_{xx}}})_i$ denotes the ith row or column of the matrix square root of $(n+\kappa)\boldsymbol{P_{xx}}$, found by e.g. the Cholesky factorization, and w_i is the weight which is associated with the ith sigma point $\mathcal{X}_i$.

After the sigma points SXi are passed through eq. (4.42), the mean of the random variable $\boldsymbol{y}$ can be approximated by

$$\bar{\boldsymbol{y}} = \sum_{i=0}^{2n} w_i \mathcal{Y}_i. \tag{4.44}$$

Correspondingly, the covariance matrix of $\boldsymbol{y}$ is approximated by

$$\boldsymbol{P_{yy}} = \sum_{i=1}^{2n} w_i (\mathcal{Y}_i - \bar{\boldsymbol{y}})(\mathcal{Y}_i - \bar{\boldsymbol{y}})^T + (w_0 + 1 + \beta - \alpha^2)(\mathcal{Y}_0 - \bar{\boldsymbol{y}})(\mathcal{Y}_0 - \bar{\boldsymbol{y}})^T, \tag{4.45}$$

with β being an additional weighting term, which weights the importance of the zeroth sigma point for the calculation of the covariance. By this scaled unscented transformation the mean and covariance of $\boldsymbol{y}$ are captured precisely up to the second order. In [Jul02] $\beta = 2$ is shown to be optimal for a Gaussian prior.

κ can be used as an extra degree of freedom to "fine tune" the higher order moments of the approximation. In [JU97] it is proposed that if $\boldsymbol{x}$ is assumed to be Gaussian, to choose κ such that

$$n + \kappa = 3. \tag{4.46}$$

Alternative weighting terms are propose in literature to further increase the approximation accuracy. Wan and Merwe [WM01] for instance propose to use different weighting terms for the calculation of the mean and covariance. Some alternative implementations can be found in [vdMWJ04], which also furnishes the corresponding pseudo-codes for implementing the Sigma-Point Kalman, or the unscented Kalman filter.

In order to utilize these results for recursive Bayesian estimation, the assumption is made that all densities remain Gaussian. In this case, the densities are completely described by their mean and covariance. Further,

an innovation vector $\boldsymbol{\nu}_k$ is introduced, by which the Kalman equations can alternatively be derived, by the following linear update rule

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k \boldsymbol{\nu}_k, \tag{4.47}$$

with

$$\boldsymbol{\nu}_k = \boldsymbol{z}_k - \boldsymbol{g}(\boldsymbol{x}_{k|k-1}, \boldsymbol{u}_k, \boldsymbol{0}) \tag{4.48}$$

denoting the innovation vector, which is the difference between the predicted observation vector and the actual sensor observation. The Kalman gain $\boldsymbol{K}_k$ is chosen such that it minimizes the mean square error of the estimate:

$$\boldsymbol{K}_k = \arg\min_{\boldsymbol{K}_k} E((\boldsymbol{x}_{k|k} - \boldsymbol{x}_k)^T (\boldsymbol{x}_{k|k} - \boldsymbol{x}_k) | \boldsymbol{Z}_{1:k}). \tag{4.49}$$

In can be shown (e.g. [AM79], [Kal60]) that the Kalman gain then results to

$$\boldsymbol{K}_k = \boldsymbol{P}_{\boldsymbol{x}\boldsymbol{\nu}}(k|k-1) \boldsymbol{P}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{-1}(k|k-1), \tag{4.50}$$

with

$$\boldsymbol{P}_{\boldsymbol{x}\boldsymbol{\nu}}(k|k-1) = E((\boldsymbol{x}_{k|k} - \boldsymbol{x}_k) \boldsymbol{\nu}_k^T | \boldsymbol{Z}_{1:k-1}), \tag{4.51}$$

and

$$\boldsymbol{P}_{\boldsymbol{\nu}\boldsymbol{\nu}}(k|k-1) = E((\boldsymbol{\nu}_k \boldsymbol{\nu}_k^T | \boldsymbol{Z}_{1:k-1}). \tag{4.52}$$

The filter iteration then follows to [Jul97]:

Prediction

State Prediction

$$\boldsymbol{x}_{k|k-1} = E(\boldsymbol{x}_k | \boldsymbol{Z}_{1:k-1}) \tag{4.53}$$

$$\boldsymbol{P}_{\boldsymbol{x}\boldsymbol{x}}(k|k-1) = E((\boldsymbol{x}_{k|k} - \boldsymbol{x}_k)(\boldsymbol{x}_{k|k} - \boldsymbol{x}_k)^T | \boldsymbol{Z}_{1:k}) \tag{4.54}$$

Observation Prediction

$$\hat{\boldsymbol{z}}_{k|k-1} = E(\boldsymbol{z}_k | \boldsymbol{Z}_{1:k-1}) \tag{4.55}$$

$$\boldsymbol{P}_{\boldsymbol{\nu}\boldsymbol{\nu}}(k|k-1) = E(\boldsymbol{\nu}_k \boldsymbol{\nu}_k^T | \boldsymbol{Z}_{1:k-1}) \tag{4.56}$$

$$\boldsymbol{P}_{\boldsymbol{x}\boldsymbol{\nu}}(k|k-1) = E(\boldsymbol{x}_{k|k-1} - \boldsymbol{x}_k) \boldsymbol{\nu}_k^T | \boldsymbol{Z}_{1:k-1}) \tag{4.57}$$

Update

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k \boldsymbol{\nu}_k \tag{4.58}$$

$$\boldsymbol{P}_{\boldsymbol{xx}}(k|k) = \boldsymbol{P}_{\boldsymbol{xx}}(k|k-1) - \boldsymbol{K}_k \boldsymbol{P}_{\boldsymbol{\nu\nu}}(k|k-1)\boldsymbol{K}_k^T \tag{4.59}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{\boldsymbol{x\nu}}(k|k-1)\boldsymbol{P}_{\boldsymbol{\nu\nu}}^{-1}(k|k-1), \tag{4.60}$$

$$\boldsymbol{\nu}_k = \boldsymbol{z}_k - \boldsymbol{g}(\boldsymbol{x}_{k|k-1}, \boldsymbol{u}_k, \boldsymbol{0}) \tag{4.61}$$

In order to apply the unscented transformation on these recursion steps, an augmented state vector is introduced:

$$\boldsymbol{x}_k^a = \begin{pmatrix} \boldsymbol{x}_k \\ \boldsymbol{w}_k \end{pmatrix}. \tag{4.62}$$

Since it is assumed that even after the nonlinear transformations Gaussian variables are obtained, this augmented state vector is a random variable with approximated mean $[\boldsymbol{x}_{k|k}^T, \boldsymbol{0}]^T$ and covariance matrix

$$\boldsymbol{P}_{k|k}^a = \begin{bmatrix} \boldsymbol{P}_{\boldsymbol{xx}}(k|k) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_k \end{bmatrix}. \tag{4.63}$$

Note, that it is assumed that $\boldsymbol{x}_k$ and $\boldsymbol{w}_k$ are uncorrelated. The $2n+1$ sigma points $\mathcal{X}_i^a(k|k)$ according to this density function can be calculated and the corresponding sigma points $\mathcal{X}_i^a(k+1|k)$ of $\boldsymbol{x}_{k+1|k}$ can be calculated by passing them through the nonlinear system equation (4.1). In order to initialize the unscented Kalman filter, the sigma points representing $\boldsymbol{x}_{0|0}$ are drawn from the prior density function of $\boldsymbol{x}_0 \sim \mathcal{N}(\bar{\boldsymbol{x}}_0, \boldsymbol{P}_0)$ and are denoted by

$$\mathcal{X}_i(0|0), \tag{4.64}$$

with weights w_i^x. The unscented Kalman filter then follows to

Prediction

State Prediction

$$\mathcal{X}_i^a(k|k-1) = \boldsymbol{f}\left(\mathcal{X}_i(k-1|k-1), \boldsymbol{u}_{k-1}\right) \tag{4.65}$$

$$\boldsymbol{x}_{k|k-1} = \sum_{i=0}^{2n^a} w_i^x \mathcal{X}_i(k|k-1), \tag{4.66}$$

with n^a being the dimension of the augmented state vector

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{xx}}(k|k-1) &= \sum_{i=0}^{2n^a} w_i^x \left(\mathcal{X}_i(k|k-1) - \boldsymbol{x}_{k|k-1}\right)\left(\mathcal{X}_i(k|k-1) - \boldsymbol{x}_{k|k-1}\right)^T + \\ &(w_0^x + 1 + \beta - \alpha^2)\left(\mathcal{X}_0(k|k-1) - \boldsymbol{x}_{k|k-1}\right)\cdot \\ &\cdot \left(\mathcal{X}_0(k|k-1) - \boldsymbol{x}_{k|k-1}\right)^T \end{aligned} \tag{4.67}$$

Observation Prediction For the observation prediction step only the sigma point inputs representing the state variable $\boldsymbol{x}_i$ are needed:

$$\mathcal{X}_i(k|k-1) = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \mathcal{X}_i^a(k|k-1), \tag{4.68}$$

with $\boldsymbol{I} \in \mathbb{R}^{n,n}$ being the identity matrix.

$$\mathcal{Z}_i(k|k-1) = \boldsymbol{g}(\mathcal{X}_i(k|k-1), \boldsymbol{u}_k) \tag{4.69}$$

$$\hat{\boldsymbol{z}}_{k|k-1} = \sum_{i=0}^{2n} w_i^z \mathcal{Z}_i(k|k-1) \tag{4.70}$$

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{\nu\nu}}(k|k-1) = &\sum_{i=0}^{2n} w_i^z (\mathcal{Z}_i(k|k-1) - \hat{\boldsymbol{z}}_{k|k-1})(\mathcal{Z}_i(k|k-1) - \hat{\boldsymbol{z}}_{k|k-1})^T + \\ &+ \boldsymbol{R}_k + (w_0^z + 1 + \beta - \alpha^2)(\mathcal{Z}_0(k|k-1) - \hat{\boldsymbol{z}}_{k|k-1})\cdot \\ &\cdot (\mathcal{Z}_0(k|k-1) - \hat{\boldsymbol{z}}_{k|k-1})^T \end{aligned} \tag{4.71}$$

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{x\nu}}(k|k-1) = &\sum_{i=0}^{2n} w_i^z (\mathcal{X}_i(k|k-1) - \boldsymbol{x}_{k|k-1})(\mathcal{Z}_i(k|k-1) - \hat{\boldsymbol{z}}_{k|k-1})^T + \\ &+ (w_0^z + 1 + \beta - \alpha^2)(\mathcal{X}_0(k|k-1) - \boldsymbol{x}_{k|k-1})\cdot \\ &\cdot (\mathcal{Z}_0(k|k-1) - \hat{\boldsymbol{z}}_{k|k-1})^T \end{aligned} \tag{4.72}$$

$$\tag{4.73}$$

Update

$$\boldsymbol{K}_k = \boldsymbol{P}_{\boldsymbol{x\nu}}(k|k-1)\boldsymbol{P}_{\boldsymbol{\nu\nu}}^{-1}(k|k-1) \tag{4.74}$$

$$\boldsymbol{\nu}_k = \boldsymbol{z}_k - \hat{\boldsymbol{z}}_{k|k-1} \tag{4.75}$$

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\nu}_k \tag{4.76}$$

$$\boldsymbol{P}_{\boldsymbol{xx}}(k|k) = \boldsymbol{P}_{\boldsymbol{xx}}(k|k-1) - \boldsymbol{K}_k\boldsymbol{P}_{\boldsymbol{\nu\nu}}(k|k-1)\boldsymbol{K}_k^T \tag{4.77}$$

The update stage is ended by calculating the sigma points $\mathcal{X}_i(k|k)$ corresponding to the probability density function $\mathcal{N}(\boldsymbol{x}_{k|k}, \boldsymbol{P}_{\boldsymbol{xx}}(k|k))$.

A detailed summary on the unscented transformation and the resulting unscented Kalman filter can be found in [vdM04]. It summarizes its main properties as

- Mean and covariance of state estimate is calculated accurately to at least the second order as opposed to the limited first order accuracy of the extended Kalman filter
- The unscented Kalman filter and the extended Kalman filter have an equivalent computational complexity
- In contrast to the extended Kalman filter, no analytical derivatives (Jacobian) need to be calculated. Hence, it can be applied on discontinuous functions
- The unscented Kalman filter consistently outperforms (or at least equals) the extended Kalman filter for state and parameter estimation

Julier and Uhlmann [JU97], [JU04] propose the problem of converting from polar to Cartesian coordinates, as an example for showing the superiority of the unscented transformation (unscented Kalman filter) over the linearized model estimation (extended Kalman filter). One possible application is a active sonar sensor, such as the "range-optimized" sonar sensor proposed in [LDW91].

Example 2. *A range-optimized sonar sensor can provide fairly good range measurements r, with e.g. a standard deviation of $2cm$, but extremely poor bearing measurements β, with e.g. a standard deviation of $15°$. These polar coordinates need to be converted to the Cartesian coordinates x, y by the nonlinear equation*

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos(\beta) \\ r\sin(\beta) \end{pmatrix}. \tag{4.78}$$

Suppose that the true position of an object is $[0m, 1m]$, then the true range and bearings result to $r = 1$ and $\beta = 90°$, respectively. Carrying out 100000 simulations by adding Gaussian noise with the above mentioned standard deviations, the effects of the unscented transformation, using two different weighting terms and the linearized estimation are presented in fig. 4.1 It shows the mean estimate of the position in Cartesian coordinates found by the 100000 Monte Carlo trials, by the linearized model and by the unscented transformation. Further, the 1-contour plots ($\{\boldsymbol{x} : (\boldsymbol{x} - \bar{\boldsymbol{x}})^T \boldsymbol{P}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}) = 1\}$) of the the covariance matrices $\boldsymbol{P}$ are plotted. The both, the scaled unscented transformation [Jul02], as well as the original unscented transformation [JU97] are tested and compared to the linearized estimate and the true values.

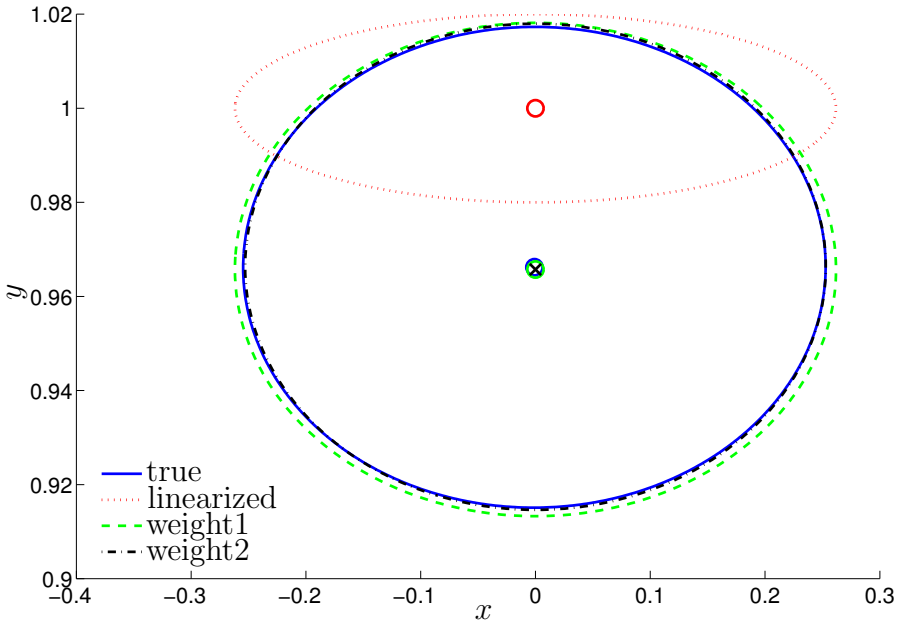

Figure 4.1: Polar to Cartesian coordinate transformation in sonar applications, weight1: $\alpha = 0.001$, $\beta = 2$ and $\kappa = 0$ (scaled unscented transformation [Jul02]), weight2: $\alpha = \beta = 0$, $\kappa = 3 - n = 1$ (classical unscented tranformation [JU97])

It can be seen, that both unscented transformations accurately estimate the true mean and the true covariance of the nonlinear system, while the linearized estimator is biased and inconsistent, in that the estimated covariance is smaller than the true covariance. In recursive estimation this inconsistency can lead to progressively worse position estimates, because the estimator believes to know the state estimate better than it really does [Jul97].

4.1.5 Nonlinear System, Non-Gaussian Noise

The extended and the unscented Kalman filter both approximate the posterior density function $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ by a Gaussian function. However, this approximation can fail, if the considered functions are strongly nonlinear, or if the considered noises are non-Gaussian. Recently particle filters have been proposed to recursively update the posterior pdf using *sequential importance sampling and resampling* [DGA00], [AMGC02]. Their principle idea is

to represent the posterior pdf by a set of weighted samples (particles) and to compute the desired state estimate based on these samples and weights. Using Monte Carlo simulation the posterior density function can be approximated by the following empirical estimate [vdM04]:

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) = \hat{p}(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i), \tag{4.79}$$

where the random samples $\boldsymbol{x}_k^i, i = 1, \ldots, N$ are drawn from $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ and $\delta(.)$ denotes the Dirac delta function. According to the law of large numbers any expectation of the form

$$E(\boldsymbol{g}(\boldsymbol{x}_k)) = \int \boldsymbol{g}(\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})d\boldsymbol{x}_k \tag{4.80}$$

can be approximated by the sum

$$E(\boldsymbol{g}(\boldsymbol{x}_k)) \approx \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}(\boldsymbol{x}_k^i). \tag{4.81}$$

Lately a large number of algorithms appeared on how to best utilize the Monte Carlo simulations for recursive Bayesian estimation. The currently most widely utilized ones are the *sampling importance resampling* (SIR), or *recursive bootstrap filter* [GSS93], and the Bayesian *sequential importance sampling* (SIS) filter. In [AMGC02] it is shown how the SIR filter can be derived from the SIS filter and a variety of alternatives, which all are based on the SIS filter are presented. In [DGA00] these estimators are compared to each other, when applied to a linear Gaussian system and a nonlinear Gaussian system. For the linear system both estimators achieve the same accuracy as the optimal Kalman filter. For the nonlinear system both estimators achieve equivalent accuracies. Similar results were found by Bergman [Ber99], who compared the two estimators on a realistic navigation system. Due to the similar performances only the SIR estimator is presented in the following lines.

In order to apply the SIR filter, it must be possible to draw samples from the prior density function $p(\boldsymbol{x}_1) = p(\boldsymbol{x}_1|\boldsymbol{z}_0)$ and from the transition density function $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^i)$. Assume that we have N samples drawn from the pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1})$, with equal weights $w_i^k = 1/N$. Then this pdf can be approximated by

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1}) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i), \tag{4.82}$$

and the update stage can be approximated by

$$\begin{aligned} p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) &= \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1})}{p(\boldsymbol{z}_k|\boldsymbol{Z}_{1:k-1})} \\ &\propto p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k-1}) \\ &\approx \sum_{i=1}^{N} \underbrace{p(\boldsymbol{z}_k|\boldsymbol{x}_k)/N}_{\tilde{w}_i^k} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i). \end{aligned} \tag{4.83}$$

Note that due to the proportionality the weights are not normalized. Normalizing these,

$$w_k^i = \tilde{w}_i^k / \sum_{i=1}^{N} \tilde{w}_i^k, \tag{4.84}$$

the update pdf approximates to

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) \approx \sum_{i=1}^{N} w_k^i \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i). \tag{4.85}$$

The minimum variance estimate of eq. (4.11) could then be approximated by combining eqs. (4.12) and (4.85)

$$\hat{\boldsymbol{x}}_k = E(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) \approx \sum_{i=1}^{N} w_k^i \boldsymbol{x}_k^i. \tag{4.86}$$

The prediction step is initialized by generating a new set of M equally weighted samples of $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$, so that it can be approximated by

$$p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k}) \approx \frac{1}{M} \sum_{i=1}^{M} \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i). \tag{4.87}$$

The prediction pdf of eq. (4.5) is then approximated by

$$\begin{aligned} p(\boldsymbol{x}_{k+1}|\boldsymbol{Z}_{1:k}) &= \int_{-\infty}^{\infty} p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})d\boldsymbol{x}_k \\ &\approx \sum_{i=1}^{M} p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^i). \end{aligned} \tag{4.88}$$

Drawing N samples, with equal weight w_i^{k+1} from eq. (4.88) allows us to return to the update stage. This sampling stage is usually carried out by sampling N/M samples from each of the M pdf

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k^i). \tag{4.89}$$

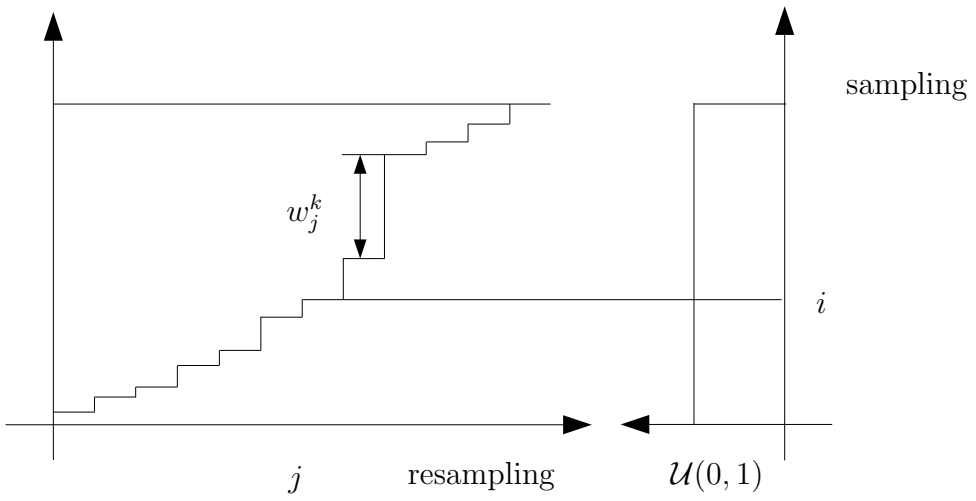

Figure 4.2: Resampling procedure for producing M equally weighted particles out of the set of N weighted particles

For reliable results M should be chosen greater than N. Rubin's guideline [Rub88] is to choose at least $M = 10N$ [Ber99].

One crucial aspect of the implementation of this recursive bootstrap filter is the resampling procedure at the beginning of the prediction step. The straightforward procedure would be to generate M independent identically distributed variables β_i of the uniform density function $\mathcal{U}(0,1)$, which has a uniform distribution on the interval $[0,1]$. The M generated variables would then be sorted in ascending order and compared to the cumulative sum of the normalized weights w_k^i. At sampling step j the particle $\boldsymbol{x}_k^j$ is chosen such that β_k is equal to the cumulate sum till weight w_k^j. Fig 4.2 gives a graphical representation of this procedure. However, the best sorting algorithm has a complexity of order $M \log M$ [Ber99], meaning that the computational time can be approximated by $t = k \cdot M \log M$, with k being some constant. It has been noted in practical applications that this severely limits the implementation possibilities. In order to decrease the computational burden, a variety of resampling procedures have been proposed. Hol et al. [HSG06] present and compare four of such resampling procedures, whose computational complexities are all linear in M. Namely those are the *multinomial resampling* [GSS93], the *stratified resampling* [Kit96], the *systematic resampling* [AMGC02], and the *residual resampling* [LC98] algorithms. Even though all these algorithms are linear in M, $t = k \cdot M$, the residual resampling algorithm has the smallest value k and therefore is fastest. It is a two step

procedure, which in the first step calculates a number of deterministically chosen children of particle i by using the floor function

$$M_i^d = \lfloor M w_i^k \rfloor. \tag{4.90}$$

The remaining

$$M^s = M - \sum_{i=1}^{M} M_i^d \tag{4.91}$$

samples are then sampled using the new weights

$$\tilde{w}_i^k = \frac{1}{M^s} \left(w_i^k M - M_i^d \right). \tag{4.92}$$

This statistical second step can be carried out by the systematic resampling algorithm, which generates M^s ordered uniformly distributed random numbers by

$$\beta_i = \frac{(i-1) + u_i}{M^s} \quad \text{with} \quad u_i \sim \mathcal{U}(0,1). \tag{4.93}$$

In order to demonstrate the strength of particle filters, the following example was proposed by van der Merwe [vdM04].

Example 3. *The state space model is given by*

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \boldsymbol{w}_k \\ \boldsymbol{z}_k &= \alpha \boldsymbol{x}_k^2 + \boldsymbol{v}_k, \end{aligned} \tag{4.94}$$

where $\boldsymbol{x}_{k+1} \in \mathbb{R}$ $\boldsymbol{w}_k$ is the additive Gaussian process noise term with a very small variance ($\sigma_x^2 = 1e{-}6$), $\alpha = 10$ is a scaling term, and $\boldsymbol{v}_k$ denotes the additive Gaussian measurement noise with variance $\sigma^2 = 1$. Due to the squared nonlinearity of the observation function, it is impossible to disambiguate the sign of the true value of the state based on the observations alone. In other words, the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ will have two maxima, one at $\boldsymbol{x}_k = x_t$ and one at $\boldsymbol{x}_k = -x_t$.

Fig. 4.3 shows the outcome of the extended Kalman filter, the unscented Kalman filter, and the bootstrap filter, when the true state value is set to $x_t = -.5$. The initial state is randomly chosen as a random vector sampled from the uniform distribution $\mathcal{U}(-1,1)$. The bottom graph PF shows the results of the particle filter. The filter quickly converges toward the two possible state outcomes $\boldsymbol{x}_k = x_t$ and $\boldsymbol{x}_k = -x_t$. Since, the extended Kalman filter, as well as the unscented Kalman fiter, approximate the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ as Gaussian functions, both estimators fail on estimating the value of the state. While the extended Kalman filter quickly converges to one of the two possible outcomes (unfortunately to the wrong one in the current case), the unscented Kalman filter results in a broad Gaussian posterior pdf, with its mean in between the two possible state estimates.

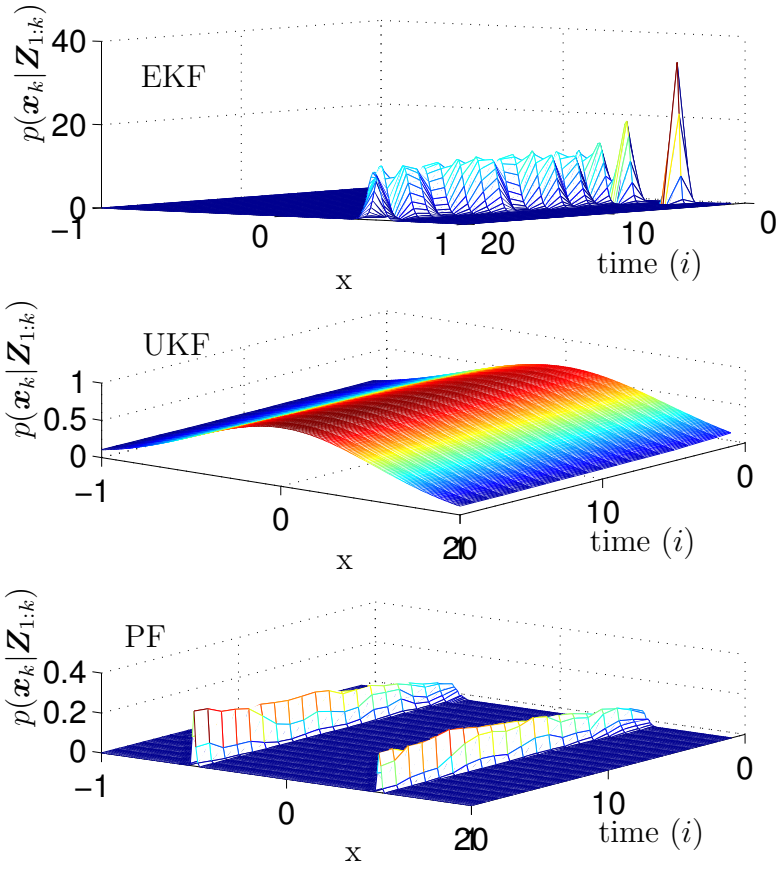

Figure 4.3: Nonlinear, non-Gaussian State Estimation

Clearly, given the posterior pdf found by the particle filter, the minimum variance estimate will be a bad choice for estimating the state value, since it is found by the mean value of the pdf, which in the current case, is around zero. Therefore, it might be favorable to use the MAP estimate in this case.

4.2 TDOA-based Passive Source Tracking

The above presented filters are now integrated in the problem of TDOA-based acoustic passive source localization. A multitude of applications for such a tracking system exist. One of the main areas of interest is the detection and localization via passive sonar of threatening vessels in coast and harbor

security and hostile vessels in military actions. A completely different application is the procedure of indoor speaker tracking for speech enhancement and teleconferences. Even though these two applications purpose completely different interests, they both utilize the same estimators and both systems can be greatly enhanced by using recursive Bayesian estimation. The observation sequence of the trajectories will be smoothed, which make further steps, such as inferring the aim of a hostile vessel, easier to implement. Future positions can more accurately be predicted, which can be used to further increase the accuracy of the estimators. For example, the estimate source position could be used to adaptively choose an appropriate sensor configuration, as proposed in chapter 2. Further, based on the estimated future position, the resulting future TDOA measures can be estimated. Often the resulting graphs of the TDOA estimators do not have a single, but multiple peaks with similar heights. In such a case the TDOA estimation becomes rather difficult, and often results in wrong position estimates. Predicting the probability density function of the future TDOA outcome, based on the already observed outcomes, this information can be used to help selecting from multiple peaks.

The key to successfully applying recursive Bayesian estimation is to establish an accurate mathematical state space description of the system.

4.2.1 System Equation

In tracking applications the source is usually considered to be a point mass and its control input is unknown. Further most often little information is available to construct the true object's dynamic model, but rather a kinematic model is developed, which describes the motion, without considering the cause of these movements. Li and Jilkov [LJ03] classify the motions into two classes: maneuver and non-maneuver. The non-maneuver motion describes a straight and level motion at a constant velocity. All other motions are collected under the term maneuver motions. Usually the motions of the maneuvering systems are due to changes of the control input. Since this generally non-random value is usually unknown, it is approximated as a random process with certain properties, which must be carefully chosen.

A sub-classification of target tracking is the distinction between coupled and uncoupled system equations. While in coupled systems a motion in one direction (e.g. $x-$axis) may impose a motion on another direction (e.g. $y-$axis), in uncoupled systems the x, y, z motions are independent of each other. Examples of coupled systems are vehicles, vessels and aircrafts. Contrary, the movement of humans can be modeled as an uncoupled system. Li and Jilkov [LJ03] offer a detailed survey on the varying system models for

target tracking.

Depending on the application, the system equation will be linear or nonlinear. For speaker tracking an uncoupled linear system equation describing a Langevin process is well-established in literature. The model was first proposed for speaker tracking by Vermaak and Blake [VB01] and later adopted by e.g. Ward et al. [WLW03] and Fallon and Godsill [FG07]. Assuming a $2D$ source tracking problem the state vector at time instant k is defined as

$$\boldsymbol{x}_k = (x_k, \dot{x}_k, y_k, \dot{y}_k)^T, \tag{4.95}$$

where $\dot{x}_k$ and $\dot{y}_k$ denote the velocities in $x-$ and $y-$ direction, respectively. The system equation is then given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{G}_k \boldsymbol{w}_k = \begin{bmatrix} \boldsymbol{A}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_y \end{bmatrix} \boldsymbol{x}_k + \begin{bmatrix} \boldsymbol{G}_x \\ \boldsymbol{G}_y \end{bmatrix} \begin{pmatrix} F_{x_k} \\ F_{y_k} \end{pmatrix}, \tag{4.96}$$

with

$$\boldsymbol{A}_x = \begin{bmatrix} 1 & \Delta T \\ 0 & a_x \end{bmatrix}, \quad \boldsymbol{A}_x = \begin{bmatrix} 1 & \Delta T \\ 0 & a_x \end{bmatrix}, \quad \boldsymbol{G}_x = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{G}_y = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{4.97}$$

and $a_i = \exp(-\beta_i \Delta T)$, $b_x = \bar{v}_i \sqrt{1 - a_i^2}$. β_i denotes a rate constant in direction i and $\bar{v}_i$ is the steady-state root-mean-square velocity [VB01].

4.2.2 TDOA Measure Tracking

The common measurement equation of TDOA-based acoustic passive source tracking is the relation between the TDOA vector $\boldsymbol{\tau}$ and the source position $\boldsymbol{x}_s$, given by

$$\boldsymbol{\tau}_1 = \begin{pmatrix} \tau_{21} \\ \vdots \\ \tau_{N1} \end{pmatrix} = \begin{pmatrix} (|\boldsymbol{x}_2 - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2)/c \\ \vdots \\ (|\boldsymbol{x}_2 - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2)/c \end{pmatrix} + \boldsymbol{n}_1, \tag{4.98}$$

if only the TDOAs w.r.t. the reference sensor $\boldsymbol{x}_1$ are available, or given by

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_{21} \\ \vdots \\ \tau_{N,N-1} \end{pmatrix} = \begin{pmatrix} (|\boldsymbol{x}_2 - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2)/c \\ \vdots \\ (|\boldsymbol{x}_N - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_{N-1} - \boldsymbol{x}_s|_2)/c \end{pmatrix} + \boldsymbol{n}, \tag{4.99}$$

if all $N(N-1)/2$ TDOA measures of the N microphones are accessible. Note that both equations are of the same form as eq. (4.2). Both are clearly nonlinear and hence, the extended Kalman filter, the unscented Kalman filter,

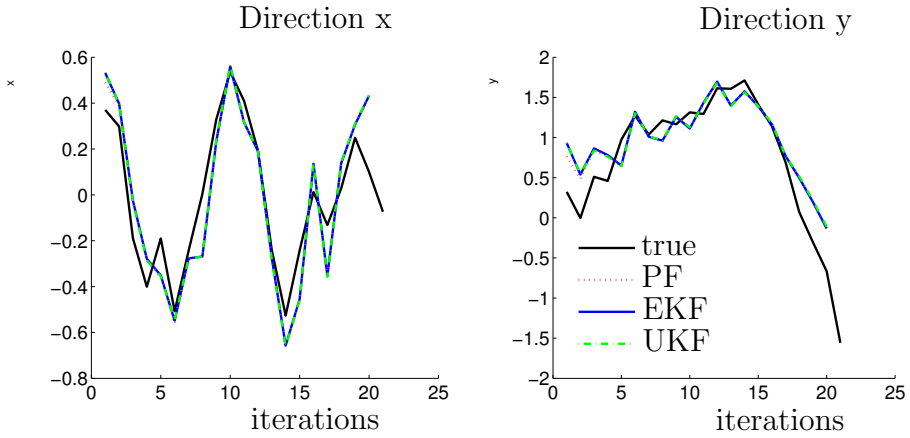

Figure 4.4: TDOA Based Tracking Estimates in x and y direction

or the Bayesian bootstrap filter must be applied for the tracking algorithm. The measurement vector is then either $\boldsymbol{z}_k = \boldsymbol{\tau}_1$, or $\boldsymbol{z}_k = \boldsymbol{\tau}$ and the considered additive measurement noise of eq.(4.2) corresponds to $\boldsymbol{v}_k = \boldsymbol{n}_1$, or $\boldsymbol{v}_k = \boldsymbol{n}$.

Example 4. *The three presented recursive estimators for nonlinear problems are tested on the following two dimensional source localization problem. The state of dynamic system is chosen to be the source position $\boldsymbol{x}_k = \boldsymbol{x}_s(k)$ and is updated by*

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{w}_k, \tag{4.100}$$

with $\boldsymbol{w}_k$ denoting zero-mean additive Gaussian noise with covariance matrix $\boldsymbol{Q}_k = 0.1m^2\boldsymbol{I}$. The initial state is randomly chosen from a Gaussian density with covariance $\boldsymbol{P}_0 = \boldsymbol{Q}_k$. Three microphones are set to positions $\boldsymbol{x}_1 = [5m, 0]$, $\boldsymbol{x}_2 = [-5m, 0]$, $\boldsymbol{x}_3 = [0, 5m]$ and only the time delays w.r.t. the reference sensor $\boldsymbol{x}_1$ are considered (eq. (4.99)). Further it is assumed that these measures are subject to zero-mean additive noise with covariance $\boldsymbol{R}_k = 0.1m^2\boldsymbol{I}$. The bootstrap filter is implemented using $N = 5000$ particles.

Figs. 4.4 compares the estimates of the extended Kalman, the unscented Kalman and the bootstrap filter, which for the presented problem perform identically. This becomes even more obvious looking at fig. 4.5, which plots the estimated density functions of the three filters. Even though the measurement equation is nonlinear, the posterior density function $p(\boldsymbol{x}_k|\boldsymbol{Z}_{1:k})$ estimated by the particle filter strongly resembles a Gaussian density, with identical mean as the extended Kalman filter and the unscented Kalman filter. Also the covariance matrices of the estimates found by the three filters are quite similar, as can be seen in table. 4.1.

Consequently, it seems that either one of the three nonlinear estimators can be used for the problem of passive acoustic source tracking based on non-

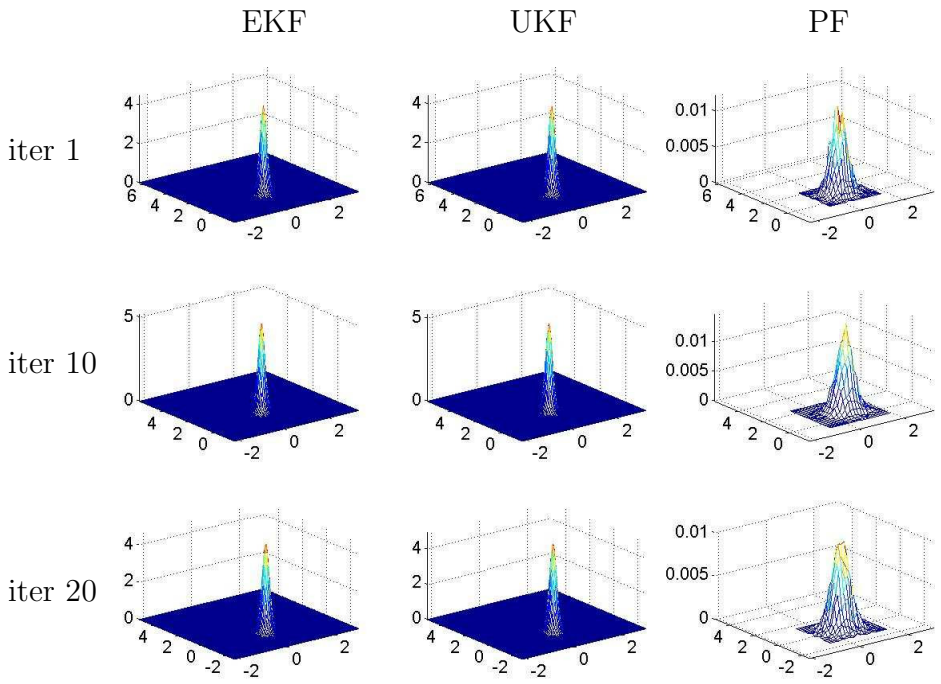

Figure 4.5: TDOA Based Tracking Densities

linear TDOA measures. In this case either the extended Kalman or the unscented Kalman filter are favorable, due to their smaller computational complexity.

Alternatively one of the linear approximation estimators or the linearized estimator of chapter 3 can be utilized to obtain a linear measurement equation, for which the optimal linear Kalman filter can be applied.

4.2.3 Position Measure Tracking

Instead of using eqs. (4.98), or (4.99) as the measurement equation of our nonlinear state space model, a TDOA-based position estimator of chapter 3 can be used to obtain a linear measurement equation. If the system equation is linear, too, the optimal Kalman filter can be applied for the problem of TDOA-based passive source localization. The measurement vector $\boldsymbol{z}_k$ is then equal to the outcome of the selected estimator

$$\boldsymbol{z}_k = \hat{\boldsymbol{x}}_s, \tag{4.101}$$

	EKF	UKF	PF
$i = 1$	$\begin{bmatrix} 0.021 & 0.015 \\ 0.015 & 0.072 \end{bmatrix}$	$\begin{bmatrix} 0.021 & 0.015 \\ 0.015 & 0.072 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.011 \\ 0.011 & 0.046 \end{bmatrix}$
$i = 10$	$\begin{bmatrix} 0.021 & 0.013 \\ 0.013 & 0.053 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.012 \\ 0.012 & 0.056 \end{bmatrix}$	$\begin{bmatrix} 0.023 & 0.017 \\ 0.017 & 0.061 \end{bmatrix}$
$i = 20$	$\begin{bmatrix} 0.019 & 0.013 \\ 0.013 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 0.017 & 0.012 \\ 0.012 & 0.07 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.014 \\ 0.014 & 0.078 \end{bmatrix}$

Table 4.1: Covariances of the EKF, UKF and particle filter for acoustic passive source localization at iteration steps $i = 1, 10, 20$

and the measurement equation follows to

$$\boldsymbol{z}_k = \boldsymbol{C}_k \boldsymbol{x}_k + \tilde{\boldsymbol{v}}_k, \tag{4.102}$$

with $\tilde{\boldsymbol{v}}_k$ representing zero-mean, additive noise with covariance $\tilde{\boldsymbol{R}}_k$ on the position estimate. Knowing the statistics of the additive noise on the TDOA measures, the statistics of the position estimate noise is dependent of the chosen estimator.

A natural choice is the linearized estimator of section 3.4.2, using the predicted outcome

$$\boldsymbol{z}_{k|k-1} = \boldsymbol{C}_{k-1} \boldsymbol{x}_{k|k-1} \tag{4.103}$$

as the initial guess for the estimator. From now on it is assumed that only the TDOAs w.r.t. the reference sensor $\boldsymbol{\tau}_1$ are available. In this case the estimator results in

$$\boldsymbol{z}_k = \boldsymbol{z}_{k|k-1} + \boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{\tau}_1 - \boldsymbol{g}(\boldsymbol{z}_{k|k-1})), \tag{4.104}$$

with

$$\boldsymbol{g}(\boldsymbol{z}_{k|k-1}) = \begin{pmatrix} \tau_{21} \\ \vdots \\ \tau_{N1} \end{pmatrix} = \begin{pmatrix} (|\boldsymbol{x}_2 - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2)/c \\ \vdots \\ (|\boldsymbol{x}_2 - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2)/c \end{pmatrix}, \tag{4.105}$$

and $\boldsymbol{\mathcal{J}}$ being the Jacobian of $\boldsymbol{g}(\boldsymbol{x})$ w.r.t. $\boldsymbol{x}$, evaluated at $\boldsymbol{x} = \boldsymbol{z}_{k|k-1}$.

If zero-mean, $\boldsymbol{\mu} = \boldsymbol{0}$, additive noise on the TDOA measures is considered, the mean and covariance of the linearized estimator result to (see: eqs. (2.56) and (2.57))

$$E(\boldsymbol{z}_k) = \boldsymbol{\theta} + \boldsymbol{\mathcal{J}}^{\dagger}\boldsymbol{\mu} = \boldsymbol{\theta}, \tag{4.106}$$

and

$$\tilde{\boldsymbol{R}}_k = \boldsymbol{\mathcal{J}}^{\dagger} \boldsymbol{R}_k (\boldsymbol{\mathcal{J}}^{\dagger})^T, \tag{4.107}$$

with $\boldsymbol{R}_k$ denoting the covariance matrix of the additive noise on the TDOA measures. The state estimate of the linear system of eq. (4.21) then follows to

$$\begin{aligned} \boldsymbol{x}_{k|k} &= \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{C}_k\boldsymbol{x}_{k|k-1}) \\ &= \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_{k|k-1} + \boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{\tau}_1 - \boldsymbol{g}(\boldsymbol{z}_{k|k-1})) - \boldsymbol{C}_k\boldsymbol{x}_{k|k-1}) \\ &= \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{\tau}_1 - \boldsymbol{g}(\boldsymbol{z}_{k|k-1})), \end{aligned} \tag{4.108}$$

with corresponding covariance eq. (4.23)

$$\boldsymbol{\Sigma}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}_k)\boldsymbol{\Sigma}_{k|k-1}. \tag{4.109}$$

The Kalman gain of eq. (4.24) results to

$$\begin{aligned} \boldsymbol{K}_k &= \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T(\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T + \tilde{\boldsymbol{R}}_k)^{-1} \\ &= \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T(\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T + \boldsymbol{\mathcal{J}}^{\dagger}\boldsymbol{R}_k(\boldsymbol{\mathcal{J}}^{\dagger})^T)^{-1}. \end{aligned} \tag{4.110}$$

Assuming that the Jacobian is a square matrix of full rank, eq. (4.110) can be written as

$$\boldsymbol{K}_k = \underbrace{\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T\boldsymbol{\mathcal{J}}^T(\boldsymbol{\mathcal{J}}\boldsymbol{C}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{C}_k^T\boldsymbol{\mathcal{J}}^T + \boldsymbol{R}_k)^{-1}}_{\tilde{\boldsymbol{K}}_k}\boldsymbol{\mathcal{J}}, \tag{4.111}$$

and eq. (4.108) follows to

$$\begin{aligned} \boldsymbol{x}_{k|k} &= \boldsymbol{x}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\mathcal{J}}^{\dagger}(\boldsymbol{\tau}_1 - \boldsymbol{g}(\boldsymbol{z}_{k|k-1})) \\ &= \boldsymbol{x}_{k|k-1} + \tilde{\boldsymbol{K}}_k(\boldsymbol{\tau}_1 - \boldsymbol{g}(\boldsymbol{z}_{k|k-1})). \end{aligned} \tag{4.112}$$

Accordingly, the covariance of the estimate follows from eq. 4.109 to

$$\boldsymbol{\Sigma}_{k|k} = (\boldsymbol{I} - \tilde{\boldsymbol{K}}_k\boldsymbol{\mathcal{J}}\boldsymbol{C}_k)\boldsymbol{\Sigma}_{k|k-1}. \tag{4.113}$$

Note $\boldsymbol{\mathcal{J}}$ is the Jacobian of $\boldsymbol{g}$ w.r.t. to the measurement vector $\boldsymbol{z}_k$, and not w.r.t. the state estimate $\boldsymbol{x}_k$. Let us the denote the Jacobian w.r.t. the state estimate $\boldsymbol{x}_k$ by $\boldsymbol{\mathcal{J}}_s$. Then, with $\boldsymbol{z}_k = \boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{v}_k$,

$$\boldsymbol{\mathcal{J}}_s = \frac{\delta\boldsymbol{g}}{\delta\boldsymbol{x}_k} = \frac{\delta\boldsymbol{g}}{\delta\boldsymbol{z}_k}\frac{\delta\boldsymbol{z}_k}{\delta\boldsymbol{x}_k} = \boldsymbol{\mathcal{J}}\boldsymbol{C}_k. \tag{4.114}$$

Replacing the terms $\boldsymbol{\mathcal{J}}\boldsymbol{C}_k$ by $\boldsymbol{\mathcal{J}}_s$ in eqs. (4.112), and (4.113), the update stage results to

$$\tilde{\boldsymbol{K}}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{\mathcal{J}}_s^T(\boldsymbol{\mathcal{J}}_s\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{\mathcal{J}}_s^T + \boldsymbol{R}_k)^{-1} \tag{4.115}$$

$$\boldsymbol{x}_{k|k} = \boldsymbol{x}_{k|k-1} + \tilde{\boldsymbol{K}}_k(\boldsymbol{z}_k - \boldsymbol{g}(\boldsymbol{x}_{k|k-1}, \boldsymbol{u}_k)) \tag{4.116}$$

$$\boldsymbol{\Sigma}_{k|k} = (\boldsymbol{I} - \tilde{\boldsymbol{K}}_k\boldsymbol{\mathcal{J}}_s)\boldsymbol{\Sigma}_{k|k-1} \tag{4.117}$$

Note that this update stage is identical to the update stage of the extended Kalman filter (eqs. (4.36), (4.37), and (4.38)). Obviously, for linear system equations the update stage of the two filters is identical, too.

4.3 Optimal Sensor Configuration

In the previous section it was shown that for linear system equations and squared, full rank Jacobians, the Kalman filter for a linear measurement equation based on the linearized estimator is identical to the extended Kalman filter. Further, it was shown in chapters 2 and 3 that the performance of the linearized estimator could be greatly increased by minimizing the condition number of the Jacobian $\boldsymbol{\mathcal{J}}$. Consequently, the performance of the extended Kalman filter will be improved, too.

Example 5. *Assume that a source has a rather chaotic movement represented by the system equation*

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{w}_k, \tag{4.118}$$

with $\boldsymbol{x}_k$ being the state of the system, which has the source position as its entries. The system is initialized by $\boldsymbol{x}_0 = [0, 0]$ and the zero-mean white Gaussian noise has a covariance matrix of $\boldsymbol{Q}_k = 0.1m^2\boldsymbol{I}$.

The linear Kalman filter using eq. (4.104), which results in a linear measurement equation of the form of eq. (4.103), with $\boldsymbol{C}_k = \boldsymbol{I}$, is compared to the recursive estimators for the nonlinear estimator resulting from the reference sensor based approach. The TDOA measures are assumed to be perturbed by zero-mean white noise with covariance matrix $\boldsymbol{R}_k = 0.1m^2\boldsymbol{I}$.

Ten trajectories were randomly constructed, each making measurements at 20 different time instances. The recursive estimators were applied to estimate the true source position using two different sensor networks, with four microphones each. The first configuration has the microphones placed at positions $[5m, 0m]$, $[-5m, 0m]$, $[0m, 5m]$, and $[0m, -5m]$. When the source is assumed to be at the origin, this corresponds to a condition number of around 1.7. The second sensor configuration is composed of four microphones placed at the positions $[5m, 0.25m]$, $[5m, -0.25m]$, $[-5m, 0.25m]$, and $[-5m, -0.25m]$. This corresponds to a condition number of around 23.

Fig. (4.6) presents the mean square error of the estimates of each trajectory. The upper bar plot shows the results for the well-posed configuration ($\kappa_2(\boldsymbol{\mathcal{J}}) \approx 1.7$): all four recursive estimators behave quite similar. The lower bars correspond to the results found by the "ill-posed" sensor network. As expected the Kalman filter as well as the extended Kalman filter now perform worse. In the previous section it was shown that these two filters are identical if the Jacobian is squared and of full rank. Note that the Jacobian in this example is not squared, but in $\mathbb{R}^{3,2}$. Still, both estimators perform nearly identical.

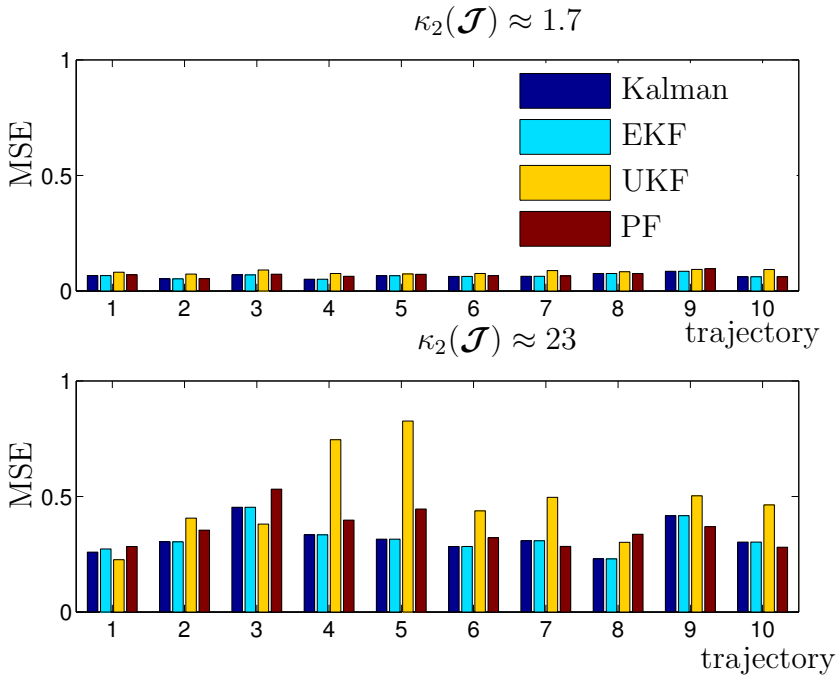

Figure 4.6: TDOA Based Tracking with varying condition numbers for the initial source position $\boldsymbol{x}_s = [0,0]^T$. The configuration with $\kappa_2(\mathcal{J}) \approx= 1.7$ uses 4 microphones located at $\boldsymbol{x}_1 = [5,0]^T$, $\boldsymbol{x}_2 = [-5,0]^T$, $\boldsymbol{x}_3 = [0,5]^T$, and $\boldsymbol{x}_4 = [0,-5]^T$.The configuration with $\kappa_2(\mathcal{J}) \approx= 23$ uses 4 microphones located at $\boldsymbol{x}_1 = [5,0.25]^T$, $\boldsymbol{x}_2 = [5,-0.25]^T$, $\boldsymbol{x}_3 = [-5,0.25]^T$, and $\boldsymbol{x}_4 = [-5,-0.25]^T$. Kalman: classical Kalman filter applied on the outcome of the linearized estimator, EKF: extended Kalman filter on the nonlinear measurement model, UKF: unscented Kalman filter, PF: particle filter based on 1000 samples

Further, note that the unscented Kalman filter and the bootstrap filter also show a clear dependency on the sensor configuration, even though they do not make use of the Jacobian of the measurement equation.

Consequently, the performance of TDOA-based acoustic passive source tracking can also be increased by choosing sensor configurations, which are optimal w.r.t. one of the proposed cost functions. Based on the above example the condition number seems to be a reasonable choice. It is dependent on the Jacobian of the measurement equation evaluated at the predicted state $\boldsymbol{x}_{k|k-1}$. *Obviously this prediction step can be integrated in the procedure for obtaining the next measurement, as discussed in section 2.4.3: e.g. if enough microphones are available and somehow distributed in the search region, only*

a subset of those can be chosen to carry out the estimation procedure, which fulfills that the condition number for the predicted state is minimized.

However, the adaptive optimization of the sensor network is only one possible advantage of the TDOA-based source tracking procedure. Obviously, another fundamental advantage of the tracking procedure is that the observation sequence becomes smoother, and the estimation outcomes become more reliable. A third important aspect is investigated in the following section.

4.4 Time Delay Estimations

In section 3.2 three general procedures for TDOA estimation were presented. Namely, those were the methods based on the cross-correlation function, the adaptive eigenvalue decomposition and the linear mean square estimator. All of these estimators choose the maximum peak of some discrete time limited sequence as the time difference of arrival estimate of one sensor pair. However, the received signals are often subject to large reverberations, or large ambient noise. In such cases, the resulting time sequence may have multiple local maxima and the peak corresponding to the true TDOA is not necessarily the largest. Obviously such a sequence will result in a wrong TDOA estimate and consequently in a wrong position estimate.

In order to avoid these wrong estimates, the recursive filters can be used to estimate the probability density function of the next estimation. Assuming a linear system equation with additive Gaussian noise, the state prediction at time instant $k-1$ denoted by $\boldsymbol{x}_{k|k-1}$ is Gaussian distributed, too, with corresponding covariance matrix $\boldsymbol{\Sigma}_{k|k-1}$ found by any of the presented recursive filters. Obviously, if the predicted source position $\boldsymbol{x}_s(k|k-1)$ is linearly dependent on the state, such that

$$\boldsymbol{x}_s(k|k-1) = \boldsymbol{L}\boldsymbol{x}_{k|k-1}, \tag{4.119}$$

the predicted source position will be a Gaussian distributed random vector with mean $\boldsymbol{L}\boldsymbol{x}_{k|k-1}$ and covariance $\mathrm{Cov}(\boldsymbol{x}_s(k|k-1) = \boldsymbol{L}\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{L}^T$. Calculating the corresponding sigma points and weights by eq. (4.43), the mean and covariance of either of the measurement equations (4.98) or (4.99) can be estimated using the unscented transformation composed of eqs. (4.44) and (4.45). Considering the microphone pair i, j the predicted mean of the TDOA $\tau := \tau_{ij}(k|k-1)$ is denoted by $\mu(\tau)$ and the corresponding covariance matrix by $\mathrm{Cov}(\tau)$. The probability density function for the TDOA measurement can

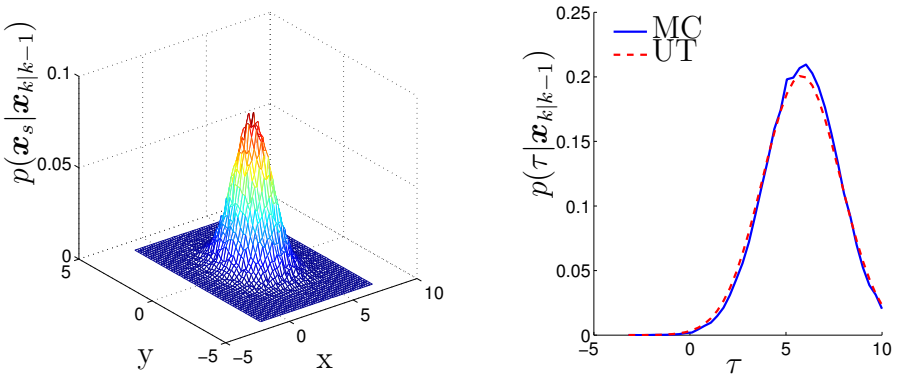

Figure 4.7: left: approximated pdf of the predicted source position $p(\boldsymbol{x}_s|\boldsymbol{x}_{k|k-1})$ found by 100,000 samples with true source position at $[3m, 0m]$ and unity covariance matrix,, microphone 1 at $[5m, 0m]$, microphone 2 at $[-5m, 0m]$, right: corresponding pdf of the range difference $(c{\cdot}\tau)$ $p(c{\cdot}\tau|\boldsymbol{x}_{k|k-1})$ found by passing the 100,000 samples through the nonlinear equation (MC) and the approximation by the unscented transformation (UT) and eq. (4.120)

then be approximated by

$$p(\tau|\boldsymbol{x}_{k|k-1}) \approx \frac{\mathcal{N}(\tau, \mu(\tau), \mathrm{Cov}(\tau)) \cdot \mathrm{rect}(\tau, -\tau_{max}, \tau_{max})}{\int_{-\tau_{max}}^{\tau_{max}} \mathcal{N}(\tau, \mu(\tau), \mathrm{Cov}(\tau)) d\tau}, \tag{4.120}$$

with

$$\mathrm{rect}(\tau, -\tau_{max}, \tau_{max}) = \begin{cases} 1, & \tau \in [-\tau_{max}, \tau_{max}] \\ 0 & \text{else} \end{cases}, \tag{4.121}$$

which assures the probabilities of $\tau \geq \tau_{max}$, or $\tau \leq \tau_{max}$, with $\tau_{max} = |\boldsymbol{x}_i - \boldsymbol{x}_j|_2$, are zero. The denominator of eq. (4.120) normalizes the integral of this product.

Fig. 4.7 compares the approximation of the pdf of the range difference (TDOA times propagation speed) by the unscented transformation followed by eq. (4.120) and the approximation of this pdf by 100,000 Monte Carlo trials: two microphones are placed at the positions $[5m, 0m]$ and $[-5m, 0m]$. 100,000 samples are then drawn from a Gaussian distribution of the predicted source position with mean $[3m, 0m]$ and covariance matrix

$$\mathrm{Cov}(\boldsymbol{x}_s(k|k-1) = 1\boldsymbol{I}. \tag{4.122}$$

The left plot of figure 4.7 shows the approximation of $p(\boldsymbol{x}_s|\boldsymbol{x}_{k|k-1})$ by the 100,000 samples. The corresponding pdf of the range difference measure is

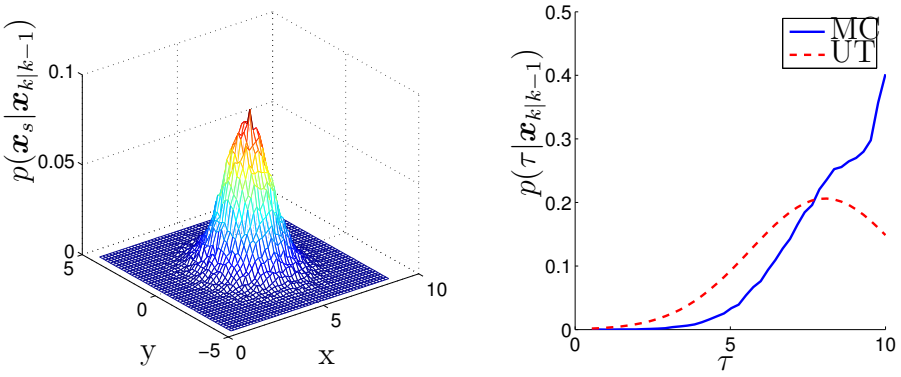

Figure 4.8: left: approximated pdf of the predicted source position $p(\boldsymbol{x}_s|\boldsymbol{x}_{k|k-1})$ found by 100,000 samples with true source position at $[4.5m, 0m]$ and unity covariance matrix, microphone 1 at $[5m, 0m]$, microphone 2 at $[-5m, 0m]$, right: corresponding pdf of the range difference $(c{\cdot}\tau)$ $p(c{\cdot}\tau|\boldsymbol{x}_{k|k-1})$ found by passing the 100,000 samples through the nonlinear equation (MC) and the approximation by the unscented transformation (UT) and eq. (4.120)

plotted on the right side. The MC graph represent the approximation of the pdf by the 100,000 samples, while the UT graph indicates the approximation found by the unscented transformation, followed by eq. (4.120). Note that the two graphs closely match. Based on the law of large numbers, the Monte Carlo approximation, and consequently the unscented transformation approximation will be quasi-alike with the true pdf. This is true as long as the true source position is not too close to one of the microphones. Fig. 4.8 shows the same example, only now the true source position is set to $[4.5m, 0m]$. Since now the drawn samples are in the neighborhood of microphone one, the resulting range differences will be close to the maximum possible range difference. The resulting pdf found by the Monte Carlo approximation no longer has a Gaussian shape, and hence, is no longer well approximated by the unscented transformation approach.

However, if the sensor configuration is adaptively updated, such that the condition number of the Jacobian is small, the found sensor positions are generally far from the predicted source position and consequently, the unscented transformation approach can be utilized.

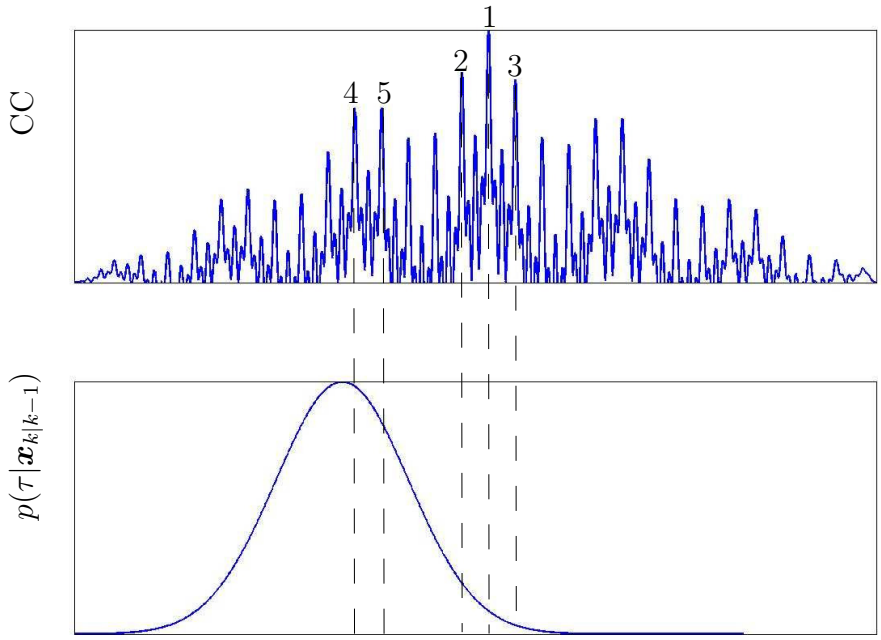

Figure 4.9: Most Likely Local Maximum approach demonstrated on unfiltered cross-correlation (CC) TDOA estimate. The $K = 5$ highest peaks of the CC function are detected and their corresponding probabilities are compared. The possible TDOA with highest probability is then selected: in this case peak 4

4.4.1 Most Likely Local Maximum

Having a reasonable approximation for the conditional probability density function of τ, this knowledge can be used to support the estimation procedure of the true TDOA. Assume that time sequence, which is used for the TDOA estimate, has multiple local maxima, of similar magnitude. Possible scenarios for such a multi-maxima observation sequence, include periodic signals, signals perturbed by directed noise, and reverberant signals. The choice of which peak to select for the TDOA estimate then becomes rather difficult.

We propose to select the K highest peaks as possible TDOA estimates and compare their corresponding conditional probabilities. The peak having the highest probability is then selected as the TDOA estimate.

Fig. 4.9 shows an example of this procedure. As the time sequence

the outcome of the unfiltered cross-correlation algorithm is chosen. For this example the source signal is a superposition of multiple sinusoidal waves with different frequencies. The simulated microphone inputs are windowed, delayed versions of this source signal. The unfiltered cross-correlation of the two resulting signals is plotted in the upper graph of fig. 4.9. Due to the period nature of the source signal, the outcome has multiple local maxima. The $K = 5$ highest peaks are detected and their corresponding conditional probability densities are obtained from $p(\tau|\boldsymbol{x}_{k|k-1})$. In this example the 4th highest peak would then be selected as the resulting TDOA estimate.

4.4.2 Weighted Probability Density Function

The most likely local maximum approach does not consider the magnitude of the individual peaks. However, if some of the peaks are large compared to all the others, this information should be taken into account. One possibility would be to adjust the number of considered peaks K adaptively: e.g. consider all the peaks whose value is at least 60% of the value of the highest peak.

Alternatively, the height of the peaks of the time sequence could be weighted by the pdf $p(\tau|\boldsymbol{x}_{k|k-1})$. Either one of the TDOA estimators provides a time series, denoted by $h(\tau)$, whose maximum is usually considered to be the TDOA. Now the additional information given by the pdf $p(\tau|\boldsymbol{x}_{k|k-1})$ could be used to rescale this sequence in the following way:

$$h_m(\tau) = h(\tau)^\beta \cdot p(\tau|\boldsymbol{x}_{k|k-1}), \tag{4.123}$$

where β denotes an additional scaling parameter. The TDOA estimate is then defined as the argument which maximizes $h_m(\tau)$:

$$\hat{\tau} = \arg\max_\tau h_m(\tau). \tag{4.124}$$

Normalizing eq. (4.123) by its area, it can be seen as a probability density function

$$p_h(\tau|\boldsymbol{x}_{k|k-1}) = \frac{h(\tau)^\beta \cdot p(\tau|\boldsymbol{x}_{k|k-1})}{\int_{-\tau_{max}}^{\tau_{max}} h(\tau)^\beta \cdot p(\tau|\boldsymbol{x}_{k|k-1}) d\tau}, \tag{4.125}$$

which is a rescaled version of the pdf $p(\tau|\boldsymbol{x}_{k|k-1})$. Probabilities at time instances with high peaks in the time series $h(\tau)$, will be amplified, while probabilities corresponding to small values in the time series will be attenuated.

4.5 Chapter Summary

The problem of recursive Bayesian estimation was presented in this chapter. The classic Kalman filter for linear state space systems and three possible extensions, namely the extended Kalman filter and the unscented Kalman filter and the bootstrap filter, for nonlinear systems were introduced. Their strengths and weaknesses were pointed out by some representative examples.

The recursive estimators were applied for the TDOA-based acoustic passive source localization procedure. Assuming a linear system equation, two possibilities for the application of these estimators result. Either, the nonlinear measurement equation, which relates the source position to the measured TDOAs, is utilized, or one of the source position estimators of chapter 3 is used to render the measurement equation linear. In the case, one of the nonlinear recursive estimators, such as the extended Kalman filter, the unscented Kalman filter, or the bootstrap filter, needs to be applied. The second procedure leads to a linear measurement equation, and hence, the optimal Kalman filter can be applied. If the linearized estimator is chosen to carry out the position estimation, it can be seen that in certain cases, the extended Kalman filter considering the nonlinear measurement equation is identical to the linear measurement approach.

Consequently, we argue that one of the optimal sensor configurations proposed in chapter 3 should be considered, even if the recursive estimators based on the nonlinear measurement equation, do not make use of the estimators proposed in chapter 3. This claim was supported by an example, which compared the results found with a well-conditioned sensor configuration with those found by an ill-conditioned configuration.

Two procedures for supporting the estimation of the TDOAs were proposed, based on the recursive estimators. Those are particular useful if the time series obtained by the TDOA estimators have multiple local maxima. In this case the conditional probability of the TDOA outcome at time instant k is estimated by the unscented transformation, knowing the observations till time instant $k-1$.

Chapter 5

DSP-based Passive Source Localization

The presented acoustic passive source localization system is intended to offer a low cost alternative to existing systems. A variety of setups can be found in literature, which tend to be quite expensive. A central part of such systems is the simultaneous analog to digital conversion (ADC) of the multiple microphones. Most often audio specific AD converters are used, which are mainly used by the music industry. These external converters are then connected to a personal computer (PC), which carries out the actual estimation of the position. While the minimum cost for such a system is currently around 1000€, an upper bound can hardly be defined. In [CL04] and [BRA04] two possible setups of this classical approach are presented. J. C. Chen et al. [CYW+03] as well as M. Chen et al. [CLH+07] propose wireless localization systems, which use multiple pocket PC [CYW+03], or multiple notebooks [CLH+07]. Such systems might be interesting for the localization of the talker in a conference room. Nowadays, most of the participants of conferences have either a pocket PC or a notebook with them. The integrated microphones can then be used to localize the talker. While the problem of synchronization is addressed in [CYW+03], M. Chen et al. avoid it by using an energy-based, rather then a TDOA-based localization algorithm.

The need for a low cost system results from its possible applications. An autonomous camera pointing system is commercially only attractive, if the price is kept small. The more money is saved on the localization/steering system, the more money can be spent on the camera. An alternative application for passive source localization is a handsfree set for telephony. The direction to the speaker is estimated by the presented algorithms and spatial filtering, also known as beamforming, is carried out to amplify the sound of the speaker's direction, and to attenuate the other directions. Such a system

could find its applications in the automobile industry. Most countries forbid the use of mobile phones for the driver of the car and only allow handsfree systems. However, the ambient noise of e.g. the motor, the tires, the aerodynamics, and other vehicles, is usually quite loud and makes the use of handsfree systems in a car rather bothersome. The use of the source localization algorithm could permit to suppress these noises and to amplify the speaker's voice. Such a system might also be interesting for virtual reality applications, and computer games. Multiple microphones could then e.g. be mounted on or integrated in the screen and the voice of the player could be used for his/her virtual immersion into the game. All these systems become commercially attractive if the position estimation can be carried out at a reasonable price.

While the above presented systems are beyond 1000 Euros, it is the goal of this chapter to present a system realizable for far less than that.

Further, the presented system will be used to evaluate the accuracy of the algorithms presented in the first part of this thesis.

5.1 System Hardware

The central component of the presented system is a digital signal processor (DSP). A DSP is a microprocessor specially designed for digital signal processing. It usually has multiple analog and digital in- and outputs. The analog input ports are converted by a ADC, while the analog output ports result from a digital to analog conversion (DAC) of the digital DSP signals. It usually offers a rather small size memory, which is optimized for real-time implementation and signal processing algorithms. Further, a DSP offers special arithmetic operations, such as fast multiply-accumulates, which optimize the calculation of signal processing based algorithms, such as the fast Fourier transform, its inverse and finite impulse responses. These properties make a DSP a powerful component for passive source localization based on microphone arrays.

The overall hardware system is presented in fig. 5.1. Multiple microphones are connected to the analog inputs of the DSP, which will simultaneously discretize the received signals. Those signals will then either be directly treated by the DSP, or will be transmitted to a PC, which can carry out the remaining part of the localization algorithm, such as the calculation of the position, or the recursive Bayesian filtering, for source tracking applications. In the following subsections the DSP-PC communication, and a variety of DSP implementations are presented.

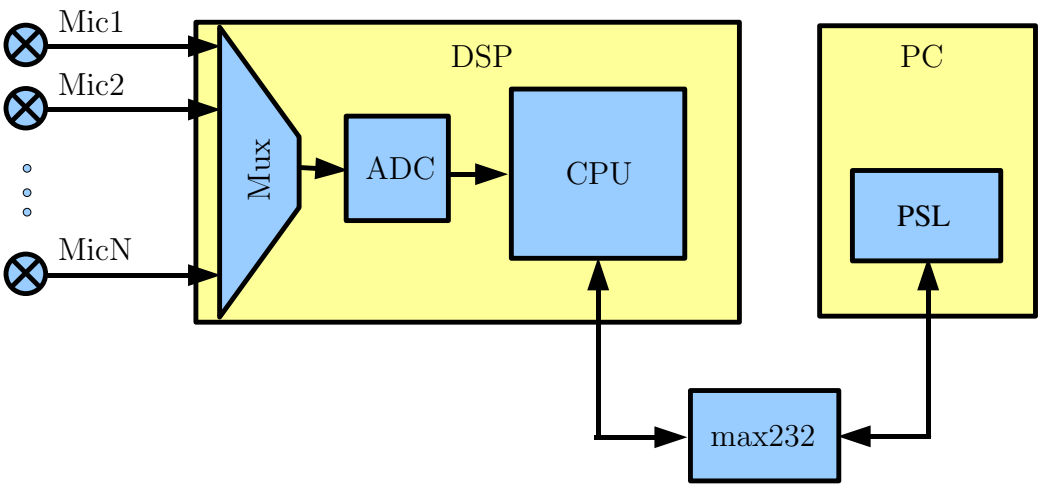

Figure 5.1: Hardware Setup of the TDOA-based passive source localization

5.1.1 PC-DSP communication

Modern microcontrollers usually have one or multiple universal asynchronous receiver/transmitter (UART) ports, which can be used for serial communication with peripheral devices, or PCs. The standard protocol used by the serial port of a PC is the recommended standard (RS) 232. It serially transmits and receives 8 bit words, which are preceded by a start bit and followed by a stop bit. A parity bit might be set between the last bit of the data and the stop bit, but is usually not utilized. The transmission speed for this serial communication is usually limited by the serial ports of standard personal computers to 115,200 bauds/s.

While the serial port of a PC requires negative logic, i.e. logic '1' is -3V to -12V and logic '0' is +3V to +12V, microcontrollers are usually using transistor-transistor logic (TTL), i.e. logic '1' is 2.2V-5V and logic '0' is 0V-0.8V. In order to establish a serial communication between those two, a level converter must be used, which converts from TTL logic to RS 232 levels, and vice versa. Commonly used integrated circuits for such a conversion are the "MAX 232" and the "DS 275". While the MAX 232 requires 4 external electrolytic capacitors of around $1\mu F$ and must be supplied with a 5V voltage, the DS 275 does not need any external components and steals power from the signal lines. However, it is a bit more expensive and guaranties a baud-rate of only 19,200 bauds/s [DS295].

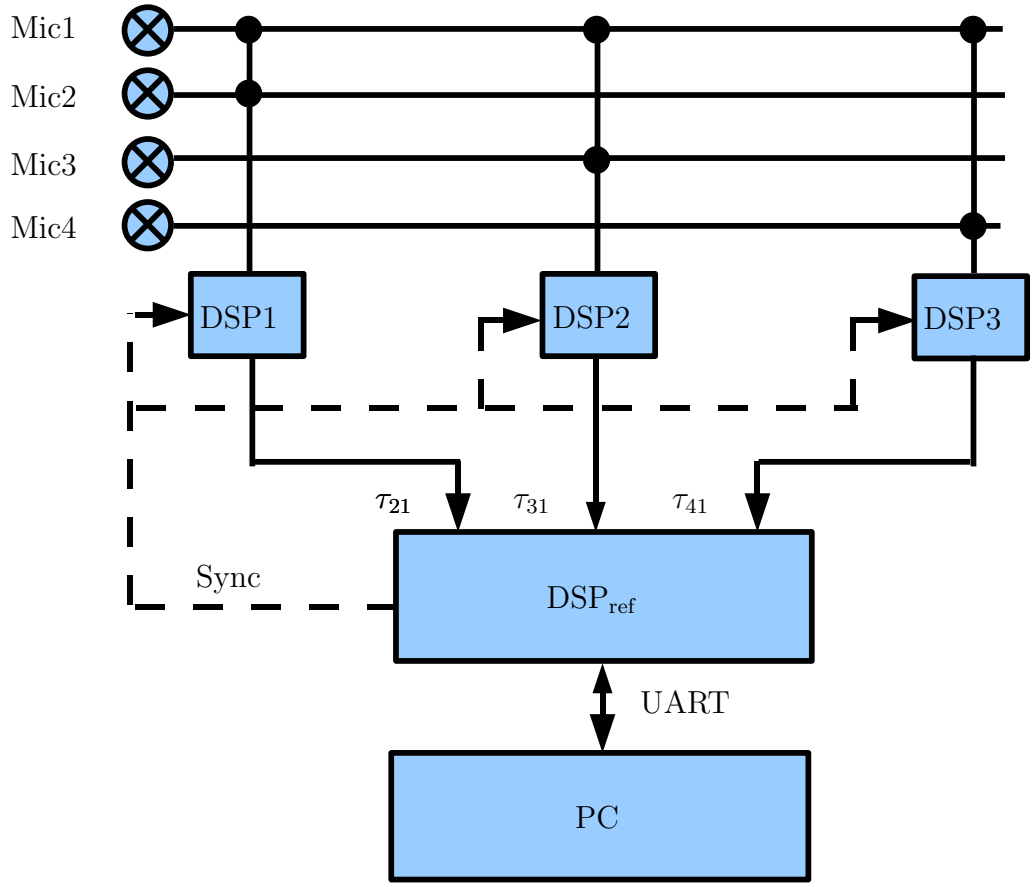

Figure 5.2: Multi-DSP setup for linear approximation estimators: TDOAs are measured w.r.t. the reference sensor Mic1. by individual DSPs and are collected by a reference DSP, denoted by DSP_{ref}, which also synchronizes the others and communicates with the PC

5.1.2 Multi-DSP System

The utilized microcontroller is a DSP, which is equipped with a single ADC. This ADC is connected to the output of a multiplexer, which multiplexes up to 32 analog input ports of the DSP. The maximum AD conversion rate of $f_{max} = 80kHz$ is hence divided by the number of selected input ports N. Thus, the maximum possible sampling frequency of an individual microphone is

$$f_{s,max} = f_{max}/N. \tag{5.1}$$

For a large number of microphones the sampling frequency could thus become insufficient. Therefore, it might be interesting to use one DSP for

each sensor pair. Consequently, the maximum possible sampling frequency would result to $f_{s,max} = 40kHz$. Fig. 5.2 shows such a system. The individual DSPs are synchronized by a reference DSP, which also collects all the estimated TDOAs. Those are then transmitted via the UART to the PC, which carries out the position estimation. Another advantage of this configuration is that only relatively little data needs to be stored and treated by one DSP. While the amount of stored data is limited by the available memory, the limited processor speed of the DSP makes it necessary to treat as little as possible data in order to assure a real-time implementation.

This multi-DSP system is well-suited for the linear approximation estimators, or the linear intersection estimator presented in sections 3.5 and 3.6. For the latter, one DSP could either be used for a sensor pair or a sensor quadruple, which could then transmit either the two TDOAs, or the resulting bearing line.

5.1.3 Single-DSP System

The alternative to the multi-DSP system is a single-DSP system, which handles all the microphones at the same time and is only supported by the PC. Its main advantages are the lower cost and circuit complexity. No synchronization of multiple DSP must be carried out. Further, no reference DSP and no communication between this reference DSP and the remaining DSP is necessary.

However, due to the relation presented by equation (5.1), the number of microphones must be kept small. If the sampling frequency drops below two times the maximum sound frequency, aliasing will affect the TDOA estimations. Further, the computational aspect becomes more important. As the number of TDOA estimates increases, so will the number of FFTs and IFFTs. This computational load might become too large for real-time applications.

5.2 Algorithm Implementation

The above presented hardware solutions for acoustic passive source localization leave a number of options for the actual position estimation. Either the TDOAs will directly be estimated by the DSPs, or the collected data can be transmitted via the serial port to the PC, which then carries out the actual estimation process. In both cases however it is intended that the TDOA-based position estimation is done by the PC.

5.2.1 Data Acquisition

Three different TDOA estimation procedures are presented in this section for the single-DSP estimator. While the first one carries out the TDOA estimation on the DSP, the two other use the PC.

Version 1

Version 1 uses the DSP to carry out the TDOA estimations, based on the unfiltered cross-correlator. The system operator chooses the parameters, such as the sampling frequency, the window size, the discretization accuracy (8,10, or 12 bits), and the chosen microphones and enters them into the PC. The PC forwards these via the serial port to the DSP. Once, the DSP has obtained these data, it starts collecting the data of the chosen microphones and stores this data in its internal memory. Afterwards, it calculates the corresponding TDOAs using the unfiltered cross-correlator. The estimated TDOAs are then transmitted to the PC, which can integrate these in the position estimation procedure.

The main reason for using the unfiltered cross-correlation estimator, is that the DSP is well-adapted for this operation. As a consequence the unfiltered cross-correlation is the computationally fastest of the presented estimators, when implemented on the DSP. Using N microphones, it might be necessary to calculate all the $N(N-1)/2$ TDOA estimates. While the calculation of the unfiltered cross-correlation is easily implemented in the chosen DSP, the alternative algorithms are more difficult to run on the DSP and above all are computationally slow.

Version 2

The second version collects the data the same way as the first version does. However, once all the collected data is stored on the internal memory it is transmitted via the serial port to the PC. The PC then carries out the TDOA estimation procedure. This leaves the door open for using alternative TDOA estimators. E.g. in case of large reverberations, the unfiltered GCC might be insufficient, and rather the adaptive eigenvalue decomposition estimator should be used, which runs faster on a modern PC than on the presented DSP.

Version 3

Version three does not store the collected data on the internal memory, but directly transmits it over the serial port to the PC. In this case the sampling

frequency is mainly defined by the baud-rate of the serial interface, which in our application is set to the maximum speed of 115,200 bauds/s. While the ADC converts with 10 or 12 bits, the serial port only transmits 8 bit words with additional start and stop bits. This leaves two options for the transmission, either only the 8 most representative bits of the data stream are transmitted, or the data is first grouped together to a single data string of length L

$$L = N \cdot B, \tag{5.2}$$

with N being the number of used microphones and B being the resolution (10 or 12 bits). The total string is then regrouped into 8 bit words, which are then individually transmitted over the serial port. The total number of bits Q to be transmitted for one conversion cycle then follows to

$$Q = N \cdot B + \left\lceil \frac{N \cdot B}{8} \right\rceil \cdot 2, \tag{5.3}$$

where $\lceil . \rceil$ denotes the rounding operator to the next higher integer. The second term on the right side of eq. (5.3) corresponds to the number of transmitted start and stop bits.

The transmission time T_t for one cycle then follows to

$$T_e = Q/115200. \tag{5.4}$$

Since the DSP does also need to convert the analog inputs, additional conversion time T_c must be considered. Sampling with 80 kHz, the conversion time follows to $T_c = 1/80kHz = 12.5\mu s$. The total sampling frequency F_s then follows to

$$F_s = \frac{1}{T_e + T_c}. \tag{5.5}$$

Table 5.1 shows the corresponding sampling frequencies for varying sampling resolutions and varying number of microphones.

If the system is to be applied for speaker localization or tracking, Shannon's theorem, saying that the sampling frequency must be at least two times the maximum frequency of the sound, might not be fulfilled. Human speech ranges from approximately 300Hz to 3400Hz [Wik08]. Thus, the sampling frequency would need to be above $2 \cdot 3400Hz = 6800Hz$. None of the presented configurations in table 5.1 validate this minimum frequency. In order to avoid aliasing, it might thus be necessary to integrate a low-pass filter in the amplification circuit of the microphones.

	8 bit	10 bit	12 bit
2 Mics	5.37	4.198	3.66
3 Mics	3.66	2.92	2.428
4 Mics	2.78	2.24	1.875
5 Mics	2.24	1.76	1.49
6 Mics	1.875	1.49	1.26

Table 5.1: maximum sampling frequencies in kHz for the single-DSP implementation using version 3

5.2.2 Microphone Circuit

Electret microphones are chosen to be implemented into the system. Those low cost microphones are a type of condenser microphone, which need no external power supply, but use the available voltages from the integrated circuit they are connected to. A metal disc and a conductible foil being separated by only a few micrometers usually build the core of those, which basically work like a plate capacitor. Applying a voltage to the plate and the foil results in a tension, which is build up between them. Arriving sound will result in a vibration of the foil, by which the distance between the plate and the foil will change, and hence the capacity of the condenser. These capacity variations lead to variations of the tension, which are proportional to the arriving sound waves. The magnitude of these voltage variations are usually only a few millivolts and must therefore be amplified in order to be captured by the DSP, which usually allows input signals in the range of 0V to 3.3V or 5V.

Fig 5.3 shows a possible amplification circuit for such electret microphones. The capsule is powered by around 10V which are taken from the supply voltage and are passed through the resistance R1. The operation amplifier is powered by a single supply voltage V_{cc}, while its negative supply is connected to ground. The input voltage must thus be around $V_{cc}/2$, which is achieved by the potential divider R_2, R_3, with R_2 and R_3 having the same resistance. The microphone signal is injected by the capacity C_1. The amplifier input is thus a voltage with a constant component of $V_{cc}/2$ plus a varying millivolt voltage, which is proportional to the received sound signal. Since the operation point of the amplifier is at $V_{cc}/2$, only the varying voltage component is treated by it. The amplification factor of the circuit is determined by the feedback resistor $R_4 = 180k\Omega$ and the serial connection of the resistor $R_5 = 180\Omega$, the capacity $C_2 = 10\mu F$ and the potentiometer

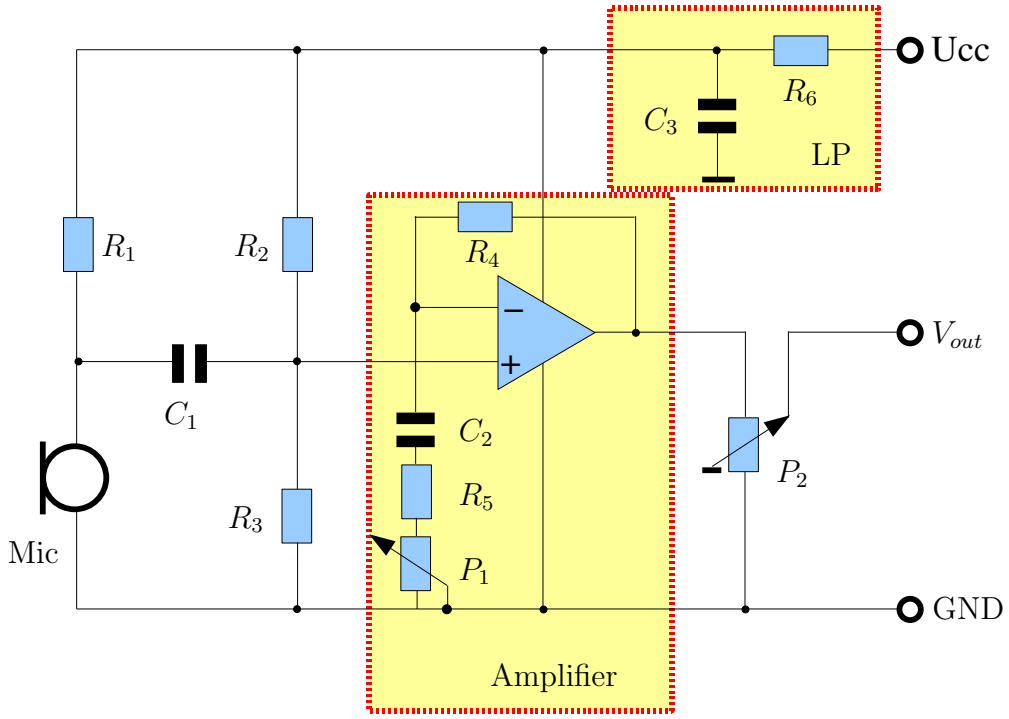

Figure 5.3: Microphone Amplification Ciccuit

$P_1 \in [0, 1k\Omega]$. The transfer function of the amplifier results to

$$H(f) = 1 + \frac{j2\pi R_4 C_2 f}{1 + j2\pi(R_5 + P_1)C_2 f}, \tag{5.6}$$

which basically is a high-pass amplifier. Low frequencies will not be amplified $H(f \to 0) = 1$, while high frequencies will be amplified by the factor

$$\nu = H(f \to \infty) = 1 + R_4/(R_5 + P_1). \tag{5.7}$$

If the potentiometer is set to $P_1 = 0\Omega$ the maximum possible amplification results to $\nu_{max} = R_4/R_5 + 1 = 1001$. The cutoff frequency of any filter is defined as the frequency at which the output signal is $1/\sqrt{2}$ times the input signal, or equivalently, at which the output energy is half the input energy. For a first-order passive RC filter, using one resistance R and one capacity C, such as the one used in the amplification circuit, the cutoff frequency f_{cut} is given by $f_{cut} = 1/(2\pi RC)$. Correspondingly, the cutoff frequency of the presented non-inverting active high-pass is given by

$$f_{cut} = \frac{1}{2\pi(R_5 + P_1)C_2}. \tag{5.8}$$

For this frequency the amplitude of the amplifier's transfer function results to

$$|H(f_{cut})| = \sqrt{\frac{1+\nu^2}{2}} + \frac{1}{\sqrt{2}}. \tag{5.9}$$

Assuming that $\nu >> 1$, this amplitude becomes approximately

$$|H(f_{cut})| \approx \frac{\nu}{\sqrt{2}}, \tag{5.10}$$

which can be seen as $1/\sqrt{2}$ times the input signal amplified by ν. Since the operation amplifier uses $0V$ and $10V$ as the power supply and has a mean-free input signal (due to capacity C2), the output of the amplifier is composed of a constant voltage of $V_{cc}/2$ and a varying voltage, which might now range from $-V_{cc}/2$ to $V_{cc}/2$, depending on the signal strength and the amplification factor ν. Potentiometer P2 is used as an adjustable potential divider, which reduces the output voltage of the amplifier to the desired voltage range of the DSP.

The low-pass filter LP eliminates potential high frequency oscillations from the power supply, which would negatively influence the behavior of the operational amplifier.

Band-pass Filter

The proposed system using the third version of the TDOA estimation, has a very limited sampling rate (see table 5.1). Considering the 2 dimensional source localization estimators based on the linear approximation estimators of section 3.5 at least four microphones are required. Using a 10 bit quantification the maximum sampling frequency results to $f_s = 2.24kHz$. If sound sources, such as human speech, are considered which have frequency components larger than $f_{max} = f_s/2 = 1.12kHz$, aliasing will affect the measurement of the TDOAs. In order to avoid these problems, it is proposed to replace the high-pass amplifier of the circuit presented in figure 5.3, by an amplifying band-pass filter. Fig. 5.4 shows an inverting active first order band-pass. Its transfer function results to

$$H(f) = -\frac{R_4}{1 + j2\pi R_4 C_4 f} \cdot \frac{j2\pi C_2}{1 + j2\pi(R_5 + P_1)C_2 \cdot f}. \tag{5.11}$$

In order to increase the flexibility for choosing the right cutoff frequencies, resistor R_4 might be replaced by another potentiometer. Fig. 5.5 shows the magnitude of the transfer functions for $R_5 = 100\Omega$, $P_1 = 0\Omega$, $R_4 = 100k\Omega$, and $C_2 = 32\mu F$. An inverting high-pass is obtained, when the capacity C_4 is

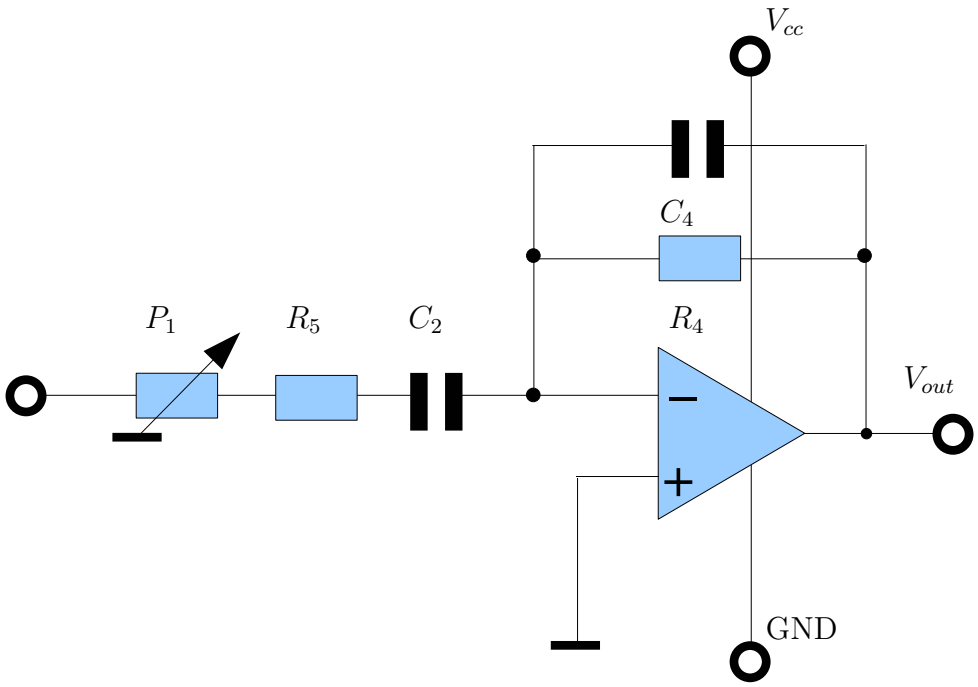

Figure 5.4: Active Band-pass Filter

removed, otherwise, if it is chosen to be $C_4 = 16nF$, a band-pass with cutoff frequencies $f_1 = 1/(2\pi(R_5+P_1)C_2) = 50Hz$ and $f_2 = 1/(2\pi R_4 C_4) = 1000Hz$ results. Hence, such a configuration would assure the validation of Shannon's sampling theorem for a 10 bit quantification of four microphones.

5.2.3 System Realization

The presented system is realized using Microchip's "Explorer 16 Development Board", mounted by the DSP "dsPIC33FJ256GP710". The development board offers a power supply of $3.3V$, $5.0V$, and $9V$. Further, it is equipped with one UART using a DB9 connector. A "PICTail plus" connector allows the connection to additional electronic circuits, and is used as the connector to the individual microphones, which are AD converted by the internal ADC of the dsPIC33. The conversion is carried out by multiplexing over up to 32 analog inputs.

The electric circuit is a slightly modified version of the microphone preamplifier circuit sold by "Conrad Electronics" [Con99]: while the commercially available circuit has an additional capacity at the output of the op-

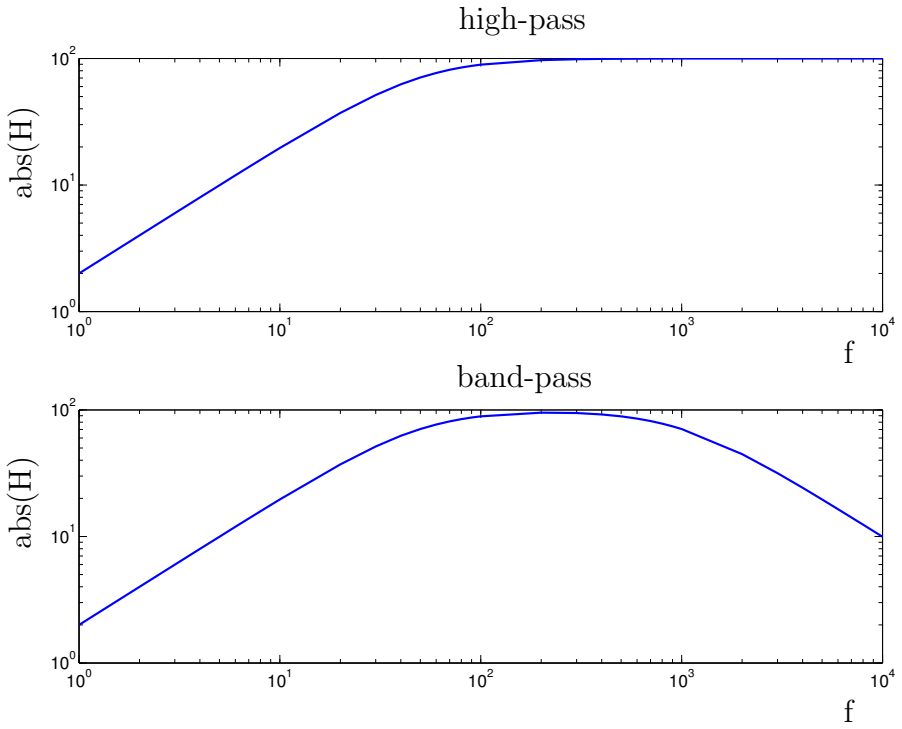

Figure 5.5: Transfer functions of the active high-pass filter with cutoff frequency $f_c = 50Hz$ and of the active band-pass filter with cutoff frequencies $f_1 = 50Hz$ and $f_2 = 1000Hz$

erational amplifier, this capacity is bypassed in our circuit. Its purpose is mainly to remove the mean of the output signal. However, in order to sample the signals by the analog-digital converter of the DSP, the signals maximum amplitude should match the maximum possible input voltage of the ADC.

The connection of the development board and the microphone circuits is realized by RJ9 connectors, which are used for wired telephones. This way commercially available telephone cables can be used for the connection.

The used microphone capsule is MCE-101, which is inexpensive and has a frequency range of 50Hz-12000Hz.Consequently, it covers the whole range of voice frequency. It is omni-directional and can be supplied with 1V-10V, which makes it compatible with the utilized development board.

5.3 Time Delay Estimation

In chapter 4 possible implementation issues for the TDOA estimations were presented for scenarios in which the future positions could be estimated by Bayesian filtering. Further, it was proposed to search the peaks only in the admissible region, which is defined by the sensor positions, i.e.

$$\tau_{ij} \in [-\tau_{max,ij}, +\tau_{max,ij}], \tag{5.12}$$

with

$$\tau_{max,ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|_2/c. \tag{5.13}$$

While these supports can help to achieve better estimation results, when large reverberations are present, or when the signal to noise ratio is low, the proposed hardware solution introduces another handicap. The resolution of the TDOA $\Delta\tau_{ij}$ and range difference estimate Δd_{ij} is limited by the sampling frequency f_s

$$\Delta\tau_{ij} = 1/f_s, \tag{5.14}$$

and

$$\Delta d_{ij} = c/f_s, \tag{5.15}$$

respectively. Assume that four microphones are used with a quantification of 10 bits the maximum possible sampling frequency is $f_s = 2.24kHz$. Assuming a propagation speed of $c = 343m/s$, the range difference resolution then follows to $\Delta d_{ij} \approx 15.3cm$, which severely limits the performance of the position estimation. In order to decrease this uncertainty, we propose to carry out an interpolation of the signals either before or after the time delay estimation.

5.3.1 Interpolation

The process of converting a discrete signal from a given sampling rate $f_s = 1/T_s$ to a different rate $f_n = 1/T_n$ is called sampling rate conversion. When the new sampling rate is higher than the original rate $f_n > f_s$, the process is generally called interpolation. The process of converting the sampling rate of a signal from a given rate f_s to a lower rate f_n is called decimation.

If the band-pass filter of section 5.2.2 is included in the microphone circuit, and the components are chosen such that the Shannon theorem is fulfilled, the received signals are band-limited and the sampling frequency is at least twice as high as the Nyquist frequency f_c, then Shannon's theorem states that the real continuous signals can be perfectly recovered. Assume that

the true, continuous signal of one of the microphones, denoted by $x(t)$, is sampled by a dirac comb $\Delta_{T_s} = T_s \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$

$$x_n(nT_s) = x(t) \cdot \Delta_{T_s} = T_s \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s). \tag{5.16}$$

The spectrum of the sampled signal $X_n(f)$ is then the periodic continuation of the spectrum of the original signal $X(f)$ with periodicity $f_s = 1/T_s$

$$\mathcal{F}(x_n[nT_s)]) = X_n(f) = \sum_{n=-\infty}^{\infty} X(f - n \cdot f_s). \tag{5.17}$$

Since the sampling frequency was chosen to be larger than the Nyquist frequency, no aliasing occurs, and the true signal can be obtained by an appropriate low-pass filter. In theory the ideal low-pass filter is often considered:

$$H(f) = \begin{cases} 1 & \text{for } |f| \leq f_c \\ 0 & \text{otherwise.} \end{cases} \tag{5.18}$$

The inverse Fourier transform of $X_n(f) \cdot H(f)$ then results in the so-called Whittaker-Shannon interpolator

$$x(t) = \sum_{n=-\infty}^{\infty} x_n[nT_s] \cdot sinc\left(\frac{t - nT_s}{T_s}\right), \tag{5.19}$$

with $sinc(t)$ denoting the sinus cardinalis.

Theoretically a perfect interpolation, meaning that the interpolated signals correspond to the true signal at the selected time instances, can be carried out by first recovering the true continuous signal by application of eq. (5.19) and then sampling the found signal by the new sampling frequency f_n. In practice however this approach is neither very convenient (passing from a discrete signal to a continuous signal in order to obtain another discrete signal), nor physically possible, since the sinus cardinalis is a non causal function. However, the digital implementation of the interpolation problem is closely related to the above presented continuous interpolation. The perfect rectangular window of eq. (5.18) is replaced by some physically realizable low-pass filter, such as the Kaiser window [Smi08]. Further, it is assumed that the ratio of the sampling rates can be expressed as a rational fraction

$$f_s/f_n = M/L, \tag{5.20}$$

	Range Differences	Source Position
Real Values	$[2.828, 2, 2]^T$	$[1, 0]^T$
$f = 100kHz$	$[2.823, 1.9997, 1.9997]^T$	$[0.996, 0]^T$
DSP: $f_s = 1875Hz$	$[2.9086, 1.9997, 1.9997]^T$	$[0.931, 0]^T$
1)Correlation, 2)Interpolation, $f_n = 20kHz$	$[2.826, 1.9997, 1.9997]^T$	$[0.9986, 0]^T$
1)Interpolation, 2)Correlation, $f_n = 20kHz$	$[2.826, 1.9997, 1.9997]^T$	$[0.9986, 0]^T$

Table 5.2: Evaluation of the interpolation procedure for passive source localization

where M and L are integers. The interpolation is than carried out by first inserting $L-1$ new sample values between each pair of sample values of $x_n[nT_s]$ and then decimating the new signal by the factor M. The interpolation step is done by first inserting $L-1$ zeros between the sample values such that the new signal $w_m[mT_n]$ is given by

$$w_m[mT_n] = \begin{cases} x_n[mT_n/L], & \text{for } m = 0, \pm L, \pm 2L, \dots \\ 0 & \text{otherwise.} \end{cases} \tag{5.21}$$

The Fourier transform of the new found signal $w_m[mT_n]$ is equal to the spectrum of the originally sampled signal $x_n[nT_s]$

$$W_m(f) = \mathcal{F}(w_m[mT_n]) = X_n(f). \tag{5.22}$$

Using an appropriate FIR low-pass filter h_m, the interpolated signal can be found by

$$x_m[mT_n] = w_m[mT_n] * h_m, \tag{5.23}$$

where $*$ denotes the convolution operator. A detailed tutorial on the digital implementation can be found in e.g. [CR81], which also presents and compares a variety of FIR low-pass filters.

5.3.2 Evaluation

Fig. 5.6 shows an evaluation of the interpolation procedure for passive source localization. A test signal, composed of an overlap of multiple sinus waves with frequencies up to 500Hz, sampled at $f = 100kHz$ is constructed and

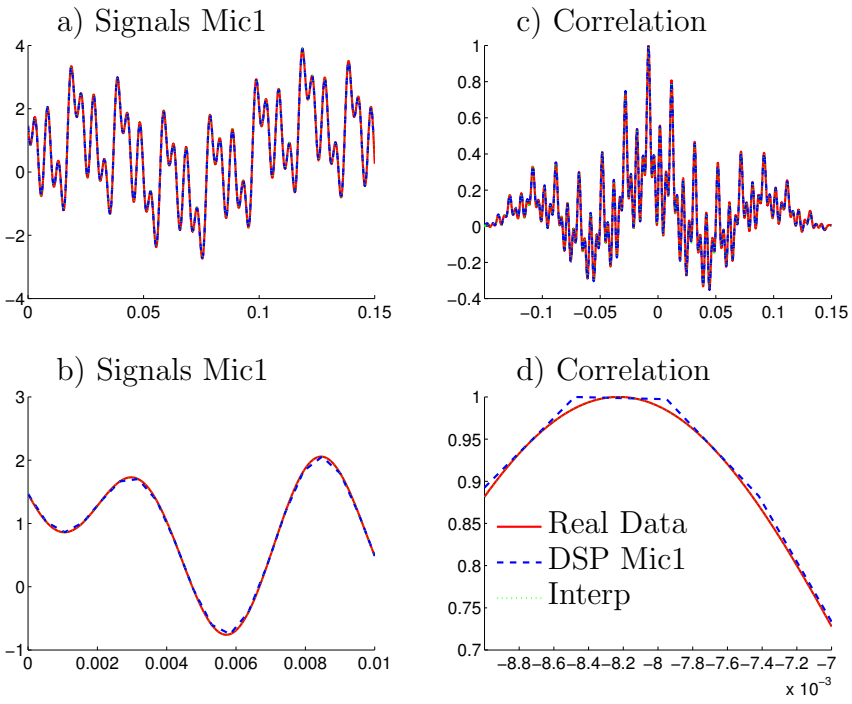

Figure 5.6: Evaluation of the Interpolation procedure, Real Data corresponds to the "continuous" signal (at $100kHz$), DSP Mic1 represents the data obtained by microphone one (at $f_s = 1875Hz$), "Interp" corresponds to the interpolated signal of microphone one ($f_n = 20kHz$), a) signal received by microphone 1, c) corresponding correlation between microphone one and two, b) detailed view on signal of microphone one, d) detailed view on correlation between microphone one and two

is used to simulate the "continuous" sound of the source. Its position is set to $[1m, 0m]$, and an optimal four sensor configuration according to the results of section 3.7.3 is presumed. The reference sensor is set at the origin $\boldsymbol{x}_1 = [0m, 0m]$ and hence the slave sensor positions result to $[-\sqrt{8}m, 0m]$, $[2m, \sqrt{8}m]$, and $[-2m, \sqrt{8}m]$. In order to not influence the results by present noise and reverberations, the real microphone system is not used for this evaluation procedure, but the received signals are simulated, as the delayed version of the true signal. The DSP sampling rate is set to the 12bit quantification rate using four microphones $f_s = 1875Hz$ (see table 5.1). It is desired to achieve a sampling rate of $f_n = 20kHz$, by using the presented interpolation procedure to obtain a total of 3000 samples. Fig. 5.6 a) shows

the true continuous signal arriving at microphone one, the sampled signal obtained by the DSP, and the interpolated signal at $f_n = 20kHz$. Looking at figure b) a detailed view is given on the obtained signals. It can be seen that the interpolated signal perfectly matches the true signal, while the DSP signal at the low frequency shows some degeneration. Figures c) and d) show the corresponding correlations of microphones two and one. In d) the problem of the low sampling frequency becomes apparent. The maximum of the DSP signal at $f_s = 1875Hz$ clearly different from the true TDOA. However, the TDOA found after the interpolation process approximates the true TDOA quite precisely, which can also be seen in table 5.2. The range difference, found by the standard cross-correlation, are indicated, as are the corresponding estimated source positions using the linear least-squares estimator presented in sec. 3.5.3. It is due to the chosen sensor configuration and the assumption of no present noise, nor any reverberations, that the range differences two and three are the same for all four estimations. Further, it is a "lucky accident" that the interpolated estimators obtain a closer estimate, than the one found by the true "continuous" signals. The fact that the range difference estimates, and hence the position estimate, are slightly different from the true values is because the "continuous" signal is not really a continuous signal, but a sampled version with high sampling frequency $f = 100kHz$. However, this high frequency corresponds to a range difference uncertainty of $\Delta_d = c/f = 3.43 \cdot 10^{-3}m$, which explains this slight offset.

Note, that it makes no difference if the interpolation is carried out before, or after the calculation of the cross-correlation. This was also observed, when random sensor positions were chosen. Using this observation, the computational burden can be decreased. Instead of carrying out two interpolations of two initial signals of length N and then calculating the correlation of the augmented signals, the correlation can be carried out with the original signals of length N. Afterwards an interpolation of the found result can be carried out in the region of possible range differences, only, instead of interpolating the total correlation outcome.

5.4 System Evaluation

The presented system was tested in a standard office with dimensions $6.5m \times 4.5m \times 2.6m$. The office was equipped with a number of tables, bookshelves, two notebooks, and a ceiling light, responsible for the major part of the ambient noise. The optimal four microphone configuration presented in chapter 3.7.3 was installed in the center of the room, and the coordinate frame for the localization procedure was set to the position of the reference

sensor $\boldsymbol{x}_1 = [0m, 0m]$. The three slave sensors were set to the positions $\boldsymbol{x}_2 = [-1m, 0m]$, $\boldsymbol{x}_3 = [0.71m, 1m]$ and $\boldsymbol{x}_3 = [0.71m, -1m]$, which results in a condition number $\kappa_2(\boldsymbol{A}) = 1$ for the source position $\boldsymbol{x}_s = [0.35m, 0m]$. The four microphones were connected to a single DSP, which sequentially sampled them with an eight bit quantification. According to table 5.2.1 this results in a sampling frequency of $f_s = 2780Hz$.

The estimation of the three time delay estimates w.r.t. the reference sensor were carried out by the unfiltered cross-correlation function as presented in section 3.2.1. The actual position estimate based on these TDOAs was then carried out by the linear least-squares estimator of section 3.5.3.

A standard computer loudspeaker, playing the sound produced by a motorcycle, was selected as the target. In a first test this loudspeaker was set to the optimal source position $\boldsymbol{x}_s = [0.35m, 0m]$, playing the sound at a moderate level. For this setup, the signal to noise ratios of the individual microphone inputs resulted to $SNR(\boldsymbol{x}_1) = 21dB$, $SNR(\boldsymbol{x}_2) = 14sdB$, $SNR(\boldsymbol{x}_3) = 10dB$, $SNR(\boldsymbol{x}_4) = 12dB$.

Fig. 5.7 shows the estimation of the true source position. The time delays of the three microphone pairs were estimated, using four different window lengths ($N = 127, 255, 511, 1023$), 100 times each and were then interpolated with a factor eight, resulting in a new sampling frequency of $f_n = 22.08kHz$. The position was then estimated using the original data, as well as the interpolated data. Consequently, for each window length, 100 estimates using the original data and 100 estimates using the interpolated data, were obtained.

Fig. 5.8 indicates the success rate of the position estimates: estimates within $5cm$ of the true location were considered to be right. Note that even with only $N = 127$ (which corresponds to only $46ms$) data samples the position estimate is accurate more than 90% of the time, and that the interpolated data estimates slightly outperform the estimates found by the original data. For the presented problem, considering the source at the optimal position, the success rate could be increased to 100% by increasing the number of data samples.

However, the position estimates are slightly biased. In the case of the unbiased data this bias can result from the low resolution of the time delay estimates. The time delay of a sensor $\boldsymbol{x}_i$ and the reference sensor $\boldsymbol{x}_1$ given in samples $k_{i1} = f_s \cdot \tau_{i1}$ is

$$k_{i1} = f_s \cdot (|\boldsymbol{x}_i - \boldsymbol{x}_s|_2 - |\boldsymbol{x}_1 - \boldsymbol{x}_s|_2)/c. \tag{5.24}$$

For the given problem these sample time delays result to $k_{21} \approx -8.05$, and $k_{31} = k_{41} \approx -5.73$. However, these values can only be estimated by integer

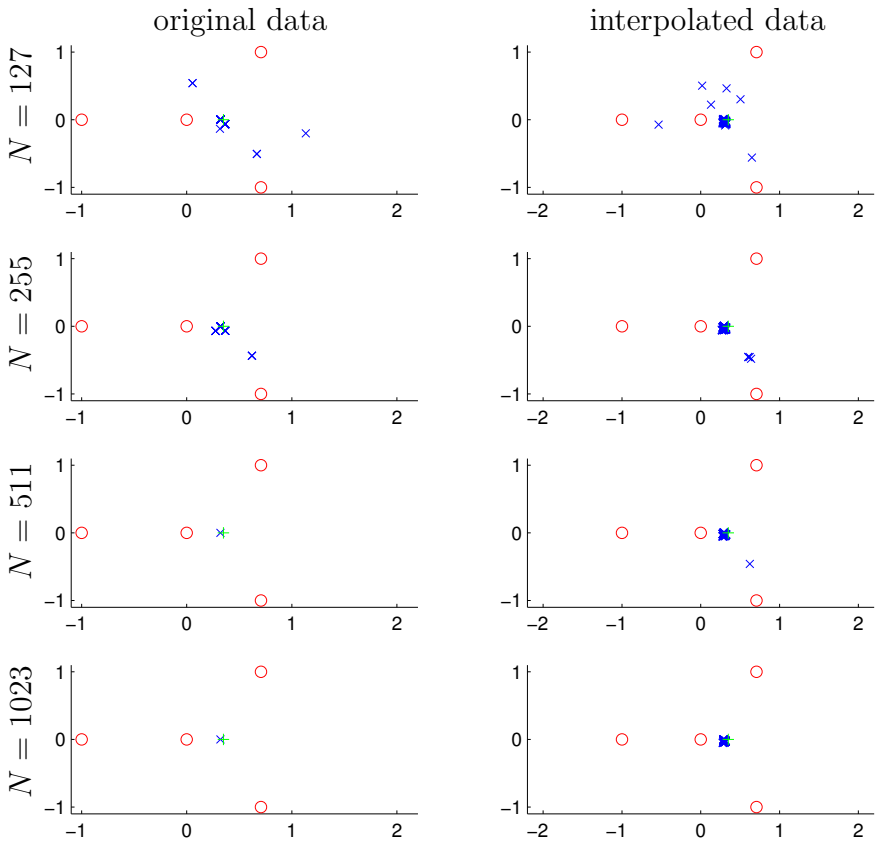

Figure 5.7: position estimation using the optimal four microphone sensor setup. true position: $\boldsymbol{x}_s = [0.35m, 0m]$. left column: estimates found by original data with sampling frequency $f_s = 2780Hz$, right column: estimates found by interpolated data with interpolation factor $M = 8$ ($f_n = 8 \cdot f_S = 22.08kHz$). rows: window lengths of the original data set

values. Consequently, if k_{31} would be estimated to 5 a range error of

$$0.73 \cdot c/f_s \approx 9cm \tag{5.25}$$

s results, and even if k_{31} is estimated to 6 the range error still is $3.3cm$.

In case of the interpolated data, the resolution of the time delay estimates increases by a factor eight. Nevertheless, a slight bias is still observed. While the true time delays w.r.t. the interpolated sampling frequency f_n calculate to $k_{21} \approx -64.4$, $k_{31} = k_{41} \approx -45.8$, the mean of the estimated time delays was

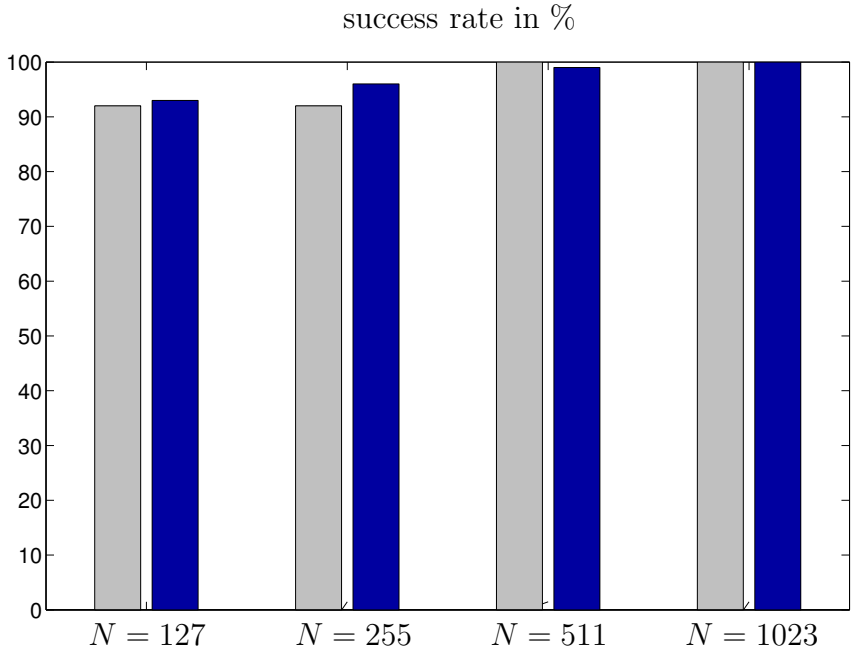

Figure 5.8: success rate of position estimate within $5cm$ of the true location in %, left bars: based on original data, right bars: based on interpolated data

observed to be $\mu(k_{21}) = -61$, $\mu(k_{31}) = -48.68$, and $\mu(k_{41}) = -44.88$. These inaccuracies can probably be explained by an inaccurate sensor installation and the fact, that the loudspeakers are not a point source, but of dimension $7cm \times 7cm \times 17cm$.

Figs 5.9 and 5.10 show the outcomes of the same tests for source positions $\boldsymbol{x}_s = [0.7m, -0.4m]$, and $\boldsymbol{x}_s = [1.6m, -0.4m]$, respectively. Both figures clearly show that the estimation procedure becomes more difficult, when the source is further from the origin, which results in a larger variation of the estimates. This is partially due to the larger distance of the source to the sensors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, resulting in a lower signal to noise ratio, which in chapter 3.2.4 was shown to degrade the time delay estimates. But this is only one reason for the more inaccurate position estimation. It can be observed, that the time delays for the position $\boldsymbol{x}_s = [0.7m, -0.4m]$, using the interpolated data, were constantly ($\geq 70\%$, considering all N) estimated to

$$k_{21} \in [-57, -54], \quad k_{31} \in [-47, -44], \quad k_{41} \in [4, 6],$$

even though the true sampled time delays for this position are $k_{21} = -60.5$, $k_{31} = -38.2$, and $k_{41} = 13.3$. Consequently, there is a bias on the time delay

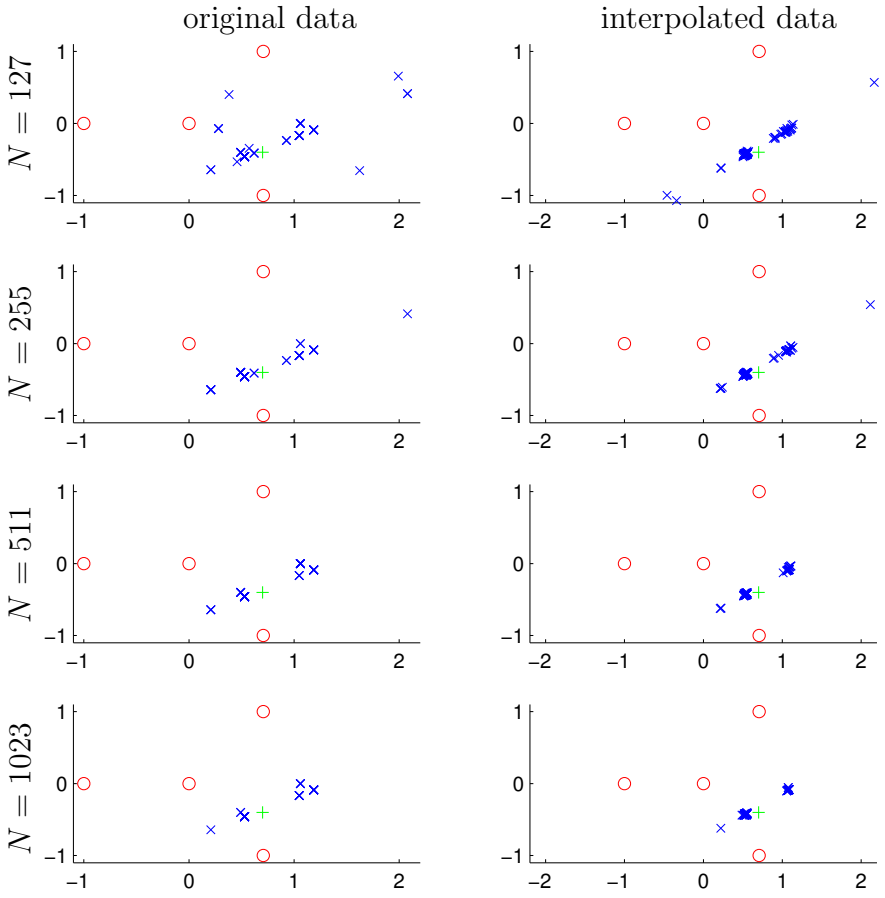

Figure 5.9: position estimation using the optimal four microphone sensor setup. true position: $\boldsymbol{x}_s = [0.7m, -0.4m]$. left column: estimates found by original data with sampling frequency $f_s = 2780Hz$, right column: estimates found by interpolated data with interpolation factor $M = 8$ ($f_n = 22.08kHz$). rows: window lengths of the original data set

estimates ranging from 3 to 8 samples for this test, corresponding to a range error between $4.7cm$ and $12.4cm$. The same observation can be made for the position estimate of $\boldsymbol{x}_s = [1.6m, -0.4m]$. Surprisingly, for the most part of the estimates the errors are smaller than 4 samples, even though the source is now even further away from the sensors. Still, looking at figs. 5.9 and 5.10, one sees that the corresponding estimates are worse than the estimates

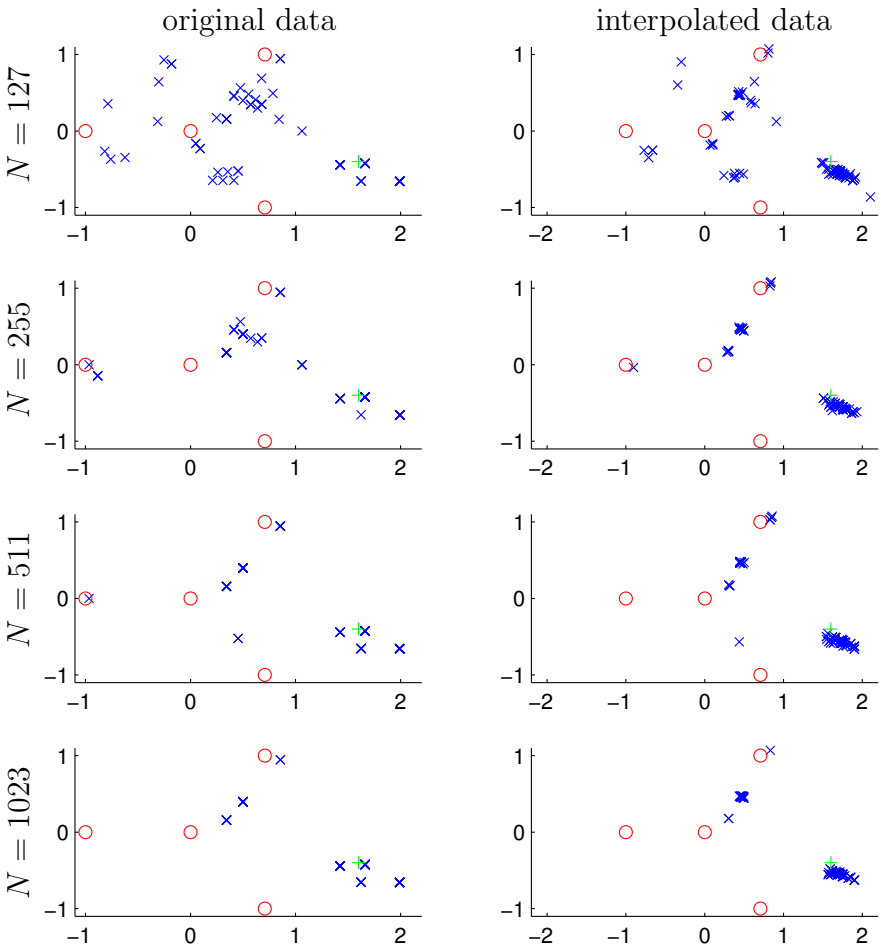

Figure 5.10: position estimation using the optimal four microphone sensor setup. true position: $\boldsymbol{x}_s = [1.6m, -0.4m]$. left column: estimates found by original data with sampling frequency $f_s = 2780Hz$, right column: estimates found by interpolated data with interpolation factor $M = 8$ ($f_n = 22.08kHz$). rows: window lengths of the original data set

of the source position $\boldsymbol{x}_s = [0.7m, -0.4m]$.

This can be explained by the more ill-conditioned situation: while the source location $\boldsymbol{x}_s = [0.3536, 0]$ results in a condition number of the LLS estimator of $\kappa_2(\boldsymbol{A}) = 1$, the source locations $\boldsymbol{x}_s = [0.7m, -0.4m]$ and $\boldsymbol{x}_s = [1.6m, -0.4m]$ result in condition numbers $\kappa_2(\boldsymbol{A}) \approx 2.3$ and $\kappa_2(\boldsymbol{A}) \approx 5.45$,

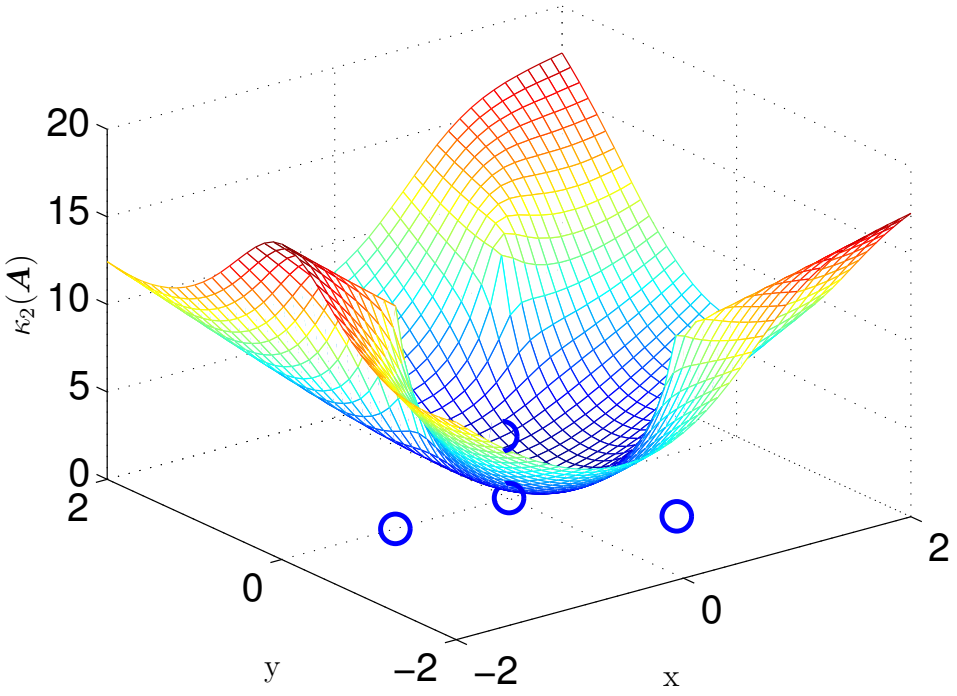

Figure 5.11: Condition number distribution as a function of the source position $\boldsymbol{x}_s = [x, y]$ for the given sensor configuration, circles: sensor positions

respectively. In order to stress the influence of the conditioning of the problem, let us assume that the sample time delays of the three situations are all biased by -5, corresponding to a range error of $-7.7cm$. A source in the optimal location would then be estimated to the position $\hat{\boldsymbol{x}}_s = [0.346m, 0m]$, which would result in an absolute error of around $8mm$, only. The same bias on the sample time delay estimates corresponding to the two other source positions would result in the estimates $\hat{\boldsymbol{x}}_s = [0.65m, -0.36m]$ and $\hat{\boldsymbol{x}}_s = [1.31 - 0.29]$, resulting in the absolute errors of $6.4cm$ and $31cm$, respectively. Clearly, the least well-conditioned estimator, with the source being located at $\boldsymbol{x}_s = [1.6m, -0.4m]$, has the greatest problems of estimating the source position.

Fig. 5.11 illustrates the importance of source position relative to the sensor network. Inside the circle spanned by the three slave sensors, the condition number is small, and consequently a position estimate will be quite accurate, if reasonable time delay estimates are available. However, outside

this circle $\kappa_2(\boldsymbol{A})$ strongly increases, which usually results in a poor estimate.

5.5 Chapter Summary

This chapter presented a low cost hardware configuration for a TDOA-based passive source localization system. Two possibilities were presented. In order to decrease the calculation time and increase the sampling frequency, the first system uses multiple DSPs. The second system uses only one DSP and is therefore easier to implement and less expensive.

The second system has been realized using Microchip's "Explorer 16 Development Board". This board has an integrated level converter, such that the serial communication with the PC is readily implemented. Further advantages are its power supply, which can be used for the electret microphone circuits, and the extension slots for easy connection with external cards.

The system was tested on a two dimensional indoor estimation problem using a four microphone optimal sensor configuration. The system was able to accurately estimate the source position, when the source was set to the reference position, and offered reasonable results, when the source was positioned at other locations. The dependency of the estimate's accuracy on the condition number was pointed out.

The main limitation of the system is the very limited sampling rate, resulting in a low resolution of the time delay estimates and severe constraint on the bandwidth of the target's sound. In order to increase the transmission rate between the DSP and the PC alternative DSPs, offering a USB communication, could be used. While the RS232 serial port only allows a maximum baud rate of 115200bauds/s, modern USB connections transmit with up to 480Mbits/s. Such a system would then fulfill Shannon's sampling theorem for the entire frequency rage of human voice. Using a low cost DSP the common sampling rate of telephone applications $f_s = 8000Hz$ could then easily be attained for multiple sensors.

Chapter 6

Conclusion and Outlook

6.1 Conclusion

The performance of TDOA-based passive acoustic source localization can be greatly increased by arranging the sensors in some optimal sense.

TDOA-based passive source localization is a two step estimation problem. In the first step time differences of arrival of multiple spatially separated microphone pairs are estimated. The second step is then composed of estimating the source position by considering the estimated time difference and utilizing the knowledge of the microphone positions and the propagation speed. Assuming a time delay estimator at hand, the question of how to best place the sensors for optimizing the performance of the existing position estimators, carrying out the second step, was addressed in this work. It was argued that in this case, the available time delay estimates can be seen as the sum of the true time differences and an additive noise term.

Depending on the available statistical information on this noise term the actual localization problem can either be stated as a statistical estimation problem, or a non-statistical estimation problem, based on the least-squares approach. If the problem is described statistically, the Cramer Rao lower bound (CRLB) and the geometric dilution of precision (GDOP) are suitable measures, not only for delivering information on the accuracy of the estimator at hand, but also for optimizing the sensor geometry.

While the CRLB, defining a lower limit on the variance of all unbiased estimators, is identical for all estimators, the GDOP is estimator specific. Consequently, an optimal sensor configuration w.r.t. the GDOP for one estimator, will usually not be the optimal configuration for others. However, considering a linearized estimator and assuming Gaussian noise, it is shown that the optimal sensor configuration w.r.t. the GDOP is identical to the

configuration found w.r.t. the CRLB.

Considering the TDOA-based localization problem as a non-statistical estimation problem has two main advantages over the statistical approach. Firstly, the statistical information does not need to be known for constructing the estimators. Quite often the ambient noise is constantly changing and then becomes difficult to obtain. Secondly, the nonlinear estimation problem can be posed as a constrained linear estimation problem, if only the time differences w.r.t. a reference sensor are considered. A number of closed-form estimators exist, which approximate the solution to this constrained problem. It was shown that these estimators are all somehow dependent on an accurate calculation of the Moore-Penrose pseudo-inverse of the LLS matrix, whose entries are composed of the sensor positions and the time difference estimates. If this matrix is close to singular, the position estimates are of poor quality. Therefore, we proposed to choose a sensor configuration minimizing the condition number (singularity measure) of the LLS matrix.

An analytical optimal sensor configurations for the linear approximation estimators are presented in this work, which all result in condition numbers equal to one.

For the above mentioned linearized estimator the condition number based optimal configuration becomes identical to the configurations found by the CRLB and the GDOP.

An optimal sensor configuration for bearing line intersection estimators (e.g. the TDOA-based linear intersection estimator) w.r.t. the condition number are obtained, if the bearing lines become orthogonal to each other.

Assuming an optimal sensor configuration for the linear approximation estimators, the constrained linear least-squares problem can be solved analytically. The found estimator outperforms the other TDOA-based linear approximation estimators, if the condition number is kept below around 1.2. However, if the condition number becomes larger, this estimator is no longer adequate.

One option for assuring a small condition number for non static sources is to use a large sensor network and to activate only those microphones which render the condition number for a predicted source position close to one. The prediction of the source's near future position can be carried out by recursive Bayesian filtering. Since, the estimation problem is generally nonlinear, the classical Kalman filter is usually not applicable, and in literature is often replace by particle filters.

In this thesis the extended Kalman filter, the unscented Kalman filter and the recursive bootstrap filter are evaluated for the TDOA-based tracking problem. The close connection of the extended Kalman filter to the above mentioned linearized estimator, makes it perform quite well compared to the

two other methods, when an optimal sensor configuration for the linearized estimator is utilized. At the same time the two other procedures also seem to increase their performance, if the microphones are installed in such an optimal setup.

6.2 Outlook

Optimal sensor configurations for the second step of the two step TDOA-based localization estimators have been proposed. Currently, these configurations do not consider the time delay estimation procedures and are invariant to scaling and rotation. Consequently, the same sensor network geometry in small scale (close to the source) and large scale (microphones far from the source) are currently considered to be equal. However, since the energy of sound reduces quadratically with increasing distance, the signal to noise ratio for a large scale configuration will be drastically smaller than the SNR of the small scale configuration, resulting in poorer time delay estimates.

Therefore, the first step should be integrated into the sensor geometry optimization. In this thesis, we presented an upper bound on the relative position error of the linear least-squares estimator, which is basically a function of the condition number, the norm of the LLS matrix and the norm of an error term on this LLS matrix. In case of TDOA-based passive source localization this error term is due to the inaccurate time delay estimates. Consequently this function upper bounds the position estimation error, considering both steps of the TDOA-based position estimators: if the LLS matrix is ill-conditioned due to an inappropriate sensor configuration, the upper bound will be large, and the quality of the estimation procedure is likely to be poor. At the same time, if the sensors are too far from the source, the time delay estimates will be poor, resulting in a large error term on the LLS matrix, also increasing the bound.

The error on the LLS matrix will mainly be a function of the SNRs and consequently, on the distance between the source and the microphones. In order to optimize the configuration, considering both estimation steps, the dependency of the error matrix on this distance must be established. An optimal sensor configuration will then probably be found by numerical minimization of this upper bound.

Appendix A

Matrix Algebra

A.1 Induced Matrix Norms

$$\|\boldsymbol{A}\|_2 = \sup_{|\boldsymbol{x}|_2=1} |\boldsymbol{A}\boldsymbol{x}|_2. \tag{A.1}$$

$$\begin{aligned} \|\boldsymbol{A}\|_2 &= \sup_{|\boldsymbol{x}|_2=1} \sqrt{\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}} \\ &= \sqrt{\sup_{|\boldsymbol{x}|_2=1} \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}} \end{aligned} \tag{A.2}$$

This constrained optimization can be solved using the Lagrangian multiplier approach:

$$\mathcal{L}(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{x}^T\boldsymbol{x} \tag{A.3}$$

$$\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{x}}\right)^T = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\lambda\boldsymbol{x} \stackrel{!}{=} \boldsymbol{0} \tag{A.4}$$

$$\Rightarrow \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \tag{A.5}$$

Hence the Lagrangian multiplier λ is an eigenvalue of $\boldsymbol{A}^T\boldsymbol{A}$.

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}^T\boldsymbol{x} = \lambda \tag{A.6}$$

$$\Rightarrow \|\boldsymbol{A}\|_2 = \sqrt{\sup\lambda} = \sqrt{\lambda_{\max}} \tag{A.7}$$

Since $\sigma = \sqrt{\lambda}$ is the singular value of $\boldsymbol{A}$, the induced 2-norm of a matrix is equal to its largest singular value:

$$\|\boldsymbol{A}\|_2 = \sigma_{\max} \tag{A.8}$$

A.2 Condition Number

Theorem 6. *Suppose $\boldsymbol{A}$ is of full rank. Then, provided that $\|\boldsymbol{A}^\dagger\|\|\Delta\boldsymbol{A}\| < 1$, with $\Delta\boldsymbol{A}$ describing an uncertainty about matrix $\boldsymbol{A}$, the relative error of the solution vector $\boldsymbol{x}$ is bounded by*

$$\frac{|\Delta\boldsymbol{x}|}{|\boldsymbol{x}|} \leq \frac{\kappa(\boldsymbol{A})}{1-\kappa(\boldsymbol{A})\dfrac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}}\left(\frac{|\Delta\boldsymbol{b}|}{|\boldsymbol{b}|}+\frac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}\right), \tag{A.9}$$

with

$$\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\|\|\boldsymbol{A}^\dagger\| \tag{A.10}$$

being the condition number of $\boldsymbol{A}$.

A system of linear equations with unknown vector $\boldsymbol{x}$ can be written as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \tag{A.11}$$

Under the presence of small perturbations in the parameters $\boldsymbol{A}, \boldsymbol{b}$ this nominal system becomes disturbed and can be represented as

$$(\boldsymbol{A}+\Delta\boldsymbol{A})(\boldsymbol{x}+\Delta\boldsymbol{x}) = \boldsymbol{b}+\Delta\boldsymbol{b}. \tag{A.12}$$

In order to quantify the effect of these perturbations on the solution, this equation can be brought into the following explicit form for the absolute error of the solution $\Delta\boldsymbol{x}$:

$$\Delta\boldsymbol{x} = (\boldsymbol{I}+\boldsymbol{A}^\dagger\Delta\boldsymbol{A})^{-1}\boldsymbol{A}^\dagger(\Delta\boldsymbol{b}-\Delta\boldsymbol{A}\boldsymbol{x}). \tag{A.13}$$

Taking the norm of this absolute error results in

$$\begin{array}{rcl} |\Delta\boldsymbol{x}| & \leq & \|(\boldsymbol{I}+\boldsymbol{A}^\dagger\Delta\boldsymbol{A})^{-1}\|\|\boldsymbol{A}^\dagger\||(\Delta\boldsymbol{b}-\Delta\boldsymbol{A}\boldsymbol{x})| \\ & \leq & \|(\boldsymbol{I}+\boldsymbol{A}^\dagger\Delta\boldsymbol{A})^{-1}\|\|\boldsymbol{A}^\dagger\|(|\Delta\boldsymbol{b}|+\|\Delta\boldsymbol{A}\||\boldsymbol{x}|). \end{array} \tag{A.14}$$

But

$$\begin{array}{rcl} |\Delta\boldsymbol{b}| & = & |\boldsymbol{b}|\dfrac{|\Delta\boldsymbol{b}|}{|\boldsymbol{b}|} \\ & \leq & \|\boldsymbol{A}\||\boldsymbol{x}|\dfrac{|\Delta\boldsymbol{b}|}{|\boldsymbol{b}|}, \end{array} \tag{A.15}$$

and hence,

$$|\Delta\boldsymbol{x}| \leq \|(\boldsymbol{I}+\boldsymbol{A}^\dagger\Delta\boldsymbol{A})^{-1}\|\|\boldsymbol{A}^\dagger\|\|\boldsymbol{A}\||\boldsymbol{x}|\left(\frac{|\Delta\boldsymbol{b}|}{|\boldsymbol{b}|}+\frac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}\right). \tag{A.16}$$

The relative error of the solution then follows to be

$$\frac{|\Delta \boldsymbol{x}|}{|\boldsymbol{x}|} \leq \|(\boldsymbol{I} + \boldsymbol{A}^\dagger \Delta \boldsymbol{A})^{-1}\| \|\boldsymbol{A}\| \|\boldsymbol{A}^\dagger\| \left(\frac{|\Delta \boldsymbol{b}|}{|\boldsymbol{b}|} + \frac{\|\Delta \boldsymbol{A}\|}{\|\boldsymbol{A}\|} \right). \tag{A.17}$$

Substituting $\boldsymbol{A}^\dagger \Delta \boldsymbol{A}$ by $-\boldsymbol{B}$, the term of the first norm of the right hand side of eq. (A.17) leads to

$$(\boldsymbol{I} - \boldsymbol{B})^{-1} = (\boldsymbol{I} + \boldsymbol{A}^\dagger \Delta \boldsymbol{A})^{-1}. \tag{A.18}$$

It is further assumed that $\|\boldsymbol{B}\| < 1$, which is equivalent to

$$\|\boldsymbol{A}^\dagger \Delta \boldsymbol{A}\| < 1. \tag{A.19}$$

As a consequence the inverse of $(\boldsymbol{I} - \boldsymbol{B})$ can be written as the Cauchy series

$$(\boldsymbol{I} - \boldsymbol{B})^{-1} = \sum_{i=0}^{\infty} \boldsymbol{B}^i \tag{A.20}$$

and its norm is bounded by the geometric series

$$\|(\boldsymbol{I} - \boldsymbol{B})^{-1}\| \leq \sum_{i=0}^{\infty} \gamma^i = \frac{1}{1-\gamma}, \tag{A.21}$$

with $\gamma = \|\boldsymbol{B}\| < 1$. Resubstitution leads to

$$\|(\boldsymbol{I} - \boldsymbol{B})^{-1}\| \leq \frac{1}{1 - \|\boldsymbol{A}^\dagger \Delta \boldsymbol{A}\|}, \tag{A.22}$$

and finally

$$\frac{|\Delta \boldsymbol{x}|}{|\boldsymbol{x}|} \leq \frac{1}{1 - \|\boldsymbol{A}^\dagger \Delta \boldsymbol{A}\|} \|\boldsymbol{A}\| \|\boldsymbol{A}^\dagger\| \left(\frac{|\Delta \boldsymbol{b}|}{|\boldsymbol{b}|} + \frac{\|\Delta \boldsymbol{A}\|}{\|\boldsymbol{A}\|} \right). \tag{A.23}$$

Defining the condition number of the linear least-squares problem as

$$\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\| \|\boldsymbol{A}^\dagger\| \tag{A.24}$$

and using the sub-multiplicative property of the matrix norm

$$\|\boldsymbol{A}^\dagger \Delta \boldsymbol{A}\| \leq \|\boldsymbol{A}^\dagger\| \|\Delta \boldsymbol{A}\| \tag{A.25}$$

proofs the theorem.

Appendix B

Geometric Dilution of Precision for GPS

Table B.1 gives an interpretation of the GDOP for the GPS. While position estimates with a GDOP value between zero and six can be trusted to be quite accurate, estimates with values greater than ten should better be discarded.

DOP Value	Rating	Description
0	Ideal	this is the highest possible confidence level to be used for applications demanding the highest possible precision at all times
1-3	Excellent	At this confidence level, positional measurements are considered accurate enough to meet all but the most sensitive applications
4-6	Good	Represents a level that marks the minimum appropriate for making business decisions. Positional measurements could be used to make reliable in-route navigation suggestions to the user
7-8	Moderate	Positional measurements could be used for calculations, but the fix quality could still be improved. A more open view of the sky is recommended
9-20	Fair	Represents a low confidence level. Positional measurements should be discarded or used only to indicate a very rough estimate of the current location
21-50	Poor	At this level, measurements are inaccurate by as much as 50 metres and should be discarded

Table B.1: Dilution of precision taken from [Per05]

Bibliography

[AB79] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the American Society of Acoustics*, 65(4):943–950, 1979.

[Abe90] Jonathan S. Abel. Optimal sensor placement for passive source localization. *Proceedings of the IEEE ICASSP*, 5:2927–2930, 1990.

[AM79] Brian D.O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.

[AMB05] Sulema Aranda, Sonia Martinez, and Francesco Bullo. On optimal senor placement and motion coordination for target tracking. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 4544– 4549, April 2005.

[AMGC02] Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, February 2002.

[AR00] Sanjeev Arulampalam and Branko Ristic. Comparison of the particle filter with range-parameterized and modified polar ekfs for angle-only tracking. *Proceedings of SPIE*, 4048:288–299, 2000.

[BAK07] Herbert Buchner, Robert Aichner, and Walter Kellermann. Trinicon-based blind system identification with application to multiple-source localization and separation. In S. Makino, T.W. Lee, and S. Sawada, editors, *Blind Speech Separation*. Springer-Verlag, Berlin, 2007.

[BAS97] Michael S. Brandstein, John E. Adcock, and Harvey F. Silverman. A closed-form location estimator for use with room envi-

ronment microphone arrays. *IEEE Transactions of Speech and Audio Processing*, 5(1):45–50, January 1997.

[BC07] Thomas Bréhard and Jean-Pierre Le Cadre. Hierarchical particle filter for bearings-only tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 43(4):1567–1585, 2007.

[BCS94] Stéphane Bédard, Benoit Champagne, and Alex Stéphenne. Effects of room reverberation on time-delay estimation performance. *Proceedings of the IEEE ICASSP, Adelaide, Australia*, pages II–261–II–264, 1994.

[Ben00] Jacob Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *Journal of the Acoustical Society of America*, 107(1):384–391, 2000.

[Ber99] Niclas Bergman. *Recursive Bayesian Estimation, Navigation and Tracking Applications*. PhD thesis, Linköping University, 1999.

[Bra95] Michael S. Brandstein. *A Framework for Speech Source Localization Using Sensor Arrays*. PhD thesis, Brown University, May 1995.

[BRA04] Xuehai Bian, James M. Rehg, and Gregory D. Abowd. *Sound Source Localization in Domestic Environment*, 2004. GVU Technical Report;GIT-GVU-04-06.

[BSMM99] Ilija N. Bronstein, Konstantin A. Semedjajew, Gerhard Musiol, and Heiner Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 1999.

[Cad98] Jean-Pierre Le Cadre. Properties of estimability criteria for target motion analysis. *IEE Proceedings- Radar, Sonar and Naviagation*, 145(2):92–99, 1998.

[CBH07] Jingdong Chen, Jacob Benesty, and Yiteng Huang. Time delay estimation in room acoustic environments: An overview. *EURASIP Journal on Applied Signal Processing*, 2006:1–19, January 2007.

[CBS96] Benoit Champagne, Stéphane Bédard, and Alex Stéphenne. Performance of time-delay estimation in the presence of room reverberation. *IEEE Transactions on Speech and Audio Processing*, 4(2):148–152, March 1996.

[CH94] Y.T. Chan and K.C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Signal Processing*, 42(8):1905–1915, August 1994.

[CL04] Yuan Cheng and Bing Li. Dsp-based acoustic source localization. Master's thesis, Hochschule für Angewandte Wissenschaften Hamburg, 2004.

[CLH+07] Minghua Chen, Zicheng Liu, Li-Wei He, Phil Chou, and Zhengyou Zhang. Energy-based position estimation of microphones and speakers for ad hoc microphone arrays. *Proceedings of the IEEE Waspaa*, pages 22–25, 2007.

[CNC73] G.C. Carter, A.H. Nuttall, and P.G. Cable. The smoothed coherence transform. *Proceedings of the IEEE*, 61(10):1497–1498, 1973.

[Con99] Conrad. *Mono-Mikrofon Vorverstärker*, 03/99 edition, 1999. Best.Nr.: 19 76 88.

[CR81] Ronald E. Crochiere and Lawrence R. Rabiner. Interpolation and decimation of digital signals - a tutorial review. *Proceedings of the IEEE*, 69(3):300–331, 1981.

[CT98] Jean-Pierre Le Cadre and Olivier Trémois. Bearings-only tracking for maneuvering sources. *IEEE Transactions on Aerospace and Electronic Systems*, 34(1):179–193, 1998.

[CYH02] Joe C. Chen, Kung Yao, and Ralph E. Hudson. Source localization and beamforming. *IEEE Signal Processing Magazine*, pages 30–39, March 2002.

[CYW+03] Joe C. Chen, L. Yip, H. Wang, D. Maniezzo, R.E. Hudson, J. Elson, and D. Estrin K. Yao. Dsp implementation of a distributed acoustical beamformer on a wireless sensor platform. *Proceedings of the IEEE ICASSP*, 2:597–600, 2003.

[DBS01] Joseph H. DiBiase, Michael S. Brandstein, and Harvey F. Silverman. Robust localization in reverberant rooms. In M. Bandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, Berlin, 2001.

[DGA00] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10:197–28, 2000.

[DLR77] Arthur Dempster, Nan Laird, and Donald Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[DS295] DS275. *DS275, Line-Powered RS-232 Transceiver Chip*. Dallas Semiconductor, 1995.

[ECR98] Bruno Emile, Pierre Comon, and Joël Le Roux. Estimation of time delays with fewer sensors than sources. *IEEE transactions on Signal Processing*, 46(7):2012–2015, July 1998.

[Fan90] Bertrand T. Fang. Simple solutions for hyperbolic and related position fixes. *IEEE Transactions on Aerospace and Electronic Systems*, 26(5):748–753, 1990.

[Fea98] Paul Fearnhead. *Sequential Monte-Carlo Methods in Filter Theory*. PhD thesis, University of Oxford, 1998.

[FG07] Maurice Fallon and Simon Godsill. Multi target acoustic source tracking using track before detect. *Proceedings of the IEEE WASPAA*, pages 102–105, October 2007.

[Fle00] Roger Fletcher. *Practical Methods of Optimization*. Whiley VCH, 2000.

[Foy76] Wade H. Foy. Position-location solutions by taylor-series estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 12(2):187–194, 1976.

[GG95] Robert M. Gray and Joseph M. Goodman. *Fourier Transforms. An Introduction for Engineers*. Springer, 1st edition, 1995.

[GRT02] Tony Gustafsson, Bhaskar D. Rao, and Mohan Trivedi. Analysis of time-delay estimation in reverberant environments. *Proceedings of the IEEE ICASSP, Orlando, FL*, pages 2097–2100, 2002.

[GRT03] Tony Gustafsson, Bhaskar D. Rao, and Mohan Trivedi. Source localization in reverberant environments: Modeling and statistical analysis. *IEEE Transactions on Speech and Audio Processing*, 11(6):791–803, November 2003.

[GSS93] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimate. *IEE Proceedings*, 140(2):107–113, 1993.

[HBC06] Yiteng Huang, Jacob Benesty, and Jingdong Chen. *Acoustic MIMO Signal Processing*. Springer-Verlag, 2006.

[HBE00] Yiteng Huang, Jacob Benesty, and Gary W. Elko. Passive acoustic source localization for video camera steering. *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pages 909–912, 2000.

[HBEM01] Yiteng Huang, Jacob Benesty, Gary W. Elko, and Russell M. Mersereau. Real-time passive source localization: A practical linear-correction least-squares approach. *IEEE Transactions on Speech and Audio Processing*, 9(8):943–956, November 2001.

[HL64] Yu Chi Ho and R.C.K. Lee. A bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9:333–339, 1964.

[HSG06] Jeroen D. Hol, Thomas B. Schön, and Fredrik Gustafsson. On resampling algorithms for particle filters. *Proceedings of the IEEE Nonlinear Statistical Signal Processing Workshop*, pages 79–82, 2006.

[HV03] Tamir Hegazy and George Vachtsevanos. Sensor placement for isotropic source localization. *Proceedings of the Second International Workshop on Information Processing in Sensor Networks*, April 2003.

[JR06] Damien B. Jourdan and Nicholas Roy. Optimal sensor placement for agent localization. *Proceeding IEEE/ION PLANS*, pages 128–139, April 2006.

[JU97] Simon J. Julier and Jeffrey K. Uhlmann. A new extension of the kalman filter to nonlinear systems. *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Control*, 1997.

[JU04] Simon J. Julier and Jeffrey K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[Jul97] Simon J. Julier. *Process Models for the Navigation of High-Speed Land Vehicles*. PhD thesis, Wadham College, Oxford, 1997.

[Jul02] Simon J. Julier. The scaled unscented transformation. *Proceedings of the American Control Conference*, 6:4555–4559, May 2002.

[Kal60] Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineeering*, 82:34–45, 1960.

[Kay93] Steven M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[KC76] Charles H. Knapp and G. Clifford Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 4(4):320–327, 1976.

[KE95] James Kennedy and Russell Eberhart. Particle swarm optimization. *Proceedings IEEE Intl. Conf. on Neural Networks*, pages 1942–1948, 1995.

[Kit96] Genshiro Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[LC98] Jun S. Liu and Rong Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.

[LDW91] John J. Leonard and Hugh F. Durrant-Whyte. *Directed Sonar Sensing for Mobile Robot Navigation*. Kluwer, Boston, MA, 1991.

[Lev00] N. Levanon. Lowest gdop in 2-d scenarios. *IEE proceedings - Radar, Sonar and Navigation*, 147:149–155, June 2000.

[LJ03] X. Rong Li and Vesselin P. Jilkov. Survey of maneuvering target tracking. part i: Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1333–1364, 2003.

[LJ07] Eric A. Lehmann and Anders M. Johansson. Particle filter with integrated voice activity detection for acoustic source tracking. *EURASIP Journal on Advances in Signal Processing*, 2007:1–11, 2007.

[LJN07] Eric A. Lehmann, Anders M. Johansson, and Sven Nordholm. Modeling of motion dynamics and its influence on the performance of a particle filter for acoustic speaker tracking. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 98–102, 2007.

[Mat06] Mathworks. *Optimization Toolbox, for use with Matlab, User's Guide, Version 3*. The MathWorks, Inc., 2006.

[MP97] J.B. McKay and M. Pachter. Geometry optimization for gps navigation. *Proceedings of the 36th Conference on Decision and Control*, pages 4695–4699, 1997.

[NBM07] Jan Neering, Marc Bordier, and Nadia Maïzi. Optimal passive source localization. *SensorComm 2007*, October 2007.

[NE92] P.A. Nelson and S.J. Elliott. *Active Control of Sound*. Academic Press, 1992.

[NFBM08] Jan Neering, Christian Fischer, Marc Bordier, and Nadia Maïzi. Optimal sensor configuration for passive position estimation. *IEEE/ION Position, Location and Navigation Symposium*, pages 951–960, May 2008.

[Per05] Jon Person. Writing your own gps applications: Part 2. *DeveloperFusion.com*, 2005. http://www.developerfusion.com/ article/ 4652/ writing-your-own-gps-applications-part-2/.

[PFTV02] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipies in C: The Art of Scientific Computing*. Cambridge University Press, 2002.

[Pie81] A.D. Pierce. *Acoustics: An Introduction to its Physical Properties and Applications*. McGraw-Hill, New York, 1981.

[Rab98] Daniel V. Rabinkin. *Optimum Sensor Placement for Microphone Arrays*. PhD thesis, Rutgers, 1998.

[RFB81] Francis A. Reed, Paul L. Feintuch, and Neil J. Bershad. Time delay estimation using the lms adaptive filter - static behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(3):561–571, June 1981.

[Rot71] Peter R. Roth. Effective measurements using digital signal analysis. *IEEE Spectrum*, 8(4):62–72, 1971.

[Rub88] D.B. Rubin. Using the sir algorithm to simulate posterior distributions. In J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, editors, *Bayesian Statistics*, pages 395–402. Oxford University Press, 1988.

[SA87] Julius O. Smith and Jonathan S. Abel. Closed-form least-square source location estimation from range-difference measurements. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(12):1661–1669, December 1987.

[Sch86] Ralph O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, March 1986.

[SMAM05] Hiroshi Sawada, Ryo Mukai, Shoko Araki, and Shoji Makino. Multiple source localization using independent component analysis. *IEEE AP-S International Symposium and USNC/URSI National Radio Science Meeting*, July 2005.

[Smi08] Julius O. Smith. Digital audio resampling home page, 2008. http://ccrma.stanford.edu/ jos/resample/.

[SMR99] Norbert Strobel, Thomas Meier, and Rudolf Rabenstein. Speaker localization using a steered filter-and-sum beamformer. In *In Proc. Erlangen Workshop on Vision, Modeling, and Visualization*, pages 195–202, 1999.

[SR87] H. C. Schau and A. Z. Robinson. Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(8):1223–1225, August 1987.

[SY06] Jan Scheuing and Bin Yang. Disambiguation of tdoa estimates in multi-path multi-source environments (datemm). *Proceedings of the IEEE Inernational Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4:837–840, 2006.

[SY07] Jan Scheuing and Bin Yang. Efficient synthesis of approximately consistent graphs for acoustic multi-source localization. *Proceedings of the IEEE Inernational Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 501–504, 2007.

[Tor84] Don J. Torrieri. Statistical theory of passive location systems. *IEEE Transactions on Aerospace and Electronic Systems*, 20:183–198, 1984.

[VB01] Jaco Vermaark and Andrew Blake. Nonlinear filtering for speaker tracking in noisy and reverberant environments. *Proceedings of the IEEE ICASSP*, 5:3021–3024, 2001.

[vdM04] Rudolph van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, Oregon Health and Science University, April 2004.

[vdMWJ04] Rudolph van der Merwe, Eric A. Wan, and Simon I. Julier. Sigma-point kalman filters nonlinear estimation and sensor fusion - applications in integrated navigation. *AIAA Guidance Navigation and Control Conference*, March 2004.

[Wik08] Wikipedia.org. Voice frequency, 2008. http://en.wikipedia.org/wiki/Voice_frequency.

[WLW03] Darren B. Ward, Eric A. Lehmann, and Robert C. Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *IEEE Transactions on Speech and Audio Processing*, 11(6):826–836, 2003.

[WM01] Eric A. Wan and Rudolph Van Der Merwe. Chapter 7: The unscented kalman filter. In *Kalman Filtering and Neural Networks*, pages 221–280. Wiley, 2001.

[YS05] Bin Yang and Jan Scheuing. Cramer-rao bound and optimum sensor array for source localization from time differences of arrival. *Proceedings of the IEEE ICASSP*, 4:961–964, 2005.

[Zha95] Hong Zhang. Two-dimensional optimal sensor placement. *IEEE Transaction on Systems, Man, and Cybernetics*, 25(5):781–792, 1995.

[Zog06] Jean-Marie Zogg. *Grundlagen der Satellitennavigation*, 2006. available at http://u-blox.com.