Andreas Richter

Estimation of Radio Channel Parameters:
Models and Algorithms

isle

2005

Estimation of Radio Channel Parameters:
Models and Algorithms

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur
(Dr.-Ing.)

vorgelegt der
Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Ilmenau

von

Dipl.-Ing. Andreas Richter
geboren am 9. Januar 1969 in Weimar

Gutachter:	Prof. Dr.-Ing. habil. Reiner S. Thomä (TU Ilmenau)
	o.Univ.Prof. i.R. Dipl.-Ing. Dr.techn. Ernst Bonek (TU Wien)
	Prof. Dr.-Ing. Martin Haardt (TU Ilmenau)

To my parents

Abstract

The theme of this thesis is the estimation of model parameters of a radio channel snapshot. The main focus was the development of a general data model for the measured radio channel, suitable for both high resolution channel parameter estimation on the one hand, and the development of a robust parameter estimator for the parameters of the designed parametric radio channel model, in line with this logical work flow is this thesis.

In the first part of this work an algebraic representation of observed propagation paths is developed using a ray-optical model known from literature. The algebraic framework is suitable for the description of SISO (single-input-single-output) radio transmission systems. A SISO system uses one antenna as the transmitter (Tx) and one antenna as the receiver (Rx). The derived expression for the propagation paths is also suitable to describe SIMO (single-input-multiple-output), MISO (multiple-input-single-output), and MIMO (multiple-input-multiple-output) radio channel measurements. In contrast to other models used for high resolution channel parameter estimation the derived model makes no restriction regarding the structure of the antenna array used throughout the measurement. This is important since the ultimate goal in radio channel sounding is the complete description of the spatial (angular) structure of the radio channel at Tx and Rx. The flexibility of the data model is a prerequisite for the optimisation of the antenna array structure with respect to the measurement task. Such an optimised antenna structure is a stacked uniform circular beam array, i.e., a cylindrical arrangement of antenna elements. This antenna array configuration is well suited for the measurement of the spatial structure of the radio channel at Tx and/or Rx in outdoor-scenarios. Furthermore, a new component of the radio channel model is introduced in the first part of this work. It describes the contribution of distributed (diffuse) scattering to the radio transmission. The new component is key for the development of a robust radio channel parameter estimator, which is derived in the main part of this work. The ignorance of the contribution of distributed scattering to radio propagation is one of the main reasons why high-resolution radio channel parameter estimators fail in practice. Since the underlying data model is wrong the estimators produce erroneous results. The improved model describes the so called dominant propagation paths by a deterministic mapping of the propagation path parameters to the channel observation. The contribution of the distributed scattering is modelled as a zero-mean circular Gaussian process. The parameters of the distributed scattering process determine the structure of the covariance matrix of this process. Based on this data model current concepts for radio channel sounding devices are discussed.

In the second part of this work expressions for the accuracy achievable by a radio channel sounder are derived. To this end the lower bound on the variance of the measurements i.e. the parameter estimates is derived. As a basis for this evaluation the concept of the Cramér-Rao lower bound is employed. On the way to the Cramér-Rao lower bound for all channel model parameters, important issues for the development of an appropriate parameter estimator are discussed. Among other things the coupling of model parameters is also discussed.

v

In the third part of this thesis, an estimator, for the propagation path parameters is derived. For the estimator the 'maximum-likelihood' approach is employed. After a short overview of existing high-resolution channel parameter estimators the estimation problem is classified. It is shown, that the estimation of the parameters of the propagation paths can be understood as a nonlinear weighted least squares problem, provided the parameters of the distributed scattering process are known. Based on this observation a general algorithm for the estimation of raw parameters for the observed propagation paths is developed. The algorithm uses the concept of structured-least-squares (SLS) and compressed maximum likelihood to reduce the numerical complexity of the estimation problem. A robust estimator for the precise estimation of the propagation path parameters is derived. The estimator is based on concepts well known from nonlinear local optimisation theory. In the last part of this chapter the application of subspace based parameter estimation algorithms for path parameter estimation is discussed. A memory efficient estimator for the signal subspace needed by, e.g., R-D unitary ESPRIT is derived. This algorithm is a prerequisite for the application of signal subspace based algorithms to MIMO-channel sounding measurements. Standard algorithms for signal subspace estimation (economy size SVD, singular value decomposition) are not suitable since they require an amount of memory which is too large. Furthermore, it is shown that ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) based algorithms can also be employed for parameter estimation from data having hidden rotation invariance structure. As an example an ESPRIT algorithm for angle estimation using circular uniform beam arrays (circular multi-beam antennas) is derived.

In the final part of this work a maximum likelihood estimator for the new component of the channel model is developed. Starting with the concept of iterative maximum likelihood estimation, an algorithm is developed having a low computational complexity. The low complexity of the algorithm is achieved by exploiting the Toeplitz-structure of the covariance matrix to estimate. Using the estimator for the (concentrated, dominant, specular-alike) propagation paths and the parametric estimator for the covariance matrix of the process describing the distributed diffuse scattering a joint estimator for all channel parameter is derived (RIMAX). The estimator is a 'maximum likelihood' estimator and uses the genuine SAGE concept to reduce the computational complexity. The estimator provides additional information about the reliability of the estimated channel parameters. This reliability information is used to determine an appropriate model for the observation. Furthermore, the reliability information i.e. the estimate of the covariance matrix of all parameter estimates is also an important parameter estimation result. This information is a prerequisite for further processing and evaluation of the measured channel parameters.

Kurzfassung

Diese Dissertation behandelt die Schätzung der Modellparameter einer Momentanaufnahme des Mobilfunkkanals. Das besondere Augenmerk liegt zum einen auf der Entwicklung eines generischen Datenmodells für den gemessenen Funkkanal, welches für die hochauflösende Parameterschätzung geeignet ist. Der zweite Schwerpunkt dieser Arbeit ist die Entwicklung eines robusten Parameterschätzers für die Bestimmung der Parameter des entworfenen Modells aus Funkkanalmessdaten. Entsprechend dieser logischen Abfolge ist auch der Aufbau dieser Arbeit.

Im ersten Teil wird ausgehend von einem aus der Literatur bekannten strahlenoptischen Modell eine algebraisch handhabbare Darstellung von beobachteten Wellenausbreitungspfaden entwickelt. Das mathematische Modell erlaubt die Beschreibung von SISO (single-input-single-output)-Übertragungssystemen, also von Systemen mit einer Sendeantenne und einer Empfangsantenne, als auch die Beschreibung von solchen Systemen mit mehreren Sende- und/oder Empfangsantennen. Diese Systeme werden im Allgemeinen auch als SIMO- (single-input-multiple-output), MISO- (multiple-input-single-output) oder MIMO-Systeme (multiple-input-multiple-output) bezeichnet. Im Gegensatz zu bekannten Konzepten enthält das entwickelte Modell keine Restriktionen bezüglich der modellierbaren Antennenarrayarchitekturen. Dies ist besonders wichtig in Hinblick auf die möglichst vollständige Erfassung der räumlichen Struktur des Funkkanals. Die Flexibilität des Modells ist eine Grundvoraussetzung für die optimale Anpassung der Antennenstruktur an die Messaufgabe. Eine solche angepasste Antennenarraystruktur ist zum Beispiel eine zylindrische Anordnung von Antennenelementen. Sie ist gut geeignet für die Erfassung der räumlichen Struktur des Funkkanals (Azimut und Elevation) in so genannten Outdoor-Funkszenarien. Weiterhin wird im ersten Teil eine neue Komponente des Funkkanaldatenmodells eingeführt, welche den Beitrag verteilter (diffuser) Streuungen zur Funkübertragung beschreibt. Die neue Modellkomponente spielt eine Schlüsselrolle bei der Entwicklung eines robusten Parameterschätzers im Hauptteil dieser Arbeit. Die fehlende Modellierung der verteilten Streuungen ist eine der Hauptursachen für die begrenzte Anwendbarkeit und die oft kritisierte fehlende Robustheit von hochauflösenden Funkkanalparameterschätzern, die in der Literatur etabliert sind. Das neue Datenmodell beschreibt die so genannten dominanten Ausbreitungspfade durch eine deterministische Abbildung der Pfadparameter auf den gemessenen Funkkanal. Der Beitrag der verteilten Streuungen wird mit Hilfe eines zirkularen mittelwertfreien Gaußschen Prozesses beschrieben. Die Modellparameter der verteilten Streuungen beschreiben dabei die Kovarianzmatrix dieses Prozesses. Basierend auf dem entwickelten Datenmodell wird im Anschluss kurz über aktuelle Konzepte für Funkkanalmessgeräte, so genannte Channel-Sounder, diskutiert.

Im zweiten Teil dieser Arbeit werden in erster Linie Ausdrücke zur Bestimmung der erzielbaren Messgenauigkeit eines Channel-Sounders abgeleitet. Zu diesem Zweck wird die untere Schranke für die Varianz der geschätzten Modellparameter, das heißt der Messwerte, bestimmt. Als Grundlage für die Varianzabschätzung wird das aus der Parameterschätztheorie bekannte Konzept der Cramér-Rao-Schranke angewandt. Im Rahmen der Ableitung der Cra-

mér-Rao-Schranke werden außerdem wichtige Gesichtspunkte für die Entwicklung eines effizienten Parameterschätzers diskutiert.

Im dritten Teil der Arbeit wird ein Schätzer für die Bestimmung der Ausbreitungspfadparameter nach dem Maximum-Likelihood-Prinzip entworfen. Nach einer kurzen Übersicht über existierende Konzepte zur hochauflösenden Funkkanalparameterschätzung wird die vorliegende Schätzaufgabe analysiert und in Hinsicht ihres Typs klassifiziert. Unter der Voraussetzung, dass die Parameter der verteilten Streuungen bekannt sind, lässt sich zeigen, daß sich die Schätzung der Parameter der Ausbreitungspfade als ein nichtlineares gewichtetes kleinstes Fehlerquadratproblem auffassen lässt. Basierend auf dieser Erkenntnis wird ein generischer Algorithmus zur Bestimmung einer globalen Startlösung für die Parameter eines Ausbreitungspfades vorgeschlagen. Hierbei wird von dem Konzept der Structure-Least-Squares (SLS)-Probleme Gebrauch gemacht, um die Komplexität des Schätzproblems zu reduzieren. Im folgenden Teil dieses Abschnitts wird basierend auf aus der Literatur bekannten robusten numerischen Algorithmen ein Schätzer zur genauen Bestimmung der Ausbreitungspfadparameter abgeleitet. Im letzten Teil dieses Abschnitts wird die Anwendung unterraumbasierter Schätzer zur Bestimmung der Ausbreitungspfadparameter diskutiert. Es wird ein speichereffizienter Algorithmus zur Signalraumschätzung entwickelt. Dieser Algorithmus ist eine Grundvoraussetzung für die Anwendung von mehrdimensionalen Parameterschätzern wie zum Beispiel des R-D unitary ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) zur Bestimmung von Funkkanalparametern aus MIMO-Funkkanalmessungen. Traditionelle Verfahren zur Signalraumschätzung sind hier im Allgemeinen nicht anwendbar, da sie einen zu großen Speicheraufwand erfordern. Außerdem wird in diesem Teil gezeigt, dass ESPRIT-Algorithmen auch zur Parameterschätzung von Daten mit so genannter versteckter Rotations-Invarianzstruktur eingesetzt werden können. Als Beispiel wird ein ESPRIT-basierter Algorithmus zur Richtungsschätzung in Verbindung mit multibeam-Antennenarrays (CUBA) abgeleitet.

Im letzten Teil dieser Arbeit wird ein Maximum-Likelihood-Schätzer für die neue Komponente des Funkkanals, welche die verteilten Streuungen beschreibt, entworfen. Ausgehend vom Konzept des iterativen Maximum-Likelihood-Schätzers wird ein Algorithmus entwickelt, der hinreichend geringe numerische Komplexität besitzt, so dass er praktisch anwendbar ist. In erster Linie wird dabei von der Toeplitzstruktur der zu schätzenden Kovarianzmatrix Gebrauch gemacht. Aufbauend auf dem Schätzer für die Parameter der Ausbreitungspfade und dem Schätzer für die Parameter der verteilten Streuungen wird ein Maximum-Likelihood-Schätzer entwickelt (RIMAX), der alle Parameter des in Teil I entwickelten Modells der Funkanalmessung im Verbund schätzt. Neben den geschätzten Parametern des Datenmodells liefert der Schätzer zusätzlich Zuverlässigkeitsinformationen. Diese werden unter anderem zur Bestimmung der Modellordnung, das heißt zur Bestimmung der Anzahl der dominanten Ausbreitungspfade, herangezogen. Außerdem stellen die Zuverlässigkeitsinformationen aber auch ein wichtiges Schätzergebnis dar. Die Zuverlässigkeitsinformationen machen die weitere Verarbeitung und Wertung der Messergebnisse möglich.

Acknowledgement

Contents

List of Tables

List of Figures

1   Introduction

As high-speed data services for multimedia Internet access are brought to focus, huge date rates per user are anticipated for future 3G and 4G mobile radio systems. The most likely method of increasing capacity for wireless transmission is to exploit smart antennas. Space-time processing using multiple antennas is supposed to enhance the performance in terms of coverage, capacity and quality of service (QoS) considerably. The signal processing methods widely discussed are beam forming, joint space-time equalisation, multi-user detection, interference cancellation, and spatial diversity. Link adaptation methods are applied to control the modulation and coding according to the available radio link performance in order to meet the required QoS. Highest link capacity is expected if multiple antennas are used at both the receiver and the transmitter site. In this case, the radio propagation channel as accessed by the antennas constitutes a multiple-input-multiple-output (MIMO) system. Profound knowledge of the multipath propagation channel impulse response (CIR) is a prerequisite for system planning, algorithm design, and link level simulation of advanced radio systems using adaptive multiple antenna modems. A lot of spatial channel models and propagation simulation tools have been developed (see [1] for an overview). In order to keep the simulation tractable, these models must extremely simplify the complicated electromagnetic transmission processes of reflection, scattering, diffraction, shadowing etc. In labyrinthine radio environments such as densely built-up areas, indoor and especially industrial areas and factory halls, car-to-car links in heavy traffic, these models cannot reproduce the reality of wave propagation. To complicate matters further, random user mobility and possible movement of parts of the environment have to be considered. Consequently, the MIMO radio channel has to be treated as a time-variant system. Moreover, adjacent-channel and co-channel interference from other subscribers or even from coexisting systems have to be considered. Therefore, advanced measurement systems are required, which deliver CIR data for realistic link level simulation and even for system level performance assessment. These data can also be used to deduce measurement based channel models.

1.1   Background

The design of future mobile radio networks (beyond 3G) requires research towards new air interfaces which are characterized by highest bandwidth efficiency and unprecedented flexibility. It is commonly understood that radio systems equipped with multiple antennas at both the mobile station (MS) and the base station (BS) have a huge potential to increase the available capacity for high bit rate wireless links, which results from a simultaneous transmission of multiple data streams from different antenna elements [2]. This multi-antenna transmission technique is called MIMO transmission. It can optimally exploit the spatial diversity of multiple propagation paths existing in a rich scattering environment. Conceptually, the multipath propagation of the radio channel gives rise to different spatio-temporal signatures for the different transmit data streams, which permits a receiver equipped with multiple antennas to separate those data streams from the received signal mixture, that are otherwise not orthogo-

nal in any of the conventional communication signal dimensions, i.e., by time, frequency, or code. With this information, it is not surprising that the performance of a MIMO system strongly depend on the radio channel conditions. A key question for a system design and implementation is therefore, do we find practically feasible schemes that are sufficiently robust for this task? On the other hand, somewhat related, what specific features are required for a practical MIMO system to work reliably under a wealth of various propagation conditions? To give an example, the upcoming IEEE 802.11n standard applies the MIMO transmission concept to increase the data rate of WLAN-systems (Wireless Local Area Networks). Furthermore, the new systems will use a bandwidth of up to 40MHz. The aim is to reach a throughput of at least 100Mbps.
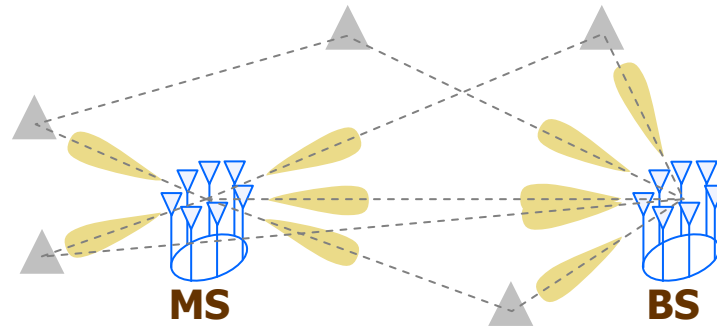
Figure 1-1:   Double directional (direction of departure and direction of arrival) structure of a multipath channel.

A thorough investigation of the multidimensional wave propagation mechanisms is a prerequisite for understanding the spatial and temporal structure of the channel transfer matrix, and thus, for optimum design and realistic performance evaluation of multiple antenna systems. There are many attempts to simulate the input-output behaviour of the channel. One approach is physically motivated, it is based on electromagnetic wave propagation analysis and uses a ray optical model. In case of ray tracing or ray launching, a detailed database describing the propagation environment is required. Other models, although ray based as well, use statistical assumptions on the distribution of reflectors (e.g. COST 259 [3], COST273 [4]). There are also models trying to reproduce the input/output behaviour in a statistical sense by formal assumptions of correlation coefficients and distributions resulting at transmit and receive antenna ports disregarding the geometrical distribution of the reflectors. A drawback of non-geometric models is that they are inherently specific for a certain antenna characteristic. For antenna independent modelling (which allows antenna de-embedding and embedding), it seems that geometry-based models are necessary [5], [6].

As the complexity of wave reflection, scattering, diffraction, etc. in real propagation environments can never be completely reproduced by electromagnetic simulation and because of the strong simplifications of the statistical approaches, all models have to be verified and parameterised by propagation measurements. Moreover, channel models can be directly deduced from measurements in real propagation environments by estimating the geometric path parameters from the recorded data [7]. Given a ray-optical path model, the parameters of a suitably defined propagation path model are DoA (Direction of Arrival) at the receiver array, DoD (Direction of Departure) at the transmitter array, TDoA (Time Delay of Arrival), Doppler shift, and the complex, polarimetric path weight matrix.

A multidimensional channel sounder is a measuring device, which allows the observation of the time-varying multipath channel impulse response in its relevant multiple dimensions. These dimensions may be temporal and spatial in nature and must contain information

on all model parameters described above. To this end, we need a broadband excitation signal to "sound" the channel in the frequency range of interest and antenna arrays, which sense the wave field in a properly defined spatial aperture. A sounder system typically consists of a mobile transmitter (Tx) which plays the role of the MS (mobile station) and a fixed receiver (Rx) acting as the BS (base station). Since the channel is reciprocal, it makes no difference if the results are interpreted as uplink or as downlink. The receiver may also be moving if we consider a peer-to-peer communication between two mobile platforms. Both excitation and recording must be repetitive with a period short enough to make the temporal variation statistics according to Doppler shift and fast fading visible. Moreover, the data recording must be continuous along a precisely defined trajectory to reproduce slow channel parameter variation. This all is called real-time MIMO channel sounding and makes the measured data usable for simulating the MIMO transceiver performance including link- and system-level aspects. Figure 1-1 highlights the double directional structure of the multipath channel. Specifically, double directional measurement, which includes joint DoA/DoD estimation, allows the influence of the measurement antennas to be separated from the channel measurements. This is a prerequisite of antenna independent channel characterisation.

The antenna array arrangement is of crucial importance to represent a certain system scenario. This applies to the typical BS or access point (AP) location in a cellular or WLAN specific deployment scenario. The MS antenna array location should resemble the characteristic user roaming behaviour. This may include almost stationary user terminals but also high mobility user platforms such as cars, aircrafts, or trains. For ad-hoc and multi-hop networking, the situation changes completely since there is no dedicated BS. Instead, both sides of the link have to represent the terminal morphology and mobility. This influences the antenna array architecture, which consists of the array size, shape and of the number, orientation and characteristics of the individual antenna elements. Both BS and AP, e.g., may have a limited viewing sector. The MS, acting as the user, should have a full angular coverage in order to represent arbitrary user antenna orientation. Moreover, advanced network specific scenarios such as multiple users including known and unknown interference, cooperative downlinks from multiple BS or APs, multi-hop networking and relaying, etc. have to be emulated by the measurement setup. It is only if the measurement scenario is properly defined that the recorded CIR data can be used for realistic link- and system level simulation. The advantages of this measurement based off-line approach in comparison with the prototype hardware demonstration are higher flexibility, lower costs, and an improved perception of the transceiver's operation. The latter is primarily due to more effective analysis techniques, which allow the observed transceiver performance to be traced back to the actual time variant space-time structure and physical propagation phenomena.

Even more specific design roles for antenna arrays apply if we have in mind high-resolution estimation of the ray-optical multipath model. The channel response can, in general, be observed only within a limited aperture volume which is somewhat related to the array size, frequency bandwidth and temporal observation window. This strictly limits the achievable parameter resolution and accuracy when classical nonparametric estimation algorithms are applied. Therefore, high-resolution parameter estimation algorithms have to be envisaged to enhance the resolution by fitting an appropriate data model to the measured data. In this case, the resolution is only limited by the signal to noise ratio (SNR), antenna and device imperfections, calibration quality, and the limited validity of the data model. The resolution performance mainly depends on the antenna array architecture and its manufacturing quality, which includes low electromagnetic coupling, high electrical and mechanical stability and precise calibration. In the context of high-resolution channel parameter estimation, also the definition of the data model is crucial for parameter estimation. It has to represent the reality of wave propagation and the influence of the measurement device. A proper choice can

dramatically reduce the algorithmic complexity and enhance the accuracy and resolution as well as the reliability of the results. Moreover, a sufficient accurate continuous device data model is required, which includes the precise knowledge of the complex polarimetric antenna array response.

1.2   Overview and Contributions

Chapter 2 - Radio Channel and System Model describes the viewpoint of the mobile radio channel used throughout this work. A model for concentrated (specular alike) propagation paths is derived, which can be expressed entirely by means of linear algebra. The model is a continuous function of the propagation path parameters. Furthermore, the data model commonly used in high-resolution channel parameter estimation, which describes the radio channel observation (measurement) by a finite sum of concentrated propagation paths and measurement noise, is extended. A third component of the channel model is introduced to account for so-called dense multipath, which is caused by distributed diffuse scattering.

Chapter 3 - Radio Channel Measurement gives a short overview of radio channel sounding measurement devices. Furthermore, general issues in radio propagation measurements related to channel parameter estimation are discussed.

In **Chapter 4 - Limits on Channel Parameter Estimation,** closed form expressions for the Cramér-Rao lower bound on the variance of any channel parameter estimator are derived. At first, expressions for the Cramér-Rao lower bound on the parameters of the propagation paths are established. The structure of the Fisher information matrix and its influence on path identifiability as well as parameter estimator structure is discussed. In the second part, a closed form expressions for the Cramér-Rao lower bound on the parameters of the stochastic part of radio channel observation (dense multipath) is derived. Finally, an expression for the joint Cramér-Rao lower bound of all channel model parameters is given.

In **Chapter 5 - Estimation of Path Parameters,** the general structure of the path parameter estimation problem is discussed. A maximum likelihood estimator based on a global search strategy to determine raw initial path parameter estimates, and a subsequent local maximization algorithm is proposed in this chapter. In both algorithms, the new channel model component (dense multipath) is taken into account. Furthermore, it is discussed how the variance of the estimated path parameters can be estimated, and how the parameter estimates are approximately distributed. Using the variance estimates, a new algorithm to determine the model order, i.e., the number of assessable propagation paths is proposed.

In the second part of this chapter some algorithms, which are important if subspace based algorithms are applied to path parameter estimation are summarized. An algorithm for economy size signal subspace estimation is described. In the new algorithm for signal subspace estimation, multidimensional smoothing is carried out implicitly thus significantly reducing the memory requirements. Furthermore, the algorithm for economy size signal subspace estimation is extended to account also for the contribution of dense multipath components, i.e., to handle coloured noise.

In **Chapter 6 - Estimation of DMC Parameters,** an approach to maximum likelihood estimation of the DMC (Dense Multipath Components) parameters is described. An algorithm based on iterative maximum likelihood is developed. Furthermore, it is shown that the direct implementation of the iterative 'maximum likelihood' estimator leads to an algorithm, which is computationally expensive. To resolve this problem an estimator is proposed, which embeds the parametric covariance matrix to estimate in a circulant matrix leading to a significant reduction in computational complexity. Finally, the RIMAX algorithm for joint estimation of all channel parameters is outlined.

In **Chapter 7 - Antenna Array Calibration,** an algorithm for the estimation of calibration matrices for uniform linear, rectangular, and circular arrays as well as circular uniform beam arrays is described. The algorithm is a significant improvement in terms of computational complexity compared with existing algorithms.

In **Chapter 8 – Summary and Conclusion,** the key issues of this thesis are summarised and its main conclusions are repeated. Finally, an overview of open issues related to mobile radio channel parameter estimation, which are of further research interest, are discussed.

The parameter estimation algorithm developed within this work has been implemented at Technische Universität Ilmenau at the Electronic Measurement Lab. It is also available as a commercial product from MEDAV GmbH. The name of the parameter estimator "RIMAX" has been proposed by MEDAV GmbH.

1.3   Notation

A capital boldfaced letter denotes a matrix, and a small boldfaced letter a vector. Furthermore, $\mathbf{A}(\mathbf{b})$ denotes a matrix-valued function of the variable $\mathbf{b}$, and $\mathbf{a}(\mathbf{b})$ denotes a vector valued function of $\mathbf{b}$. Unless otherwise noted, are all vectors column vectors. A boldfaced $\mathbf{1}$ denotes a column vector with 1-elements. The superscripts $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $(\cdot)^{*}$ denotes transposition, complex conjugate (Hermitian) transposition and element-wise complex conjugate, respectively. The element i,k of a matrix is called with $\{\mathbf{A}\}_{ik}$. The small letter j denotes the square root of -1, i.e., $j = \sqrt{-1}$. The operators $\Re\{\cdot\}$, $\Im\{\cdot\}$, and $\mathrm{E}\{\cdot\}$ denotes real part, imaginary part, and the expectation operator, respectively. The symbol $^\wedge$ over a variable denotes its estimate. Additionally, important equations are marked with the symbol ∎. Appendix A, B, and C contains frequently used symbols, acronyms, and often used relations. The symbol $*_{x}$ denotes convolution over the variable (domain) x.

2 Radio Channel and System Model

The term "radio channel" does not have a clearly defined notion. It rather depends on the specific scientific area such as communications, wave propagation, Radar and so on what will be describe using the term "radio channel". Since the radio channel is a distributed system, a critical issue is the definition of the input and the output ports of the system. In this work the term "radio channel" will be used to describe the physical wave propagation between two points in space. In other words, the antennas or antenna arrays do not belong to the radio channel. This is mainly motivated by the fact that we can generally choose the antenna array within certain limits, but not the radio channel. This is true for radio channel measurements but also applies to mobile communications.

The focus of this chapter is the derivation of a mathematical framework to describe an observation of the radio channel. The algebraic data model will be the basis for all radio channel parameter estimators derived throughout this work. There are two contradicting design criteria for the data model. On one hand, it is desirable to make the data model as precise as possible since we want to investigate the radio channel. On the other hand, every measurement contains finite information. That means we have to choose a channel model that can be estimated from the radio channel measurements. The amount of information contained in a radio channel observation must be sufficient to determine the number of unknowns in the model. This is a significant difference to radio channel modelling in general. If we have to generate a realisation of the radio channel, e.g., for simulations, we can make the data model arbitrarily precise.

A generally accepted model for the radio channel describes the channel as a superposition of a finite number of rays. This model has been used for a long time for ray-tracing based radio channel simulations, but with the growing interest in the spatial structure of the radio channel, it has been also applied to channel modelling and channel measurement. The model is applicable for the synthesis of radio channel realisations. The only drawback is its high complexity if wideband radio channels have to be synthesized. As we shall see later, this model is not applicable as a data model for parameter estimation since it violates the aforementioned requirement on model complexity. In [8] has been shown that the radio channel contains concentrated propagation paths and so-called dense multipath components caused by *distributed* diffuse scattering. The ray model is appropriate to model the contribution of the concentrated (dominant) propagation paths. However, it is not suitable to model the contribution of the dense multipath components. In the next sections, we will derive an algebraic model for the concentrated propagation paths based on the ray-model. We extend this model with a new component describing the dense multipath components later on in this chapter.

2.1 Definition of a Ray

The smallest entity we will use to describe the radio channel is a *ray*. From ray-optical modelling it is well known that even complicated wave propagation phenomena in a continuum of

reflecting, diffracting, and scattering objects can be approximately modelled by a superposition of discrete waves.
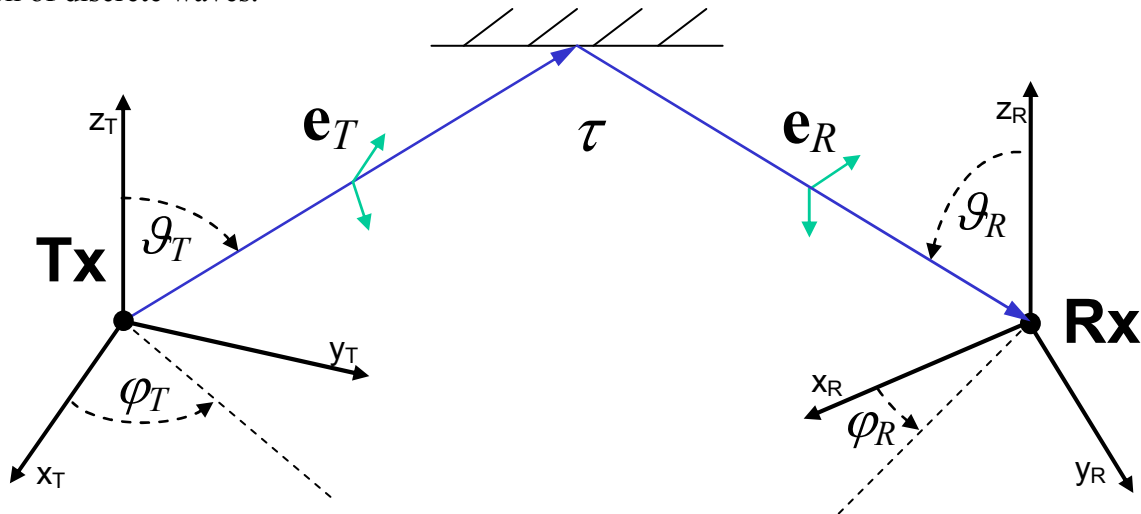
Figure 2-1:   Definition of a ray

The method is based on geometric optics and the number of rays used determines the accuracy. The five parameters φ_T, ϑ_T, τ, φ_R, and ϑ_R of a ray have the following meaning. If we generate a spherical wave at the transmitter position, only the part of the wave moving in the direction determined by the angle pair φ_T, ϑ_T will reach the receiver along this ray. We call these two angles the transmit-azimuth and -elevation. Due to the geometrical length of the ray and the propagation speeds within the mediums along the ray, the transmitted wave needs a certain time to reach the receiver over this ray. The total travelling time is called the time-delay of arrival (TDoA) τ. Finally, the angles φ_R, ϑ_R determine the azimuth and elevation of the approaching wave reaching the receiver along this ray.

Observe that the angles at the transmitter and at the receiver are defined in the local coordinate systems. There exist various advantages and disadvantages for the definition of the angles in the local coordinate systems and a global coordinate system. From a radio channel measurement point of view, the definition in the local Tx- and Rx-coordinate systems is preferable. This is mainly due to the fact, that the angle information is gathered by the *local* antenna arrays at Tx and Rx. Consequently, the angles of the rays are related to the local coordinate system. Furthermore, the ray parameters can always be projected into the global coordinate system if the absolute positions of Tx and Rx are known. A discussion about coordinate system definitions in radio channel modelling can be found in [5].
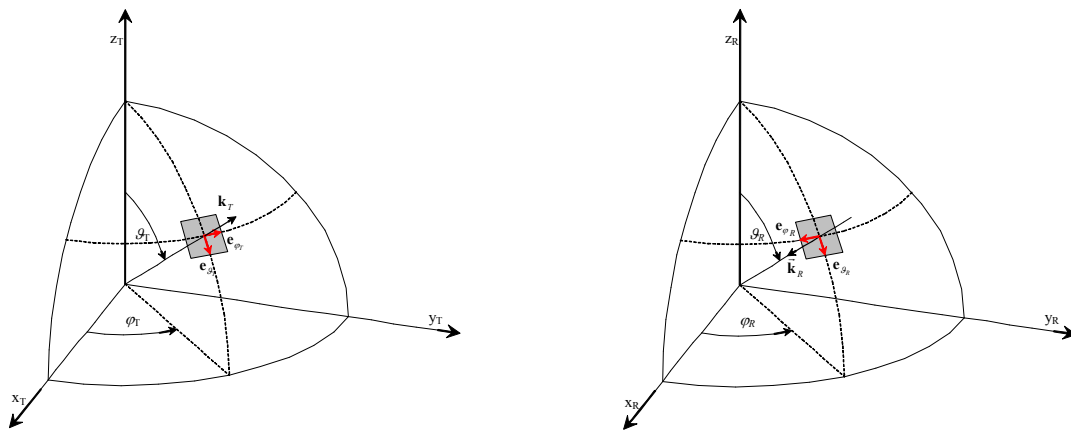
Figure 2-2:   Polarisation vectors on the transmit (left) and the receive site (right)

Using the transmit- and receive-angle pairs we define the Pointing-vectors at the transmitter and the receiver position, $\mathbf{k}_T$, and $\mathbf{k}_R$. Due to the energy transport, the Pointing-vector points away from the transmitter and towards the receiver. Using the definitions of the azimuth and elevation angles, as well as the Pointing-vectors, we define the two linear polarised components $\mathbf{e}_{\varphi_T}, \mathbf{e}_{\vartheta_T}$ and $\mathbf{e}_{\varphi_R}, \mathbf{e}_{\vartheta_R}$ of the wave leaving the transmitter and of the wave reaching the receiver, respectively. Figure 2-2 illustrates the definition of the linear polarised wave components. The ray projects the two transmitted linear polarised components onto the received linear polarised components.

2.2   Definition of a Propagation Path

The ray model derived in the previous section is suitable for the generation of a realisation of the stochastic process radio channel in simulations. It is also useful to understand wave propagation mechanisms. However, the ray model is not suitable to interpret the parameter estimation results of a channel parameter estimator. This is due to the limited resolution of any radio channel measurement system. We have to accept that the resolution of rays, which are close together, is always limited. The reason is as already stated before that we gather only finite information about the underlying physical phenomena while observing a system. Therefore, we introduce the *propagation path*. A propagation path is a superposition of multiple rays, which are close together, i.e., they from a cluster. Observe that we introduce the cluster concept in the context of channel parameter estimation, as a concept to interpret the estimation result. The parameterisation of a ray and a propagation path is identical. However, the ray model cannot be used to explain the observed time variance of the path weight (ray weight) if the structure of a cluster is changing. The time variance of the path weight of a propagation path can reveal, whether it is a superposition of multiple rays (cf. [9]).
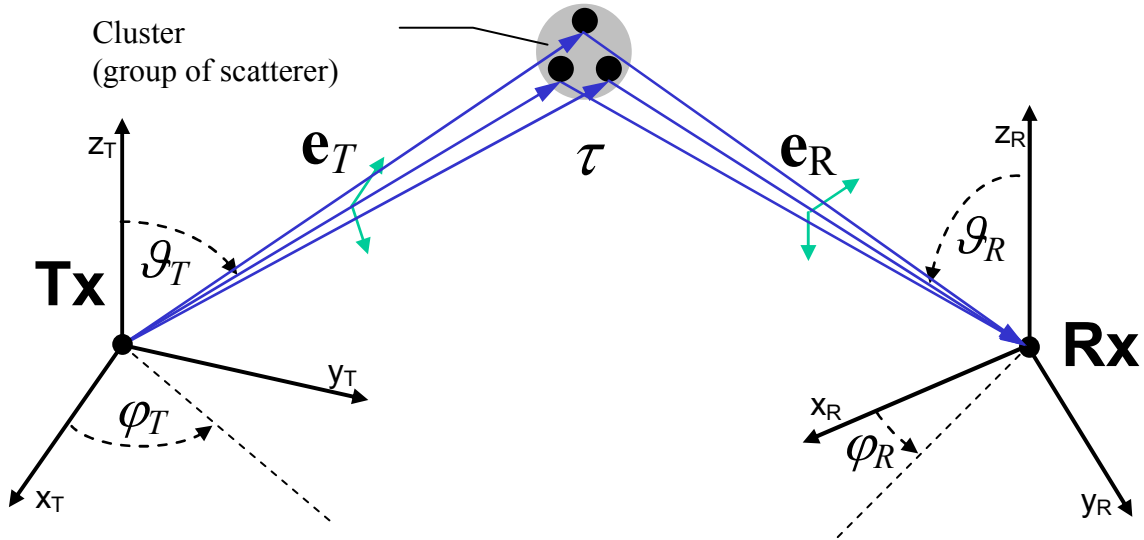
Figure 2-3:   Definition of a propagation path

The parameters of a propagation path are the mean values of the parameters of the rays combined in the propagation path. As discussed in the previous section, the radio channel can be approximated by the superposition of multipath components. Let us suppose we transmit a narrowband signal with the baseband representation $x(t)$ over a single propagation path with an electrical length of l_p at a carrier frequency f_c, using a transmitter with the impulse response $g_T(t)$, and a receiver having a impulse response of $g_R(t)$. Then[1]

---

[1] The symbol $*_t$ denotes convolution in the time domain.

$$y(t) = \gamma_p \cdot g_R(t) *_t g_T(t) *_t x\left(t - \frac{l_p}{c_0}\right) \cdot e^{-j2\pi f_c \frac{l_p}{c_0}}$$

is a baseband representation of the received signal $y(t)$, where c_0 denotes the velocity of light and γ_p describes all effects which can be treated as frequency independent such as free space loss, complex antenna gains, loss on scattering or reflection points, and so on. One should note that the term narrowband refers to a system with a small relative bandwidth. A radio system with a bandwidth of $B_m = 100\,\text{MHz}$ operating at a carrier frequency of $f_c = 5\,\text{GHz}$ is a narrowband system, and a radio system with a bandwidth of 1 GHz at a carrier frequency of 60 GHz is a narrowband system too. The equivalent frequency-domain representation of the received signal $y(t)$ is given by

$$Y(f) = X(f) \cdot \gamma_p \cdot G_{T_f}(f) \cdot G_{R_f}(f) \cdot e^{-j2\pi f \tau_p} \cdot e^{-j2\pi f_c \frac{l_p}{c_0}},$$

where $\tau_p = \frac{l_p}{c_0}$ denotes the time a signal needs for the transmission over the propagation path p. From a system theoretic point of view the term

$$H(f) = \gamma_p \cdot G_{T_f}(f) \cdot G_{R_f}(f) \cdot e^{-j2\pi f \tau_p} \cdot e^{-j2\pi f_c \frac{l_p}{c_0}}$$

is a frequency domain representation of a time invariant single-input single-output (SISO) system.

If we assume that at least one, the transmitter, the receiver, a reflector or scatterer related to this propagation path is moving, the SISO system becomes time-variant. We assume the effective speed v_p of the moving objects related to the propagation path including transmitter and receiver is constant within a short observation interval. Consequently, the time-variant frequency response can be expressed for a small change in the electrical length l_p by

$$H(f,t) = \gamma_p \cdot G_{T_f}(f) \cdot G_{R_f}(f) \cdot e^{-j2\pi \cdot f \cdot \tau_p} \cdot e^{-j2\pi \frac{(l_p + v_p t)}{\lambda_c}}, \tag{2.1}$$

where $\lambda_c = \frac{c_0}{f_c}$ denotes the wavelength of the carrier-frequency. This model is an approximation insofar as it neglects any change in the parameter τ_p and assumes a constant complex path-weight γ_p. Hence, it is only valid as long as $v_p \cdot t << \frac{c_0}{B_m}$, where B_m denotes the observation bandwidth of the measurement system. We can derive a more common expression for (2.1) if we introduce the Doppler-shift $\alpha_p = -\frac{f_c}{c_0} \cdot v_p$ and merge the phase shift at time $t = 0$ into the complex path weight

$$\gamma_p' = \gamma_p \cdot e^{-j2\pi f_c \frac{l_p}{c_0}},$$

yielding for (2.1) the form

$$H(f,t) = \gamma_p' \cdot G_{T_f}(f) \cdot G_{R_f}(f) \cdot e^{-j2\pi \cdot f \cdot \tau_p} \cdot e^{-j2\pi \cdot \alpha_p \cdot t}. \tag{2.2}$$

So far, we have derived a radio channel model for a SISO system with omni-directional antennas at both the transmitter and the receiver site, now we extend this model to directional antennas at both sides. To this end we introduce the complex beam pattern of the transmit antenna

$$\mathbf{b}_T(\varphi_T, \vartheta_T, f) = \left[b_{T_H}(\varphi_T, \vartheta_T, f) \; b_{T_V}(\varphi_T, \vartheta_T, f) \right] \in \mathbb{C}^{1 \times 2}$$

and the receive antenna

$$\mathbf{b}_R(\varphi_R,\vartheta_R,f)=\left[b_{R_H}(\varphi_R,\vartheta_R,f)\ b_{R_V}(\varphi_R,\vartheta_R,f)\right]\in\mathbb{C}^{1\times2},$$

where $b_{T_H}(\varphi_T,\vartheta_T,f)$, $b_{R_H}(\varphi_R,\vartheta_R,f)$ and $b_{T_V}(\varphi_T,\vartheta_T,f)$, $b_{R_V}(\varphi_R,\vartheta_R,f)$ are the complex frequency dependent beam pattern for horizontal and vertical polarisation, respectively. Assuming that the signal bandwidth is equal or smaller as the antenna bandwidth at f_c, we can write for the complex beam pattern of the transmit and the receive antenna

$$\mathbf{b}_T(\varphi_T,\vartheta_T)=\mathbf{b}_T(\varphi_T,\vartheta_T,f_c)\ \text{,and}\ \mathbf{b}_R(\varphi_R,\vartheta_R)=\mathbf{b}_R(\varphi_R,\vartheta_R,f_c).$$

Using these definitions the single-path SISO model (2.2) can be extended to directional antennas at both link ends

$$H(f,\varphi_T,\vartheta_T,\varphi_R,\vartheta_R,t)=G_{R_f}(f)\cdot\mathbf{b}_R(\varphi_R,\vartheta_R)\cdot\mathbf{\Gamma}_p\cdot\mathbf{b}_T(\varphi_T,\vartheta_T)^{\mathrm{T}}\cdot G_{T_f}(f)\cdot e^{-j2\pi\cdot f\cdot\tau_p}\cdot e^{-j2\pi\cdot\alpha_p\cdot t},\quad(2.3)$$

where the path weight matrix $\mathbf{\Gamma}_p$ contains the four polarimetric transmission coefficients

$$\mathbf{\Gamma}_p=\begin{bmatrix}\gamma_{HH,p}&\gamma_{VH,p}\\\gamma_{HV,p}&\gamma_{VV,p}\end{bmatrix}\in\mathbb{C}^{2\times2}.$$

Since the single-path SISO model is a linear model, it can be easily extended to the multipath SISO model using the superposition principle. For P propagation paths the multipath SISO radio channel model is given by

$$H(f,t)=G_{R_f}(f)\cdot G_{T_f}(f)\cdot\sum_{p=1}^{P}\left\{\mathbf{b}_R(\varphi_{R,p},\vartheta_{R,p})\cdot\mathbf{\Gamma}_p\cdot\mathbf{b}_T(\varphi_{T,p},\vartheta_{T,p})^{\mathrm{T}}\cdot e^{-j2\pi\cdot f\cdot\tau_p}\cdot e^{-j2\pi\cdot\alpha_p\cdot t}\right\}.\quad(2.4)$$

If we employ multiple antennas at both sides of the radio link, and use a multi-channel transmitter as well as a multi-channel receiver, to feed the transmit antennas and to receive multiple signals, we yield a MIMO communication system. We will denote the number of transmit antenna array ports of such a MIMO system with M_T and the number of the receive array ports with M_R. The MIMO channel is the entirety of all $M_T\times M_R$ SISO channels. The extension of the SISO radio channel model to the MIMO radio channel model is straightforward. To this end, let us introduce the matrix-valued functions

$$\mathbf{B}_T(\varphi_T,\vartheta_T)=\begin{bmatrix}\mathbf{b}_{T,1}(\varphi_T,\vartheta_T)\\\vdots\\\mathbf{b}_{T,M_T}(\varphi_T,\vartheta_T)\end{bmatrix},\quad\mathbb{R}^2\mapsto\mathbb{C}^{M_T\times2},\quad(2.5)$$

and

$$\mathbf{B}_R(\varphi_R,\vartheta_R)=\begin{bmatrix}\mathbf{b}_{R,1}(\varphi_R,\vartheta_R)\\\vdots\\\mathbf{b}_{R,M_R}(\varphi_R,\vartheta_R)\end{bmatrix},\quad\mathbb{R}^2\mapsto\mathbb{C}^{M_R\times2}.\quad(2.6)$$

$\mathbf{B}_T(\varphi_T,\vartheta_T)$, and $\mathbf{B}_R(\varphi_R,\vartheta_R)$ describe the far field beam pattern of all antenna array ports at the transmit array and the receive array, respectively. Using the definitions (2.5) and (2.6) in the model for the SISO channel (2.4) yields an expression for the MIMO radio channel $\mathbf{H}(f,t)\in\mathbb{C}^{M_R\times M_T}$ (2.7).

$$\mathbf{H}(f,t)=G_{R_f}(f)\cdot G_{T_f}(f)\cdot\sum_{p=1}^{P}\left\{\mathbf{B}_R(\varphi_{R,p},\vartheta_{R,p})\cdot\mathbf{\Gamma}_p\cdot\mathbf{B}_T(\varphi_{T,p},\vartheta_{T,p})^{\mathrm{T}}\cdot e^{-j2\pi\cdot f\cdot\tau_p}\cdot e^{-j2\pi\cdot\alpha_p\cdot t}\right\}\quad(2.7)$$

There exist two additional systems namely the SIMO radio system (single-input multiple-output = SIMO) and the MISO radio system (multiple-input single-output = MISO). From a systematic point of view the SISO, the SIMO, and the MISO radio system are only special MIMO systems. For example, one can always reduce the channel model for the MIMO system (2.7) to the channel model $\mathbf{h}_{SIMO}(f,t) \in \mathbb{C}^{M_R \times 1}$ of a SIMO system yielding

$$\mathbf{h}_{SIMO}(f,t) = G_{R_f}(f) \cdot G_{T_f}(f) \cdot \sum_{p=1}^{P} \left\{ \mathbf{B}_R(\varphi_{R,p}, \vartheta_{R,p}) \cdot \mathbf{\Gamma}_p \cdot \mathbf{b}_T(\varphi_{T,p}, \vartheta_{T,p})^{\mathrm{T}} \cdot e^{-j2\pi \cdot f \cdot \tau_p} \cdot e^{-j2\pi \cdot \alpha_p \cdot t} \right\}$$

or equivalently to the model $\mathbf{h}_S(f,t) \in \mathbb{C}^{1 \times M_T}$ of a MISO system

$$\mathbf{h}_{MISO}(f,t) = G_{R_f}(f) \cdot G_{T_f}(f) \cdot \sum_{p=1}^{P} \left\{ \mathbf{b}_R(\varphi_{R,p}, \vartheta_{R,p}) \cdot \mathbf{\Gamma}_p \cdot \mathbf{B}_T(\varphi_{T,p}, \vartheta_{T,p})^{\mathrm{T}} \cdot e^{-j2\pi \cdot f \cdot \tau_p} \cdot e^{-j2\pi \cdot \alpha_p \cdot t} \right\}.$$

2.3 Frequency and Temporal Domain Sampling

To use the MIMO radio channel model, derived in Section 2.2, as a model for radio channel parameter estimation or as a channel model for digital communications, we have to sample the continuous model (2.4) in the temporal as well as in the frequency domain. To this end, we introduce the sampling interval in the frequency domain f_0, and in the temporal domain t_0. Equidistant frequency domain sampling of (2.7) at M_f frequency points yields a set of M_f matrices $\mathbf{H}(m_f \cdot f_0, t) \in \mathbb{C}^{M_R \times M_T}$ where $m_f = -\frac{M_f - 1}{2}, \ldots, +\frac{M_f - 1}{2}$. Using these matrices the broadband MIMO radio channel model can be expressed in matrix form as follows

$$\mathbf{H}_f(t) = \begin{bmatrix} \mathbf{H}\left(-\frac{M_f-1}{2} f_0, t\right) & 0 & \cdots & & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & 0 \\ 0 & \cdots & 0 & \mathbf{H}\left(+\frac{M_f-1}{2} f_0, t\right) \end{bmatrix} \in \mathbb{C}^{(M_f M_R) \times (M_f M_T)}. \tag{2.8}$$

In the same way temporal domain sampling of $\mathbf{H}_f(t)$ at M_t time instances yields a set of matrices

$$\mathbf{H}_f(m_t \cdot t_0) \in \mathbb{C}^{(M_R M_f) \times (M_T M_f)}.$$

It should be observed that (2.8) covers a bandwidth of $B_m = f_0 \cdot M_f$, and also that the set of M_t matrices $\mathbf{H}_f(m_t \cdot t_0)$ represent the radio channel within the time interval $T_m = M_t \cdot t_0$.

2.3.1 Expressions for the MIMO-Channel

In the previous section, we have shown that the sampled radio channel model can be expressed at one time instance using a block diagonal matrix (2.8). To derive a more explicit expression let us introduce some matrices containing basic elements of the radio channel model. First we define the parameter vectors containing the model parameters summarised in Table 2-1.

Using those parameter vectors we define four matrix valued functions,

$$\mathbf{B}_{T_H}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) = \begin{bmatrix} b_{T_H,1}(\varphi_{T,1}, \vartheta_{T,1}) & \cdots & b_{T_H,1}(\varphi_{T,P}, \vartheta_{T,P}) \\ \vdots & & \vdots \\ b_{T_H,M_T}(\varphi_{T,1}, \vartheta_{T,1}) & \cdots & b_{T_H,M_T}(\varphi_{T,P}, \vartheta_{T,P}) \end{bmatrix}, \quad \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_T \times P}, \tag{2.9}$$

Table 2-1: Definition of Propagation Path Parameter Vectors

Angles at the transmitter site:	$\boldsymbol{\varphi}_T = \begin{bmatrix} \varphi_{T,1} & \cdots & \varphi_{T,P} \end{bmatrix}^{\mathrm{T}}, \; \boldsymbol{\vartheta}_T = \begin{bmatrix} \vartheta_{T,1} & \cdots & \vartheta_{T,P} \end{bmatrix}^{\mathrm{T}}$
Angles at the receiver site:	$\boldsymbol{\varphi}_R = \begin{bmatrix} \varphi_{R,1} & \cdots & \varphi_{R,P} \end{bmatrix}^{\mathrm{T}}, \; \boldsymbol{\vartheta}_R = \begin{bmatrix} \vartheta_{R,1} & \cdots & \vartheta_{R,P} \end{bmatrix}^{\mathrm{T}}$
Time delays:	$\boldsymbol{\tau} = \begin{bmatrix} \tau_1 & \cdots & \tau_P \end{bmatrix}^{\mathrm{T}}$
Doppler-shifts:	$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_P \end{bmatrix}^{\mathrm{T}}$
Complex path weights:	$\boldsymbol{\gamma}_{HH} = \begin{bmatrix} \gamma_{HH,1} & \cdots & \gamma_{HH,P} \end{bmatrix}^{\mathrm{T}}$ $\boldsymbol{\gamma}_{HV} = \begin{bmatrix} \gamma_{HV,1} & \cdots & \gamma_{HV,P} \end{bmatrix}^{\mathrm{T}}$ $\boldsymbol{\gamma}_{VH} = \begin{bmatrix} \gamma_{VH,1} & \cdots & \gamma_{VH,P} \end{bmatrix}^{\mathrm{T}}$ $\boldsymbol{\gamma}_{VV} = \begin{bmatrix} \gamma_{VV,1} & \cdots & \gamma_{VV,P} \end{bmatrix}^{\mathrm{T}}$

$$\mathbf{B}_{T_V}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) = \begin{bmatrix} b_{T_V,1}(\varphi_{T,1}, \vartheta_{T,1}) & \cdots & b_{T_V,1}(\varphi_{T,P}, \vartheta_{T,P}) \\ \vdots & & \vdots \\ b_{T_V,M_T}(\varphi_{T,1}, \vartheta_{T,1}) & \cdots & b_{T_V,M_T}(\varphi_{T,P}, \vartheta_{T,P}) \end{bmatrix}, \; \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_T \times P}, \qquad (2.10)$$

$$\mathbf{B}_{R_H}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) = \begin{bmatrix} b_{R_H,1}(\varphi_{R,1}, \vartheta_{R,1}) & \cdots & b_{R_H,1}(\varphi_{R,P}, \vartheta_{R,P}) \\ \vdots & & \vdots \\ b_{R_H,M_T}(\varphi_{R,1}, \vartheta_{R,1}) & \cdots & b_{R_H,M_T}(\varphi_{R,P}, \vartheta_{R,P}) \end{bmatrix}, \; \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_R \times P}, \qquad (2.11)$$

and

$$\mathbf{B}_{R_V}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) = \begin{bmatrix} b_{R_V,1}(\varphi_{R,1}, \vartheta_{R,1}) & \cdots & b_{R_V,1}(\varphi_{R,P}, \vartheta_{R,P}) \\ \vdots & & \vdots \\ b_{R_V,M_T}(\varphi_{R,1}, \vartheta_{R,1}) & \cdots & b_{R_V,M_T}(\varphi_{R,P}, \vartheta_{R,P}) \end{bmatrix}, \; \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_R \times P}. \qquad (2.12)$$

to describe the mapping of the parameter vectors $\boldsymbol{\varphi}_T$, $\boldsymbol{\vartheta}_T$ to the transmit array responses as well as the mapping of the parameter vectors $\boldsymbol{\varphi}_R$, $\boldsymbol{\vartheta}_R$ to the receive array responses for vertical and horizontal polarisation. Using (2.9) - (2.12) in representation (2.7) yields for $f = 0$ and $t = 0$

$$\begin{aligned} \mathbf{H}(0,0) = G_{T_f}(0) \cdot G_{R_f}(0) \cdot \big(&\mathbf{B}_{R_H}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) \cdot \mathrm{diag}\{\boldsymbol{\gamma}_{HH}\} \cdot \mathbf{B}_{T_H}^{\mathrm{T}}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) + \\ &\mathbf{B}_{R_H}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) \cdot \mathrm{diag}\{\boldsymbol{\gamma}_{VH}\} \cdot \mathbf{B}_{T_V}^{\mathrm{T}}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) + \\ &\mathbf{B}_{R_V}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) \cdot \mathrm{diag}\{\boldsymbol{\gamma}_{HV}\} \cdot \mathbf{B}_{T_H}^{\mathrm{T}}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) + \\ &\mathbf{B}_{R_V}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) \cdot \mathrm{diag}\{\boldsymbol{\gamma}_{VV}\} \cdot \mathbf{B}_{T_V}^{\mathrm{T}}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) \big). \end{aligned} \qquad (2.13)$$

For notational convenience let us drop the dependency of the matrices $\mathbf{B}$ from the respective parameters $\boldsymbol{\varphi}, \boldsymbol{\vartheta}$. To express the frequency dependence of (2.7) we define another fundamental matrix valued function $\mathbf{A}_\tau(\boldsymbol{\tau})$, mapping the time delays $\boldsymbol{\tau}$ to the related complex exponentials $e^{-j2\pi \cdot m f_0 \cdot \tau_P}$ in the frequency domain

$$\mathbf{A}_\tau(\boldsymbol{\tau}) = \begin{bmatrix} e^{-j2\pi\left(-\frac{M_f-1}{2}\right)\cdot f_0 \cdot \tau_1} & \cdots & e^{-j2\pi\left(-\frac{M_f-1}{2}\right)\cdot f_0 \cdot \tau_P} \\ \vdots & & \vdots \\ e^{-j2\pi\left(+\frac{M_f-1}{2}\right)\cdot f_0 \cdot \tau_1} & \cdots & e^{-j2\pi\left(+\frac{M_f-1}{2}\right)\cdot f_0 \cdot \tau_P} \end{bmatrix} \in \mathbb{C}^{M_f \times P}. \tag{2.14}$$

Again, we drop the dependence of $\mathbf{A}_\tau(\boldsymbol{\tau})$ on $\boldsymbol{\tau}$ and use the abbreviation $\mathbf{A}_\tau$. Furthermore, let us introduce two diagonal matrices

$$\mathbf{G}_{T_f} = \mathrm{diag}\left\{\left[G_{T_f}\left(-\tfrac{M_f-1}{2}f_0\right) \quad \cdots \quad G_{T_f}\left(+\tfrac{M_f-1}{2}f_0\right)\right]\right\} \in \mathbb{C}^{M_f \times M_f} \tag{2.15}$$

and

$$\mathbf{G}_{R_f} = \mathrm{diag}\left\{\left[G_{R_f}\left(-\tfrac{M_f-1}{2}f_0\right) \quad \cdots \quad G_{R_f}\left(+\tfrac{M_f-1}{2}f_0\right)\right]\right\} \in \mathbb{C}^{M_f \times M_f} \tag{2.16}$$

containing the sampled frequency responses of the transmitter and the receiver respectively. Using the definitions (2.14) - (2.16) in the broadband MIMO radio channel model (2.8) yields for time $t = 0$ the expression

$$\begin{aligned} \mathbf{H}_f(0) = &\left(\mathbf{G}_{R_f} \otimes \mathbf{B}_{R_H}\right) \cdot \left(\mathbf{I} \otimes \mathrm{diag}\{\boldsymbol{\gamma}_{HH}\}\right) \cdot \mathrm{diag}\{\mathrm{vec}\{\mathbf{A}_\tau^{\mathrm{T}}\}\} \cdot \left(\mathbf{G}_{T_f} \otimes \mathbf{B}_{T_H}\right)^{\mathrm{T}} + \\ &\left(\mathbf{G}_{R_f} \otimes \mathbf{B}_{R_V}\right) \cdot \left(\mathbf{I} \otimes \mathrm{diag}\{\boldsymbol{\gamma}_{HV}\}\right) \cdot \mathrm{diag}\{\mathrm{vec}\{\mathbf{A}_\tau^{\mathrm{T}}\}\} \cdot \left(\mathbf{G}_{T_f} \otimes \mathbf{B}_{T_H}\right)^{\mathrm{T}} + \\ &\left(\mathbf{G}_{R_f} \otimes \mathbf{B}_{R_H}\right) \cdot \left(\mathbf{I} \otimes \mathrm{diag}\{\boldsymbol{\gamma}_{VH}\}\right) \cdot \mathrm{diag}\{\mathrm{vec}\{\mathbf{A}_\tau^{\mathrm{T}}\}\} \cdot \left(\mathbf{G}_{T_f} \otimes \mathbf{B}_{T_V}\right)^{\mathrm{T}} + \\ &\left(\mathbf{G}_{R_f} \otimes \mathbf{B}_{R_V}\right) \cdot \left(\mathbf{I} \otimes \mathrm{diag}\{\boldsymbol{\gamma}_{VV}\}\right) \cdot \mathrm{diag}\{\mathrm{vec}\{\mathbf{A}_\tau^{\mathrm{T}}\}\} \cdot \left(\mathbf{G}_{T_f} \otimes \mathbf{B}_{T_V}\right)^{\mathrm{T}}. \end{aligned} \tag{2.17}$$

For the time being let us ignore the time dependence of $\mathbf{H}_f(t)$. As already discussed $\mathbf{H}_f$ is a sparse matrix. Since only the non-zero elements of $\mathbf{H}_f$ are of interest, we choose a different representation. We sort the elements of the M_f block matrices $\mathbf{H}(m_f \cdot f_0) \in \mathbb{C}^{M_R \times M_T}$ into a vector

$$\mathbf{s} = \left[\mathbf{z}_{1,1}, \ldots, \mathbf{z}_{M_R,1}, \mathbf{z}_{1,2}, \ldots, \mathbf{z}_{M_R,2}, \mathbf{z}_{1,3}, \ldots, \mathbf{z}_{M_R,M_T-1}, \mathbf{z}_{1,M_T}, \ldots, \mathbf{z}_{M_R,M_T}\right]^{\mathrm{T}} \in \mathbb{C}^{M_f M_R M_T \times 1}, \tag{2.18}$$

where

$$\mathbf{z}_{jk} = \left[\mathbf{H}\left(-\left(\tfrac{M_f-1}{2}\right)\cdot f_0\right)_{j,k} \quad \cdots \quad \mathbf{H}\left(+\left(\tfrac{M_f-1}{2}\right)\cdot f_0\right)_{j,k}\right]$$

is a vector containing all frequency domain samples for one transmit port - receive port combination of the MIMO system. Using the elementary functions $\mathbf{A}_\tau$, $\mathbf{B}_{T_H}, \mathbf{B}_{T_V}$, and $\mathbf{B}_{R_H}, \mathbf{B}_{R_V}$ the vector defined in (2.18) can be calculated from the model parameters as follows[2]

---

[2] Observe the following properties of the $\mathrm{vec}\{\bullet\}$ operator, the Schur $\circ$ and the Khatri-Rao $\Diamond$ product:
$\mathrm{vec}\{\mathbf{B} \cdot \mathrm{diag}\{\mathbf{c}\} \cdot \mathbf{A}^{\mathrm{T}}\} = (\mathbf{A} \Diamond \mathbf{B}) \cdot \mathbf{c}$, $(\mathbf{D} \Diamond \mathbf{B} \Diamond \mathbf{a}^{\mathrm{T}}) \cdot \mathbf{c} = (\mathbf{D} \Diamond \mathbf{B}) \cdot (\mathbf{a} \circ \mathbf{c})$, $\mathbf{B} \cdot \mathrm{diag}\{\mathbf{a} \circ \mathbf{c}\} \cdot \mathbf{D}^{\mathrm{T}} = \mathbf{B} \cdot \mathrm{diag}\{\mathbf{c}\} \cdot \mathrm{diag}\{\mathbf{a}\} \cdot \mathbf{D}^{\mathrm{T}}$

$$\mathbf{s} = \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right)\right) \cdot \boldsymbol{\gamma}_{HH} + \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right)\right) \cdot \boldsymbol{\gamma}_{HV} +$$
$$\left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right)\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right)\right) \cdot \boldsymbol{\gamma}_{VV} , \tag{2.19}$$

where $\mathbf{G}_{S_f} = \mathbf{G}_{T_f} \cdot \mathbf{G}_{R_f}$ is a diagonal matrix and $\Diamond$ denotes the Khatri-Rao product (column wise Kronecker product). Equation (2.19) can be rewritten to

$$\mathbf{s} = \left(\mathbf{I} \otimes \mathbf{G}_{S_f}\right) \cdot \left\{ \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{A}_\tau\right) \cdot \boldsymbol{\gamma}_{HH} + \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{A}_\tau\right) \cdot \boldsymbol{\gamma}_{HV} + \right.$$
$$\left. \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{A}_\tau\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{A}_\tau\right) \cdot \boldsymbol{\gamma}_{VV} \right\}.$$

One should observe, that the ordering of the elements in $\mathbf{s}$ is arbitrary, we can as well define a vector

$$\mathbf{s}' = \left[\text{vec}\left\{ \mathbf{H}\left(-\tfrac{(M_f-1)}{2} f_0\right) \right\}^{\mathrm{T}} \cdots \text{vec}\left\{ \mathbf{H}\left(+\tfrac{(M_f-1)}{2} f_0\right) \right\}^{\mathrm{T}} \right]^{\mathrm{T}},$$

which may be written as

$$\mathbf{s}' = \left(\left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right) \Diamond \mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H}\right) \cdot \boldsymbol{\gamma}_{HH} + \left(\left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right) \Diamond \mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V}\right) \cdot \boldsymbol{\gamma}_{HV} +$$
$$\left(\left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right) \Diamond \mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H}\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau\right) \Diamond \mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V}\right) \cdot \boldsymbol{\gamma}_{VV} . \tag{2.20}$$

This result can actually be generalised, a reordering of the basis functions $\mathbf{A}_\tau$, $\mathbf{B}_{T_H}, \mathbf{B}_{T_V}$, and $\mathbf{B}_{R_H}, \mathbf{B}_{R_V}$ lead to a permutation of the elements in $\mathbf{s}$ only. Equations (2.19), (2.20) motivate the definition of another basis function

$$\mathbf{B}_f(\boldsymbol{\tau}) = \mathbf{G}_{S_f} \cdot \mathbf{A}_\tau(\boldsymbol{\tau}). \tag{2.21}$$

Again, we drop the dependency of $\mathbf{B}_f(\boldsymbol{\tau})$ on the parameter vector $\boldsymbol{\tau}$, writing only $\mathbf{B}_f$. Using the definition (2.21) in equation (2.19) yields

$$\mathbf{s} = \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f\right) \cdot \boldsymbol{\gamma}_{HH} + \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f\right) \cdot \boldsymbol{\gamma}_{HV} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f\right) \cdot \boldsymbol{\gamma}_{VV} . \tag{2.22}$$

To complete the MIMO system model, we introduce the matrix valued function

$$\mathbf{A}_\alpha(\boldsymbol{\alpha}) = \begin{bmatrix} e^{-\mathrm{j}2\pi\left(-\frac{M_t-1}{2}\right)t_0 \cdot \alpha_1} & \cdots & e^{-\mathrm{j}2\pi\left(-\frac{M_t-1}{2}\right)t_0 \cdot \alpha_P} \\ \vdots & & \vdots \\ e^{-\mathrm{j}2\pi\left(+\frac{M_t-1}{2}\right)t_0 \cdot \alpha_1} & \cdots & e^{-\mathrm{j}2\pi\left(+\frac{M_t-1}{2}\right)t_0 \cdot \alpha_P} \end{bmatrix} \in \mathbb{C}^{M_t \times P}, \tag{2.23}$$

mapping the Doppler-shifts $\boldsymbol{\alpha}$ to the phase-shifts of the path weights at each time instance $m_t \cdot t_0, \forall m_t = -\frac{M_t-1}{2}, \ldots, +\frac{M_t-1}{2}$. For the sake of completeness, we furthermore introduce the related matrix valued function

$$\mathbf{B}_t(\boldsymbol{\alpha}) = \mathbf{G}_t \cdot \mathbf{A}_\alpha(\boldsymbol{\alpha}). \tag{2.24}$$

One should note that if the transmitter as well as the receiver is a time invariant system what is usually the case, the matrix $\mathbf{G}_t$ is an identity matrix $\mathbf{G}_t = \mathbf{I} \in \mathbb{R}^{M_t \times M_t}$. From (2.7) we can deduce that for time $m_t \cdot t_0, \forall m_t = -\frac{M_t-1}{2}, \ldots, +\frac{M_t-1}{2}$ the expression for the MIMO system

$$\mathbf{s}(m_t \cdot t_0) = \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f\right) \cdot \left(\mathbf{z}(m_t) \circ \boldsymbol{\gamma}_{HH}\right) + \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f\right) \cdot \left(\mathbf{z}(m_t) \circ \boldsymbol{\gamma}_{HV}\right) +$$
$$\left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f\right) \cdot \left(\mathbf{z}(m_t) \circ \boldsymbol{\gamma}_{VH}\right) + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f\right) \cdot \left(\mathbf{z}(m_t) \circ \boldsymbol{\gamma}_{VV}\right),$$

with

$$\mathbf{z}(m_t) = \left[e^{-j2\pi \cdot m_t \cdot t_0 \cdot \alpha_1} \cdots e^{-j2\pi \cdot m_t \cdot t_0 \cdot \alpha_P} \right]^{\mathrm{T}} \in \mathbb{C}^{P \times 1}$$

holds. Observing, that $\mathbf{A}_\alpha$ and all vectors $\mathbf{z}(m_t)$, $m_t = -\frac{M_t-1}{2}, \ldots, +\frac{M_t-1}{2}$ are related in the following way

$$\mathbf{A}_\alpha = \begin{bmatrix} \mathbf{z}\left(-\frac{M-1}{2}\right) \\ \vdots \\ \mathbf{z}\left(-\frac{M+1}{2}\right) \end{bmatrix}$$

we extend the MIMO system model (2.22) to[3]

$$\begin{aligned}
\mathbf{s} = & \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{HH} + \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{HV} + \\
& \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{VV} .
\end{aligned} \tag{2.25}$$

If the antenna arrays at both link ends can only radiate and receive waves with horizontal polarisation, the MIMO system model (2.25) reduces to

$$\mathbf{s}_{HH} = \mathbf{B}_{R_H} \Diamond \mathbf{B}_{T_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH} . \tag{2.26}$$

If the transmit antenna array radiates vertical polarised waves only, we have to reduce the MIMO system model (2.25) to

$$\mathbf{s}_{VX} = \mathbf{B}_{R_H} \Diamond \mathbf{B}_{T_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \cdot \boldsymbol{\gamma}_{VH} + \mathbf{B}_{R_V} \Diamond \mathbf{B}_{T_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \cdot \boldsymbol{\gamma}_{VV} , \quad X \in [H\ V]. \tag{2.27}$$

Noticing that equation (2.25) describes essentially the mapping of the parameter vectors $\boldsymbol{\tau}$, $\boldsymbol{\varphi}_T$, $\boldsymbol{\vartheta}_T$, $\boldsymbol{\varphi}_R$, $\boldsymbol{\vartheta}_R$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}_{HH}$, $\boldsymbol{\gamma}_{HV}$, $\boldsymbol{\gamma}_{VH}$, and $\boldsymbol{\gamma}_{VV}$ to $\mathbf{s}$, we define the parameter vector

$$\boldsymbol{\theta}_{sp} = \left[\boldsymbol{\alpha}^{\mathrm{T}}\ \boldsymbol{\tau}^{\mathrm{T}}\ \boldsymbol{\varphi}_T^{\mathrm{T}}\ \boldsymbol{\vartheta}_T^{\mathrm{T}}\ \boldsymbol{\varphi}_R^{\mathrm{T}}\ \boldsymbol{\vartheta}_R^{\mathrm{T}}\ \Re\{\boldsymbol{\gamma}_{HH}^{\mathrm{T}}\}\ \Im\{\boldsymbol{\gamma}_{HH}^{\mathrm{T}}\}\ \Re\{\boldsymbol{\gamma}_{HV}^{\mathrm{T}}\}\ \Im\{\boldsymbol{\gamma}_{HV}^{\mathrm{T}}\}\ \Re\{\boldsymbol{\gamma}_{VH}^{\mathrm{T}}\}\ \Im\{\boldsymbol{\gamma}_{VH}^{\mathrm{T}}\}\ \Re\{\boldsymbol{\gamma}_{VV}^{\mathrm{T}}\}\ \Im\{\boldsymbol{\gamma}_{VV}^{\mathrm{T}}\}\right]^{\mathrm{T}} \tag{2.28}$$

and use hereinafter the notation

$$\begin{aligned}
\mathbf{s}(\boldsymbol{\theta}_{sp}) = & \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{HH} + \left(\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{HV} + \\
& \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right) \cdot \boldsymbol{\gamma}_{VV} ,
\end{aligned} \tag{2.29}$$

to state explicitly the dependence of $\mathbf{s}$ on the model parameters $\boldsymbol{\theta}_{sp}$. The data model for the concentrated propagation paths can also be expressed as

$$\mathbf{s}(\boldsymbol{\theta}_{sp}) = \left[\mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \quad \mathbf{B}_{T_H} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \quad \mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \quad \mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t\right] \cdot \begin{bmatrix} \boldsymbol{\gamma}_{HH} \\ \boldsymbol{\gamma}_{HV} \\ \boldsymbol{\gamma}_{VH} \\ \boldsymbol{\gamma}_{VV} \end{bmatrix} \tag{2.30}$$

without loss of generality. We can conclude that the structure of the data model has always the same structure. I.e. the expressions (2.25), (2.26), (2.27), (2.29) fit into the same model

---

[3] Observe the following properties of the Schur $\circ$ and the Khatri-Rao $\Diamond$ product: $\left(\mathbf{D} \Diamond \mathbf{B} \Diamond \mathbf{a}^{\mathrm{T}}\right) \cdot \mathbf{c} = \left(\mathbf{D} \Diamond \mathbf{B}\right) \cdot \left(\mathbf{a} \circ \mathbf{c}\right)$

$$\mathbf{s}(\boldsymbol{\theta}_{sp}) = \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma} \, , \qquad\qquad\qquad (2.31)$$

with $\mathbf{B}(\boldsymbol{\mu}) \in \mathbb{C}^{M \times 4P}$ and $\boldsymbol{\gamma} \in \mathbb{C}^{4P \times 1}$

The matrix valued function $\mathbf{B}(\boldsymbol{\mu})$ is a description of the structure of the radio channel. Hence, we denote the nonlinear parameters $\boldsymbol{\mu}$ also as structural parameters. The coefficients $\boldsymbol{\gamma}$ are linear weights in the model. We will refer to them also as linear parameters. In the following, we use expression (2.31) as a representative for all models of the concentrated propagation paths if possible.

2.4   Data Models for Antenna Arrays

The structure of the data model expressing the beam pattern of the elements of an antenna array restricts the applicable parameter estimation algorithms. In the following section, general concepts to model the beam patterns of antenna arrays as a function of the angle (azimuth and elevation) of incoming waves are summarised. Unless noted otherwise, we assume plane waves. In principle, incoming waves can be treated as plane waves if the approximation error of the antenna array model, due to the wave curvature, is smaller than other errors and uncertainties of the channel model as a whole.

From a parameter estimation point of view, a data model for antenna arrays should have a low computational complexity. The data model should provide valid results for the complete angular domain, i.e., it should express the beam pattern as a continuous function of azimuth and elevation. Finally, the data model should be continuously differentiable. This is a prerequisite for the computation of the Cramér-Rao lower bound of channel model parameters, or for the application of gradient-based parameter estimation algorithms.

One physical antenna element may have two ports to measure both horizontal and vertical polarisations. We will use the term array element for the physical antenna element and the term antenna element port for the electrical antenna element. The array response is a two dimensional function $\mathbf{B}(\varphi, \vartheta)$ of the variables φ-azimuth and ϑ-elevation. Furthermore, if we excite an antenna array port we generate two field components in the far field, namely a horizontal $\mathbf{e}_{\varphi}$ and vertical $\mathbf{e}_{\vartheta}$ polarised component. Since an antenna array is a passive system, we can use the same function to describe the array response to a far field source with a horizontal as well as a vertical polarised component. There exist three different approaches to describe the relationship between the array response (beam pattern, array manifold) and a far field point source.

2.4.1   Stored Beam-Pattern

A straightforward way to model an antenna array response to a far field source is to measure the directional characteristics of the antenna array in an anechoic chamber and to store the data. This model has two advantages, is its simple and it can be used to describe all available antenna arrays. However, this model has also some drawbacks. The function $\mathbf{B}(\varphi, \vartheta)$ is continuous in the parameters φ, and ϑ, but we can only store a sampled version of $\mathbf{B}(\varphi, \vartheta)$. The storage of the sampled array responses requires a large amount of memory and if we need an array response for an angle pair that has not been measured, we have to interpolate it using some interpolation algorithm on the stored data. Furthermore, since the array response is not expressed in an algebraic form the derivatives with respect to the parameters φ and ϑ

$$\tfrac{\partial}{\partial \varphi} \mathbf{B}(\varphi, \vartheta), \text{ and } \tfrac{\partial}{\partial \vartheta} \mathbf{B}(\varphi, \vartheta),$$

necessary for some parameter estimation algorithms (see Section 5.2) and for the calculation of the Cramér-Rao lower bound (cf. Section 4.1), cannot be calculated directly from the stored data.

2.4.2 Array Response Factorisation

The second class of antenna array models decompose the array beam pattern $\mathbf{b}_{com}(\varphi, \vartheta)$ into an element beam pattern shared by all elements and a phase vector $\mathbf{a}_{struct}(\varphi, \vartheta)$ relating the position of the individual elements within the array to the phases of the array response. The model describes the complete array response of an ideal antenna array by

$$\mathbf{b}_{ideal}(\varphi, \vartheta) = \mathbf{a}_{struct}(\varphi, \vartheta) \cdot \mathbf{b}_{com}(\varphi, \vartheta). \tag{2.32}$$

Since real antenna arrays are subjected to errors such as mutual antenna element coupling, we introduce a square matrix $\mathbf{K}$ and approximate the real antenna array response with

$$\mathbf{b}_{real}(\varphi, \vartheta) = \mathbf{K} \cdot \mathbf{a}_{struct}(\varphi, \vartheta) \cdot \mathbf{b}_{com}(\varphi, \vartheta). \tag{2.33}$$

The most prominent representative belonging to this class of array models is the uniform linear array (ULA) model. The phase vector for an ULA row, having element spacing d_r, is given by

$$\mathbf{a}_{row}(\mu^{(1)}) = \mathbf{a}_{ULA}(\mu^{(1)}) = \left[e^{+j\left(\frac{M_1-1}{2}\right)\mu^{(1)}} \quad \cdots \quad e^{-j\left(\frac{M_1-1}{2}\right)\mu^{(1)}} \right]^{\mathrm{T}}. \tag{2.34}$$

The spatial frequency $\mu^{(1)}$ is defined as $\mu^{(1)} = 2\pi \frac{d_r}{\lambda} \cos(\varphi)\sin(\vartheta)$. The phase vector for a ULA column having element spacing d_c can be expressed equivalently with

$$\mathbf{a}_{col}(\mu^{(2)}) = \mathbf{a}_{ULA}(\mu^{(2)}) = \left[e^{+j\left(\frac{M_2-1}{2}\right)\mu^{(2)}} \quad \cdots \quad e^{-j\left(\frac{M_2-1}{2}\right)\mu^{(2)}} \right]^{\mathrm{T}}, \tag{2.35}$$

where $\mu^{(2)} = 2\pi \frac{d_c}{\lambda} \sin(\vartheta)$ is again the related spatial frequency. Since a uniform rectangular array (URA), can be understood as a set of uniform linear arrays in a row or a column the related phase vector can be easily derived from (2.34) and (2.35) yielding
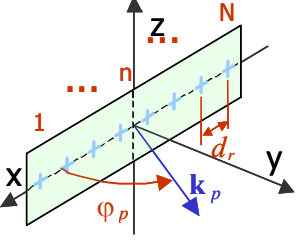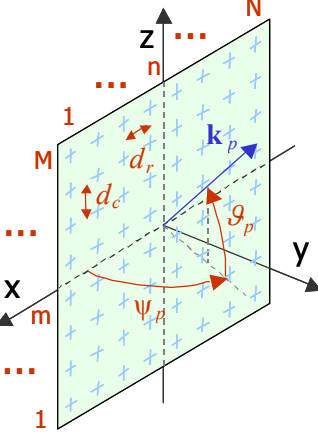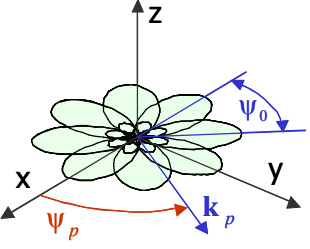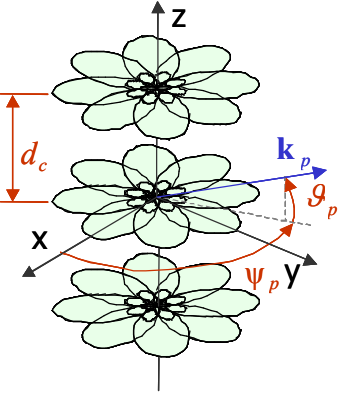
$$\mathbf{a}_{ura}(\mu^{(1)}, \mu^{(2)}) = \mathbf{a}_{col}(\mu^{(2)}) \otimes \mathbf{a}_{row}(\mu^{(1)}). \tag{2.36}$$

Using (2.36) in (2.33) we yield an approximation of the real array response for an uniform rectangular array as

$$\mathbf{b}_{ura}(\varphi, \vartheta) = \mathbf{K} \cdot \mathbf{a}_{ura}(\varphi, \vartheta) \cdot \mathbf{b}_{com}(\varphi, \vartheta).$$

Table 2-1 gives an overview of antenna arrays whose array responses can be approximated using the factorisation approach (2.33) and the related expressions.

Table 2-2: Planar and circular antenna architectures

Uniform Linear Array (ULA)	
	ULA – Row: $\quad \mu^{(1)} = 2\pi \frac{d_r}{\lambda} \cos(\varphi)\sin(\vartheta)$ $\mathbf{a}_{row}\left(\mu^{(1)}\right) = \left[e^{+j\left(\frac{M_1-1}{2}\right)\mu^{(1)}} \quad \cdots \quad e^{-j\left(\frac{M_1-1}{2}\right)\mu^{(1)}} \right]^{\mathrm{T}}$ ULA-Column: $\quad \mu^{(2)} = 2\pi \frac{d_c}{\lambda} \sin(\vartheta)$ $\mathbf{a}_{col}\left(\mu^{(2)}\right) = \left[e^{+j\left(\frac{M_2-1}{2}\right)\mu^{(2)}} \quad \cdots \quad e^{-j\left(\frac{M_2-1}{2}\right)\mu^{(2)}} \right]^{\mathrm{T}}$ $\mathbf{b}_{real}(\varphi,\vartheta) = \mathbf{K} \cdot \mathbf{a}_{struct}(\varphi,\vartheta) \cdot \mathbf{b}_{com}(\varphi,\vartheta)$
Uniform Rectangular Array (URA)	
	$\mu^{(1)} = 2\pi \frac{d_r}{\lambda} \cos(\varphi)\sin(\vartheta), \ \mu^{(2)} = 2\pi \frac{d_c}{\lambda} \sin(\vartheta)$ $\mathbf{a}_{row}\left(\mu^{(1)}\right) = \left[e^{+j\left(\frac{M_1-1}{2}\right)\mu^{(1)}} \quad \cdots \quad e^{-j\left(\frac{M_1-1}{2}\right)\mu^{(1)}} \right]^{\mathrm{T}}$ $\mathbf{a}_{col}\left(\mu^{(2)}\right) = \left[e^{+j\left(\frac{M_2-1}{2}\right)\mu^{(2)}} \quad \cdots \quad e^{-j\left(\frac{M_2-1}{2}\right)\mu^{(2)}} \right]^{\mathrm{T}}$ $\mathbf{a}_{ura}\left(\mu^{(1)},\mu^{(2)}\right) = \mathbf{a}_{col}\left(\mu^{(2)}\right) \otimes \mathbf{a}_{row}\left(\mu^{(1)}\right)$ $\mathbf{b}_{ura}(\varphi,\vartheta) = \mathbf{K} \cdot \mathbf{a}_{ura}(\varphi,\vartheta) \cdot \mathbf{b}_{com}(\varphi,\vartheta)$
Circular Uniform Beam Array (CUBA)	
	$\varphi_p = \psi_p$ $\mathbf{a}_a\left(\varphi_p\right) = \left[1 \ e^{-j\varphi_p} \ ... \ e^{-j(N-1)\varphi_p} \right]^{\mathrm{T}}$ $\mathbf{b}\left(\psi_p\right) = \mathbf{b}\left(n\psi_0 - \psi_p\right)$ $\mathbf{a}_a\left(\varphi_p\right) = \mathbf{G}_0^{-1}\mathbf{F}\mathbf{b}\left(\psi_p\right)$
Stacked Circular Uniform Beam Array (SCUBA)	
	$\theta_p = \frac{d_c}{\lambda}\sin(\vartheta_p), \ \varphi_p = \psi_p$ $\mathbf{a}_c\left(\theta_p\right) = \left[1 \ e^{-j2\pi\theta_p} \ ... \ e^{-j2\pi(M-1)\theta_p} \right]^{\mathrm{T}}$ $\mathbf{a}_a\left(\varphi_p\right) = \left[1 \ e^{-j\varphi_p} \ ... \ e^{-j(N-1)\varphi_p} \right]^{\mathrm{T}}$ $\mathbf{b}\left(\psi_p\right) = \mathbf{b}\left(n\psi_0 - \psi_p\right)$ $\mathbf{a}_a\left(\varphi_p\right) = \mathbf{G}_0^{-1}\mathbf{F}\mathbf{b}\left(\psi_p\right)$ $\mathbf{A}\left(\theta_p,\varphi_p\right) = \mathbf{a}_c\left(\theta_p\right) \cdot \mathbf{a}_a\left(\varphi_p\right)^T$

2.4.3   Effective Aperture Distribution Function

The third model is related to the first model. In general we restrict the azimuth angle to the interval $[-\pi,+\pi)$ and the elevation angle to the interval $[0,\pi]$. Although redundant we can, without loss of generality, extend the beam pattern of an arbitrary antenna to cover the interval $[-\pi,+\pi)$ in azimuth and elevation. One has simply to map the function from the elevation range $[0,\pi]$, to the range $(-\pi,0)$. Consequently, we yield a 2-dimensional function that is under mild restrictions periodic in φ and ϑ with period 2π in both dimensions. Hence, the beam pattern can be expressed by a 2-dimensional complex Fourier-series expansion. We will refer to this 2-dimensional Fourier series as the effective aperture distribution function (EADF).

The expression, relating the EADF and the beam pattern of an antenna array port for horizontal polarisation is

$$b_H(\varphi,\vartheta) = \mathbf{a}(\varphi)^{\mathrm{T}} \cdot \mathbf{G}_{H2D} \cdot \mathbf{a}(\vartheta) \tag{2.37}$$

with the phase vectors

$$\mathbf{a}(\varphi) = \left[\mathrm{e}^{-\mathrm{j}\cdot\left(-\frac{N_1-1}{2}\right)\varphi} \ \ldots \ \mathrm{e}^{-\mathrm{j}\cdot\left(\frac{N_1-1}{2}\right)\varphi} \right]^{\mathrm{T}} \in \mathbb{C}^{N_1 \times 1}, \tag{2.38}$$

$$\mathbf{a}(\vartheta) = \left[\mathrm{e}^{-\mathrm{j}\cdot\left(-\frac{N_2-1}{2}\right)\vartheta} \ \ldots \ \mathrm{e}^{-\mathrm{j}\cdot\left(\frac{N_2-1}{2}\right)\vartheta} \right]^{\mathrm{T}} \in \mathbb{C}^{N_2 \times 1}. \tag{2.39}$$

Here, $\mathbf{G}_{H2D} \in \mathbb{C}^{N_1 \times N_2}$ denotes the EADF for horizontal polarisation. Since expression (2.37) is only appropriate to describe a single array port, we apply the vector operator yielding

$$\begin{aligned} b_H(\varphi,\vartheta) &= \mathrm{vec}\left\{ \mathbf{a}(\varphi)^{\mathrm{T}} \cdot \mathbf{G}_{H2D} \cdot \mathbf{a}(\vartheta) \right\} \\ &= \left(\mathbf{a}(\vartheta)^{\mathrm{T}} \otimes \mathbf{a}(\varphi)^{\mathrm{T}} \right) \cdot \mathrm{vec}\left\{ \mathbf{G}_{H2D} \right\} \\ &= \mathrm{vec}\left\{ \mathbf{G}_{H2D} \right\}^{\mathrm{T}} \cdot \left(\mathbf{a}(\vartheta) \otimes \mathbf{a}(\varphi) \right). \end{aligned} \tag{2.40}$$

Now collecting the EADFs of all antenna array ports for horizontal polarisation in the matrix

$$\mathbf{G}_H = \begin{bmatrix} \mathrm{vec}\left\{ \mathbf{G}_{H2D,1} \right\}^T \\ \vdots \\ \mathrm{vec}\left\{ \mathbf{G}_{H2D,M} \right\}^T \end{bmatrix} \in \mathbb{C}^{M \times N_1 N_2} \tag{2.41}$$

the vector valued function mapping the azimuth and elevation angles to the antenna array response for horizontal polarisation is given by

$$\mathbf{b}_H(\varphi,\vartheta) = \mathbf{G}_H \cdot \left(\mathbf{a}(\vartheta) \otimes \mathbf{a}(\varphi) \right), \quad \mathbb{R}^2 \mapsto \mathbb{C}^M. \tag{2.42}$$

The mapping of the azimuth and the elevation angle to the antenna array response for vertical polarisation can be expressed in the same fashion using

$$\mathbf{b}_V(\varphi,\vartheta) = \mathbf{G}_V \cdot \left(\mathbf{a}(\vartheta) \otimes \mathbf{a}(\varphi) \right), \quad \mathbb{R}^2 \mapsto \mathbb{C}^M \tag{2.43}$$

where the matrix

$$\mathbf{G}_V = \begin{bmatrix} \mathrm{vec}\left\{ \mathbf{G}_{V2D,1} \right\}^T \\ \vdots \\ \mathrm{vec}\left\{ \mathbf{G}_{V2D,M} \right\}^T \end{bmatrix} \in \mathbb{C}^{M \times N_1 N_2} \tag{2.44}$$

contains all EADFs of the antenna array ports for vertical polarisation. One should observe that the redundancy caused by the extension of the elevation angle range from $[0, \pi]$ to $[-\pi, +\pi)$ can be exploited to reduce the operations in numerical implementations.

2.4.4   Separating Radio Channel Model and Antenna Array Model

It is often discussed whether or not the influence of the antenna arrays used throughout channel sounding measurements can be separated from the observed radio channel. In this section we will investigate whether or not this can be accomplished for the data model derived so far. To this end, we introduce four matrices containing the EADFs (2.41), (2.44) of the transmit antenna array and the receive antenna array for horizontal and vertical polarisation:

$$\mathbf{G}_{T_H} \in \mathbb{C}^{M_T \times N_{1,T} N_{2,T}}, \ \mathbf{G}_{T_V} \in \mathbb{C}^{M_T \times N_{1,T} N_{2,T}}, \tag{2.45}$$

and

$$\mathbf{G}_{R_H} \in \mathbb{C}^{M_R \times N_{1,R} N_{2,R}}, \text{ and } \mathbf{G}_{R_V} \in \mathbb{C}^{M_R \times N_{1,R} N_{2,R}}. \tag{2.46}$$

Furthermore, we define four matrix valued functions to express the mapping of all angles $\boldsymbol{\varphi}_T$, $\boldsymbol{\vartheta}_T$, and $\boldsymbol{\varphi}_R$, $\boldsymbol{\vartheta}_R$ to all P vectors $\mathbf{a}(\varphi_{T,p})$, $\mathbf{a}(\vartheta_{T,p})$, and $\mathbf{a}(\varphi_{R,p})$, $\mathbf{a}(\vartheta_{R,p})$:

$$\mathbf{A}_{\varphi_T}(\boldsymbol{\varphi}_T) = [\mathbf{a}(\varphi_{T,1}) \cdots \mathbf{a}(\varphi_{T,P})] \in \mathbb{C}^{N_{1,T} \times P}, \ \mathbf{A}_{\vartheta_T}(\boldsymbol{\vartheta}_T) = [\mathbf{a}(\vartheta_{T,1}) \cdots \mathbf{a}(\vartheta_{T,P})] \in \mathbb{C}^{N_{2,T} \times P}, \tag{2.47}$$

and

$$\mathbf{A}_{\varphi_R}(\boldsymbol{\varphi}_R) = [\mathbf{a}(\varphi_{R,1}) \cdots \mathbf{a}(\varphi_{R,P})] \in \mathbb{C}^{N_{1,R} \times P}, \ \mathbf{A}_{\vartheta_R}(\boldsymbol{\vartheta}_R) = [\mathbf{a}(\vartheta_{R,1}) \cdots \mathbf{a}(\vartheta_{R,P})] \in \mathbb{C}^{N_{2,R} \times P}. \tag{2.48}$$

Now we can represent the mapping of the angles $\boldsymbol{\varphi}_T$, $\boldsymbol{\vartheta}_T$ of all P propagation paths at the transmitter site to the transmit antenna array responses using the EADF with

$$\mathbf{B}_{T_H}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) = \mathbf{G}_{T_H} \cdot (\mathbf{A}_{\vartheta_T}(\boldsymbol{\vartheta}_T) \Diamond \mathbf{A}_{\varphi_T}(\boldsymbol{\varphi}_T)), \quad \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_T \times P}, \tag{2.49}$$

$$\mathbf{B}_{T_V}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T) = \mathbf{G}_{T_V} \cdot (\mathbf{A}_{\vartheta_T}(\boldsymbol{\vartheta}_T) \Diamond \mathbf{A}_{\varphi_T}(\boldsymbol{\varphi}_T)), \quad \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_T \times P}. \tag{2.50}$$

The mapping of the angles $\boldsymbol{\varphi}_R$, $\boldsymbol{\vartheta}_R$ of all P propagation paths at the receiver site to the receive antenna array responses can be expressed in the same way as

$$\mathbf{B}_{R_H}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) = \mathbf{G}_{R_H} \cdot (\mathbf{A}_{\vartheta_R}(\boldsymbol{\vartheta}_R) \Diamond \mathbf{A}_{\varphi_R}(\boldsymbol{\varphi}_R)), \quad \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_R \times P}, \tag{2.51}$$

$$\mathbf{B}_{R_V}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R) = \mathbf{G}_{R_V} \cdot (\mathbf{A}_{\vartheta_R}(\boldsymbol{\vartheta}_R) \Diamond \mathbf{A}_{\varphi_R}(\boldsymbol{\varphi}_R)), \quad \mathbb{R}^{P \times 2} \mapsto \mathbb{C}^{M_R \times P}. \tag{2.52}$$

Again, let us use the abbreviations $\mathbf{A}_{\varphi_T}$, $\mathbf{A}_{\vartheta_T}$ for the functions defined in (2.47), and $\mathbf{A}_{\varphi_R}$, $\mathbf{A}_{\vartheta_R}$ for the functions defined in (2.48). Using the relations (2.49) to (2.52) in the general expression for the MIMO system model (2.29) yields

$$\begin{aligned}
\mathbf{s}(\boldsymbol{\theta}_{sp}) = &\left(\left(\mathbf{G}_{T_H} \cdot (\mathbf{A}_{\vartheta_T} \Diamond \mathbf{A}_{\varphi_T}) \right) \Diamond \left(\mathbf{G}_{R_H} \cdot (\mathbf{A}_{\vartheta_R} \Diamond \mathbf{A}_{\varphi_R}) \right) \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau \right) \Diamond (\mathbf{I} \cdot \mathbf{A}_\alpha) \right) \cdot \boldsymbol{\gamma}_{HH} + \\
&\left(\left(\mathbf{G}_{T_H} \cdot (\mathbf{A}_{\vartheta_T} \Diamond \mathbf{A}_{\varphi_T}) \right) \Diamond \left(\mathbf{G}_{R_V} \cdot (\mathbf{A}_{\vartheta_R} \Diamond \mathbf{A}_{\varphi_R}) \right) \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau \right) \Diamond (\mathbf{I} \cdot \mathbf{A}_\alpha) \right) \cdot \boldsymbol{\gamma}_{HV} + \\
&\left(\left(\mathbf{G}_{T_V} \cdot (\mathbf{A}_{\vartheta_T} \Diamond \mathbf{A}_{\varphi_T}) \right) \Diamond \left(\mathbf{G}_{R_H} \cdot (\mathbf{A}_{\vartheta_R} \Diamond \mathbf{A}_{\varphi_R}) \right) \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau \right) \Diamond (\mathbf{I} \cdot \mathbf{A}_\alpha) \right) \cdot \boldsymbol{\gamma}_{VH} + \\
&\left(\left(\mathbf{G}_{T_V} \cdot (\mathbf{A}_{\vartheta_T} \Diamond \mathbf{A}_{\varphi_T}) \right) \Diamond \left(\mathbf{G}_{R_V} \cdot (\mathbf{A}_{\vartheta_R} \Diamond \mathbf{A}_{\varphi_R}) \right) \Diamond \left(\mathbf{G}_{S_f} \cdot \mathbf{A}_\tau \right) \Diamond (\mathbf{I} \cdot \mathbf{A}_\alpha) \right) \cdot \boldsymbol{\gamma}_{VV}.
\end{aligned} \tag{2.53}$$

Now we are finally able to decompose the MIMO system model into a radio channel model and a transmission or measurement system model, respectively

$$
\begin{aligned}
\mathbf{s}(\boldsymbol{\theta}_{sp}) = \quad & \mathbf{G}_{HH} && \cdot(\mathbf{A}_{\vartheta_T}\Diamond\mathbf{A}_{\varphi_T}\Diamond\mathbf{A}_{\vartheta_R}\Diamond\mathbf{A}_{\varphi_R}\Diamond\mathbf{A}_{\tau}\Diamond\mathbf{A}_{\alpha})\cdot\boldsymbol{\gamma}_{HH} + \\
& \mathbf{G}_{HV} && \cdot(\mathbf{A}_{\vartheta_T}\Diamond\mathbf{A}_{\varphi_T}\Diamond\mathbf{A}_{\vartheta_R}\Diamond\mathbf{A}_{\varphi_R}\Diamond\mathbf{A}_{\tau}\Diamond\mathbf{A}_{\alpha})\cdot\boldsymbol{\gamma}_{HV} + \\
& \mathbf{G}_{VH} && \cdot(\mathbf{A}_{\vartheta_T}\Diamond\mathbf{A}_{\varphi_T}\Diamond\mathbf{A}_{\vartheta_R}\Diamond\mathbf{A}_{\varphi_R}\Diamond\mathbf{A}_{\tau}\Diamond\mathbf{A}_{\alpha})\cdot\boldsymbol{\gamma}_{VH} + \\
& \underbrace{\mathbf{G}_{VV}}_{\text{measurment system model}} && \cdot\underbrace{(\mathbf{A}_{\vartheta_T}\Diamond\mathbf{A}_{\varphi_T}\Diamond\mathbf{A}_{\vartheta_R}\Diamond\mathbf{A}_{\varphi_R}\Diamond\mathbf{A}_{\tau}\Diamond\mathbf{A}_{\alpha})\cdot\boldsymbol{\gamma}_{VV}}_{\text{radio channel model}} ,
\end{aligned}
\tag{2.54}
$$

using the following definitions for the radio channel measurement system:

$$
\mathbf{G}_{HH} = \mathbf{G}_{T_H}\otimes\mathbf{G}_{R_H}\otimes\mathbf{G}_{S_f}\otimes\mathbf{I}, \quad \mathbf{G}_{HV} = \mathbf{G}_{T_H}\otimes\mathbf{G}_{R_V}\otimes\mathbf{G}_{S_f}\otimes\mathbf{I},
$$

and

$$
\mathbf{G}_{VH} = \mathbf{G}_{T_V}\otimes\mathbf{G}_{R_H}\otimes\mathbf{G}_{S_f}\otimes\mathbf{I}, \quad \mathbf{G}_{VV} = \mathbf{G}_{T_V}\otimes\mathbf{G}_{R_V}\otimes\mathbf{G}_{S_f}\otimes\mathbf{I}.
$$

Similarly, we can decompose expression (2.17) into a radio channel model and a MIMO transmission system model. To this end, we define two system matrices describing the MIMO transmission system separately at both link ends. For the transmitter we define the system matrices

$$
\mathbf{G}_{T_{fH}} = (\mathbf{G}_{T_f}\otimes\mathbf{G}_{T_H}), \; \mathbf{G}_{T_{fV}} = (\mathbf{G}_{T_f}\otimes\mathbf{G}_{T_V})
\tag{2.55}
$$

using (2.15), (2.45). In addition, we express the system matrices for the receiver

$$
\mathbf{G}_{R_{fH}} = (\mathbf{G}_{R_f}\otimes\mathbf{G}_{R_H}), \; \mathbf{G}_{R_{fV}} = (\mathbf{G}_{R_f}\otimes\mathbf{G}_{R_V}),
\tag{2.56}
$$

using (2.16), (2.46). All four matrices (2.55), (2.56) together describe the narrow band MIMO transmission system entirely. Using them in the MIMO system model (2.17) yield the representation for horizontal polarisation

$$
\mathbf{H}_{HH_S} = \underbrace{\mathbf{G}_{R_{fH}}}_{\text{Receiver}} \cdot \underbrace{\mathbf{H}_{HH}}_{\text{Channel}} \cdot \underbrace{\mathbf{G}_{T_{fH}}^{\mathrm{T}}}_{\text{Transmitter}} ,
$$

where

$$
\mathbf{H}_{HH} = \underbrace{(\mathbf{I}\otimes(\mathbf{A}_{\vartheta_R}\Diamond\mathbf{A}_{\varphi_R}))}_{\text{angles at receiver}}\cdot\underbrace{(\mathbf{I}\otimes\mathrm{diag}\{\boldsymbol{\gamma}_{HH}\})}_{\substack{\text{frequency independent}\\\text{path weights}}}\cdot\underbrace{\mathrm{diag}\{\mathrm{vec}\{\mathbf{A}_{\tau}^{\mathrm{T}}\}\}}_{\text{time-delay}}\cdot\underbrace{(\mathbf{I}\otimes(\mathbf{A}_{\vartheta_T}\Diamond\mathbf{A}_{\vartheta_T}))^{\mathrm{T}}}_{\text{angles at transmitter}}
\tag{2.57}
$$

denotes the related radio channel matrix for horizontal-horizontal polarisation. The other three channel matrices for the remaining polarisation combinations can be easily derived by replacing the path weights as well as the system matrices $\mathbf{G}_{R_{fH}}$ and $\mathbf{G}_{T_{fH}}$. Combining the transmitter system matrices (2.55) to

$$
\mathbf{G}_T = \begin{bmatrix} \mathbf{G}_{T_{fH}} & \mathbf{G}_{T_{fV}} \end{bmatrix} \in \mathbb{C}^{(M_f\cdot M_T)\times(2\cdot N_f\times N_T)}
$$

and the receiver system matrices (2.56) to

$$
\mathbf{G}_R = \begin{bmatrix} \mathbf{G}_{R_{fH}} & \mathbf{G}_{R_{fV}} \end{bmatrix} \in \mathbb{C}^{(M_f\cdot M_R)\times(2\cdot N_f\times N_R)}
$$

yields an representation for the complete polarimetric MIMO system model

$$
\mathbf{H}_S(\boldsymbol{\theta}_{sp}) = \mathbf{G}_R\cdot\mathbf{H}(\boldsymbol{\theta}_{sp})\cdot\mathbf{G}_T^{\mathrm{T}}
\tag{2.58}
$$

where

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}_{HH} & \mathbf{H}_{VH} \\ \mathbf{H}_{HV} & \mathbf{H}_{VV} \end{bmatrix} \tag{2.59}$$

denotes the polarimetric narrow band radio channel model. Using representation (2.58) one can evaluate the influence of the antenna arrays, applied at the transmitter and the receiver, as well as the influence of the radio channel onto the performance of an arbitrary MIMO transmission system independently from each other.

Both, expression $\mathbf{s}(\boldsymbol{\theta}_{sp})$ (2.54) and expression $\mathbf{H}(\boldsymbol{\theta}_{sp})$ (2.59), are valid representations of the parametric radio channel. In the context of parameter estimation, the measured radio channel is an observation of the system "radio channel" and the propagation path parameters $\boldsymbol{\theta}_{sp}$ are hidden variables. It primarily depends on the application whether equation (2.59) or equation (2.54) is the better choice to formulate a problem. The input-output representation (2.59) is mainly useful for the analysis of communication systems, whereas (2.54), i.e., the mapping of the parameters $\boldsymbol{\theta}_{sp}$ to the vector $\mathbf{s}$

$$\boldsymbol{\theta} \mapsto \mathbf{s},$$

is the best choice for the formulation of radio channel parameter estimation problems. One obvious advantage of equation (2.54) from a parameter estimation point of view is its regular structure. To conclude, the algebraic model for the observed concentrated propagation paths can be separated into a component describing the radio channel and a component describing the system used to observe the channel. However, this decomposition requires that the observed radio channel can be approximated using concentrated propagation paths only. This does not hold in practice since the radio propagation is partially carried by distributed diffuse scattering as mentioned before. Consequently, the antenna influence can only be removed from channel sounding measurements with respect to the concentrated propagation paths.

2.4.5   Definition of Data- and Parameter-Dimensions

Beside the two algebraic forms discussed above, which express the radio channel as a vector- or matrix-valued function of the propagation path parameters, there exist another instructive representation. We express the sampled radio channel by means of a 4-dimensional data array. Here, the Khatri-Rao product in equation (2.25) motivates the four dimensions. The vector-valued function $\mathbf{s}(\boldsymbol{\theta}_{sp})$ and the 4-dimensional data array are related in the following way[4]:

1. Reshape the $M_t \cdot M_f \cdot M_R \cdot M_T$ vector $\mathbf{s}(\boldsymbol{\theta}_{sp})$ into a matrix $\mathbf{S}_1(\boldsymbol{\theta}_{sp})$ of size $M_t \times M_f \cdot M_R \cdot M_T$
2. Reshape the matrix $\mathbf{S}_1(\boldsymbol{\theta}_{sp})$ into a three-dimensional data array $\mathbf{S}_2(\boldsymbol{\theta}_{sp})$ having size $M_t \times M_f \times M_R \cdot M_T$
3. Reshape the matrix $\mathbf{S}_2(\boldsymbol{\theta}_{sp})$ into a four-dimensional data array $\mathbf{S}_3(\boldsymbol{\theta}_{sp})$ having size $M_t \times M_f \times M_R \times M_T$.

Every data dimension of $\mathbf{S}_3(\boldsymbol{\theta}_{sp})$ corresponds to a domain where the entity transmit antenna array, radio channel, and receive antenna array has been sampled. The four domains are time, frequency, receive antenna array port, and transmit antenna array port. The representation $\mathbf{S}_3(\boldsymbol{\theta}_{sp})$ is only a different way to understand the sampled radio channel $\mathbf{s}(\boldsymbol{\theta}_{sp})$. It is reasonable to say, that the observation of the radio channel $\mathbf{s}(\boldsymbol{\theta}_{sp})$ has four data dimensions. Observe, that the term *data dimensions* refers clearly *not* to the algebraic dimensionality of $\mathbf{s}(\boldsymbol{\theta}_{sp})$, which is one as it is a vector.

---

[4] In MatLab© or Octave the mapping of $\mathbf{s}$ to $\mathbf{S}_3$ is carried out by means of: S3=reshape(s,[Mt Mf MR MT]) and the mapping of $\mathbf{S}_3$ to $\mathbf{s}$ can be done by means of s=S3(:).

In many practical cases, one observes the radio channel with a measurement system, which is only able to measure some of the four data dimensions. Hence, we introduce the variable R_D to count the measured data dimensions. Sometimes the ports of an antenna array can be further separated into two independent data dimensions. An example is the ideal URA. We can separate the antenna array ports of the ideal URA into a column- and row-element domain. Consequently the maximum number of data domains is $R_D = 6$. The range of data domains is

$$1 \leq R_D \leq 6.$$

Closely related to the number of data domains is the dimensionality of the parameter domains. From a signal-processing point of view, the number of free parameters determines the dimensionality of the parameter estimation problem. That means the length of the parameter vector $\boldsymbol{\theta}_{sp}$ (2.28) gives the dimensionality of the parameter estimation problem. However, in the context of radio-channel parameter estimation the term multi-dimensional is often used in a different way. Here, the dimensionality of the problem determines the number of free structural (nonlinear) parameters of a *single* propagation path. The structural parameters are time-delay, Doppler-shift, receive-azimuth and -elevation, and transmit-azimuth and -elevation. To be precise, the dimensionality of the parameter estimation problem in the context of channel parameter estimation refers to the number of structural parameter domains. Hence, the estimation of all structural parameters is referred to as a 6-dimensional parameter estimation problem. Since the number of parameter dimensions can vary, we introduce the model parameter R_P describing the number of structural parameters of a single propagation path, which can be determined from a given channel sounding measurement.

The structure of the radio channel model (2.25) implies that the data dimensions (data-domains) and the parameter dimensions (domains of structural parameter) are closely related. The frequency-domain contains the information about the time-delay, the time-domain carries the information about the Doppler-shifts, and the transmit- and receive-port domains contain the information about the transmit- and receive-angles (azimuth and elevation), respectively. Therefore, the following relation between the number of data dimensions and the number of parameter dimensions

$$R_D \leq R_P$$

holds. The valid range of the parameter dimensions is the same as the range of data dimensions, i.e., the range is

$$1 \leq R_D \leq 6.$$

2.5 Dense Multipath Components

On every crossing between two propagation media, where the relative electric or magnetic permeability changes, a propagating wave is subdivided into parts. A part of the waves travels into the other medium, and another part is reflected or scattered. The probability that a scattered wave reaches the receiver is generally higher than the probability that a reflected wave reaches it. This is because a reflection requires a sufficiently large object with a reflecting surface, and if the reflection occurs it can only reach the receiver if the angles of incidence and the angle of reflection are appropriate to reach the next reflector or the receive antenna. Altogether the amount of specular (discrete/concentrated) propagation paths in a scenario is relatively small but their contribution to the total power transferred from the transmit antenna to the receive antenna is usually dominating the transmission. Although the contribution to the received power of a single scattered wave is small compared to the contribution of a reflected

wave, the contribution of all scattered waves together reaching the receiver is significant due to their large number. There exist scenarios where the dense multipath components dominate the transmission. One example is the industry scenario. Hence, we cannot ignore the contribution of the distributed scattering to the radio transmission, if we want to model the radio channel and estimate the model parameters from channel observations. This section is devoted to the parametric modelling of the dense multipath components (DMC) of the observed radio channel in the time-delay domain.

2.5.1   Model for Dense Multipath Components in the Time Delay Domain

Due to the finite measurement apertures used to observe the radio channel, it is not possible to resolve the large amount of scattered waves reaching the receiver. We can only resolve them within the Rayleigh-resolution determined by the apertures used during the measurements. The waves caused by distributed diffuse scattering reaching the receiver within a time interval, which is the reciprocal of the measurement bandwidth cannot be resolved. The power of the individual components reaching the receiver within this time interval is largely determined by the free space attenuation, which is approximately constant over this interval. Therefore, we can assume that one complex delay-bin of the impulse response representing an observed time interval of $\frac{1}{M_f f_0}$ contains the superposition of some "propagation paths" caused by distributed scattering having approximately the same power. However, the phase of this "propagation paths" is due to the large difference in terms of the wavelength between their path lengths, approximately uniformly distributed within the interval. To clarify this, let us discuss an example. We assume the radio channel is observed with a bandwidth of 120MHz at a carrier frequency of 5.2GHz. Then a single delay-bin in the impulse-response represents the superposition of all propagation paths with an electrical length difference of approximately 2.5m. Since the wavelength at 5.2GHz is approximately 5.77cm, this interval represents propagation paths with an electrical length difference of up to 43 wavelengths.

Based on the central limit theorem we can assume that one complex delay-bin in the impulse response can be modelled as a complex circular normal distribution with zero mean, provided the relative bandwidth of the measurements is small $B_m << f_c$. Here (complex) circular Normal (Gaussian) distributed means that the real and imaginary parts of a complex random variable are Normal distributed. The real part and the imaginary part are realizations of two independent processes and have the same variance. What we need furthermore, is a parametric model for the statistics of all delay-bins of the observed impulse response. To this end let us discuss a continuous model, which has been used to describe the power delay profile of the radio channel by various researchers [10], [11], [12]. The model is based on the observation that the power delay profile has an exponential decay over the time-delay and a base delay which is of course related to the distance between the transmit and the receive antenna. Except for the infinite bandwidth assumed, the proposed model (2.60) describes the power delay profile or more precisely the variance as a function of the time-delay of the dense multipath components reasonably well. The bandwidth B_d is related to the coherence bandwidth of the dense multipath components according to $B_3 = \frac{\sqrt{3}}{\pi} B_d$, and α_1 denotes the maximum variance (power).

$$\psi_h(\tau) = \mathrm{E}\left\{|h(\tau)|^2\right\} = \begin{cases} 0 & \tau < \tau_d' \\ \alpha_1 \cdot \frac{1}{2} & \tau = \tau_d' \\ \alpha_1 \cdot \mathrm{e}^{-B_d(\tau - \tau_d')} & \tau > \tau_d' \end{cases} \qquad (2.60)$$

The related power spectrum density, i.e., the Fourier-transform of (2.60) is

$$\Psi_H(\Delta f) = \frac{\alpha_1}{\beta_d + \mathrm{j}2\pi\Delta f} \cdot \mathrm{e}^{-\mathrm{j}2\pi\Delta f \tau_d}, \tag{2.61}$$

where $\beta_d = \frac{B_d}{B_m} = \frac{B_d}{M_f \cdot f_0}$ is the normalized coherence bandwidth of the DMC.

Table 2-3: Parameter Definitions for the Dense Multipath Components

B_d	-	coherence bandwidth of the diffuse components
τ_d'	-	base TDoA of the diffuse components
$\beta_d = \frac{B_d}{B_m} = \frac{B_d}{M_f \cdot f_0}$	-	coherence bandwidth of the diffuse components normalised to the measurement bandwidth
M_f	-	number of frequency points measured within the measurement bandwidth
$\tau_d = \frac{\tau_d'}{t_m}$	-	base TDoA of the diffuse components normalised to the total length of the observed impulse-response
α_1	-	power of the diffuse components at $\tau = \tau_d$
$B_3 = \frac{\sqrt{3}}{\pi} B_d$	-	3dB coherence bandwidth of the diffuse components

The model (2.60) for the dense multipath components in the delay domain is incomplete insofar as it neglects the covariance between components at different time-delays. Let us assume the Fourier-transform of the channel impulse response $h(\tau)$, i.e., channel transfer function, exist and is $H(f)$

$$h(\tau) \circ\!\!\!-\!\!\!\bullet\, H(f),$$

where the symbol $\circ\!\!\!-\!\!\!\bullet$ denotes the Fourier integral transformation. Then assuming that $\mathrm{E}\{h(\tau_1)h(\tau_2)\} = 0, \quad \forall \tau_1 \neq \tau_2$ the covariance function of the channel impulse response $\psi_{hh}(\tau_1, \tau_2)$ and the covariance function of the channel transfer function $\Psi_{HH}(f_1, f_2)$ for infinite bandwidth are related in the following way

$$\psi_{hh}(\tau_1, \tau_2) \quad = \psi_h(\tau_2) \cdot \delta(\tau_1 - \tau_2) = \mathrm{E}\{h(\tau_1) \cdot h(\tau_2)^*\}$$

$$= \psi_h(\tau_2) \cdot \mathrm{e}^{-\mathrm{j}2\pi f_1 \tau_2} \tag{2.62}$$

$$\Psi_{HH}(f_1, f_2) = \Psi_H(f_1 - f_2) \qquad = \mathrm{E}\{H(f_1) \cdot H(f_2)^*\}.$$

Hence, the function $\Psi_H(f)$ (2.61) describes a spectral correlation between components of the channel transfer function having a distance of $f_1 - f_2$. In other words, the process $h(\tau)$ is stationary in the frequency domain and we can use the WSSUS concept outlined in [13] in this particular domain.

Since we observe the radio channel usually with a limited bandwidth B_m we introduce the frequency response of the measurement system as $G(f)$ and the related impulse response $g(\tau)$.

Using the frequency response $G(f)$ in (2.62) yields

$$\Psi_{HHgg}(f_1,f_2) = \Psi_H(f_1-f_2)\cdot G(f_1)\cdot G(f_2)^* \qquad = \mathrm{E}\left\{H(f_1)G(f_1)\cdot H(f_2)^*G(f_2)^*\right\}$$

$$= \left(\psi_h(\tau_2)\cdot \mathrm{e}^{-\mathrm{j}2\pi f_1\tau_2}\cdot G(f_1)\right)*_{\tau_2} g(\tau_2)^*$$

$$\psi_{hhgg}(\tau_1,\tau_2) = \psi_h(\tau_2)\cdot\delta(\tau_1-\tau_2)*_{\tau_1} g(\tau_1)*_{\tau_2} g(\tau_2)^* = \mathrm{E}\left\{h(\tau_1)h(\tau_2)^*\right\}*_{\tau_1} g(\tau_1)*_{\tau_2} g(\tau_2)^*,$$

$$(2.63)$$

where $*_{\tau_1}$ and $*_{\tau_2}$ denotes convolution over τ_1 and τ_2, respectively. If we measure the channel using a measurement system with a bandwidth of B_m and a rectangular frequency response the covariance function of the observed channel impulse response becomes[5]

$$\psi_{hhgg}(\tau_1,\tau_2) = \psi_{xx}(\tau_1,\tau_2)*_{\tau_1} \mathrm{si}(\pi B_m\tau_1)*_{\tau_2} \mathrm{si}(\pi B_m\tau_2). \qquad (2.64)$$

One should observe that the convolution with the sinc-function in equation (2.64) guaranties that our assumption of a circular complex normal distribution of a single delay-bin is valid.

Equation (2.63) reveals that the often-used time domain representation (2.60) is not a valid representation for the distribution of the dense multipath components for real measurements, since it ignores the effect of bandwidth limitations of real systems. The frequency domain representation (2.61) in conjunction with (2.63) is a better representation for the second order statistics of the DMC.

Figure 2-4 shows the PDP (Power Delay Profile) of the dense multipath components and measurement noise, i.e., the elements of the variance-covariance function $\psi_{xxgg}(\tau_1,\tau_2)$ for $\tau=\tau_1=\tau_2$ and the related model parameters. Here α_0 represents the noise variance of i.i.d. (independent identical distributed) Gaussian measurement noise. The effect of bandwidth limitation is clearly visible around τ_d in the example. In channel parameter estimation, we deal usually with sampled versions of the channel transfer function or the channel impulse response. Therefore, a parametric model for the covariance matrix of the DMC is derived in the next section. Recall that the covariance matrix is effectively a sampled version of the continuous variance-covariance function (2.63).

---

[5] The si-function is defined as $\mathrm{si}(x)=\frac{\sin(x)}{x}$.
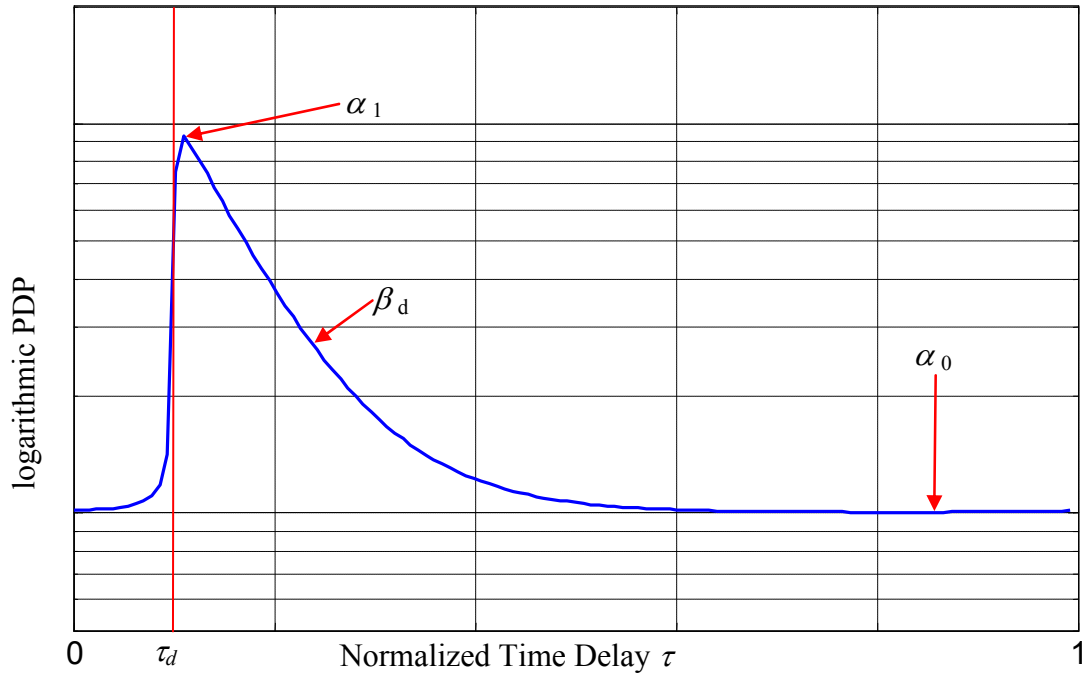
Figure 2-4: Dense multipath distribution model in the time delay domain $\psi_{xxgg}(\tau,\tau)$.

2.5.2 Data Model for Parameter Estimation

Based on the model for the dense multipath components (DMC) (2.63) we will now derive a statistic for a sampled version of the observed channel transfer function containing dense multipath components only. Let us assume that we measure the channel transfer function $\mathbf{x}$ at M_f frequency points equidistantly over the measurement bandwidth B_m. With this assumption and recalling the discussion about the nature of the band limited channel observations in the previous section, we model the distribution of $\mathbf{x}$ as a multivariate circular normal distribution

$$p(\mathbf{x}|\boldsymbol{\theta}_{dmc}) = \frac{1}{\pi^{M_f} \det(\mathbf{R}_f(\boldsymbol{\theta}_{dmc}))} e^{-\mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_f(\boldsymbol{\theta}_{dmc})^{-1} \cdot \mathbf{x}}, \tag{2.65}$$

having zero mean. Here, the covariance matrix $\mathbf{R}_f(\boldsymbol{\theta}_{dmc})$ is effectively a sampled version of the variance-covariance function (2.63), i.e.,

$$\mathbf{R}_f = \begin{bmatrix} \Psi_H(0) & \Psi_H(-f_0) & \cdots & \Psi_H((M_f-1)f_0) \\ \Psi_H(f_0) & \Psi_H(0) & \ddots & \\ \vdots & \ddots & \ddots & \Psi_H(-f_0) \\ \Psi_H((M_f-1)f_0) & \cdots & \Psi_H(f_0) & \Psi_H(0) \end{bmatrix}.$$

Since an observation contains usually multiple realisations of this process $\mathbf{x}_1,\ldots,\mathbf{x}_N$, taken over time (sequence of observed transfer functions) or space (space sampling using an antenna array) we collect multiple observations in the matrix $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_N]$ and write for the distribution of the N independent realisations

$$p(\mathbf{X}_f|\boldsymbol{\theta}_{dmc}) = \frac{1}{\pi^{M_f N} \det(\mathbf{R}_f(\boldsymbol{\theta}_{dmc}))^N} e^{-tr(\mathbf{X}_f^{\mathrm{H}} \cdot \mathbf{R}_f(\boldsymbol{\theta}_{dmc})^{-1} \cdot \mathbf{X}_f)}. \tag{2.66}$$

Equation (2.62) implies that the covariance matrix of $\mathbf{x}$ has a Toeplitz structure (in the frequency domain). Hence, assuming the measurement system used to observe the channel has a flat transfer function within the measurement bandwidth, we express the covariance matrix as

$$\mathbf{R}_f(\boldsymbol{\theta}_{dmc}) = \mathrm{toep}\!\left(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}), \boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc})^{\mathrm{H}}\right), \qquad (2.67)$$

where $\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}) \in \mathbb{C}^{M_f \times 1}$ is a sampled version of the power spectrum density $\Psi_x(f)$, and *toep*( ) is the Toeplitz-operator (see Appendix C). Since the measurement noise can also be modelled as a multivariate circular normal distribution, we include it in our model, assuming it is independent identical distributed (i.i.d.) and has a covariance matrix of $\alpha_0 \mathbf{I}$. Consequently $\boldsymbol{\kappa}(\boldsymbol{\theta}_{dan})$ is given by

$$\boldsymbol{\kappa}(\boldsymbol{\theta}_{dan}) = \frac{\alpha_1}{M_f}\left[\frac{1}{\beta_d} \quad \frac{e^{-j2\pi\tau_d}}{\beta_d + j2\pi\frac{1}{M_f}} \quad \cdots \quad \frac{e^{-j2\pi(M_f-1)\tau_d}}{\beta_d + j2\pi\frac{M_f-1}{M_f}}\right]^{\mathrm{T}} + \alpha_0\mathbf{e}_0, \qquad (2.68)$$

where α_0 is the variance of the circular (i.i.d.) Normal distributed measurement noise, and $\mathbf{e}_0 = [1\ 0\ \cdots\ 0]^{\mathrm{T}}$ is a unit vector. Here, $\boldsymbol{\theta}_{dan}$ is the parameter vector describing the distribution of the DMC and the measurement noise. The vector valued function $\boldsymbol{\kappa}(\boldsymbol{\theta}_{dan})$ is the sampled auto-correlation function of the channel transfer function without the contribution of concentrated propagation paths.

The PDF (probability density function) of the DMC in the time-delay domain can be derived from (2.66) using the discrete Fourier-transform (DFT). With the DFT-matrix

$$\mathbf{F} = \frac{1}{\sqrt{M}}\begin{bmatrix} w^0 & w^0 & w^0 & \cdots & w^0 \\ w^0 & w^1 & w^2 & \cdots & w^{M-1} \\ w^0 & w^2 & w^4 & \cdots & w^{2(M-1)} \\ \vdots & \vdots & & & \vdots \\ w^0 & w^{M-1} & w^{2(M-1)} & \cdots & w^{(M-1)^2} \end{bmatrix}, \quad w = e^{-\frac{j2\pi}{M}},$$

the relation between the covariance matrix of the DMC in the time-delay domain and the frequency domain can be expressed by

$$\mathbf{R}_\tau(\boldsymbol{\theta}_{dan}) = \mathbf{F}^{\mathrm{H}} \cdot \mathbf{R}_f(\boldsymbol{\theta}_{dan}) \cdot \mathbf{F}.$$

The relation between the observed channel impulse responses $\mathbf{X}_\tau$ and the observed channel transfer functions $\mathbf{X}_f$ is given by $\mathbf{X}_\tau = \mathbf{F}^{\mathrm{H}} \cdot \mathbf{X}_f$. Therefore, the PDF of the DMC in the time-delay domain is

$$p(\mathbf{X}_f|\boldsymbol{\theta}_{dan}) = \frac{1}{\pi^{M_f N}\det\!\left(\mathbf{F}^{\mathrm{H}} \cdot \mathbf{R}_f(\boldsymbol{\theta}_{dan}) \cdot \mathbf{F}\right)^N}\,e^{-\mathrm{tr}\left(\mathbf{X}_f^{\mathrm{H}} \cdot \mathbf{F} \cdot \mathbf{F}^{\mathrm{H}} \cdot \mathbf{R}_f^{-1}(\boldsymbol{\theta}_{dan}) \cdot \mathbf{F} \cdot \mathbf{F}^{\mathrm{H}} \cdot \mathbf{X}_f\right)}$$

$$\Downarrow$$

$$p(\mathbf{X}_\tau|\boldsymbol{\theta}_{dmc}) = \frac{1}{\pi^{M_f N}\det(\mathbf{R}_\tau(\boldsymbol{\theta}_{dan}))^N}\,e^{-\mathrm{tr}\left(\mathbf{X}_\tau^{\mathrm{H}} \cdot \mathbf{R}_\tau^{-1}(\boldsymbol{\theta}_{dan}) \cdot \mathbf{X}_\tau\right)},$$

since $\det(\mathbf{F}) = 1$. One should observe that the covariance matrix in the time domain is diagonal dominant. Nevertheless, it is not a diagonal matrix due to the convolution with the $\frac{\sin(x)}{x}$ functions (2.64) and the periodicity of the DFT. Hence, from a signal processing point of view a representation of the observed radio channel in the frequency domain is preferable. In the

frequency domain, the Toeplitz structure of the covariance matrix $\mathbf{R}_f(\boldsymbol{\theta}_{dan})$ can be exploited to reduce the computational complexity of channel parameter estimation algorithms.

Figure 2-5 and Figure 2-6 illustrates the relationship between the channel impulse response, the channel transfer function, and the respective covariance matrices. The channel impulse responses as well as the channel transfer functions are derived from channel sounding measurements. The radio channel was measured with an 8-element uniform linear array four times in a street micro cell scenario. The concentrated propagations paths have been removed from the measurements. Therefore, the channel transfer functions and the related impulse responses shown in Figure 2-5 contain only contributions from distributed diffuse scattering and measurement noise. Observe the complex structure of the covariance matrix in the time-domain. In contrast, the structure of the covariance matrix is very simple in the frequency domain due to the Toeplitz structure. One should note that the Toeplitz structure is only valid for narrow band systems. That means, the model has to be refined if ultra wideband channels are investigated. The same applies also to the model of the concentrated propagation paths derived in this chapter.
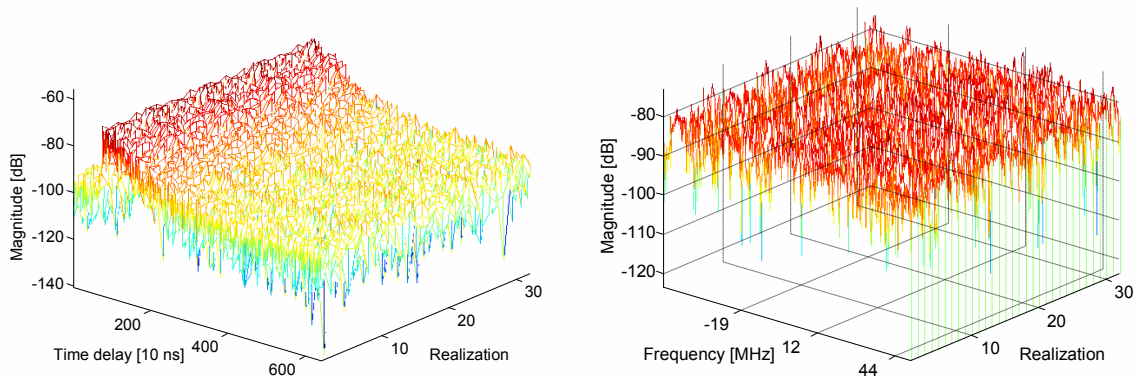
Figure 2-5:   PDP of 32 measured impulse responses after removing the contributions of the concentrated propagation paths (left) and the magnitude of the related channel transfer functions (right).
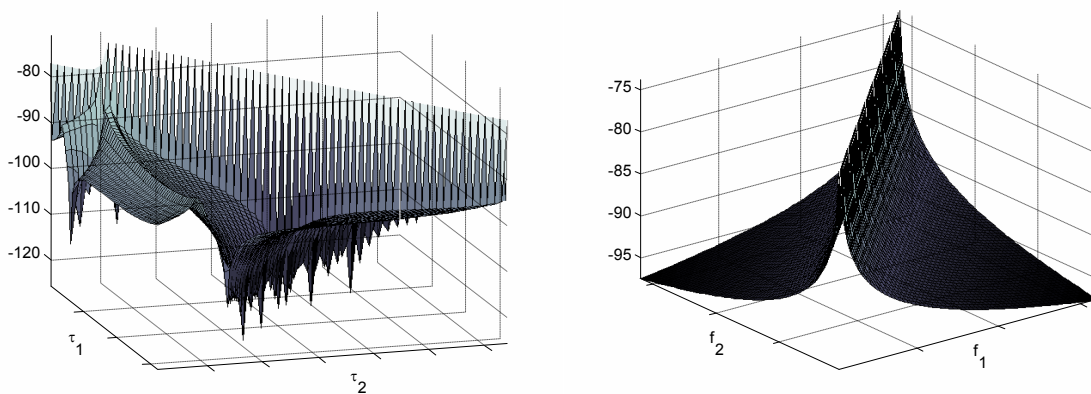
Figure 2-6:   Structure of the covariance matrix of the DMC $\mathbf{R}_\tau(\boldsymbol{\theta}_{dmc})$ in the time delay domain (left) and the structure of the covariance matrix of the DMC $\mathbf{R}_f(\boldsymbol{\theta}_{dmc})$ in the frequency domain (right).

2.5.3 Modeling the DMC in the Spatial and the Time Domain

The covariance function derived in section 2.5.1 describing the DMC is restricted to the frequency domain and time-delay domain, respectively. In general, we have to represent the second order statistics of the dense multipath components with a 12-dimensional function, i.e.,

$$\psi\left(f_1, f_2, \varphi_{T,1}, \varphi_{T_2}, \vartheta_{T,1}, \vartheta_{T,2}, \varphi_{R,1}, \varphi_{R,2}, \vartheta_{R,1}, \vartheta_{R,2}, t_1, t_2\right).$$

Provided the WSSUS assumption [13] can be applied in the temporal and angular domains as well, the correlation function can be expressed by distances in the respective domains. This leads to a 6-dimensional correlation function

$$\psi(\Delta f, \Delta\varphi_T, \Delta\vartheta_T, \Delta\varphi_R, \Delta\vartheta_R, \Delta t).$$

The full covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{dmc})$ of the dense multipath components is of size $M_R M_T M_f M_t \times M_R M_T M_f M_t$. Since the available measurement apertures in the time (Doppler) domain as well as the spatial (angular) domain are relatively small, no satisfactory parametric models for the complete covariance matrix could be developed so far. Hence, we assume that the DMC are i.i.d. in the remaining domains. Nevertheless, let us outline some important points, which should be considered if refinement is attempted.

From a signal processing point of view it is desirable to factorize the full covariance matrix, to keep the numerical complexity of the signal processing algorithms low. A promising approach, although not explicitly proposed for the DMC, is the so-called Kronecker model for narrow band MIMO channels [14]. Here, the idea is to factorize the full covariance matrix into Kronecker products. To this end, we introduce the covariance matrices $\mathbf{R}_R(\boldsymbol{\theta}_{dmc}) \in \mathbb{C}^{M_R \times M_R}$ and $\mathbf{R}_T(\boldsymbol{\theta}_{dmc}) \in \mathbb{C}^{M_T \times M_T}$ describing the angular distribution of the dense multipath components at the transmitter and the receiver position, respectively. In addition, we define the covariance matrix $\mathbf{R}_t(\boldsymbol{\theta}_{dmc})$ to model the correlation of the DMC for a short time interval. Now, a straightforward extension of the Kronecker model yields the following approximation for the covariance matrix

$$\mathbf{R}(\boldsymbol{\theta}_{dmc}) = \mathbf{R}_R(\boldsymbol{\theta}_{dmc}) \otimes \mathbf{R}_T(\boldsymbol{\theta}_{dmc}) \otimes \mathbf{R}_f(\boldsymbol{\theta}_{dmc}) \otimes \mathbf{R}_t(\boldsymbol{\theta}_{dmc}).$$

In the Chapter 4 and 5, it is shown how the Kronecker model can be exploited to reduce the computational complexity of parameter estimation algorithms. Here the Kronecker model is effectively based on the assumption of uncorrelated scattering between the transmit angle, time delay, Doppler shift, and receive angle domains of the DMC. Consequently, the covariance matrix of the DMC-process has the structure

$$\mathbf{R}(\boldsymbol{\theta}_{dmc}) = \mathbf{I}_{M_R} \otimes \mathbf{I}_{M_T} \otimes \mathbf{R}_f(\boldsymbol{\theta}_{dmc}) \otimes \mathbf{I}_{M_t}. \tag{2.69}$$

One should observe that the assumption of a Kronecker model for the DMC does not imply that the whole MIMO channel can be expressed by a Kronecker model. Strictly speaking, the assumption of a Kronecker model for a wideband channel is unlikely to hold due to the deterministic components in the channel, i.e., due to the existence of well-separated dominant propagation paths. For a discussion of the validity of the Kronecker-Model in general, see [14].

2.5.4 Examples for DMC and Discussion

Figure 2-7 and Figure 2-8 show power delay profiles of radio channels measured in a street micro-cell scenario. The measurements have been carried out with an omni-directional transmit antenna and an 8-element uniform linear receive array at a carrier frequency of 5.2GHz. The measurement bandwidth used to observe the channel was 100MHz.
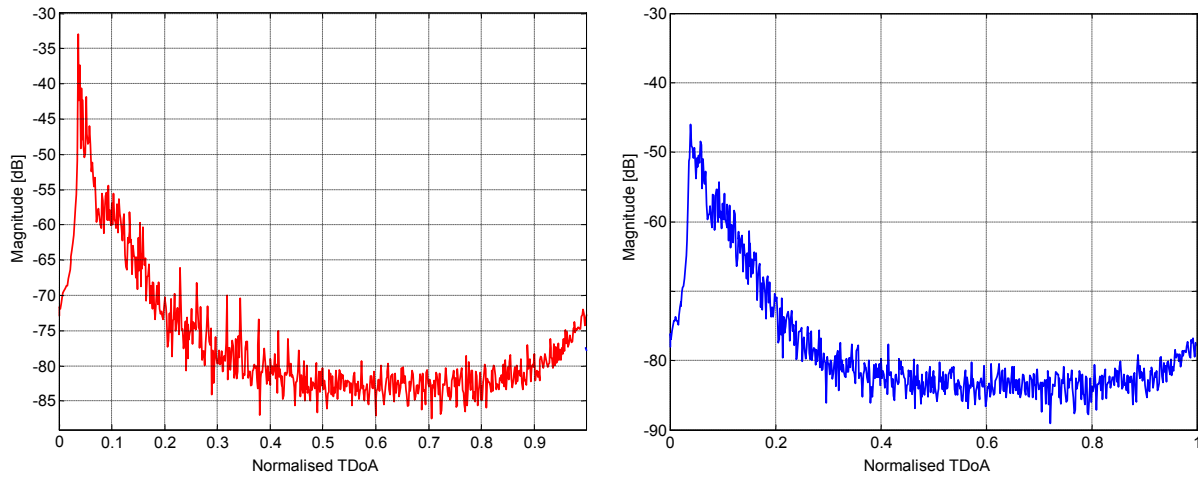
Figure 2-7:   PDP of a complex SIMO-impulse response (LOS) averaged over all receive antennas (left) and the same impulse response after removing the concentrated propagation paths (right).
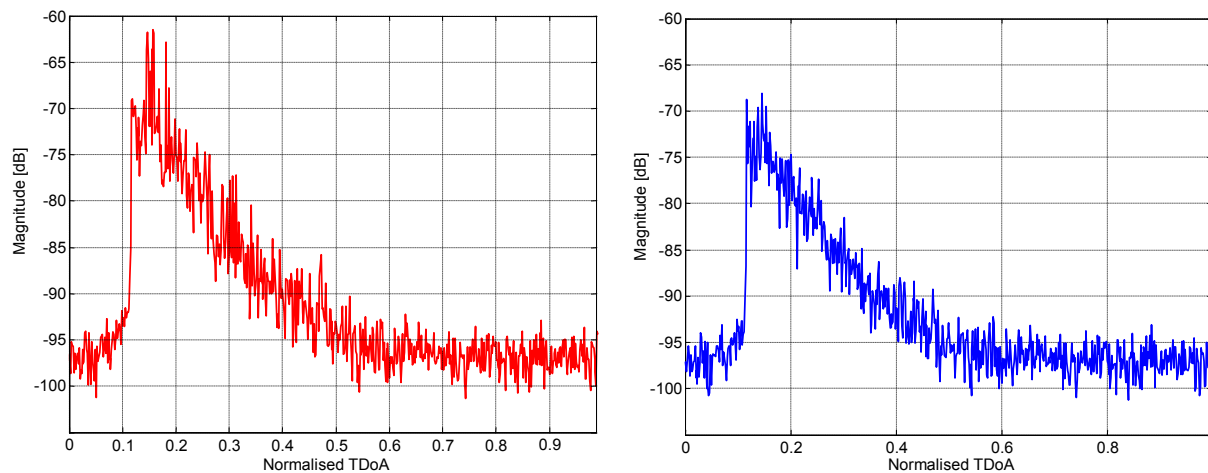
Figure 2-8:   PDP of a SIMO-impulse response (NLOS) averaged over all receive antennas (left) and the same impulse response after removing the concentrated propagation paths (right).

The first example is a line of sight (LOS) scenario, whereas the second example is a non-line of sight (NLOS) scenario. Both figures show the power delay profile (PDP) averaged over the SIMO channels on the left hand side. The right hand side of the figures shows the PDP after removing the concentrated propagation paths from the channel observations. In both cases, the remainder resembles the PDP described in (2.60). Furthermore, it is important to realise that this PDP contains by no means dominant propagation paths anymore. Instead, it resembles a dense stochastic process. Therefore, the contribution of the distributed diffuse scattering to the channel impulse response is denoted as dense multipath.

Since the DMC process is a strongly correlated process in the frequency domain, every parameter estimator ignoring its contribution to the observed channel impulse response will inevitably fail. The contributions of the distributed diffuse scattering to the radio transmission is also the reason why model order selection methods fail if the underlying data model take only white noise and concentrated propagation paths into account. Therefore, having discovered the influence of the DMC to the radio channel observations it is understandable why the author of [15] encountered problems with the application of the Minimum Description Length (MDL) method [16], [17]. Strictly speaking, if time-delay information has to be estimated

from channel sounding measurements, the application of model order selection criteria is only reasonable if the contribution by distributed diffuse scattering to wave propagation is taken into account. Otherwise, model order selection criteria will fail since the set of models we are choosing from will not fit to the observation. This issue is more deeply discussed in Section 5.2.7. However, this fact is not restricted to channel parameter estimation. It applies to all parameter estimation problems.

So far, we have only discussed the importance of the DMC for channel parameter estimation. However, the contribution of the DMC to the radio transmission must also be considered in radio channel modelling. Figure 2-9 shows a measured PDP in a LOS scenario and the PDP reconstructed using the parameter estimates of 62 concentrated propagation paths.
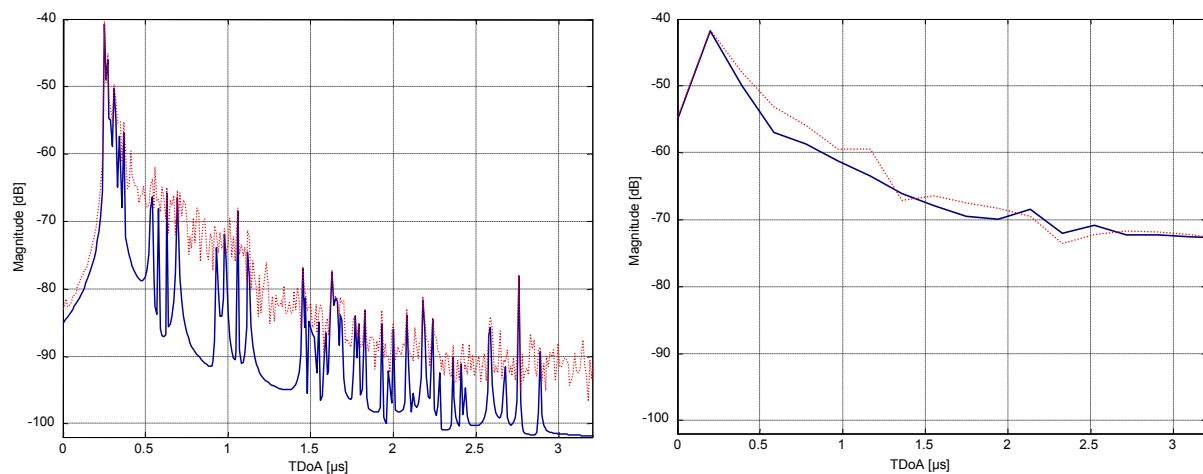
Figure 2-9:   A measured PDP (dotted line) and the PDP (straight line) reconstructed using 62 concentrated propagation paths estimated from the measurement. The left hand side shows the PDPs for an observation bandwidth of 100MHz and the right hand side the same PDPs after reducing the bandwidth to 5MHz. The parameters of the propagation paths have been estimated from the measurement having 100MHz bandwidth.

The graph on the left hand side of Figure 2-9 shows the power delay profile reconstructed with a bandwidth of 100MHz whereas the right hand side shows the same channel with a reconstruction bandwidth of 5MHz. The approximation of the channel having a bandwidth of 100MHz with 62 propagation paths is not acceptable. On the other hand, the approximation of the channel having bandwidth 5MHz with 62 propagation paths is acceptable. To achieve the same approximation of the radio channel at 100MHz probably more than 500 concentrated propagation paths are necessary. However, the additional propagation paths are only necessary to approximate the DMC since the concentrated propagation paths are already described by the 62 propagation paths in the example. Hence, the complexity of the radio channel generator can be significantly decreased if the same model for the DMC proposed for channel parameter estimation is used. That means one has to generate some propagation paths, to describe the contribution of the concentrated propagation paths to the transmission and a realisation of a coloured circular normal distributed process to describe the contribution of the DMC. A numerically efficient generator for such a process is described in Section 6.1.11.

In Section 2.5.1 it is assumed that the DMC can be modelled as a circular Gaussian process. This assumption is based on physical considerations. To support this assumption the estimated distribution, i.e., the relative frequency of the DMC samples in a LOS scenario after whitening, and a Normal distribution with variance $\frac{1}{\sqrt{2}}$ are shown in Figure 2-10. For the computation of the histogram only the delay bins between 250ns and 2000ns has been used.

The SIMO observation was taken from the same measurement campaign described in the beginning of this section. The parameters of the distribution of the DMC and the parameters of the specular propagation paths have been jointly estimated with the maximum-likelihood estimator described in Section 6.2.1. The contribution of the specular propagation paths has been removed from the observation using their estimated parameters. For the whitening of the DMC the estimated covariance matrix $\mathbf{R}_f\left(\hat{\boldsymbol{\theta}}_{dmc}\right)$ has been employed. The total number of real valued samples used to compute the histogram was 2816 and the number of classes shown is 81. The second example (Figure 2-11) shows the relative frequency of the DMC samples in a non line of sight scenario after whitening. The PDP on the left hand side shows again the PDP of the observed DMC after removing the estimates of the specular propagation paths. In this example the delay bins between 950ns and 2300ns has been used to compute the histogram shown on the right hand side. The total number of real valued samples used, to compute this histogram, was 2176 and the number of classes in the histogram was again 81. The Normal distribution is in both cases a reasonable hypothesis for the observed distribution of the DMC.
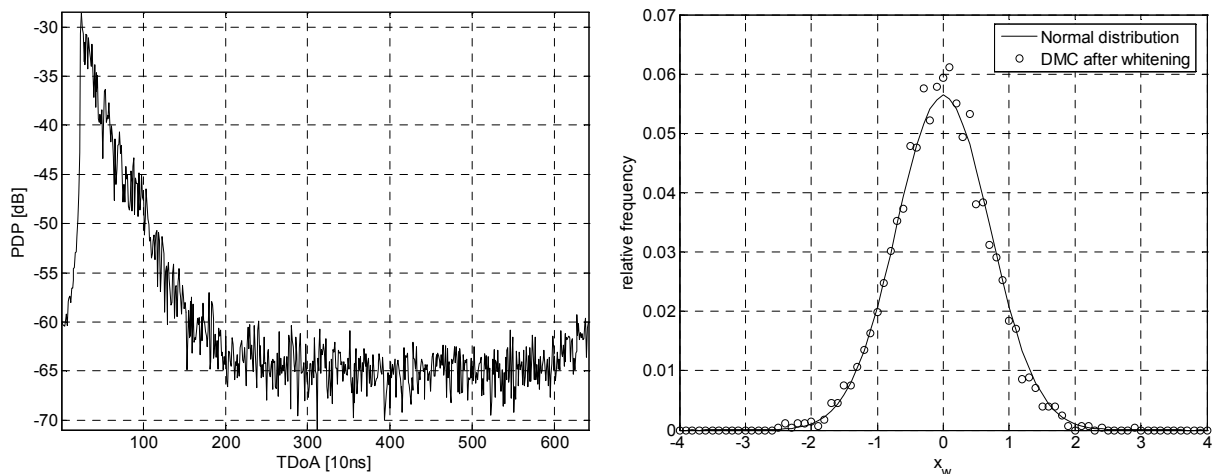
Figure 2-10: PDP of DMC in a line of sight scenario (left hand side) and the distribution of the real and imaginary parts of the delay bins between (250ns-2000ns) after whitening of the impulse response (right hand side).
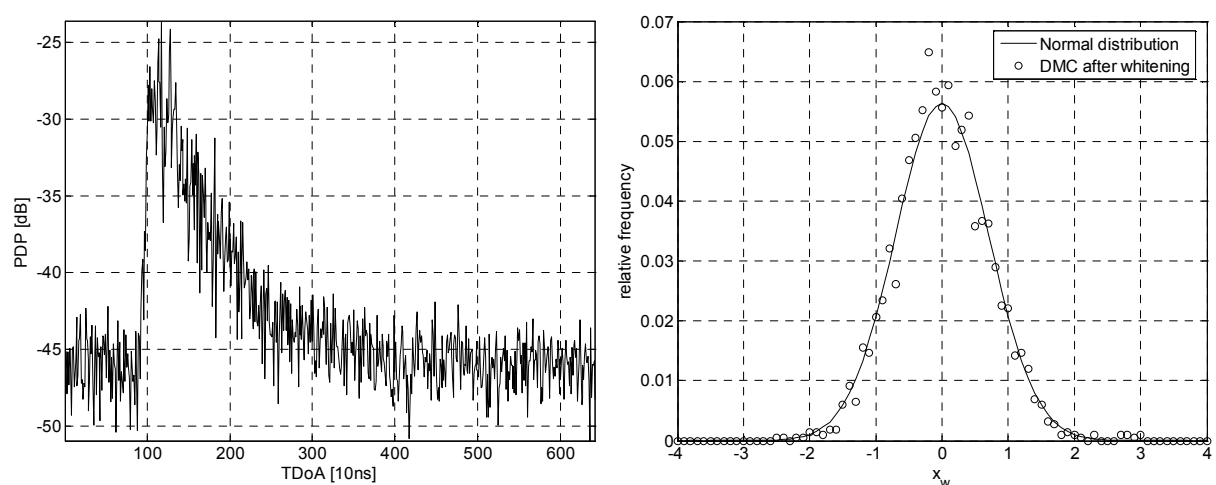
Figure 2-11: PDP of DMC in a non line of sight scenario (left hand side) and the distribution of the real and imaginary parts of the delay bins between (950ns-2300ns) after whitening of the impulse response (right hand side).

Note that the contribution of dense multipath components can also be found in ultra-wideband (UWB) radio channel measurements [18]. Hence, the outlined concept to model the radio channel for channel parameter estimation can also be adopted to describe upcoming UWB channel sounding measurements. Finally, the modelling of diffuse scattering is also discussed in the literature dealing with ray-tracing channel simulations. For example, the authors of [19] describe how to improve ray-tracing simulation accuracy by including diffuse scattering in the radio propagation model.

2.6   Complete Radio Channel Model

The complete model for the sampled radio channel $\mathbf{h} \in \mathbb{C}^{M \times 1}$, taking concentrated propagation paths as well as dense multipath components into account, is given by the superposition of the contribution of the concentrated propagation paths $\mathbf{s}(\boldsymbol{\theta}_{sp}) \in \mathbb{C}^{M \times 1}$ and the contribution of the dense multipath components $\mathbf{d}_{dmc}$ as

$$\mathbf{h} = \mathbf{s}(\boldsymbol{\theta}_{sp}) + \mathbf{d}_{dmc} . \tag{2.70}$$

The realisation $\mathbf{d}_{dmc} \in \mathbb{C}^{M \times 1}$ of the stochastic process describing the dense multipath $\mathcal{D}(\boldsymbol{\theta}_{dmc})$ is distributed according to

$$\mathbf{d}_{dmc} \sim \mathcal{D}(\boldsymbol{\theta}_{dmc}) = \mathcal{N}_C(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dmc})) \in \mathbb{C}^{M \times 1} ,$$

where $\mathcal{N}_C(\mathbf{m}, \mathbf{R})$ is a complex circular symmetric Gaussian process with mean $\mathbf{m} \in \mathbb{C}^{M \times 1}$ and covariance matrix $\mathbf{R}$. A numerically efficient way to generate a realisation of the process $\mathcal{D}(\boldsymbol{\theta}_{dmc})$ for covariance matrices $\mathbf{R}(\boldsymbol{\theta}_{dmc})$ having Toeplitz-structure (cf. equation (2.67)) is described in Section 6.1.11. Equation (2.70) suggest that the sampled radio channel can be understood as a realisation of the process

$$\mathbf{h} \sim \mathcal{N}_C(\mathbf{s}(\boldsymbol{\theta}_{sp}), \mathbf{R}(\boldsymbol{\theta}_{dmc})) \in \mathbb{C}^{M \times 1} , \tag{2.71}$$

so $\mathbf{s}(\boldsymbol{\theta}_{dmc})$ can be interpreted as the mean of $\mathbf{h}$.

One should keep in mind that all parameters of the sampled radio channel

$$\boldsymbol{\theta}_{chn} = \begin{bmatrix} \boldsymbol{\theta}_{sp} \\ \boldsymbol{\theta}_{dmc} \end{bmatrix}$$

are stochastic by nature. The distribution of $\boldsymbol{\theta}_{chn}$ depends on the radio scenario type, e.g., Macro-, Micro-, Pico-cell, whereas their (time) variance depends on the movement of the objects in the scenario. The same applies also to the realisation $\mathbf{d}_{dmc}$, as long as no object having influence on the channel is moving the realisation $\mathbf{d}_{dmc}$ is time invariant. If an object is moving, some or all parameters become time variant. The structural parameters such as time-delays of arrival, angles, Doppler-shifts, coherence bandwidth, and base-delay are typically slow time variant stochastic processes. The linear parameters γ_{HH}, γ_{HV}, γ_{VH}, and γ_{VV} are typically fast time variant processes.

2.7   On Radio Channel Statistics

The data model for the concentrated propagation paths is a deterministic model mapping the propagation path parameters to the channel observation. The propagation path parameters itself are stochastic by nature. The complete statistic of the radio channel can be written as

$$p(\mathbf{h}, \boldsymbol{\theta}_{chn}) = p(\mathbf{h}|\boldsymbol{\theta}_{chn}) \cdot p(\boldsymbol{\theta}_{chn}) .$$

The data model (2.71) corresponds to the conditional probability density function $p(\mathbf{h}|\boldsymbol{\theta}_{chn})$. The focus of this work is to solve the inverse problem $\hat{\boldsymbol{\theta}}_{chn,i} = f(\mathbf{h}_i)$ given the ith observation of the *instantaneous impulse response* $\mathbf{h}_i$ taken at time t_i. The radio channel parameter estimator $\hat{\boldsymbol{\theta}}_{chn,i} = f(\mathbf{h}_i)$ will provide samples of the function $\boldsymbol{\theta}_{chn}(t)$. The time variance of the process $\boldsymbol{\theta}_{chn}(t)$ depends on the motion speed of the objects contributing to the radio propagation. Consequently, the processes $\boldsymbol{\theta}_{chn}(t)$ and $\mathbf{h}(t)$ might not be time variant at all, if no objects in the observed radio propagation scenario are moving. To summarise, the variable $\boldsymbol{\theta}_{chn}(t) \in \mathbb{R}^{L \times 1}$ has to be described by a conditional probability density function $p(\boldsymbol{\theta}_{chn}(t)|v_{chn})$, where v_{chn} describes the position and movement of the objects in the observed radio scenario. A main goal of channel sounding and subsequent channel parameter estimation is to provide sufficient data for the modelling of $p(\boldsymbol{\theta}_{chn}(t)|v_{chn})$, and $p(v_{chn})$.

3 Radio Channel Measurement

From a historical perspective, the first sounding experiments were carried out by using single tone CW (continuous wave) signals. This was sufficient as long as only the narrow-band channel behavior was of interest. Single tone CW sounding, however, gives us no information to resolve path time delays. To resolve propagation paths regarding their time delay, we need a frequency domain bandwidth, which is roughly the inverse of the desired delay resolution. Sequential sounding at a number of different frequencies is the easiest approach to achieve time-delay resolution. The achievable resolution may be very high since standard vector network analyser can be applied. The drawback is the huge measurement time needed to sample the whole frequency response. This inhibits measurement of mobile radio channels. The only solution is to keep the environment fixed during one series of frequency sampling measurements. Frequency domain sampling has its equivalent in sequential sampling of the antenna array geometry. It may be considered as the synthetic antenna aperture approach, applied in the frequency domain. Sustained measurement along some longer trajectory is clearly prohibitive.

3.1   Broadband Radio Channel Sounding Techniques

Short duration repetitive pulses, together with envelope detectors, have been used in early broadband real-time sounding experiments. The main drawback of this method is the high peak-to-mean power ratio at the transmitter and only power delay profiles can be measured. To achieve the maximum signal-to-noise ratio at the receiver, excitation signals are required having a minimum crest factor. The crest factor is given by the ratio of the peak value of the signal to its root mean square (r.m.s.) amplitude. Minimum crest factor signals are distinguished by a constant envelope in the time domain. At the same time, they must have a constant spectrum, which leads to a short autocorrelation function. This pulse compression approach is well known from spread spectrum technology. It makes these signals very useful for real-time identification of time delay systems since all frequencies are instantaneously excited and a considerable SNR processing gain is achieved in the time domain by correlation processing.

Pulse compression requires noise-like structured signals. Periodic pseudo-random excitation signals are of special importance as they can be processed in integer periods. The time-period must be at least as long as the maximum path excess time-delay τ_{max} to avoid TDoA ambiguities. With a maximum delay-Doppler spreading factor $S = \tau_{max} B_{max}$ of a typical mobile radio channel well below 0.01, the period of the received time-variant channel response signal is still almost the same as of the excitation signal. This presumes that the minimum signal time period is chosen. Then the channel output can be transformed to the frequency domain by DFT/FFT (discrete Fourier transform, fast Fourier transform) processing without any significant leakage variance.

Probably the best-known examples of those excitation signals are periodic pseudo-random binary signals (PRBS). PRBS can be very easily generated by a shift register since only digital circuits are required. This makes it possible to generate broadband excitation signals, even suitable for ultra-wideband sounding [20]. Another advantage of PRBS is that they can be repeated in the receiver with a slightly slower clock rate. This is applied in the classical swept time-delay cross-correlation sounder implementation as originally proposed by Cox [21]. This "sliding correlation" sounder requires only slow AD converters (Analogue to Digital converter). The disadvantage of this principle, working sequentially in delay, is again the long measurement time, which prohibits real-time operation.

The power spectrum of PRBS has the typical sinc-shape. For system identification purposes it can only be used up to a frequency of about $0.4f_c$, where f_c is the clock rate [22]. Nevertheless, since the spectrum decays only slowly, a very high sampling rate or a suitable anti-aliasing filter at the receiver is required to avoid aliasing. In addition, the system under test is excited in a frequency band, which is not used. This effectively throws away transmit power. Moreover, most experimental transmit spectrum permissions given by regulation authorities will require strictly band-limited spectra. Then the signal must be filtered at the transmitter to stay within a finite bandwidth. Any filtering and phase slope modification, however, will increase the crest factor of the PRBS, which is supposed to be unity in the ideal case.

A much more flexible excitation signal concept is known as the "periodic multi-sine signal". This approach is well known from frequency domain system identification in measurement engineering [22]. In communication engineering terms, this signal may be refereed to as a multi-carrier spread spectrum signal (MCSSS). The MCSSS is defined by its complex discrete Fourier coefficients $X(\mu f_0)$:

$$x(nt_0) = \sum_{\mu=0}^{N-1} X(\mu f_0) e^{j2\pi\mu/N} \qquad (3.1)$$

with $t_p = Nt_0 = 1/f_0$. Once designed in the frequency domain, the corresponding time domain waveform $x(nt_0)$ is stored in an arbitrary waveform generator memory and periodically repeated at the Tx. It possesses all the advantages which are discussed above for periodic signals. Unlike the case of PRBS, the phases and magnitudes of $X(\mu f_0)$ can be arbitrarily chosen in order to optimize the system performance. As an example of this signal design flexibility, in Figure 3-1 a MCSSS excitation signal with uniform power spectrum is shown. The phases of the Fourier coefficients are chosen to minimize the crest factor of the signal waveform. Although a quadratic phase slope typically results in a crest factor below 2, numerical optimisation can even further reduce the crest factor to about 1.4. Eventually analogue hardware phase distortion (e.g. from the filters) and even nonlinear distortion (from the power amplifier) can be mitigated. This means that a predefined ideal transmit signal is iteratively pre-distorted throughout a calibration procedure where the real output signal is measured and optimized.

Regarding the overall spectral shape, the main advantage of MCSSS is its "brick wall" shape, which allows concentrating the signal energy exactly to the band of interest. This can even be multiple bands when spectral magnitudes are set to zero. One example application is FDD (frequency division duplex) sounding which means that the sounder simultaneously excites both the up- and the down-link band. To meet the UTRA FDD specifications, we need a total bandwidth of more than 200 MHz. Note that the desired full flexibility of the excitation signal requires quadrature up-conversion at the transmitter.

At the receiver side the signal is filtered, down converted, and demodulated by a quadrature demodulator. An efficient architecture is based on low IF (intermediate frequency) ana-

logue down conversion, IF sampling and final digital down conversion. A measurement-bandwidth of 240 MHz, requires an IF frequency of 160 MHz and an ADC (analogue to digital converter) sampling rate of 640 MHz. For real-time processing Nyquist sampling at the receiver is necessary in most cases. One integer period of the received time-variant channel response $y(t, nt_0)$ signal is sampled and transformed to the frequency domain by FFT processing. The final quadrature down-conversion is accomplished by cyclic FFT-shifting of the result, which finally gives the baseband representation $Y(t, \mu f_0)$ of the received signal. Frequency selective fading as shown in Figure 3-1, shapes the power spectrum of the received signal. (bottom row, right).
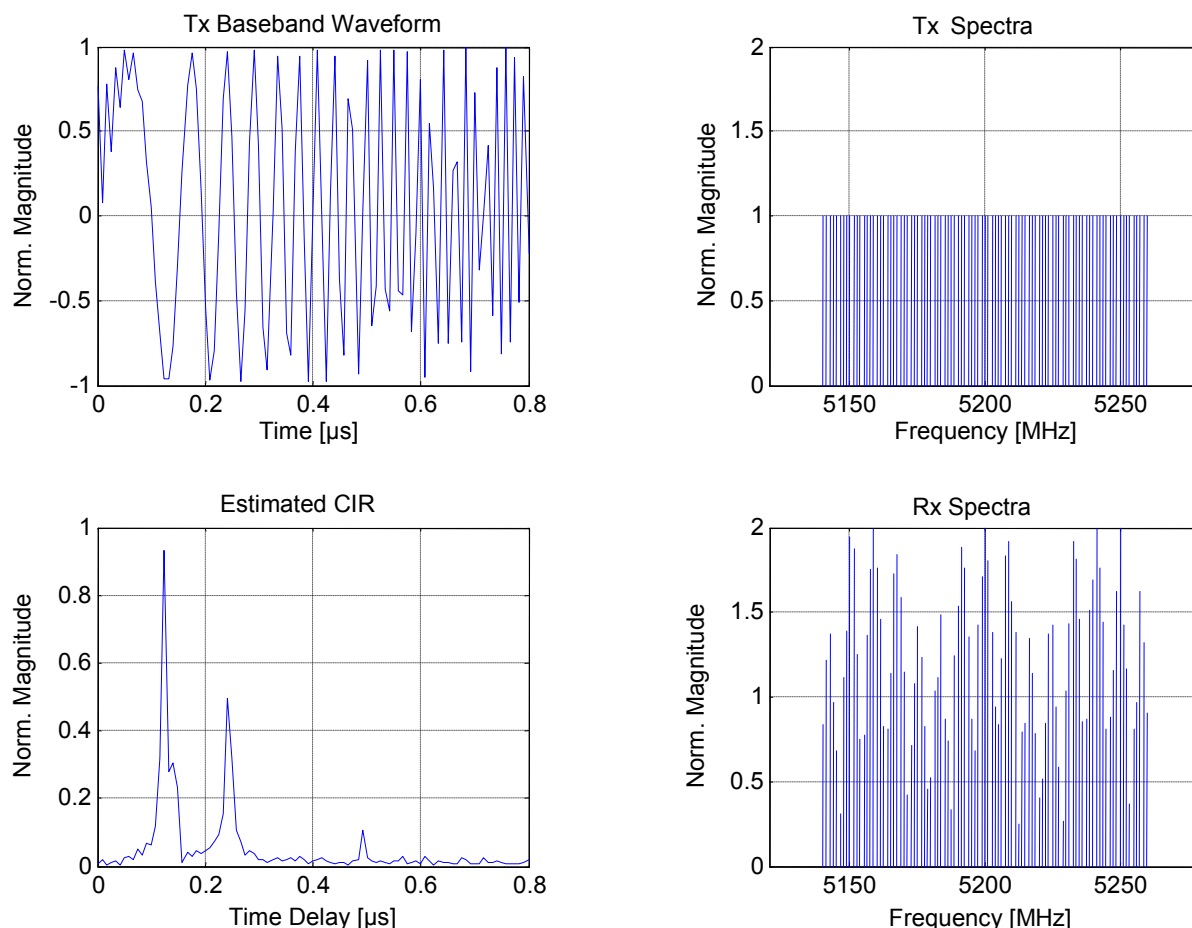
Figure 3-1:   Broadband multi-carrier spread spectrum signal (MCSSS) in the time and magnitude frequency domain (top row) and estimated CIR and received signal spectrum (bottom row).

An estimate of the time-variant channel frequency response is calculated from input-output cross correlation as

$$H(t, \mu f_0) = \frac{Y(t, \mu f_0) X^*(\mu f_0)}{|X(\mu f_0)|^2} = \frac{Y(t, \mu f_0)}{X(\mu f_0)} \tag{3.2}$$

The uniform shape of the excitation signal spectrum and its low crest factor at the transmitter maximizes the SNR. With integer period data acquisition, there is no additional estimation variance resulting from leakage noise [22]. Therefore, the required data acquisition time is minimal and the estimation variance is as small as possible. With Nyquist sampling at the receiver, the highest possible measurement repetition rate for a channel with a maximum

excess time-delay τ_{max} can be achieved, which is $1/\tau_{max}$. The lower limit is given by the Doppler bandwidth B_{max}. It results from the Nyquist sampling criterion of the fast fading channel response. However, since the delay-Doppler spreading factor $S = \tau_{max} \cdot B_{max}$ of a typical mobile radio channels is well below 0.01, there are large gaps allowed between successive measured channel response functions without sacrificing the Nyquist criterion. Normally, there is no need to measure any more quickly since additional CIRs (which may be required for link level simulation) can always be calculated by band-limited interpolation. Nevertheless, faster measurement speed may be desirable if further noise reduction by synchronous averaging of a temporal sequence $y(t, nt_0)$ is aimed at. Only if the averaging window approaches or exceeds $1/B_{max}$ this would act as a Doppler low-pass filter and potentially suppress fast fading.

Figure 3-1 also shows the impulse response which would result from the inverse Fourier transform of $H(t, \mu f_0)$. Calculating the impulse response in this way requires a tapering window function in the frequency domain, which effectively throws away measured data and, hence, reduces SNR and limits the resolution. A better choice is to use $H(t, \mu f_0)$ as an observation vector in the frequency domain for high-resolution TDoA parameter estimation described in Chapters 5 and 6.

3.2   MIMO Channel Sounding

A MIMO channel sounder measures the channel response matrix between all M_T antenna ports at the transmit side and all M_R antenna ports at the receiver side. This could be carried out by applying a transmitter and a receiver having multiple transmit and receive chains, respectively. However, true parallel systems are not only extremely expensive, they are also inflexible (when considering changing the number of antenna channels) and susceptible to phase drift errors. Also parallel operation of the transmitter channels would cause problems since the M_T transmitted signals have to be separated at the receiver. Therefore, orthogonal transmit signals must be used. Since the time/frequency domains are the only domain available to make the transmit signals orthogonal, the only gain of a parallel transmitter is the gain in total transmit power. There is no gain in measurement time. A much more suitable sounder architecture is based on switched antenna access [23], [24], [25], [26], [27]. A switched antenna sounder contains only one physical transmitter and one receiver channel. Only the antennas and the switching channels are parallel. This reduces the sensitivity to channel imbalance.

Figure 3-2 shows the switching time frame of a sequential MIMO sounder using antenna arrays at both sides of the link [28]. Any rectangular block in the figure represents one period of the transmit/receive signal. Synchronous switching at the Rx and Tx is required in order to clearly assign the received signal periods to any input-output combination of the channel matrix. Timing and switching frame synchronization is established during an initial synchronization process prior to measurement data recording and must be maintained over the complete measurement time even in the case of remote operation of Tx and Rx. This is accomplished by rubidium reference oscillators (atomic clock) at both Rx and Tx. The total snapshot time length is now given by $t_s = 2\tau_{max} M_T M_R$, where M_T and M_R are the number of antenna array ports at the Tx and the Rx site, respectively. The factor of two comes from the one blank period, which is inserted at the receiver after every period acting as a guard interval to avoid switching transients. Similar to OFDM (orthogonal frequency division multiplexing), this CIR estimation principle relies on a periodic signal model for excitation and reception. Therefore, the guard interval has to cope with the channel and the device response. For some signal processing operations, based upon the recorded data, it may be a disadvantage that the antenna channels are not sampled at the same instant. If the maximum Doppler bandwidth for

real-time sounding is less than $1/t_s$, the antenna channels can be individually interpolated resulting in MIMO channel responses with aligned sampling time for all channels.

Further considerations concerning the hardware operation of the sounder system refer to the Tx/Rx synchronization in the remote operation mode, to the calibration, to the transmit power, and to link budget issues. Only a short overview to the more important topics will be given here.
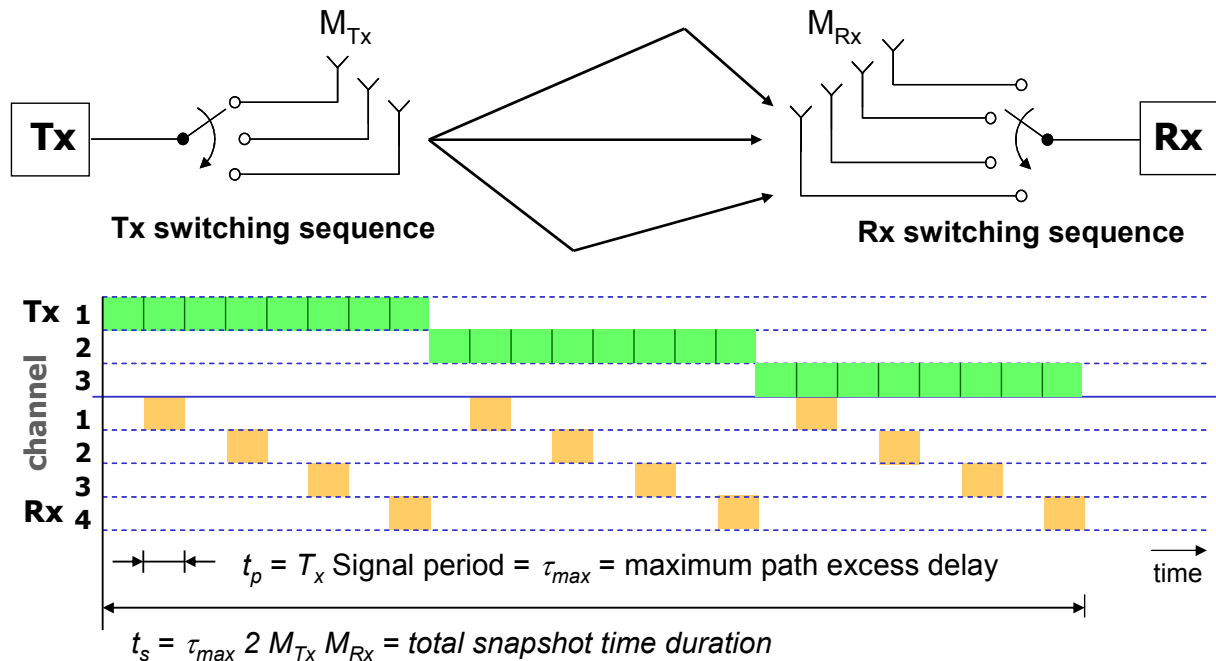
Figure 3-2:   MIMO sounder switching time frame.

Remote operation means that there is no synchronization link applied between Tx and Rx. Initial synchronization is accomplished by a back-to back calibration procedure. Hereby the overall device frequency response $\mathbf{G}_{S_f}$ is measured and stored for equalisation purposes. In addition, the frequency references are synchronized. The synchronization has to be maintained throughout the whole measurement cycle. Separate Rubidium reference sources at both Tx and Rx are required and the local oscillator (LO) signals have to be generated at both sides. This makes a sounder fundamentally different to a standard network analyser and asks for specific considerations. For DoA/DoD estimation full coherent operation is necessary during the snapshot period t_s. If Doppler estimation is required, or if a sequence of snapshots is to be averaged for SNR enhancement, the coherent operation period must extent to multiples of t_s. This sets the limits for phase noise parts having a coherence time below this time interval. However, the time-period between two calibration measurements may easily take some hours if field measurements are considered. In this case, some drift of the references cannot be avoided even if Rubidium sources are used. This can normally be accepted as long as the reference offset is markedly smaller than the specified Doppler bandwidth. A small reference frequency offset would be measured as a respective Doppler shift. Note that in the case of synthetic antenna aperture measurements and for antenna array calibration a much longer coherent operation period will be necessary, which may require a direct Tx/Rx synchronization by cable.

Calibration has to include the absolute device power gain as well. This is also achieved throughout the back-to-back calibration when operating the transmitter with its nominal output power to a reference attenuator. Nevertheless, antenna independent path loss

estimation is only possible if the antennas are calibrated in absolute gain and if the DoA/DoDs of the polarimetric wave components at the antennas are known. This means that both the DoA and DoD have to be estimated. Otherwise, the antenna influence cannot be separated from the measurements and the path attenuation can only be given including the influence of the specific antennas used throughout the measurement.

Further issues are related to automatic gain control (AGC). AGC at the receiver has to ensure maximum signal level throughout the receiver chain from the antenna to the ADC input. At the same time, it has to avoid overloading. The receiver should have a switched AGC in well-defined calibrated steps, which should cover at least a range between 50 and 60dB. The AGC setting has to be implemented on basis of instantaneous peak value estimation. To avoid uncontrolled transients, the AGC timing control must be synchronized to the MIMO switching time frame described in Figure 3-2. For very accurate angle-of-arrival estimation, the same AGC setting should be used for all antennas of the arrays. The best results are achieved if the complex frequency responses of all AGC steps are individually calibrated (including the complex frequency response, which may vary because of changing electrical length).

Regarding the arrangement of antenna switches and amplifiers there is always a trade-off in sensitivity and phase stability. Individual low noise amplifiers (LNA) at Rx antennas and/or individual power amplifiers (PA) at the Tx are mostly inadequate because of the increase in phase drift between antenna channels. However, if there is only a single PA at the Tx, the corresponding antenna switch has to handle the full output power which may exceed 10 W for broadband bad urban measurements. At the Rx the switch just adds its attenuation to the receiver noise figure.

Future steps in real-time MIMO sounding will include the usage of multiple sounding transmitters and/or receivers to emulate system specific scenarios and interference situations. Two transmitters and one receiver, e.g., can be operated in a coordinated way where the transmitters are switched on/off in a staggered temporal sequence. This allows quasi simultaneous measurement of two spatially distributed links. These links can represent a multiuser scenario as seen from a base station. Also two base stations can be emulated to represent soft handover scenarios and cooperative downlink operation from spatially distributed access points. Moreover, a dual-hop link as a part of a multi-hop or ad-hoc network or just a relay extension can be investigated. A future sounder interface will be able to handle dual-band up- and down-converters to emulate tandem air interfaces which will operate in completely different frequency bands. For ultra-wideband operation, sounders will be developed having a real-time bandwidth of some GHz, e.g., from 3 to 11 GHz. The requirements on the hardware of these sounders will be extremely demanding and will require integrated SiGe-technology [20]. This relates also to very broadband sounding at mm-Wave frequencies, e.g., at 60 GHz. To achieve enough spatial resolution of indoor propagation environments, the bandwidth has to be enhanced up to several GHz. The very high frequency will set extreme demands to phase noise if DoA/DoD has to be estimated. UWB operation, however, will shift the angle resolution paradigm from phase difference estimation to time delay estimation allowing wider antenna distances and, thus, compensates loss in accuracy.

3.3   Antenna Array Architectures for Channel Sounding Applications

The spatial dimension of the channel response is accessed by antenna arrays. This mainly relates to "true" arrays but can also include synthetic aperture arrays. Those arrays consist of a sequentially sampled spatial aperture where only one antenna (or a subset) of the respective array is physically deployed. The angular resolution capability of any array depends on the

effective aperture size as seen from the respective wave direction. So the spatial arrangement of the antenna elements has a major influence.

Sophisticated antenna architecture design is required to achieve high DoD/DoA resolution. This has to be developed with mechanically and electrically stable construction and precise calibration. Since there is always a tradeoff between various specifications including resolution, measurement time, availability and costs, there is a wide variety of useful antenna array architectures. In the following, we summarise some design considerations:

- Planar antenna arrays such as uniform linear arrays or uniform rectangular arrays always have a limited viewing angle and suffer from inherent forward/backward ambiguity. They are useful to represent a base station's view to the channel. Moreover, there is a nonlinear transformation from the geometrical DoA/DoD to the respective normalised structural parameters $\mu_k^{(i)}$. Consequently, the effective array aperture depends on the DoA/DoD and the resolution capability is not uniform. Circular antennas, on the other hand, have a full field of view. They can be used to represent the mobile station. Their angular resolution capability is uniform since the effective aperture does not change with azimuth angle.

- Double directional estimation requires arrays at both sides of the link and MIMO operation of the sounder. For cellular system consideration, a combination of planar and circular arrays is adequate, whereas for ad-hoc peer-to-peer networks identical circular arrays are most preferable.

- Mainly for micro- and pico-cell scenarios, estimation of the elevation is aspired in addition to the azimuth. This requires application of uniform rectangular, cylindrical, or spherical arrays. But three dimensional wave analysis (azimuth and elevation) is not only necessary to deduce three dimensional propagation models. It is also required for removal of the influence of the measurement antennas from the data if there are incoming waves with non-zero elevation. Moreover, this must also include polarisation resolution.

- Spherical antenna arrays may be applied for full azimuth and elevation coverage. However, there exists no geometric solution to arrange more than 20 patch antenna elements on a spherical surface with identical inter-element distances. Therefore, non-uniform inter-element distances and various relative polarisation orientations of adjacent elements will complicate the design of spherical arrays. Moreover, optimisation of the inter-element distance for circular and spherical arrays (or of the diameter in case of a fixed number of antenna elements, respectively) is required to minimize the side-lobes of the angular correlation function to reduce the probability of outliers in iterative parameters search. This typically leads to inter-element distances something smaller as half of the wavelength.

- Full polarimetric analysis of the radio channel requires not only polarimetric reception but also polarimetric excitation of the channel. This is even true for omni-directional excitation where we need a two-port antenna, which launches both orthogonal polarised waves with omni-directional characteristics and, thus, doubles the required sounder output ports.

- High and reliable resolution in terms of separation capability of closely spaced paths and low probability of outliers requires an antenna architecture, which offers a minimum of antenna array aperture size in the respective spatial dimension, including a minimum number of antenna elements, low antenna element coupling, and precise calibration. This has also to include the antenna switches and feeder cables. An instructive discussion of real antenna array architectures developed for radio channel sounding and the occurrence of outliers (virtual paths) can be found in [29], [30]. The problem of virtual paths is also discussed in connection with channel model (order) selection in Section 5.2.7.

- The characteristics of the antenna elements depend on the basic element design (dipoles, patches, slots, etc.). It has a strong influence to high-resolution performance, estimation ambiguities, probability of outliers and polarisation resolution capability, gain, bandwidth

etc. E.g., the directivity of the antenna elements is a means to mitigate the inherent forward/backward ambiguity of ULA and URA.

- For later relation of recorded data to the respective propagation scenario video cameras should be included into the antenna module. The optical viewing field of the cameras should correspond to the electromagnetic viewing field of the antennas. Also GPS (global positioning system) position recording, electronic compass, and inclination sensors helps to precisely document the measurement setup. Furthermore, a laser pointer should be contained in the antenna array to support angular adjustment.

The following figures show examples of high-resolution antennas. The URA in Figure 3-3 comprises 8x8 vertical polarised patch elements. Three peripheral dummy rows and columns are included to mitigate the fringing field effect, which distorts the beam patterns. The module also includes a 64x1 multiplexer, LNA, and filter. It can be used for joint azimuth and elevation estimation within the bore side viewing sector of 120 deg. and 60 deg., respectively. The UCA (uniform circular array) in Figure 3-4 (left) consists of 32 sleeve antennas, which do not require a ground plane. Here, a 2W power switch is included to support the application as a transmit antenna. The usage is essentially restricted to azimuth estimation only since there is no vertical aperture available for low elevation paths (which are most important for mobile radio application). The SPUCPA (stacked polarimetric uniform circular patch array) in Figure 3-4 (right) is the currently (2004) most sophisticated array available world wide. It comprises 4 stacked rings of 24 polarimetric patches yielding 192 output ports in total. The RF-multiplexer is arranged inside of the cylindrical body of the array. The cylindrical architecture gives a maximum resolution in azimuth for low elevation paths and good resolution of elevation within ± 30 deg.
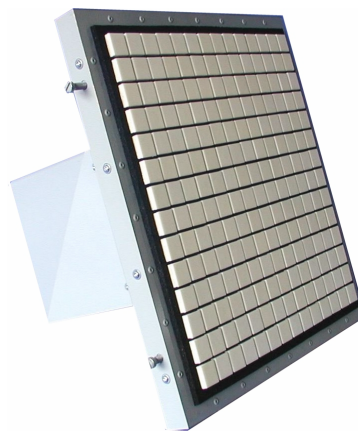
Figure 3-3:   Uniform rectangular patch array (URA8x8) [31], [32].
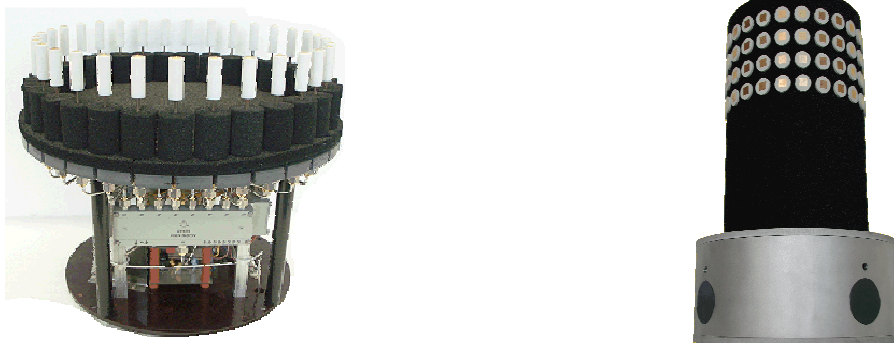
Figure 3-4:   Circular dipole array (UCA32), left, and stacked polarimetric uniform circular patch array (SPUCPA4x24), right [31], [32].

3.4   On the Choice of Reference Scenarios

The performance evaluation of channel-sounding systems requires appropriate test scenarios. The channel parameter estimator can be considered to be part of the channel sounder. Consequently, the measurement results are the estimated channel parameters. Hence, we need appropriate scenarios to determine the accuracy of the individual parameter estimates, for example the systematic and the random error of the estimated time delay related to a single propagation path. In addition, appropriate scenarios are needed to evaluate the resolution capability of the channel sounding system.

Furthermore, the test conditions must be reproducible by someone else with a reasonable effort. From our experience [33], [34], [35], [36], [37], and [38], this can only be guaranteed if the measurements for the performance evaluation are carried out in an anechoic chamber. Measurements in application scenarios are generally not reproducible by third parties and are therefore not suited to evaluate the performance of a channel sounding system.

We propose to use two scenarios in an anechoic chamber for the performance evaluation of channel sounding systems. To determine the measurement accuracy, a single path scenario is the best choice. The setup of such a scenario in an anechoic chamber is simple, since the LOS path is, within the limits of the wave absorbing material, the only path. The parameters of this propagation path can be varied by means of a positioner, which changes the position of the transmitter antenna or of the receive antenna array.

For the evaluation of the path resolution capabilities, we have to create two propagation paths in the anechoic chamber, where the parameters of one propagation path must be adjustable. One may try to use a reflector, for example a metallic plate or a metallic sphere, to create a second propagation path in addition to the direct path. However, it turns out that the parameters of these two propagation paths cannot be adjusted independently of each other. A change of the transmitter position or the transmit power will change the parameters of both paths. In addition, a variation of the reflector position will simultaneously change some and not only one of the parameters belonging to the reflected path. A better approach to create a two-path scenario is based on the observation that a receiver cannot distinguish between a scenario with a real and a virtual signal source and a scenario with two real signal sources as long as the same signal is broadcasted by them. Hence, a two path scenario can be created if we broadcast the transmit signal with two independent transmit antennas. Since every transmit antenna is related to only one propagation path in an anechoic chamber, the propagation path parameters can be adjusted independently. This setup also allows independent adjustment of the individual parameters of one propagation path.

The described test scenarios have been used to evaluate the performance of various antenna arrays [33], [34], [35], [36], [37], and [38]. It is understood, that especially the second test case with two propagation paths, i.e., two sources, is an important benchmark for the performance evaluation of any channel parameter estimation algorithm, see also [39]. Furthermore, both test scenarios should be considered while choosing an antenna array structure for radio channel measurements. An antenna array that provides sufficient information to estimate the parameters of a single propagation path may not provide sufficient information to resolve two propagation paths even if they are substantially separate. An example for such an array structure is a cross array.

3.5   A Cross Array is Not Suitable for Radio Channel Sounding

A question, often asked is. "Why is a cross array not suitable for channel sounding?" It is well known, that azimuth and elevation of a single source can be estimated if a cross array is used to receive the signal of the source. Hence, the cross array can be applied to estimate the

propagation path parameters of test case one (a single propagation path). However, the received signal contains replicas of the same signal in real world scenarios due to multipath propagation. Hence, the second test case (two propagation paths) must be considered too. If the two propagation paths have approximately the same time delay at the receive site but a phase difference of π, they will suppress each other if the magnitude of the path weights are equal. This is effectively a case of space selective fading. The standing wave field of such a scenario is shown in Figure 3-5. In the case shown on the right hand side, the vertical ULA does not receive any signal from the two sources. Consequently, the channel observations measured with a cross array in such a scenario contains no information about the elevation of the two sources. The same problem will also arise, of course, if the two sources are separated only in elevation, i.e., if they have the same azimuth angle.
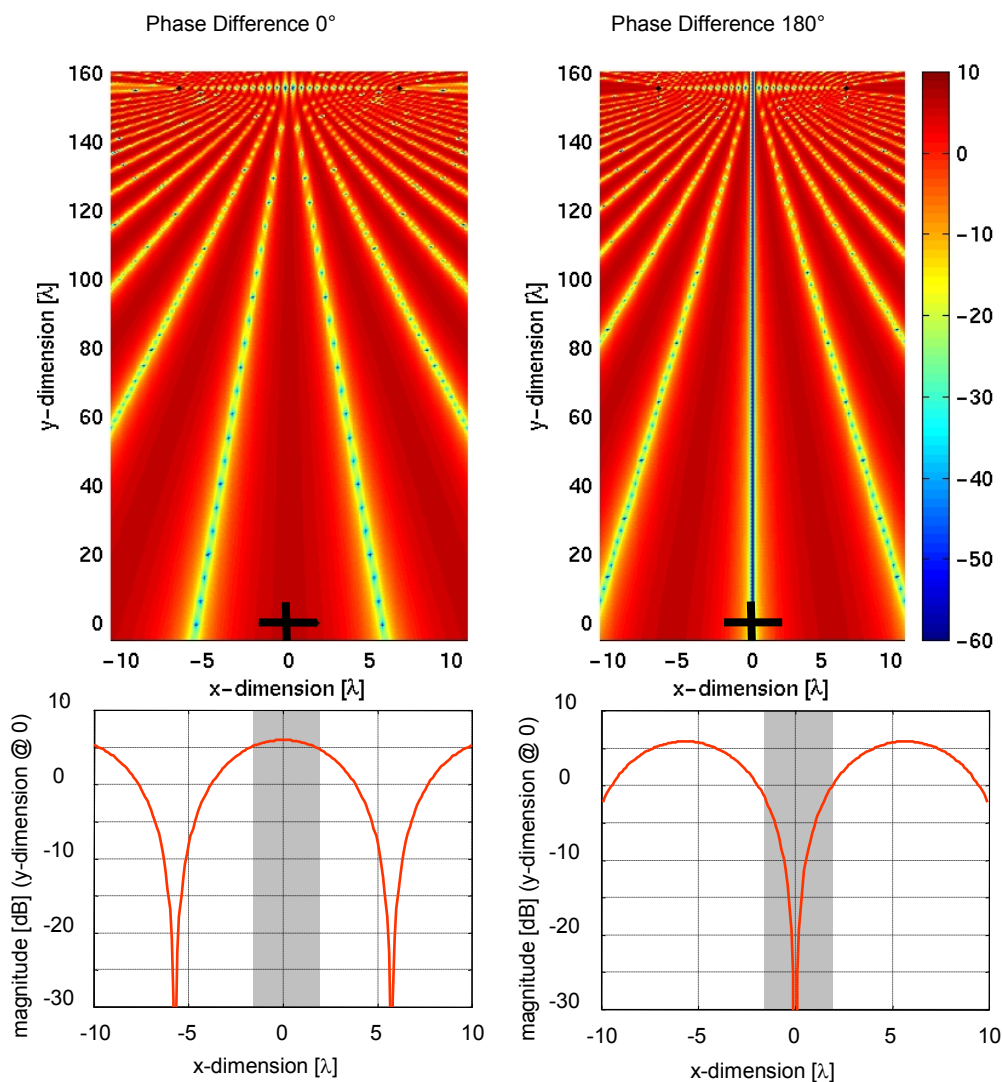
Figure 3-5:   Standing wave field and a cross array.

It will depend on the phase difference between the two propagation paths in such a scenario, whether information about their elevation angles is available in the measurement data or not. Consequently, the parameter estimator will be unable to solve the estimation problem in certain cases. An analytical way to determine, whether sufficient information are available to solve an estimation problem or not, is provided by the *Fisher information matrix* and the *Cramér-Rao lower bound* [40]. Therefore, Chapter 4 is entirely devoted to the application of this performance measures to radio channel parameter estimation.

It is important to note, that the cross array provides sufficient information for the estimation of azimuth and elevation of multiple signal sources which are uncorrelated or only partially coherent, or if they are coherent but well separated in another parameter domain, e.g., in the time delay or Doppler domain.

3.6   Parameter Normalization

All basis functions $\mathbf{B}_f$, $\mathbf{B}_{T_H}$, $\mathbf{B}_{T_V}$, $\mathbf{B}_{R_H}$, $\mathbf{B}_{R_V}$, and $\mathbf{B}_t$ of the channel model (2.29) have basically the same structure. The prototype for all functions $\mathbf{B}$ is

$$\mathbf{B} = \mathbf{G} \cdot \mathbf{A}, \tag{3.3}$$

where $\mathbf{G} \in \mathbb{C}^{M \times N}$ is a matrix describing the measurement system, and $\mathbf{A} \in \mathbb{C}^{N \times P}$ is a matrix of complex exponentials related to the radio channel parameters $\boldsymbol{\tau}$, $\varphi_T, \vartheta_T, \varphi_R, \vartheta_R$, or $\boldsymbol{\alpha}$ [see (2.14), (2.23), (2.34), (2.35), (2.38), and (2.39)]. For a signal processing, the physical meaning of the parameters is not important. Moreover, the different range of values of the physical parameters can be even disadvantageous. Therefore, we introduce the normalised parameter $\boldsymbol{\mu}^{(i)}$ and the related vector valued functions

$$\boldsymbol{\mu}^{(i)} \in \mathbb{R}^{P \times 1} \mapsto \mathbf{A}\left(\boldsymbol{\mu}^{(i)}\right) = \begin{bmatrix} e^{-j\left(-\frac{N-1}{2}\right)\mu_1^{(i)}} & \cdots & e^{-j\left(-\frac{N-1}{2}\right)\mu_P^{(i)}} \\ \vdots & & \vdots \\ e^{-j\left(\frac{N-1}{2}\right)\mu_1^{(i)}} & \cdots & e^{-j\left(\frac{N-1}{2}\right)\mu_P^{(i)}} \end{bmatrix} \in \mathbb{C}^{N \times P}, \tag{3.4}$$

and

$$\left(\boldsymbol{\mu}^{(i)} \in \mathbb{R}^{P \times 1}, \boldsymbol{\mu}^{(k)} \in \mathbb{R}^{P \times 1}\right) \mapsto \mathbf{A}\left(\boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}^{(k)}\right) = \mathbf{A}\left(\boldsymbol{\mu}^{(k)}\right) \diamond \mathbf{A}\left(\boldsymbol{\mu}^{(i)}\right) \in \mathbb{C}^{N_i N_k \times P}. \tag{3.5}$$

The definition of the complex exponentials in (3.4) is unusual, a much more common definition is

$$\boldsymbol{\mu}^{(i)} \in \mathbb{R}^{P \times 1} \mapsto \mathbf{A}'\left(\boldsymbol{\mu}^{(i)}\right) = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ e^{-j(N-1)\cdot\mu_1^{(i)}} & \cdots & e^{-j(N-1)\mu_P^{(i)}} \end{bmatrix} \in \mathbb{C}^{N \times P}, \tag{3.6}$$

There are mainly two reasons, why we should favor definition (3.4) over definition (3.6). Firstly, the reference point for the complex path weight is exactly in the center of the observed aperture, i.e., in the middle of the observed frequency band, and in the center of the antenna arrays. Secondly, the matrix $\mathbf{A}\left(\boldsymbol{\mu}^{(i)}\right)$ is conjugate-symmetric along the columns, i.e.,

$$\mathbf{A}\left(\boldsymbol{\mu}^{(i)}\right) = \boldsymbol{\Pi}\mathbf{A}^*\left(\boldsymbol{\mu}^{(i)}\right).$$

This symmetry can be exploited in implementations to reduce the computational complexity. In Table 3-1 the basis functions introduced in Sections 2.2 - 2.4, the related normalised parameters, and the new basis functions based on the normalised parameters are summarised.

3.7   Incorporation of Sequential Spatial Sampling into the Data Model

The data model for the observed propagation paths developed in Chapter 2 is based on the assumption that the complete MIMO – channel is measured at one time. However, this is not true if the SISO-channels between the transmit- and receive-antenna array ports have been measured sequentially. Since the structural parameters $\boldsymbol{\mu}$ of the radio channel are slow time variant processes, we can treat them as constant within the time needed to take one channel snapshot. In contrast, the path weights may be fast time variant. Here the term fast time vari-

ant relates to the time necessary to measure all SISO-channels, i.e., to acquire a complete snapshot of the MIMO channel. If we assume only the phase of the path weights $\boldsymbol{\gamma}$ is changing within the observation time, the influence of sequential MIMO-channel measurement principle can be incorporated into the data model easily. The phase shift is determined by the Doppler-shift $\mu_p^{(\alpha)} = 2\pi\alpha_p t_0$ of the concentrated propagations paths.

Table 3-1: Definition of Propagation Path Parameter Vectors

Basis Function	Normalisation	Denormalisation	Basis Function with Normalised Parameters
$\mathbf{B}_f = \mathbf{G}_f \cdot \mathbf{A}_\tau(\boldsymbol{\tau})$	$\mu^{(\tau)} = 2\pi f_0 \tau$	$\tau = \dfrac{\mu^{(\tau)}}{2\pi f_0}$	$\mathbf{B}_f = \mathbf{G}_f \cdot \mathbf{A}(\boldsymbol{\mu}^{(\tau)})$
$\mathbf{B}_t = \mathbf{G}_t \cdot \mathbf{A}_\alpha(\boldsymbol{\alpha})$	$\mu^{(\alpha)} = 2\pi t_0 \alpha$	$\alpha = \dfrac{\mu^{(\alpha)}}{2\pi t_0}$	$\mathbf{B}_t = \mathbf{G}_t \cdot \mathbf{A}(\boldsymbol{\mu}^{(\alpha)})$
$\mathbf{B}_{T_H} = \mathbf{G}_{T_H} \cdot (\mathbf{A}_{\vartheta_T} \Diamond \mathbf{A}_{\varphi_T})$ $\mathbf{B}_{T_V} = \mathbf{G}_{T_V} \cdot (\mathbf{A}_{\vartheta_T} \Diamond \mathbf{A}_{\varphi_T})$ $\mathbf{B}_{R_H} = \mathbf{G}_{R_H} \cdot (\mathbf{A}_{\vartheta_R} \Diamond \mathbf{A}_{\varphi_R})$ $\mathbf{B}_{R_V} = \mathbf{G}_{R_V} \cdot (\mathbf{A}_{\vartheta_R} \Diamond \mathbf{A}_{\varphi_R})$	$\mu^{(\varphi)} = \varphi$ $\mu^{(\vartheta)} = \vartheta$	$\varphi = \mu^{(\varphi)}$ $\vartheta = \mu^{(\vartheta)}$	$\mathbf{B}_{T_V} = \mathbf{G}_{T_V} \cdot \mathbf{A}(\boldsymbol{\mu}^{(\varphi_T)}, \boldsymbol{\mu}^{(\vartheta_T)})$ $\mathbf{B}_{T_V} = \mathbf{G}_{T_V} \cdot \mathbf{A}(\boldsymbol{\mu}^{(\varphi_T)}, \boldsymbol{\mu}^{(\vartheta_T)})$ $\mathbf{B}_{R_H} = \mathbf{G}_{R_H} \cdot \mathbf{A}(\boldsymbol{\mu}^{(\varphi_R)}, \boldsymbol{\mu}^{(\vartheta_R)})$ $\mathbf{B}_{R_V} = \mathbf{G}_{R_V} \cdot \mathbf{A}(\boldsymbol{\mu}^{(\varphi_R)}, \boldsymbol{\mu}^{(\vartheta_R)})$
$\mathbf{B}_{ura} = \mathbf{b}_{com} \Diamond$ $\quad \mathbf{K}_{ura} \cdot (\mathbf{A}(\boldsymbol{\mu}^{(c)}) \Diamond \mathbf{A}(\boldsymbol{\mu}^{(r)}))$ $\mathbf{B}_{ulacol} = \mathbf{b}_{com} \Diamond$ $\quad \mathbf{K}_{ulacol} \cdot \mathbf{A}(\boldsymbol{\mu}^{(c)})$	$\mu^{(c)} = \frac{2\pi d_c}{\lambda} \sin(\vartheta)$ $\mu^{(r)} = \frac{2\pi d_r}{\lambda} \cos(\varphi) \cdot$ $\qquad \cdot \sin(\vartheta)$	$\vartheta = \arcsin\left(\dfrac{\mu^{(c)}\lambda}{2\pi d_c}\right)$ $\varphi = \arccos\left(\dfrac{\mu^{(r)} d_c}{\mu^{(c)} d_r}\right)$	$\mathbf{B}_{ura} = \mathbf{b}_{com} \cdot$ $\quad \mathbf{K}_{ura} \cdot \mathbf{A}(\boldsymbol{\mu}^{(r)}, \boldsymbol{\mu}^{(c)})$ $\mathbf{B}_{ulacol} = \mathbf{b}_{com} \Diamond$ $\quad \mathbf{K}_{ulacol} \cdot \mathbf{A}(\boldsymbol{\mu}^{(c)})$

Given the path weight γ_1 at time t_1 the path weight at time t_2 can be calculated according to

$$\gamma_{p,t_2} = \gamma_{p,t_1} \cdot \mathrm{e}^{-\mathrm{j}\frac{(t_2 - t_1)}{t_0}\mu_p^{(\alpha)}} .$$

This follows directly from (2.2). Furthermore, suppose the antenna array outputs are sampled with a uniform sequence. Here uniform means that the time offset, when the SISO channel transfer function between the antenna ports m_T $(0 \le m_T < M_T)$, m_R $(0 \le m_R < M_R)$ is measured, can be expressed in the form $t_{TR} = t_{0,T} \cdot m_T + t_{0,R} \cdot m_R$. This holds for example if the switching sequence shown in Figure 3-2 is used. The sampling interval is $t_{0,T} = 2M_R \tau_{max} = M_R t_{0,R}$ at Tx and $t_{0,R} = 2\tau_{max}$ at Rx. The sampling of the MIMO channel is uniform if the antenna switching sequence at the Rx site is the same for every active Tx antenna element. This is important since in modern MIMO channel sounding devices the switching sequence can be chosen arbitrarily [32]. The data model cannot be expressed anymore using the Khatri-Rao product, if a non-regular switching sequence has been chosen throughout channel sounding measurements. This will in turn lead to an increase in computational complexity of the channel parameter estimator since some simplifications used in Section 4.1 as well as in Chapter 5 will not apply anymore.

Let us suppose the receive array output ports have been sampled at the time offsets $t_{R,1}, \ldots, t_{R,M_R}$ within the receiver switching frame, and the transmit array ports have been activated at times $t_{T,1}, \ldots, t_{T,M_T}$ within the transmitter switching period, see also Figure 3-2. To incorporate the spatial sampling into the data model we introduce the following matrices

$$\mathbf{A}_{t,R} = \begin{bmatrix} e^{-j\frac{t_{R,1}}{t_0}\mu_1^{(\alpha)}} & e^{-j\frac{t_{R,1}}{t_0}\mu_2^{(\alpha)}} & \cdots & e^{-j\frac{t_{R,1}}{t_0}\mu_P^{(\alpha)}} \\ e^{-j\frac{t_{R,2}}{t_0}\mu_1^{(\alpha)}} & e^{-j\frac{t_{R,2}}{t_0}\mu_2^{(\alpha)}} & \cdots & e^{-j\frac{t_{R,2}}{t_0}\mu_P^{(\alpha)}} \\ \vdots & \vdots & & \vdots \\ e^{-j\frac{t_{R,M_R}}{t_0}\mu_1^{(\alpha)}} & e^{-j\frac{t_{R,M_R}}{t_0}\mu_2^{(\alpha)}} & \cdots & e^{-j\frac{t_{R,M_R}}{t_0}\mu_P^{(\alpha)}} \end{bmatrix} \in \mathbb{C}^{M_R \times P} \tag{3.7}$$

and

$$\mathbf{A}_{t,T} = \begin{bmatrix} e^{-j\frac{t_{T,1}}{t_0}\mu_1^{(\alpha)}} & e^{-j\frac{t_{T,1}}{t_0}\mu_2^{(\alpha)}} & \cdots & e^{-j\frac{t_{T,1}}{t_0}\mu_p^{(\alpha)}} \\ e^{-j\frac{t_{T,2}}{t_0}\mu_1^{(\alpha)}} & e^{-j\frac{t_{T,2}}{t_0}\mu_2^{(\alpha)}} & \cdots & e^{-j\frac{t_{T,2}}{t_0}\mu_p^{(\alpha)}} \\ \vdots & \vdots & & \vdots \\ e^{-j\frac{t_{T,M_T}}{t_0}\mu_1^{(\alpha)}} & e^{-j\frac{t_{T,M_T}}{t_0}\mu_2^{(\alpha)}} & \cdots & e^{-j\frac{t_{T,M_T}}{t_0}\mu_p^{(\alpha)}} \end{bmatrix} \in \mathbb{C}^{M_T \times P}. \tag{3.8}$$

Using the (3.7) and (3.8) the model for the concentrated propagation paths (H-H component)

$$\mathbf{s}_{HH} = \mathbf{B}_{R_H} \Diamond \mathbf{B}_{T_H} \Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH}$$

can be extended to account also for the influence of sequential MIMO channel sounding measurements yielding

$$\mathbf{s}_{HH,ss} = \left(\mathbf{B}_{R_H} \circ \mathbf{A}_{t,R}\right)\Diamond\left(\mathbf{B}_{T_H} \circ \mathbf{A}_{t,T}\right)\Diamond \mathbf{B}_f \Diamond \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH}. \tag{3.9}$$

The influence of sequential MIMO channel sounding measurements can be incorporated into the other components of the path model $\mathbf{s}_{HV}, \mathbf{s}_{VH}$, and $\mathbf{s}_{VV}$ in a similar fashion. The sequential MIMO measurement principle has no influence on the model of the DMC, since the contributions of the DMC to the individual SISO channels are modelled as i.i.d. realisations (cf. Section 2.5.3).

   It is important to note, that the matrices (3.7) and (3.8) can only be applied to refine the data model for channel parameter estimation if sufficient information for Doppler-shift estimation is available. However, radio channel sounders available so far are not able to measure the radio channel continuously. That means the time needed to measure as single MIMO-channel snapshot has to be significantly smaller than the time between two consecutive MIMO-channel snapshots ($\frac{t_s}{t_0} = 0.05,...,0.1$). Consequently, the influence of the sequential measurement principle can be neglected. However, upcoming channel sounders will be able to measure the radio channel without gaps between consecutive MIMO-channel snapshots [32]. To exploit their capabilities fully the refined data model (3.9) should be applied. The influence of sequential spatial sampling on ESPRIT based parameter estimation algorithms has been studied for example in [41].

   The algorithms derived in this work are all based on the assumption that the influence of the sequential MIMO measurement principle can be neglected. I.e., the acquisition time for a single MIMO channel snapshot t_S is much smaller than the sampling interval between consecutive MIMO snapshots t_0.

3.8   Measurements with Missing Apertures

If radio channel measurements have been carried out with a measurement system, which is unable to acquire the information necessary to determine the parameters of the complete channel model (2.70), we have to reduce it. If we neglect the effect of sequential spatial sampling the six structural parameters τ, φ_T, ϑ_T, φ_R, ϑ_R, and α relate to a specific aperture, i.e., data dimension. There are four data domains in total. These are the frequency domain,

which carries the information about the time delay τ, the time domain carrying the information about the Doppler-shift α, and the spatial domains, i.e., antenna ports at the Tx- and the Rx-site carrying the information about the transmit angle pair φ_T, ϑ_T and receive angle pair φ_R, ϑ_R, respectively. Furthermore, the antenna arrays used for the measurements also determine which of the four polarimetric path weights can be estimated from the channel observations. In the following, we discuss some cases of measurements with missing apertures.

Case I: If only a single channel observation is available, the data model (2.70) reduces to

$$s(\boldsymbol{\theta}_{sp}) = (\mathbf{B}_{T_H} \lozenge \mathbf{B}_{R_H} \lozenge \mathbf{B}_{\tau}) \cdot \boldsymbol{\gamma}_{HH} + (\mathbf{B}_{T_H} \lozenge \mathbf{B}_{R_V} \lozenge \mathbf{B}_{\tau}) \cdot \boldsymbol{\gamma}_{HV} +$$
$$(\mathbf{B}_{T_V} \lozenge \mathbf{B}_{R_H} \lozenge \mathbf{B}_{\tau}) \cdot \boldsymbol{\gamma}_{VH} + (\mathbf{B}_{T_V} \lozenge \mathbf{B}_{R_V} \lozenge \mathbf{B}_{\tau}) \cdot \boldsymbol{\gamma}_{VV} \,,$$

as a result the Doppler-shifts of the propagation paths are not assessable. Additionally, the resolution of the propagation paths is reduced. However, the loss in resolution compared with other domains is comparatively small since the aperture in the time domain, i.e., the number of consecutive channel observations, must be kept small anyway to assure validity of the channel model. As discussed earlier, it is assumed that a movement of the objects in the radio scenario cause only phase shifts of the path weights. Missing information about the Doppler-shift does not lead to systematic errors in the estimates of the remaining parameters, as long as the influence of sequential spatial sampling can be neglected.

Case II: The radio channel is only measured at one frequency. Hence, the data model (2.29) reduces to

$$s(\boldsymbol{\theta}_{sp}) = (\mathbf{B}_{T_H} \lozenge \mathbf{B}_{R_H} \lozenge \mathbf{B}_{\alpha}) \cdot \boldsymbol{\gamma}_{HH} + (\mathbf{B}_{T_H} \lozenge \mathbf{B}_{R_V} \lozenge \mathbf{B}_{\alpha}) \cdot \boldsymbol{\gamma}_{HV} +$$
$$(\mathbf{B}_{T_V} \lozenge \mathbf{B}_{R_H} \lozenge \mathbf{B}_{\alpha}) \cdot \boldsymbol{\gamma}_{VH} + (\mathbf{B}_{T_V} \lozenge \mathbf{B}_{R_V} \lozenge \mathbf{B}_{\alpha}) \cdot \boldsymbol{\gamma}_{VV} \,.$$

Consequently, the time delays of the propagation paths cannot be estimated and the resolution of the propagation paths is strongly reduced. As a rule of thumb, the frequency domain is the domain with the largest resolution capability. Missing information about the time delays of the propagation paths does not lead to systematic errors in the remaining parameters estimated.

Case III: The radio channel is measured with reduced apertures in the spatial domain at the Tx-site or at the Rx-site. Since the influence of a reduced aperture is equivalent at the Tx-site and the Rx-site, we will discuss the influence at the Tx-site only. There exists a variety of possible measurement setups, since we can choose from a large amount of different antenna array structures.

The antenna array may not provide enough information about azimuth and elevation. This applies to many array structures, such as URA, UCA, ULA, and CUBA. Consequently, the azimuth and/or the elevation will contain ambiguities. The antenna array does not provide enough information to estimate both angles, this is typical for antenna arrays having line apertures, e.g., uniform linear arrays. In contrast to the previous cases, assumptions about the missing information must be made. I.e., we have to choose a prior, the most likely value for the unobserved parameters. This often leads to systematic errors, since the prior is wrong. For example, if the radio channel is measured using a uniform linear array we only have information about one angle, e.g., the elevation if a ULA-column is used. However, the beam pattern of the array elements is also a function of the unobserved angle, i.e., the azimuth in our exam-

ple. Consequently, the estimates of the path weights will be scaled versions of the true path weights, except we know the true azimuth angles of the observed propagation paths.

Finally, the antenna array may not measure the field components $\mathbf{e}_\varphi$ and $\mathbf{e}_\vartheta$ completely. This applies if the antenna elements have only one port either for "horizontal" or "vertical" polarisation. Therefore, only two of the four polarimetric path weights can be determined. If the transmit antenna array generates only field components in $\mathbf{e}_\vartheta$, i.e., $\mathbf{B}_{T_H} = \mathbf{0}$ $\forall \varphi_T, \vartheta_T$, the data model (2.29) reduces to

$$s\left(\boldsymbol{\theta}_{sp}\right) = \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_\tau \Diamond \mathbf{B}_\alpha\right) \cdot \boldsymbol{\gamma}_{VH} + \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_V} \Diamond \mathbf{B}_\tau \Diamond \mathbf{B}_\alpha\right) \cdot \boldsymbol{\gamma}_{VV}.$$

If the receive array receives only field components in $\mathbf{e}_\varphi$, i.e., $\mathbf{B}_{R_V} = \mathbf{0}$ $\forall \varphi_R, \vartheta_R$ the data model reduces further to

$$s\left(\boldsymbol{\theta}_{sp}\right) = \left(\mathbf{B}_{T_V} \Diamond \mathbf{B}_{R_H} \Diamond \mathbf{B}_\tau \Diamond \mathbf{B}_\alpha\right) \cdot \boldsymbol{\gamma}_{VH} .$$

For measurements, channel sounding with incomplete apertures is not optimal. As already discussed in Section 3.3, we acquire incomplete information about the propagation paths. Furthermore, we reduce with every missing aperture, i.e., parameter domain the ability to resolve the propagation paths in the scenario.

To summarise, for a signal processing, the variety of measurement setups is not a problem insofar as the basic data model structure is not altered if an aperture is missing. Having a parameter estimator for the model (2.29), we are from an algorithmic point of view able to cover also cases with missing apertures. However, it is important to realise that the estimator will only yield reliable results if the assumptions about the unobserved information are accurate. This is especially important if information about polarisation and/or angles are missing, since the signal at the output port of an antenna element depends on two angles and the strength of the two electromagnetic field components in $\mathbf{e}_\varphi$ and $\mathbf{e}_\vartheta$. In contrast, the frequency and the time domain only carry information about a single parameter the time delay of arrival and the Doppler-shift, respectively.

4   Limits on Channel Parameter Estimation

It is often criticized that high-resolution channel parameter estimators are not robust [14]. As an example, a robustness issue encountered regularly is the splitting of a propagation path into two. This leads to unusable parameter estimation results, since the path weights of these two paths have often a phase difference of $\sim \pi$ and consequently very large absolute path weights. Strictly speaking, the stated "low robustness" is often not a result of incapability of the parameter estimator, but rather due to the wrong underlying model. Namely, a wrong number of propagation paths or the negligence of the dense multipath components leads to completely wrong estimates.

Furthermore, since one of the main goals of radio channel sounding in general is the derivation of channel parameter statistics, i.e., the modelling of the distribution $p(\boldsymbol{\theta})$, it is important to know the limits on resolution and accuracy of $\hat{\boldsymbol{\theta}}$. In this context, accuracy means primarily the variance of the channel parameter estimates. If the variance of the estimates is ignored the derived distribution functions may partly model the statistics of the estimator $p(\hat{\boldsymbol{\theta}})$ instead of the statistics $p(\boldsymbol{\theta})$ of the true channel parameters.

Additionally, for measurement system design the reachable performance of a channel sounding system, i.e., of the entity channel sounder and parameter estimator is of interest. For example, the reachable accuracy of the estimates of the structural parameters compared with the Rayleigh resolution limit [40] is of interest. The Rayleigh resolution limit is a function of the aperture used to observe a system. For the normalised channel parameters μ_i the Rayleigh resolution is $\propto \frac{2\pi}{M_i}$, i.e., it is not a function of the signal to noise ratio. The Rayleigh resolution determines roughly the resolution of beamforming and DFT methods without further processing.

A uniform bound on the variance of any parameter estimator $\hat{\boldsymbol{\theta}}$ for the parameters of a given model is provided by the Cramér-Rao lower bound [42]. It determines the variance of an unbiased estimator, which is minimum variance unbiased (MVUB). An unbiased estimator $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{L \times 1}$ for the parameters of a given model $\boldsymbol{\theta} \in \mathbb{R}^{L \times 1}$ satisfies the following equation

$$\mathrm{E}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\} = \mathbf{0},$$

i.e. the estimation error has zero mean. Then the Cramér-Rao lower bound CRLB states that for the covariance matrix $\mathbf{C}_{\hat{\boldsymbol{\theta}}} \in \mathbb{R}^{L \times L}$ of any unbiased estimator the following inequality holds.

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \cdot (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}\} \geq \mathbf{CRB}_{\hat{\boldsymbol{\theta}}}.$$

Observe, that the inequality sign between the matrices means that the difference $\mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathbf{CRB}_{\hat{\boldsymbol{\theta}}}$ is positive definite.

Before we start with the derivation of the Cramér-Rao lower bound let us recall the parametric channel model developed in Chapter 2. It describes the sampled radio channel $\mathbf{h}$ as a circular Gaussian distributed process

$$\mathbf{h} \sim \mathscr{N}_C\big(\mathbf{s}(\boldsymbol{\theta}_{sp}), \mathbf{R}(\boldsymbol{\theta}_{dmc})\big) \in \mathbb{C}^{M \times 1}.$$

with mean $\mathbf{s}(\boldsymbol{\theta}_{sp})$ and covariance $\mathbf{R}(\boldsymbol{\theta}_{dmc})$. The stochastic part of the sampled radio channel, i.e., the contribution of the dense multipath is distributed according to

$$\mathbf{d}_{dmc} \sim \mathscr{D}(\boldsymbol{\theta}_{dmc}) \sim \mathscr{N}_C\big(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dmc})\big) \in \mathbb{C}^{M \times 1}.$$

As described in Chapter 3, every channel sounding measurement is disturbed by measurement noise. We will assume in the following that the measurement noise is i.i.d. circular Gaussian distributed according to

$$\mathbf{w} \sim \mathscr{N}_C(\mathbf{0}, \alpha_0 \mathbf{I}) \in \mathbb{C}^{M \times 1}.$$

Since both, $\mathbf{d}_{dmc}$ and $\mathbf{w}$ are circular Gaussian distributed processes and independent, we can merge them into one process

$$\mathbf{n}_{dan} \sim \mathscr{N}_C\big(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dmc}) + \alpha_0 \mathbf{I}\big) \in \mathbb{C}^{M \times 1}. \tag{4.1}$$

For a signal processing, it is unimportant that both originate from different sources, i.e., $\mathbf{d}_{dmc}$ is part of the radio channel, and $\mathbf{w}$ models the measurement noise and is therefore part of the measurement system model. The covariance matrix of the joint process $\mathbf{n}_{dan} = \mathbf{w} + \mathbf{d}_{dmc}$ is

$$\mathbf{R}_{nn} = \alpha_0 \mathbf{I} + \mathbf{R}(\boldsymbol{\theta}_{dmc}). \tag{4.2}$$

Altogether, the observed sampled radio channel $\mathbf{x}$ can be expressed with

$$\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}_{sp}) + \mathbf{n}_{dan}. \tag{4.3}$$

Consequently, the probability density function of a radio channel observation is a multivariate normal distribution

$$p\big(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}\big) = \frac{1}{\pi^M \det(\mathbf{R}_{nn}(\boldsymbol{\theta}))} e^{-(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}))^{\mathrm{H}} \cdot \mathbf{R}_{nn}(\boldsymbol{\theta})^{-1} \cdot (\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}))}. \tag{4.4}$$

The data model, i.e., the model of the observation $\mathbf{x}$ (4.3), is a deterministic function of the parameters $\boldsymbol{\theta}_{sp}$. Therefore the maximum likelihood estimation of $\boldsymbol{\theta}_{sp}$ is a so-called deterministic maximum likelihood (DML) problem. In contrast, the maximum likelihood estimation of the parameters of the stochastic process $\mathbf{n}$ constitutes a stochastic maximum likelihood (SML) problem. Here, we estimate the parameters of a covariance matrix (4.2). Clearly, the joint Cramér-Rao lower bound (CRLB) of all parameters has to be derived. However, for the sake of clarity we split the joint estimation problem into the DML and SML parts, and derive the CRLBs for both problems independently in Section 4.1 and Section 4.3. In Section 4.4 the results for the DML and SML problem are generalised to the joint maximum likelihood parameter estimation task, i.e., the joint Cramér-Rao lower bound for all model parameters $\boldsymbol{\theta}_{chn}$ is derived.

Before we start with the derivation of the Cramér-Rao Lower Bound, let us summarise some important general algebraic properties of the data model for the sampled radio channel.

Cond. 1: The data model is continuous within the parameter space $\boldsymbol{\Theta}_{chn}$ of the parameters $\boldsymbol{\theta}_{sp}$ and $\boldsymbol{\theta}_{dmc}$.

Cond. 2: The parameter space of $\boldsymbol{\Theta}_{chn}$ is compact and the true parameter vector $\boldsymbol{\theta}_{chn}$ is an interior point.

Cond. 3: The stochastic process $\mathbf{n}$ has zero mean and its covariance matrix $\mathbf{R}_{nn}$ is positive definite.

4.1 Cramér-Rao Lower Bound for the Deterministic Parameters

Similar to the derivations in [43], [44], [45], [46], [47] we derive in this section the Cramér-Rao lower bound of the deterministic parameters of the data model describing the channel observation $\mathbf{x}$. We assume for the time being, that the covariance matrix $\mathbf{R}_{nn}$ (cf. (4.2)) is known. Hence the parameters of the vector valued function $\mathbf{s}(\boldsymbol{\theta}_{sp})$ are the only model parameter we have to estimate. For notational convenience we drop for the time being the subscript $_{sp}$ of the parameter vector $\boldsymbol{\theta}_{sp}$. Furthermore, we state an additional property of the channel model in particular of the vector valued function $\mathbf{s}(\boldsymbol{\theta}_{sp})$.

Cond. 4: The vector valued function $\mathbf{s}(\boldsymbol{\theta}_{sp})$ is continuous and has bounded continuous first and second order partial derivatives in $\boldsymbol{\Theta}_{sp}$.

In addition, the vector valued function $\mathbf{s}(\boldsymbol{\theta})$ has to fulfil the following condition:

Cond. 5: For any pair of parameter vectors $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_{sp}$, $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_{sp}$ with $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_k$, the following inequality must hold

$$\frac{\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}_i)}{\sqrt{\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}_i)\mathbf{s}(\boldsymbol{\theta}_i)}} \frac{\mathbf{s}(\boldsymbol{\theta}_k)}{\sqrt{\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}_k)\mathbf{s}(\boldsymbol{\theta}_k)}} < 1,$$

i.e., the mapping $\boldsymbol{\theta} \in \mathbb{R}^{L \times 1} \mapsto \mathbf{s}(\boldsymbol{\theta})\mathbb{C}^{M \times 1}$ must be unique for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{sp}$.

Observe that condition 5 is a fundamental design criterion for a radio channel measurement device. See also Chapter 3 for a discussion of design criteria for channel sounders.

4.1.1 Fisher Information Matrix for the DML Problem

In order to derive the Cramér-Rao lower bound we have to derive the *Fisher information matrix* (FIM) [40] of the parameters $\boldsymbol{\theta}$.

Taking the logarithm of (4.4) yields the log-likelihood function as

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = \ln(p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn})) = -M \cdot \ln(\pi) - \ln(\det(\mathbf{R}_{nn})) - (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^{\mathrm{H}} \mathbf{R}_{nn}^{-1}(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})). \quad (4.5)$$

The first order partial derivative with respect to the parameters $\boldsymbol{\theta}$ of the log-likelihood function (4.5) is the so-called *score-function* [40]

$$\blacksquare \qquad \mathbf{q}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = \left[\frac{\partial}{\partial \theta_1} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) \ldots \frac{\partial}{\partial \theta_L} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn})\right]^{\mathrm{T}} \in \mathbb{C}^{L \times 1}. \quad (4.6)$$

The score-function is effectively the first gradient of the log-likelihood function at $\boldsymbol{\theta}$ given the observation $\mathbf{x}$. The partial derivative of the log-likelihood function (4.5) with respect to the parameter θ_i is

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) &= +\left(\frac{\partial}{\partial \theta_i}\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta})\right)\mathbf{R}_{nn}^{-1}\mathbf{x} + \mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1}\left(\frac{\partial}{\partial \theta_i}\mathbf{s}(\boldsymbol{\theta})\right) \\
&\quad -\left(\frac{\partial}{\partial \theta_i}\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta})\right)\mathbf{R}_{nn}^{-1}\mathbf{s}(\boldsymbol{\theta}) - \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta})\mathbf{R}_{nn}^{-1}\left(\frac{\partial}{\partial \theta_i}\mathbf{s}(\boldsymbol{\theta})\right) \\
&= +\left(\frac{\partial}{\partial \theta_i}\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta})\right)\mathbf{R}_{nn}^{-1}(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})) + (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^{\mathrm{H}}\mathbf{R}_{nn}^{-1}\left(\frac{\partial}{\partial \theta_i}\mathbf{s}(\boldsymbol{\theta})\right).
\end{aligned} \quad (4.7)$$

Since $a + a^* = 2 \cdot \Re\{a\}$ equation (4.7) can be simplified to

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\left\{\left(\frac{\partial}{\partial \theta_i}\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta})\right)\mathbf{R}_{nn}^{-1}(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))\right\}. \quad (4.8)$$

To achieve a compact expression for the score function we define the matrix valued function containing the first order partial derivatives of the data model $\mathbf{s}(\boldsymbol{\theta})$ with respect to the parameter vector $\boldsymbol{\theta}$ as

■ $$\mathbf{D}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{s}(\boldsymbol{\theta}) = \left[\left(\frac{\partial}{\partial \theta_1} \mathbf{s}(\boldsymbol{\theta}) \right) \cdots \left(\frac{\partial}{\partial \theta_L} \mathbf{s}(\boldsymbol{\theta}) \right) \right] \in \mathbb{C}^{M \times L}. \tag{4.9}$$

Observe that (4.9) is effectively the *Jacobian* of $\mathbf{s}(\boldsymbol{\theta})$. Using the definition (4.9) in (4.6) yields the following compact form of the score-function

■ $$\mathbf{q}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\{\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta}) \mathbf{R}_{nn}^{-1} (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))\}. \tag{4.10}$$

Now we recall that the observation $\mathbf{x}$ is only one realisation of a stochastic process. Consequently, the log-likelihood and the score-function become random variables too. The negative covariance matrix of the score-function is the Fisher information matrix, and is defined as follows

■ $$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = -\mathrm{E}\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) \cdot \left(\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) \right)^{\mathrm{T}} \right\}. \tag{4.11}$$

A single element of the Fisher information matrix is given by the relation

$$\mathcal{J}_{ik}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = -\mathrm{E}\left\{ \frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) \cdot \frac{\partial}{\partial \theta_k} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) \right\} = -\mathrm{E}\left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) \right\}. \tag{4.12}$$

Hence, the Fisher information matrix can be understood as the negative expected value of the second gradient of the log-likelihood function [40]. The second partial derivative of the log-likelihood function with respect to the parameters θ_i, and θ_k, i.e., the first partial derivative of the score function with respect to the parameter θ_k is

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = &+\left(\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \right) \mathbf{R}_{nn}^{-1} \mathbf{x} + \mathbf{x}^{\mathrm{H}} \mathbf{R}_{nn}^{-1} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathbf{s}(\boldsymbol{\theta}) \right) \\ &-\left(\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \right) \mathbf{R}_{nn}^{-1} \mathbf{s}(\boldsymbol{\theta}) - \left(\frac{\partial}{\partial \theta_i} \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \right) \mathbf{R}_{nn}^{-1} \left(\frac{\partial}{\partial \theta_k} \mathbf{s}(\boldsymbol{\theta}) \right) \\ &-\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \mathbf{R}_{nn}^{-1} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathbf{s}(\boldsymbol{\theta}) \right) - \left(\frac{\partial}{\partial \theta_i} \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \right) \mathbf{R}_{nn}^{-1} \left(\frac{\partial}{\partial \theta_k} \mathbf{s}(\boldsymbol{\theta}) \right). \end{aligned}$$

Again using the relation $a + a^* = 2 \cdot \Re\{a\}$ we get the expression

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\Bigg\{ &\left(\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \right) \cdot \mathbf{R}_{nn}^{-1} \cdot (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})) \\ &-\left(\frac{\partial}{\partial \theta_i} \mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta}) \right) \cdot \mathbf{R}_{nn}^{-1} \cdot \left(\frac{\partial}{\partial \theta_k} \mathbf{s}(\boldsymbol{\theta}) \right) \Bigg\}. \end{aligned} \tag{4.13}$$

Now we define a matrix-valued function containing the second order partial derivatives of $\mathbf{s}(\boldsymbol{\theta})$ as

$$\mathbf{D}''(\boldsymbol{\theta}) = \frac{\partial^2}{(\partial \boldsymbol{\theta} \otimes \partial \boldsymbol{\theta})^{\mathrm{T}}} \mathbf{s}(\boldsymbol{\theta}) = \left[\frac{\partial^2}{\partial \theta_1 \partial \theta_1} \mathbf{s}(\boldsymbol{\theta}) \cdots \frac{\partial^2}{\partial \theta_1 \partial \theta_L} \mathbf{s}(\boldsymbol{\theta}) \cdots \frac{\partial^2}{\partial \theta_L \partial \theta_1} \mathbf{s}(\boldsymbol{\theta}) \cdots \frac{\partial^2}{\partial \theta_L \partial \theta_L} \mathbf{s}(\boldsymbol{\theta}) \right]. \tag{4.14}$$

With this definition the *stochastic Fisher-Matrix* [40] $\widetilde{\mathcal{J}} \in \mathbb{R}^{L \times L}$, sometimes also called *observed Fisher Information* [48] can be written in the following compact form (cf. Appendix C regarding the definition of the mat$\{\bullet\}$ operator)

$$\widetilde{\mathcal{J}}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\{\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{D}(\boldsymbol{\theta})\} - 2 \cdot \Re\{\mathrm{mat}\{\mathbf{D}''^{\mathrm{H}}(\boldsymbol{\theta}) \cdot \mathbf{R}_{nn}^{-1} \cdot (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))\}\}. \qquad (4.15)$$

The expected value of the stochastic Fisher-Matrix is the Fisher information matrix

$$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathrm{E}\{\widetilde{\mathcal{J}}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn})\}.$$

Since we assume an unbiased estimator $\hat{\boldsymbol{\theta}}$ and due to the model property Cond. 3, the term in equation (4.15) containing the second gradient of $\mathbf{s}(\boldsymbol{\theta})$ goes to zero since

$$\mathrm{E}\{\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})\} = \mathbf{0}.$$

Consequently, the Fisher information matrix of the parameter vector $\boldsymbol{\theta}$, assuming the covariance matrix (4.2) is known, can be calculated as follows

■ $$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\{\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta}) \mathbf{R}_{nn}^{-1} \mathbf{D}(\boldsymbol{\theta})\}. \qquad (4.16)$$

If the stochastic process $\mathbf{n}$ is i.i.d. circular Gaussian with covariance matrix $\mathbf{R}_{nn} = \alpha_0 \mathbf{I}$ the expression for the Fisher information matrix can be further simplified to

$$\mathcal{J}(\boldsymbol{\theta}, \alpha_0 \mathbf{I}) = \frac{2}{\alpha_0} \cdot \Re\{(\mathbf{D}(\boldsymbol{\theta}))^{\mathrm{H}} \cdot \mathbf{D}(\boldsymbol{\theta})\} \qquad (4.17)$$

Observe, that the Fisher information matrix is a symmetric matrix, i.e.,

$$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathcal{J}^{\mathrm{T}}(\boldsymbol{\theta}, \mathbf{R}_{nn}).$$

The main diagonal elements of the Fisher information matrix indicate how informative the measurement data are with regard to the model parameters $\boldsymbol{\theta}$. The off-diagonal elements indicate the amount of mutual information between two parameters θ_i and θ_k. If the off-diagonal elements $\{\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})\}_{ik} = \{\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})\}_{ki}$ are zero, the two related model parameters θ_i and θ_k are asymptotically uncoupled. In general a data model is optimally parameterised if the Fisher information matrix is diagonal. This is an important criterion if we have to choose between two equivalent models.

Furthermore, if the Fisher information matrix is singular the information contained in an observation $\mathbf{x}$ is not sufficient to estimate all model parameters $\boldsymbol{\theta}$, i.e., the data are not sufficiently detailed enough. Consequently, there is no unique solution to the parameter estimation problem at hand. Theoretically, there exist two ways to resolve this problem. We can try to gather more information about the parameters, or we reduce the complexity of the model, i.e., we decrease the number of free parameters. Practically, the first solution requires generally a change of the measurement setup, e.g. a modification of the channel sounder. For a given measurement, this is clearly not a valid option. Hence, the only solution is the reduction of the models complexity.

Altogether we can state the following necessary condition for the unique solvability of a given parameter estimation problem $\hat{\boldsymbol{\theta}}(\mathbf{x})$:

Cond. 6: For the parameter set $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ of the model $\mathbf{s}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times 1}$, the Fisher information matrix must have full rank, i.e., $\mathrm{rank}(\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})) = L$.

The necessary condition 6 directly implies the common requirement $L < 2M$, i.e., the number of measurements must be equal to or larger than the number of unknowns.

If we have to reduce the number of parameters, to get a parameter estimation problem having a unique solution, we have at first to detect the parameters which are strongly coupled. To this end, we introduce the normalised Fisher information matrix[6]

$$\hat{\mathcal{J}}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = (\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) \circ \mathbf{I}_L)^{-\frac{1}{2}} \cdot \mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) \cdot (\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) \circ \mathbf{I}_L)^{-\frac{1}{2}}. \qquad (4.18)$$

Clearly, the normalised Fisher information matrix only exists if the following inequality for all main diagonal elements of the Fisher information matrix holds

$$\{\mathcal{J}(\boldsymbol{\theta}, \mathbf{R})\}_{ii} \neq 0. \qquad (4.19)$$

If the main diagonal element related to the parameter θ_i is $\{\mathcal{J}(\boldsymbol{\theta}, \mathbf{R})\}_{ii} \equiv 0$, the parameter θ_i is not a parameter of the model, i.e., the observation contains no information about it. Consequently, we can state that (4.18) exist for all reasonable parameter sets $\boldsymbol{\theta}$.

The normalised Fisher information matrix contains the correlation coefficients between two parameters θ_i and θ_k. That means the main diagonal elements of (4.18) are

$$\{\hat{\mathcal{J}}(\boldsymbol{\theta}, \mathbf{R}_{nn})\}_{jj} = 1 \quad \forall j,$$

and the off-diagonal elements are within the closed interval

$$-1 \leq \{\hat{\mathcal{J}}(\boldsymbol{\theta}, \mathbf{R}_{nn})\}_{jk} \leq 1 \quad j \neq k, \forall j, \forall k.$$

If the absolute value of an off-diagonal element is $|\{\hat{\mathcal{J}}(\boldsymbol{\theta})\}_{ik}| = 1$ the related parameter pair θ_i, θ_k carries identical information. Therefore, (4.18) is a first diagnosis tool to detect potential problems in the parameterisation of the model to be estimated. If two parameters θ_i, θ_k carry identical information, but are parameters of different propagation paths the parameter estimation problem can be solved if one of the propagation paths is dropped, i.e., removed from the model. However, in practice the parameter estimation problem is already ill-posed if one or several off-diagonal elements of (4.18) are close to 1 in magnitude.

As a rule of thumb, the smaller the magnitude of the off diagonal elements of the normalised Fisher information matrix (NFIM), the better the parameter estimation problem conditioned is.

4.1.2 Cramér-Rao Lower Bound for the DML Problem

The asymptotic Cramér-Rao lower bound on the covariance matrix of every unbiased estimator $\hat{\boldsymbol{\theta}}$ for the deterministic parameters $\boldsymbol{\theta}$ from the observation $\mathbf{x}$ is determined by the following inequality[7]

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathrm{E}\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \cdot (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}}\right\} \geq \mathcal{J}^{-1}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathbf{CRB}_{\boldsymbol{\theta}}. \qquad (4.20)$$

Using the result for the Fisher information matrix (4.16) in (4.20) yields the following explicit expression for the CRLB

$$\mathbf{CRB}_{\boldsymbol{\theta}} = \frac{1}{2} \cdot \left(\Re\{\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta}) \mathbf{R}_{nn}^{-1} \mathbf{D}(\boldsymbol{\theta})\}\right)^{-1}. \qquad (4.21)$$

It is often questioned whether or not this lower bound on the covariance matrix $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$ is practically relevant. In [49] Viberg et al. have shown, that the DML estimator $\hat{\boldsymbol{\theta}}$ is consistent

---

[6] Observe that the main diagonal elements of the Fisher information matrix are positive, since it is a covariance matrix.
[7] Here, the inequality of two matrices $\mathbf{A} \geq \mathbf{B}$ means, that the difference $\mathbf{A} - \mathbf{B}$ is positive semi-definite.

under the Gaussian noise assumption. Since the stochastic part of the radio channel observation $\mathbf{x}$ is Gaussian distributed, the DML estimator for the propagation path parameters is asymptotically efficient (for large M), i.e., the covariance matrix $\mathbf{C}_{\hat{\theta}}$ of the parameter estimates coincides with $\mathbf{CRB}_{\theta}$. This result has been verified for radio channel sounding measurements, using the ML (Maximum Likelihood) channel parameter estimation algorithm outlined in Chapter 5, by Landmann et al. in [36] (see also [37] and [38]).

Observe, if the Fisher information matrix of the model parameters is diagonal, i.e., the parameters are asymptotically uncoupled then the CRLB matrix is also diagonal.

$$\text{var}\{\hat{\boldsymbol{\theta}}\} \geq \text{diag}\{(\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}))^{-1}\}. \tag{4.22}$$

Now suppose all but one of the parameters of the model are known. Then the Cramér-Rao lower bound on the variance of this parameter θ_i is given by the reciprocal of the respective main diagonal entry of the Fisher information matrix $\{\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})\}_{ii}$. This implies that the Cramér-Rao lower bound of the individual parameters in the uncoupled case is asymptotically the same as the Cramér-Rao lower bound if only a single parameter is unknown. This is an important result for the development of a channel parameter estimator, implying that propagation paths with uncoupled parameters can be estimated separately.

4.1.3 Cramér-Rao Lower Bound of Physical Path Parameters

As we are usually interested in the lower bound of the variance of the estimated physical parameters, it may be tempting to use the channel model parameterised by the physical parameters as described in Chapter 2 to derive the respective CRLBs. However, this may be a time-consuming task since the model contains non-linear functions for mapping the physical parameters to the normalised parameters (see, e.g., Table 3-1). A better way to calculate the Cramér-Rao lower bound of the physical parameters is the invariance property of the Cramér-Rao lower bound under reparametrisation [40] (pp. 229), [50]. It is a chain rule for the Fisher information matrix and consequently also for the Cramér-Rao lower bound.

Suppose the vector valued function $\boldsymbol{\eta} = \boldsymbol{\rho}(\boldsymbol{\theta})$ is a nonsingular mapping of the normalised parameters $\boldsymbol{\theta} \in \mathbb{R}^{L \times 1}$ to the physical parameters $\boldsymbol{\eta} \in \mathbb{R}^{L \times 1}$, e.g., as outlined in Table 3-1. If the first order partial derivatives of $\boldsymbol{\rho}(\boldsymbol{\theta})$ exist, the Fisher information matrix of the physical parameters is related to the Fisher information matrix of the normalised parameters as follows

$$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\theta}} \, \mathcal{J}(\boldsymbol{\eta}, \mathbf{R}_{nn}) \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\theta}}^{\mathrm{T}}, \tag{4.23}$$

with

$$\mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\theta}} = \frac{\partial \boldsymbol{\rho}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \in \mathbb{R}^{L \times L}, \quad \{\mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\theta}}\}_{ik} = \frac{\partial \{\boldsymbol{\rho}(\boldsymbol{\theta})\}_k}{\partial \boldsymbol{\theta}_i}. \tag{4.24}$$

Hence, the CRLB of the physical parameters is related to the CRLB of the normalised parameters as follows

■ $$\mathbf{CRB}_{\hat{\boldsymbol{\eta}}} = \mathcal{J}^{-1}(\boldsymbol{\eta}, \mathbf{R}_{nn}) = \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\theta}}^{\mathrm{T}} \mathcal{J}^{-1}(\boldsymbol{\theta}, \mathbf{R}_{nn}) \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\theta}}. \tag{4.25}$$

Observe, that under the stated condition of a nonsingular mapping $\boldsymbol{\eta} = \boldsymbol{\rho}(\boldsymbol{\theta})$ also the following relation holds

$$\mathbf{CRB}_{\hat{\boldsymbol{\eta}}} = \mathcal{J}^{-1}(\boldsymbol{\eta}, \mathbf{R}_{nn}) = \mathbf{P}_{\boldsymbol{\theta}\boldsymbol{\eta}}^{-\mathrm{T}} \mathcal{J}^{-1}(\boldsymbol{\theta}, \mathbf{R}_{nn}) \mathbf{P}_{\boldsymbol{\theta}\boldsymbol{\eta}}^{-1} \tag{4.26}$$

with

$$\mathbf{P_{\theta\eta}} = \frac{\partial \boldsymbol{\rho}_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\mathrm{T}}} \in \mathbb{R}^{L \times L}, \quad \boldsymbol{\theta} = \boldsymbol{\rho}_i(\boldsymbol{\eta}), \tag{4.27}$$

since

$$\mathbf{P_{\theta\eta}} \mathbf{P_{\eta\theta}^{-1}} = \mathbf{I}.$$

In general, it is favourable to calculate the CRLB of the physical parameters via the CRLB of the normalised parameters. Observe also that the projection matrices (4.24) and (4.27) are sparse matrices, since one physical parameter is at most a function of two normalised parameters and vice versa.

4.1.4   Deterministic Cramér-Rao Lower Bounds for several canonical Models

Although the expression for the CRLB (4.22) is explicit it is somewhat unwieldy; to get more insight, let us discuss some examples.

Example 4-1:   One dimensional single path case, i.i.d. circular Gaussian noise, exponential model, single path weight (magnitude and phase).

Let us start with the simplest model. The observed radio channel consists of one propagation path only, i.e., $P = 1$. The stochastic process $\mathbf{n} \in \mathbb{C}^{M \times 1}$ is i.i.d. circular Gaussian with standard deviation σ. In addition, the basis function has only one data dimension with structure (1-D complex exponential model)

$$\mathbf{B}(\boldsymbol{\mu}) = \mathbf{a}(\mu) = \left[e^{-\mathrm{j}\cdot\mu\cdot\left(-\frac{M-1}{2}\right)} \; \ldots \; e^{-\mathrm{j}\cdot\mu\cdot\left(\frac{M-1}{2}\right)} \right]^{\mathrm{T}} \in \mathbb{C}^{M \times 1},$$

see also Section 3.6, Table 3-1 for a list of normalised parameters and the related basis functions $\mathbf{B}(\boldsymbol{\mu})$. Furthermore, there is only one path weight (e.g. non-polarimetric measurement) $\gamma = r\, e^{\mathrm{j}\phi}$. Consequently, the general data model (4.3) reduces to

$$\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n} = \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma} = \mathbf{a}(\mu) \cdot r\, e^{\mathrm{j}\phi} + \mathbf{n}.$$

This data model applies for example if the radio channel has been observed (i) using a ULA (μ - normalised angle), or (ii) using a SISO channel sounder having a flat transfer function (μ - normalised TDoA). The function mapping the parameter vector $\boldsymbol{\theta} = \left[\mu \; r \; \phi\right]^{\mathrm{T}} \in \mathbb{R}^{3 \times 1}$ to the observed channel is

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{a}(\mu) \cdot r\, e^{\mathrm{j}\phi}. \tag{4.28}$$

At first, we have to determine the Jacobian matrix of (4.28) being

$$\mathbf{D}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{s}(\boldsymbol{\theta}) = \left[-r\, e^{\mathrm{j}\varphi} \cdot \mathrm{j}\boldsymbol{\Xi} \cdot \mathbf{a}(\mu) \quad e^{\mathrm{j}\varphi} \cdot \mathbf{a}(\mu) \quad \mathrm{j} \cdot r\, e^{\mathrm{j}\varphi} \cdot \mathbf{a}(\mu) \; \right] \tag{4.29}$$

where $\boldsymbol{\Xi}$ is a diagonal matrix

$$\boldsymbol{\Xi} = \begin{bmatrix} -\frac{M-1}{2} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{M-1}{2} \end{bmatrix}. \tag{4.30}$$

Using the Jacobian matrix (4.29) in (4.17) yields the Fisher information matrix for Example 4-1 as

$$\mathcal{I}(\boldsymbol{\theta},\sigma^2) = \frac{2}{\sigma^2}\begin{bmatrix} r^2 \cdot \mathbf{a}^H(\mu)\cdot \boldsymbol{\Xi}^2 \cdot \mathbf{a}(\mu) & 0 & 0 \\ 0 & \mathbf{a}^H(\mu)\cdot \mathbf{a}(\mu) & 0 \\ 0 & 0 & r^2 \cdot \mathbf{a}^H(\mu)\cdot \mathbf{a}(\mu) \end{bmatrix}.$$

This shows that the Fisher information matrix is diagonal, meaning there is no coupling between parameters. Consequently, the Cramér-Rao lower bound on the variance of any unbiased estimator for $\boldsymbol{\theta}$ is a diagonal matrix and the main diagonal elements are

$$\text{CRB}_{\hat{\mu}} = \frac{\sigma^2}{r^2}\frac{1}{2M}\frac{12}{(M^2-1)} = \frac{\sigma^2}{r^2}\frac{6}{M\cdot(M^2-1)}, \tag{4.31}$$

$$\text{CRB}_{\hat{\phi}} = \frac{\sigma^2}{r^2}\frac{1}{2M}, \tag{4.32}$$

and

$$\text{CRB}_{\hat{r}} = \frac{\sigma^2}{2M}. \tag{4.33}$$

Observe, the term $\left(\frac{\sigma}{r}\right)^2$ in the expressions for the CRLB of $\hat{\mu}$, and $\hat{\phi}$. It is the inverse signal to noise ratio of the propagation path. Furthermore, the term $\frac{1}{2M}$ corresponds to the averaging gain over the $2M$ real valued measured data samples. That means the estimates of structural parameter μ and the phase of the path weight ϕ gain equally in accuracy if the SNR is increased. But whereas the variance of the parameters of the path weight decreases proportional to M, the bound on the variance decreases with M^3. The term $M\cdot(M^2-1)$ in equation (4.31) can be interpreted in the following way: The term (M^2-1) corresponds to the Rayleigh resolution, and M to the averaging gain, i.e., the noise reduction. The effective noise contribution to all parameter estimates is $\frac{\sigma^2}{2M}$.

Observe that the matrix (4.30) is an expression for the gain in accuracy with increasing aperture. The larger the apertures the larger are the outer elements in (4.30).

Example 4-2: One dimensional single-path case, i.i.d. circular Gaussian noise, exponential model, single path weight (real and imaginary components)

In the previous example the path weight was expressed by means of magnitude and phase. Since the phase is a nonlinear model parameter it is, from a signal processing point of view, better to express the weight by its real and imaginary part. Hence, we change the parameters of the model in Example 4-1 as follows

$$\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n} = \mathbf{B}(\mu)\cdot\boldsymbol{\gamma} = \mathbf{a}(\mu)\cdot(\gamma_r + j\gamma_i) + \mathbf{n}.$$

The new parameter vector is $\boldsymbol{\theta} = [\mu \ \gamma_r \ \gamma_i]^T \in \mathbb{R}^{3\times1}$, and the related channel model

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{a}(\mu)\cdot(\gamma_r + j\gamma_i).$$

The Jacobian matrix of $\mathbf{s}(\boldsymbol{\theta})$ is

$$\mathbf{D}(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}^T}\mathbf{s}(\boldsymbol{\theta}) = [-\gamma\cdot j\boldsymbol{\Xi}\cdot\mathbf{a}(\mu) \quad \mathbf{a}(\mu) \quad j\cdot\mathbf{a}(\mu) \ ]. \tag{4.34}$$

Again using the Jacobian matrix (4.34) in (4.17) yields the Fisher information matrix as

$$\mathcal{J}(\boldsymbol{\theta},\sigma^2)=\frac{2}{\sigma^2}\begin{bmatrix} |\gamma|^2 \cdot \mathbf{a}^{\mathrm{H}}(\mu)\cdot \mathbf{\Xi}^2 \cdot \mathbf{a}(\mu) & 0 & 0 \\ 0 & \mathbf{a}^{\mathrm{H}}(\mu)\cdot \mathbf{a}(\mu) & 0 \\ 0 & 0 & \mathbf{a}^{\mathrm{H}}(\mu)\cdot \mathbf{a}(\mu) \end{bmatrix}.$$

Since the Fisher information matrix is again diagonal, the CRLB matrix is also diagonal having entries on the main diagonal as follows

$$\mathrm{CRB}_{\hat{\mu}} = \frac{\sigma^2}{|\gamma|^2}\frac{1}{2M}\frac{12}{(M^2-1)}, \tag{4.35}$$

$$\mathrm{CRB}_{\hat{\gamma}_r} = \mathrm{CRB}_{\hat{\gamma}_i} = \frac{\sigma^2}{2M}. \tag{4.36}$$

It is apparent from (4.36) that the Cramér-Rao lower bounds on the linear weight parameters (real and imaginary part of the complex path weight) are equal. This means that if the observed radio channel contains only one propagation path, the estimates of the real and imaginary parts of the complex path weight have asymptotically the same variance.

Suppose the normalised parameter μ relates to the physical parameter ϑ via

$$\vartheta = \arcsin\left(\frac{\lambda\mu}{2\pi d_c}\right), \tag{4.37}$$

i.e., the radio channel has been measured with an column-ULA (cf. Table 2-2 and Table 3-1). Applying the invariance theorem of the CRLB (4.26) yields

$$CRB_{\hat{\vartheta}} = \frac{\lambda^2}{(2\pi d_c)^2\cos^2(\vartheta)}\cdot CRB_{\hat{\mu}} = \frac{\lambda^2}{(2\pi d_c)^2\sin^2(\vartheta)}\frac{\sigma^2}{|\gamma|^2}\frac{6}{M(M^2-1)}. \tag{4.38}$$

The CRLB of the parameters r and ϕ is unchanged, since the parameters are not part of the transformation (4.37). The complete projection matrix $\mathbf{P}_{\eta\theta}$ is

$$\mathbf{P}_{\eta\theta} = \begin{bmatrix} \frac{-\lambda}{2\pi d_c \sin(\vartheta)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The term $(\sin(\vartheta))^{-1}$ describes the influence of the effective ULA aperture as a function of the DoA or DoD on the variance of the estimate $\hat{\vartheta}$.

In the next example, we extend the one-dimensional data model to a two-dimensional problem.

Example 4-3:   Two-dimensional single path case, i.i.d. circular Gaussian noise, exponential model, single path weight (real and imaginary parts).

The extension of the one-dimensional single-path model in Example 4-2 is straight forward. The observation $\mathbf{x} \in \mathbb{C}^{M\times 1}$ has basically the same structure, i.e.,

$$\mathbf{x} = \mathbf{s}(\boldsymbol{\theta})+\mathbf{n} = \mathbf{B}(\boldsymbol{\mu})\cdot\boldsymbol{\gamma} = \mathbf{a}_{2D}(\boldsymbol{\mu})\cdot(\gamma_r+\mathrm{j}\gamma_i)+\mathbf{n}.$$

The function mapping the parameter vector $\boldsymbol{\theta} = \begin{bmatrix} \mu^{(1)} & \mu^{(2)} & \gamma_r & \gamma_i \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{4\times 1}$ to the observed channel is

$$\mathbf{s}(\mathbf{\theta}) = \gamma \cdot \mathbf{a}_{2D}(\mathbf{\mu}) = (\gamma_r + \mathrm{j}\gamma_i) \cdot \mathbf{a}_{2D}(\mathbf{\mu}), \tag{4.39}$$

where the basis function $\mathbf{B}(\mathbf{\mu})$ is given by (2-D complex exponential model)

$$\mathbf{B}(\mathbf{\mu}) = \mathbf{a}_{2D}(\mathbf{\mu}) = \mathbf{a}(\mu^{(2)}) \Diamond \, \mathbf{a}(\mu^{(1)}). \tag{4.40}$$

The two basis functions $\mathbf{a}(\mu^{(1)}) \in \mathbb{C}^{M_1 \times 1}$ and $\mathbf{a}(\mu^{(2)}) \in \mathbb{C}^{M_2 \times 1}$ are given by

$$\mathbf{a}(\mu^{(i)}) = \left[\mathrm{e}^{-\mathrm{j} \cdot \mu^{(i)} \cdot \left(-\frac{M_i-1}{2} \right)} \; \dots \; \mathrm{e}^{-\mathrm{j} \cdot \mu^{(i)} \cdot \left(\frac{M_i-1}{2} \right)} \right]^T \in \mathbb{C}^{M_i}. \tag{4.41}$$

The partial derivative of the data model (4.39) with respect to the parameters $\mathbf{\theta}$ is

$$\mathbf{D}(\mathbf{\theta}) = \frac{\partial}{\partial \mathbf{\theta}^T} \mathbf{s}(\mathbf{\theta}) = \left[\gamma \, \mathbf{a}(\mu^{(2)}) \Diamond \left(-\mathrm{j}\mathbf{\Xi}_1 \mathbf{a}(\mu^{(1)}) \right) \quad \left(-\mathrm{j}\gamma \, \mathbf{\Xi}_2 \mathbf{a}(\mu^{(2)}) \right) \Diamond \, \mathbf{a}(\mu^{(1)}) \quad \mathbf{a}_{2D}(\mathbf{\mu}) \quad \mathrm{j}\mathbf{a}_{2D}(\mathbf{\mu}) \right] \tag{4.42}$$

with

$$\mathbf{\Xi}_i = \begin{bmatrix} -\frac{M_i-1}{2} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{M_i-1}{2} \end{bmatrix}.$$

The Fisher information matrix is diagonal again

$$\mathcal{J}(\mathbf{\theta}, \sigma^2) = \frac{2}{\sigma^2} \begin{bmatrix} M & 0 & 0 & 0 \\ 0 & M & 0 & 0 \\ 0 & 0 & |\gamma|^2 M \frac{M_1^2-1}{12} & 0 \\ 0 & 0 & 0 & |\gamma|^2 M \frac{M_2^2-1}{12} \end{bmatrix}, \tag{4.43}$$

i.e., the lower bound on the covariance between the parameter estimates $\hat{\theta}_i, \hat{\theta}_k$ is $\mathrm{E}\{\hat{\theta}_i \hat{\theta}_k\} = 0 \quad i \neq k, \forall i, \forall k$. Consequently, the matrix $\mathbf{CRB}_{\hat{\mathbf{\theta}}}$ is diagonal too having main diagonal entries

$$\mathrm{CRB}_{\hat{\mu}^{(1)}} = \frac{\sigma^2}{|\gamma|^2} \frac{6}{M \cdot (M_1^2 - 1)}, \tag{4.44}$$

$$\mathrm{CRB}_{\hat{\mu}^{(2)}} = \frac{\sigma^2}{|\gamma|^2} \frac{6}{M \cdot (M_2^2 - 1)}, \tag{4.45}$$

and

$$\mathrm{CRB}_{\hat{\gamma}_r} = \mathrm{CRB}_{\hat{\gamma}_i} = \frac{\sigma^2}{2M}. \tag{4.46}$$

The results for the lower bound on the variance of the normalised structural parameters $\mu^{(i)}$ support the observation that the term $(M_i^2 - 1)$ corresponds to the Rayleigh resolution limit. Furthermore, it is intriguing to see that an increase of the array aperture in one dimension reduces the CRLB in the other dimension linearly. This is reasonable since we effectively increase the number of independent observations. The interesting fact is that the structure, e.g., the complex exponential model, of one dimension has no influence on the Cramér-Rao

lower bound of the other dimension. In the next example we generalize the result from the 2-D data model to the R-D data model.

Example 4-4: R-dimensional single path case, i.i.d. circular Gaussian noise, exponential model, single path weight (real and imaginary part).

The observation $\mathbf{x} \in \mathbb{C}^{M \times 1}$ has the same general structure as in the previous examples, i.e.,

$$\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n} = \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma} = \mathbf{a}_{RD}(\boldsymbol{\mu}) \cdot (\gamma_r + \mathrm{j}\gamma_i) + \mathbf{n} .$$

The function mapping the parameter vector $\boldsymbol{\theta} = \begin{bmatrix} \mu^{(1)} & \cdots & \mu^{(R)} & \gamma_r & \gamma_i \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(R+2) \times 1}$ to the observed channel is

$$\mathbf{s}(\boldsymbol{\theta}) = \gamma \cdot \mathbf{a}_{RD}(\boldsymbol{\mu}) = (\gamma_r + \mathrm{j}\gamma_i) \cdot \mathbf{a}_{RD}(\boldsymbol{\mu}), \tag{4.47}$$

here the basis function $\mathbf{B}(\boldsymbol{\mu})$ is given by (R-D complex exponential model)

$$\mathbf{a}_{RD}(\boldsymbol{\mu}) = \mathbf{a}\big(\mu^{(R)}\big) \lozenge \ldots \lozenge \mathbf{a}\big(\mu^{(1)}\big),$$

and the vector valued function mapping the parameters $\mu^{(i)}$ to the related complex exponential is defined as in equation (4.41). The total number of samples is determined by the number of samples of the individual data domains, i.e.,

$$M = \prod_{r=1}^{R_d} M_r .$$

The Jacobian matrix and the Fisher information matrix of (4.47) have the same structure as in (4.42) and (4.43), respectively. Consequently the CRLB for the normalised structural parameters $\hat{\mu}^{(r)}$ is

$$\mathrm{CRB}_{\hat{\mu}^{(r)}} = \frac{\sigma^2}{|\gamma|^2} \frac{6}{M \cdot \big(M_r^2 - 1\big)} \quad \forall r . \tag{4.48}$$

And the lower bound on the variance of any estimator for $\Re\{\hat{\gamma}\}$ and $\Im\{\hat{\gamma}\}$ is given by

$$\mathrm{CRB}_{\hat{\gamma}_r} = \mathrm{CRB}_{\hat{\gamma}_i} = \frac{\sigma^2}{2M} . \tag{4.49}$$

Naturally, the results of Example 4-2 and Example 4-3 are special cases of (4.48), and (4.49).

So far, we have not discussed the validity range of the bounds (4.48), and (4.49). In channel sounding measurements, we encounter some propagation paths with a high SNR, and many propagation paths with a low SNR. As already discussed in Section 4.1.2 the bounds (4.48) and (4.49) are reasonable in the high SNR case, i.e.,

$$\frac{|\gamma|^2}{\sigma^2} \cdot \frac{1}{2M} \gg 1 .$$

However, suppose the path weight has an absolute value of

$$|\gamma| = \frac{\sigma}{\sqrt{M}} , \tag{4.50}$$

then (4.48) becomes

$$\text{CRB}_{\hat{\mu}^{(r)}} = \frac{6}{\left(M_r^2 - 1\right)}, \tag{4.51}$$

and the CRLB on the relative variance of the magnitude estimate is due to (4.33)

$$\text{var}_{rel}\{|\gamma|\} = \frac{\text{CRB}_{|\gamma|}}{|\gamma|^2} = \frac{1}{2}.$$

In other words, the estimate of the path magnitude has an SNR of 3dB. This is the lowest limit for equations (4.48) and (4.49) to be reasonable. Moreover, let us compare (4.51) with the Rayleigh limit on resolution $\frac{2\pi}{M_r}$. For the minimum aperture $M_r = 2$ the ratio between the Cramér-Rao lower bound and the Rayleigh limit on resolution of the structural parameter is given by

$$\sqrt{\frac{\text{CRB}_{\hat{\mu}_r}}{\left(\frac{2\pi}{M_r}\right)^2}} = \frac{\sqrt{2}}{\pi} \approx 0.45, \quad M_r = 2, |\gamma| = \frac{\sigma}{\sqrt{M}}.$$

As M_r goes to infinity this ratio becomes

$$\lim_{M_r \to \infty} \sqrt{\frac{\text{CRB}_{\hat{\mu}^{(r)}}}{\left(\frac{2\pi}{M_r}\right)^2}} = \lim_{M_r \to \infty} \sqrt{\frac{6}{M_r^2 - 1} \cdot \frac{M_r}{2\pi}} = \frac{\sqrt{6}}{2\pi} \approx 0.39, \quad |\gamma| = \frac{\sigma}{\sqrt{M}}.$$

Hence, the CRLB on the variance of the structural parameter $\mu^{(r)}$ is roughly ½ of the Rayleigh limit on resolution provided the magnitude of the path weight is $\sigma/\sqrt{M}$. Furthermore observe that for detection a SNR of 3dB is a low threshold. That means there is a high probability (13.5%) that the detected path is part of the stochastic process and should therefore not be classified as a concentrated propagation path. Altogether we can state that the equations for the Cramér-Rao lower bound derived in Example 4-1 to Example 4-5 of the deterministic parameters provides reasonable results for propagation paths having a path weight of at least $|\gamma| > \sigma/\sqrt{M}$. That means they have a significant power and can be classified as dominant propagation paths.

Up to now we have only discussed cases with a single propagation path. In the next example we extend the data model to the more general multipath case.

Example 4-5: Two-dimensional multipath case, i.i.d. circular Gaussian noise, exponential model with one path weight (real and imaginary parts) for every propagation path.

The basic structure of the channel observation is unchanged, being

$$\mathbf{x} = \mathbf{s}(\mathbf{\theta}) + \mathbf{n}.$$

The function describing the contribution of the P observed concentrated propagation paths is

$$\mathbf{s}(\mathbf{\theta}) = \mathbf{B}(\mathbf{\mu}) \cdot \mathbf{\gamma} = \mathbf{A}(\mathbf{\mu}) \cdot (\mathbf{\gamma}_r + \mathrm{j}\mathbf{\gamma}_i), \tag{4.52}$$

where $\mathbf{A}(\mathbf{\mu}) = \mathbf{A}(\mathbf{\mu}^{(2)}) \Diamond \mathbf{A}(\mathbf{\mu}^{(1)})$ is a matrix containing the Khatri-Rao product of the basis matrices $\mathbf{A}(\mathbf{\mu}^{(1)})$ and $\mathbf{A}(\mathbf{\mu}^{(2)})$ (see (3.4) for a definition) and

$$\mathbf{\theta} = \left[\mathbf{\mu}^{(1)} \ \mathbf{\mu}^{(2)} \ \mathbf{\gamma}_r \ \mathbf{\gamma}_i\right]^{\mathrm{T}} = \left[\mathbf{\mu} \ \mathbf{\gamma}_r \ \mathbf{\gamma}_i\right]^{\mathrm{T}} \in \mathbb{R}^{4P \times 1}$$

is the related parameter vector. Observe that the structural parameters $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are merged into one vector $\boldsymbol{\mu} = \left[\boldsymbol{\mu}^{(1)}\ \boldsymbol{\mu}^{(2)}\right]^{\mathrm{T}} \in \mathbb{R}^{2P \times 1}$.

Now we define the matrix containing the first order partial derivatives of the matrix-valued function $\mathbf{A}(\boldsymbol{\mu}^{(r)})$ with respect to the related parameters $\boldsymbol{\mu}^{(r)}$ as

$$\mathbf{D}^{(r)} = \left[\frac{\partial}{\partial \mu_1^{(r)}}\mathbf{a}(\mu^{(r)}) \cdots \frac{\partial}{\partial \mu_P^{(r)}}\mathbf{a}(\mu^{(r)})\right] = -\mathrm{j}\boldsymbol{\Xi}_r \mathbf{A}(\boldsymbol{\mu}^{(r)}). \tag{4.53}$$

The Jacobian matrix of (4.52) can be expressed using the definitions (4.30), (4.53), and the diagonal matrix $\boldsymbol{\Gamma} = \mathrm{diag}\{\boldsymbol{\gamma}\}$ in a compact form by

$$\mathbf{D}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\mathbf{s}(\boldsymbol{\theta}) = \left[\mathbf{A}(\boldsymbol{\mu}^{(2)})\Diamond\boldsymbol{\Xi}_1\boldsymbol{\Gamma} \quad \boldsymbol{\Xi}_2\Diamond\mathbf{A}(\boldsymbol{\mu}^{(1)})\boldsymbol{\Gamma} \quad \mathbf{A}(\boldsymbol{\mu}) \quad \mathrm{j}\mathbf{A}(\boldsymbol{\mu})\right]. \tag{4.54}$$

To gain further insight into the structure of the Jacobian (4.54) we factorize it using the Khatri-Rao product of the following matrices

$$\mathbf{D}^{(0)} = \left[\boldsymbol{\gamma}^{\mathrm{T}}\ \boldsymbol{\gamma}^{\mathrm{T}}\ \mathbf{1}^{\mathrm{T}}\ \mathrm{j}\mathbf{1}^{\mathrm{T}}\right], \tag{4.55}$$

$$\mathbf{D}_1 = \left[\mathbf{D}^{(1)}\ \mathbf{A}(\boldsymbol{\mu}^{(1)})\ \mathbf{A}(\boldsymbol{\mu}^{(1)})\ \mathbf{A}(\boldsymbol{\mu}^{(1)})\right], \tag{4.56}$$

and

$$\mathbf{D}_2 = \left[\mathbf{A}(\boldsymbol{\mu}^{(2)})\ \mathbf{D}^{(2)}\ \mathbf{A}(\boldsymbol{\mu}^{(2)})\ \mathbf{A}(\boldsymbol{\mu}^{(2)})\right]. \tag{4.57}$$

Observe that the matrices (4.56) and (4.57) are both block matrices with a regular structure. The Jacobian matrix is given by the simple expression

$$\mathbf{D}(\boldsymbol{\theta}) = \mathbf{D}_2\Diamond\mathbf{D}_1\Diamond\mathbf{D}_0. \tag{4.58}$$

This expression can be easily extended to the multidimensional case due to its regular structure. This is especially important in numerical implementations, since it provides a simple means to scale the dimensionality of the model.

Since for the Gram matrix of the Khatri-Rao product (4.58) the following identity[8] (cf. Appendix C) $(\mathbf{E}\Diamond\mathbf{F})^{\mathrm{H}}(\mathbf{E}\Diamond\mathbf{F}) = (\mathbf{E}^{\mathrm{H}}\mathbf{E})\circ(\mathbf{F}^{\mathrm{H}}\mathbf{F})$ holds, the Fisher information matrix (FIM) for the parameter estimation problem at hand is

$$\mathcal{J}(\boldsymbol{\theta}, \sigma^2) = \frac{2}{\sigma^2}\cdot\Re\left\{\left(\mathbf{D}_0^{\mathrm{H}}\cdot\mathbf{D}_0\right)\circ\left(\mathbf{D}_1^{\mathrm{H}}\cdot\mathbf{D}_1\right)\circ\left(\mathbf{D}_2^{\mathrm{H}}\cdot\mathbf{D}_2\right)\right\}. \tag{4.59}$$

The main diagonal entries of (4.59) are the same as in the single source case, i.e., the elements of (4.43) are closed form expressions for the related elements in $\mathrm{diag}\{\mathcal{J}(\boldsymbol{\theta}, \sigma^2)\}$. Furthermore, the off-diagonal elements of the FIM belonging to the same propagation path are all zero due to (4.43), i.e., the parameters are asymptotically uncoupled.

The remaining elements describe the coupling of parameters of different propagation paths. As already discussed in Section 4.1.1 a model having a diagonal FIM is a good model since all parameters are orthogonal, whereas a model having strong off-diagonal elements in its FIM is a bad model since it contains strongly coupled parameters, i.e., they are not orthogonal. To understand, which parameter combinations lead to a bad parameterised and consequently ill-posed problem we have to analyze the off-diagonal elements belonging to different propagation paths further. First observe, that the Gram matrices of (4.56) and (4.57) are both real matrices, i.e., $\mathbf{D}_1^{\mathrm{H}}\mathbf{D}_1 \in \mathbb{R}^{4P \times 4P}$ and $\mathbf{D}_2^{\mathrm{H}}\mathbf{D}_2 \in \mathbb{R}^{4P \times 4P}$. This is due to the fact, that the

---

[8] The symbol $\circ$ denotes the Hadamard or Schur product (element-wise product).

sub-matrices $\mathbf{A}(\boldsymbol{\mu}^{(1)})$, $\mathbf{A}(\boldsymbol{\mu}^{(2)})$, $\mathbf{D}^{(1)}$, and $\mathbf{D}^{(2)}$ are all conjugate symmetric along the columns (cf. Appendix C). Hence (4.59) becomes

$$\mathcal{J}(\boldsymbol{\theta}, \sigma^2) = \frac{2}{\sigma^2} \cdot \Re\{(\mathbf{D}_0^{\mathrm{H}} \cdot \mathbf{D}_0)\} \circ (\mathbf{D}_1^{\mathrm{H}} \cdot \mathbf{D}_1) \circ (\mathbf{D}_2^{\mathrm{H}} \cdot \mathbf{D}_2). \tag{4.60}$$

Now supposing that the path weights of two propagation paths are orthogonal, that means their phase difference is 90° or -90°. Then all elements of the Gram matrix $\Re\{\mathbf{D}_0^{\mathrm{H}}\mathbf{D}_0\} \in \mathbb{R}^{4P \times 4P}$ contributing to the coupling of the structural parameters are zero. Hence, we can state that a two propagation path model with orthogonal path weights constitute the best case in terms of the path weights. This effect has been observed in measurements carried out for channel sounder performance evaluation [36], [37].

Finally, let us investigate the dependency of the elements of the four basic block matrices $\mathbf{A}^{\mathrm{H}}(\boldsymbol{\mu}^{(i)}) \cdot \mathbf{A}(\boldsymbol{\mu}^{(i)})$, $\mathbf{A}^{\mathrm{H}}(\boldsymbol{\mu}^{(i)}) \cdot \mathbf{D}^{(i)}$, $(\mathbf{D}^{(i)})^{\mathrm{H}} \mathbf{A}(\boldsymbol{\mu}^{(i)})$, and $(\mathbf{D}^{(i)})^{\mathrm{H}} \cdot \mathbf{D}^{(i)}$ on the structural parameters $\boldsymbol{\mu}^{(i)}$. Observe that the correlation, in the algebraic sense, of two arbitrary vectors $\mathbf{a}(\mu_1^{(i)})$, $\mathbf{a}(\mu_2^{(i)})$ does not depend on the absolute value of $\mu_1^{(i)}$ and $\mu_2^{(i)}$, but rather on their distance $\Delta\mu^{(i)} = \mu_2^{(i)} - \mu_1^{(i)}$. The same applies to the inner product of the vectors $\mathbf{d}(\mu_1^{(i)})$, $\mathbf{a}(\mu_2^{(i)})$ and $\mathbf{d}(\mu_1^{(i)})$, $\mathbf{d}(\mu_2^{(i)})$, where $\mathbf{d}(\mu^{(i)})$ is according to (4.53) defined as $\mathbf{d}(\mu^{(i)}) = -\mathrm{j}\boldsymbol{\Xi}_r \mathbf{a}(\mu^{(i)})$.
Figure 4-1 and Figure 4-2 show the dependency of the three inner products on the parameter distance $\Delta\mu$. The correlation function on the left hand side of Figure 4-1 is due to the exponential data model a periodic sinc-function $\frac{2}{\mu}\sin(\frac{M}{2}\mu)$. Observe, that no parameter distance $\Delta\mu$ exist where all vectors are orthogonal. The vectors $\mathbf{a}(\mu_1^{(i)})$, $\mathbf{a}(\mu_2^{(i)})$ are orthogonal for a parameter distance of $\Delta\mu^{(i)} = \mu_2^{(i)} - \mu_1^{(i)} = k\frac{2\pi}{M_i}$ for all $\forall|k| \in \mathbb{N}$ with $|k| \neq i \cdot M$ $\forall i \in \mathbb{N}$, whereas the related inner products $\mathbf{d}(\mu_1^{(i)})$, $\mathbf{a}(\mu_2^{(i)})$ and $\mathbf{d}(\mu_1^{(i)})$, $\mathbf{d}(\mu_2^{(i)})$ are not equal to zero at the same points. Nevertheless, all correlations are decaying with increasing parameter distance.

In general, we can state that usually some of the parameters of several propagation paths in a given scenario are strongly coupled and some propagation paths are only slightly coupled. Thereby, the Fisher information matrix or more precisely the normalised Fisher information matrix (cf. Section 4.1.1) determines which paths can be treated as uncoupled. Strictly speaking, the structure of the FIM already determines the structure of the parameter estimator for a given estimation task, i.e., whether parameters can be estimated separately or have to be processed jointly.

Equation (4.59) shows in the same context mathematically that an additionally measured parameter dimension improves the resolution of the channel sounding system, i.e., the resolution capabilities of the entity channel sounder and parameter estimator. This is due to the fact, that the off diagonal elements of the Gram matrices $\mathbf{D}_i^{\mathrm{H}}\mathbf{D}_i \in \mathbb{R}^{4P \times 4P}$ are always smaller or equal to the related main diagonal elements. Observe, that the main diagonal elements of $\mathbf{D}_i^{\mathrm{H}}\mathbf{D}_i \in \mathbb{R}^{4P \times 4P}$ are positive real values.

Furthermore, if we can measure an additional dimension, i.e., add a new aperture dimension, and have to choose between two parameter dimensions, we should choose the parameter dimension with the largest parameter spread since it provides the highest increase in resolution. The higher the parameter spread, the higher the probability that some basis vectors in $\mathbf{B}(\boldsymbol{\mu})$ are nearly orthogonal.

Hence, for parameter estimation, the optimal parameter distribution is the uniform distribution. This statement implies that the channel sounder maps the physical parameters linearly to the normalised parameters.
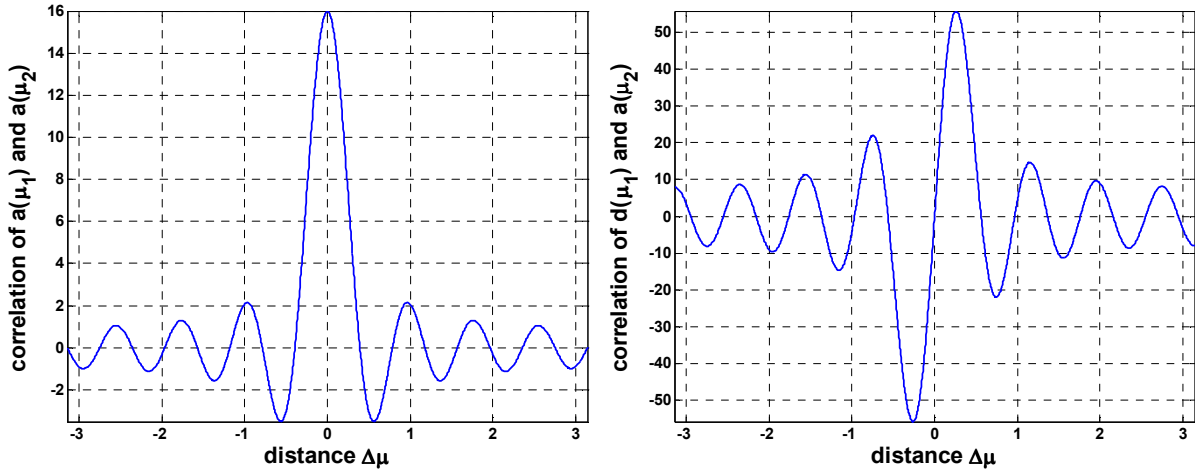
Figure 4-1: Correlation between two vectors $\mathbf{a}(\mu_1)$, $\mathbf{a}(\mu_2)$ (left hand side), and $\mathbf{a}(\mu_1)$, $\mathbf{d}(\mu_2)$ (right hand side) as a function of the parameter distance $\Delta\mu = \mu_2 - \mu_1$
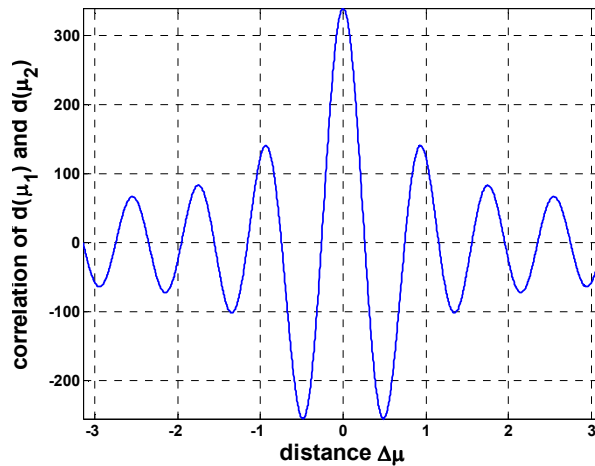
Figure 4-2: Correlation between the vectors $\mathbf{d}(\mu_1)$ and $\mathbf{d}(\mu_2)$ as a function of the parameter distance $\Delta\mu = \mu_2 - \mu_1$

This first requires that the effective aperture is not a function of the parameter itself; consequently a ULA or URA is a suboptimal antenna array structure, provided the DoAs and DoDs are uniformly distributed. Using the same argument we can state that a UCA, CUBA, or SUCPA (stacked uniform circular patch array) is a much more suitable array structure for channel sounding applications. Nevertheless, those are only general rules. As an example, a SUCPA is a better choice for outdoor applications than a spherical array. It is unlikely that we encounter a uniform angular distribution in elevation in outdoor scenarios. However, the reverse is true for indoor measurements.

4.1.5 Inherent Limits on the Variance of the Deterministic Parameters

In view of the expressions for the Cramér-Rao lower bound of the deterministic parameters in Example 4-1, Example 4-2, Example 4-3, and Example 4-5 one may be tempted to assume that the variance of the parameter estimates can be made arbitrarily small, just by reducing the measurement noise. However, this is not true since the stochastic part $\mathbf{n}$ of the observation $\mathbf{x}$ contains not only measurement noise but also the contribution of the dense multipath components. The limit of (4.2) for $\alpha_0 \to 0$ is

$$\mathbf{R}_{nn} = \mathbf{R}(\mathbf{\theta}_{dmc}).$$

Consequently, the variance of any parameter estimator for the deterministic channel parameters in the high SNR case is inherently limited by the radio channel itself and the apertures of the measurement device. Hence, the only way to improve the accuracy and resolution of the parameter estimates is to increase the number of samples, i.e., increasing the aperture in space, frequency, or time. Notice in the same context that an extension of the measurement apertures would inevitably lead to a more complex data model. This is because simplifications, such as frequency independents of the antenna array beam patterns will not hold anymore. Consequently, the data model has to be refined leading to a more complex model and therefore to a more complex estimation algorithm.

4.2 Expression for the Jacobian and the Fisher Information Matrix

In the previous section only the general form of the FIM and the Cramér-Rao lower bound has been derived. In this section we summarise the expressions for the Jacobian matrix and the FIM for the complete data-model derived in Chapter 2., i.e.,

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{B}_{R_H} \lozenge \mathbf{B}_{T_H} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH} + \mathbf{B}_{R_V} \lozenge \mathbf{B}_{T_H} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HV} +$$
$$+ \mathbf{B}_{R_H} \lozenge \mathbf{B}_{T_V} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{VH} + \mathbf{B}_{R_V} \lozenge \mathbf{B}_{T_V} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{VV} . \tag{4.61}$$

Since the variety of possible measurement setups is enormous, we put the focus on the modularity of the expressions. The expressions are structured in such a way that all possible models, i.e., measurement configurations, are covered.

First, recall that each structural parameter is, as already discussed in Section 2.4.5, directly related to a specific data dimension. Table 4-1 summarises the associations between the data dimensions, the physical parameters and the normalised parameters.

Table 4-1: Relationship between data dimensions and parameters

	Aperture / Data Domain	Parameters	
		Physical	Normalised
Temporal samples	Time t	Doppler-shift α	$\mu^{(\alpha)}$
Frequency samples	Frequency f	Time delay of arrival τ	$\mu^{(\tau)}$
Tx antenna ports	Space at Tx	Azimuth of Departure φ_T Elevation of Departure ϑ_T	$\mu^{(\varphi_T)}$ $\mu^{(\vartheta_T)}$
Rx antenna ports	Space at Rx	Azimuth of Arrival φ_R Elevation of Arrival ϑ_R	$\mu^{(\varphi_R)}$ $\mu^{(\vartheta_R)}$

All matrices $\mathbf{B}_r \in \mathbb{C}^{M_r \times P}$ have as shown in Section 3.6 $r \in [f, t, T_H, T_V, R_H, R_V]$ the same basic structure, i.e.,

$$\mathbf{B}_r = \mathbf{G}_r \cdot \mathbf{A}(\boldsymbol{\mu}^{(i)}) \tag{4.62}$$

if the data domain r contains information about one parameter. If the data domain contains information about two parameters, e.g., azimuth and elevation the structure of $\mathbf{B}_r$ is

$$\mathbf{B}_r = \mathbf{G}_r \cdot \left(\mathbf{A}\big(\boldsymbol{\mu}^{(k)}\big) \lozenge \mathbf{A}\big(\boldsymbol{\mu}^{(i)}\big) \right). \tag{4.63}$$

Since each parameter is related to one column of $\mathbf{A}(\boldsymbol{\mu})$, and thus also to one column of $\mathbf{B}_r$ the first order partial derivatives with respect to the structural parameters $\boldsymbol{\mu}$ can be expressed in compact form as

■
$$\mathbf{D}_i = -\mathrm{j}\mathbf{G}_r\boldsymbol{\Xi}_r\mathbf{A}\big(\boldsymbol{\mu}^{(i)}\big) \tag{4.64}$$

for (4.62), and as

■
$$\mathbf{D}_i = -\mathrm{j}\mathbf{G}_r\left(\mathbf{A}\big(\boldsymbol{\mu}^{(k)}\big) \lozenge \boldsymbol{\Xi}_i\mathbf{A}\big(\boldsymbol{\mu}^{(i)}\big) \right), \quad \mathbf{D}_k = -\mathrm{j}\mathbf{G}_r\left(\boldsymbol{\Xi}_k\mathbf{A}\big(\boldsymbol{\mu}^{(k)}\big) \lozenge \mathbf{A}\big(\boldsymbol{\mu}^{(i)}\big) \right) \tag{4.65}$$

for (4.63) with $i,k \in [\alpha, \tau, \varphi_T, \vartheta_T, \varphi_R, \vartheta_R]$. The matrix $\boldsymbol{\Xi}_i \in \mathbb{C}^{N_i \times N_i}$ is a diagonal weighting matrix having the same definition as (4.30).

4.2.1 Full Polarimetric Model

Let us start with derivation of expressions for the Jacobian matrix and the Fisher information matrix for the full polarimetric model (4.61). Observe that this model is only applicable if the radio channel measurements provide sufficient information about the horizontal and vertical polarisation at both link ends. First, define the following matrices,

$$\mathbf{B}_H = \mathbf{B}_{R_H} \lozenge \boldsymbol{\gamma}_{HH}^T + \mathbf{B}_{R_V} \lozenge \boldsymbol{\gamma}_{HV}^T, \tag{4.66}$$

$$\mathbf{B}_V = \mathbf{B}_{R_H} \lozenge \boldsymbol{\gamma}_{VH}^T + \mathbf{B}_{R_V} \lozenge \boldsymbol{\gamma}_{VV}^T$$

and the related partial derivatives using (4.64), (4.65) as

$$\mathbf{D}_{\varphi_R,H} = \mathbf{D}_{R_H,\varphi_R} \lozenge \boldsymbol{\gamma}_{HH}^T + \mathbf{D}_{R_V,\varphi_R} \lozenge \boldsymbol{\gamma}_{HV}^T, \quad \mathbf{D}_{\vartheta_R,H} = \mathbf{D}_{R_H,\vartheta_R} \lozenge \boldsymbol{\gamma}_{HH}^T + \mathbf{D}_{R_V,\vartheta_R} \lozenge \boldsymbol{\gamma}_{HV}^T, \tag{4.67}$$

$$\mathbf{D}_{\varphi_R,V} = \mathbf{D}_{R_H,\varphi_R} \lozenge \boldsymbol{\gamma}_{VH}^T + \mathbf{D}_{R_V,\varphi_R} \lozenge \boldsymbol{\gamma}_{VV}^T, \text{ and } \mathbf{D}_{\vartheta_R,V} = \mathbf{D}_{R_H,\vartheta_R} \lozenge \boldsymbol{\gamma}_{VH}^T + \mathbf{D}_{R_V,\vartheta_R} \lozenge \boldsymbol{\gamma}_{VV}^T.$$

With these definitions, we are prepared to form component matrices of the Jacobian matrix for (4.61). Table 4-2 summarises the elementary matrices.

Table 4-2: Component matrices of the Jacobian matrix for the complete data model.

$\boldsymbol{\theta} \rightarrow$	α	τ	φ_T	ϑ_T	φ_R	ϑ_R	$\gamma_{HH,r}$	$\gamma_{HH,i}$	$\gamma_{HV,r}$	$\gamma_{HV,i}$	$\gamma_{VH,r}$	$\gamma_{VH,i}$	$\gamma_{VV,r}$	$\gamma_{VV,i}$	
$\mathbf{D}_t =$	$[\mathbf{D}_\alpha$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$]$
$\mathbf{D}_f =$	$[\mathbf{B}_f$	$\mathbf{D}_\tau$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$]$
$\mathbf{D}_{T_H} =$	$[\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{D}_{\varphi_T,H}$	$\mathbf{D}_{\vartheta_T,H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_V}$	$\mathbf{B}_{T_V}$	$\mathbf{B}_{T_V}$	$\mathbf{B}_{T_V}$	$]$
$\mathbf{D}_{T_V} =$	$[\mathbf{B}_{T_V}$	$\mathbf{B}_{T_V}$	$\mathbf{D}_{\varphi_T,V}$	$\mathbf{D}_{\vartheta_T,V}$	$\mathbf{B}_{T_V}$	$\mathbf{B}_{T_V}$	0	0	0	0	0	0	0	0	$]$
$\mathbf{D}_{R_H} =$	$[\mathbf{B}_H$	$\mathbf{B}_H$	$\mathbf{B}_H$	$\mathbf{B}_H$	$\mathbf{D}_{\varphi_R,H}$	$\mathbf{D}_{\vartheta_R,H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_V}$	$\mathbf{B}_{R_V}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_V}$	$\mathbf{B}_{R_V}$	$]$
$\mathbf{D}_{R_V} =$	$[\mathbf{B}_V$	$\mathbf{B}_V$	$\mathbf{B}_V$	$\mathbf{B}_V$	$\mathbf{D}_{\varphi_R,V}$	$\mathbf{D}_{\vartheta_R,V}$	0	0	0	0	0	0	0	0	$]$
$\mathbf{D}_g =$	$[\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T\mathrm{j}$	$\mathbf{1}^T$	$\mathbf{1}^T\mathrm{j}$	$\mathbf{1}^T$	$\mathbf{1}^T\mathrm{j}$	$\mathbf{1}^T$	$\mathbf{1}^T\mathrm{j}$	$]$

With the definitions of the matrices $\mathbf{D}_t$, $\mathbf{D}_f$, $\mathbf{D}_{T_H}$, $\mathbf{D}_{T_V}$, $\mathbf{D}_{R_H}$, $\mathbf{D}_{R_V}$, and $\mathbf{D}_g$ the Jacobian matrix can be represented in compact form as

■
$$\mathbf{D} = \left(\mathbf{D}_{R_V} \lozenge \mathbf{D}_{T_V} + \mathbf{D}_{R_H} \lozenge \mathbf{D}_{T_H} \right) \lozenge \mathbf{D}_g \lozenge \mathbf{D}_f \lozenge \mathbf{D}_t. \tag{4.68}$$

Assuming the covariance matrix of the stochastic process $\mathbf{n}$ is $\mathbf{R}_{nn} = \sigma^2 \mathbf{I}$ the Fisher information matrix of the normalised deterministic parameters of the radio channel is given by

$$\mathcal{J}(\boldsymbol{\theta},\sigma^2)=\frac{2}{\sigma^2}\Re\left\{\mathbf{D}_t^H\mathbf{D}_t\circ\mathbf{D}_f^H\mathbf{D}_f\circ\mathbf{D}_g^H\mathbf{D}_g\circ\right.$$

$$\left.\circ\left(\mathbf{D}_{T_H}^H\mathbf{D}_{T_H}\circ\mathbf{D}_{R_H}^H\mathbf{D}_{R_H}+\mathbf{D}_{T_H}^H\mathbf{D}_{T_V}\circ\mathbf{D}_{R_H}^H\mathbf{D}_{R_V}+\mathbf{D}_{T_V}^H\mathbf{D}_{T_H}\circ\mathbf{D}_{R_V}^H\mathbf{D}_{R_H}+\mathbf{D}_{T_V}^H\mathbf{D}_{T_V}\circ\mathbf{D}_{R_V}^H\mathbf{D}_{R_V}\right)\right\}.$$
(4.69)

Observe, that (4.69) follows from (4.17), (4.68), and $(\mathbf{E}\Diamond\mathbf{F})^H(\mathbf{E}\Diamond\mathbf{F})=(\mathbf{E}^H\mathbf{E})\circ(\mathbf{F}^H\mathbf{F})$. Likewise, the Fisher information matrix involving the covariance matrix $\mathbf{R}_{nn}\in\mathbb{C}^{M\times M}$ having structure $\mathbf{R}_{nn}=\mathbf{I}\otimes\mathbf{I}\otimes\mathbf{R}_f\otimes\mathbf{I}$ (cf. Section 2.5.3), is given by

$$\blacksquare\qquad \mathcal{J}(\boldsymbol{\theta},\mathbf{R}_{nn})=2\cdot\Re\left\{\mathbf{D}_t^H\mathbf{D}_t\circ\mathbf{D}_f^H\mathbf{R}_f^{-1}\mathbf{D}_f\circ\mathbf{D}_g^H\mathbf{D}_g\circ\right.$$

$$\left.\circ\left(\mathbf{D}_{T_H}^H\mathbf{D}_{T_H}\circ\mathbf{D}_{R_H}^H\mathbf{D}_{R_H}+\mathbf{D}_{T_H}^H\mathbf{D}_{T_V}\circ\mathbf{D}_{R_H}^H\mathbf{D}_{R_V}+\mathbf{D}_{T_V}^H\mathbf{D}_{T_H}\circ\mathbf{D}_{R_V}^H\mathbf{D}_{R_H}+\mathbf{D}_{T_V}^H\mathbf{D}_{T_V}\circ\mathbf{D}_{R_V}^H\mathbf{D}_{R_V}\right)\right\}.$$
(4.70)

The structure of (4.70) results directly from the Kronecker structure of the covariance matrix $\mathbf{R}_{nn}$. Recall, that the covariance matrix $\mathbf{R}_{nn}$ describes the contribution of the dense multipath components plus measurement noise to the observation.

The computation of the FIM using (4.69) is significantly cheaper in terms of computational complexity than the computation using the direct approach (4.16), i.e., via the full Jacobian matrix (4.68). This is especially true in view of the strong redundancies of the component matrices $\mathbf{D}$ shown in Table 4-2. It is not necessary to compute the products $\mathbf{D}_r^H\mathbf{D}_r$ $r\in[t,f,T_H,T_V,R_H,R_V]$ explicitly, since all products are block matrices with redundant blocks. For example all block matrices related to the Fisher information matrix for the frequency domain only $\mathcal{J}_f=\mathbf{D}_f^H\mathbf{D}_f$, are given by

$$\mathcal{J}_f'=\begin{bmatrix}\mathbf{B}_f^H\\\mathbf{D}_\tau^H\end{bmatrix}\cdot\begin{bmatrix}\mathbf{R}_f^{-1}\mathbf{B}_f & \mathbf{R}_f^{-1}\mathbf{D}_\tau\end{bmatrix}=\begin{bmatrix}\mathbf{B}_f^H\mathbf{R}_f^{-1}\mathbf{B}_f & \mathbf{B}_f^H\mathbf{R}_f^{-1}\mathbf{D}_\tau\\\mathbf{D}_\tau^H\mathbf{R}_f^{-1}\mathbf{B}_f & \mathbf{D}_\tau^H\mathbf{R}_f^{-1}\mathbf{D}_\tau\end{bmatrix}.$$
(4.71)

Furthermore, since we face increasing numbers of large scale problems due to advances in channel sounder technology, computations involving the full Jacobian matrix are not reasonable. Channel observations with a large number of samples M and a large number of assessable propagation paths P leads to a large Jacobian matrix. Today, measurements with more than 400 000 real valued samples and more than 50 assessable paths are not uncommon. Altogether, it is in view of memory requirements favourable to avoid computations with the full matrix $\mathbf{D}$, instead the component matrices from Table 4-2 should be used directly.

If the covariance matrix $\mathbf{R}_{nn}$ contains no redundancies one has to use the expression (4.16) to compute the Fisher information matrix and the Cramér-Rao lower bound of the deterministic parameters.

Observe that equations (4.68) and (4.70) cover also cases with missing apertures in the time- and/or frequency-domain. One has to delete the matrices $\mathbf{D}_t$ or $\mathbf{D}_f$ from the equations (4.68) and (4.70), and the columns in Table 4-2 belonging to the related parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$, to get the reduced expressions. Similarly, if an antenna array has been applied at one link end, which is only able to gather information about one angle the column for the other parameter must be removed from all entries in Table 4-2.

4.2.2 Polarimetric Model for one Link End

If channel sounding measurements have been carried out using only one antenna array able to gather sufficient information about the horizontal and vertical field strength, the channel model (4.61) has to be reduced. For polarimetric measurements at the Rx-site we get

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{B}_{R_H} \lozenge \mathbf{B}_{T_H} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH} + \mathbf{B}_{R_V} \lozenge \mathbf{B}_{T_H} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HV}, \tag{4.72}$$

and for polarimetric measurements at the Tx-site we have for example the model

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{B}_{R_H} \lozenge \mathbf{B}_{T_H} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH} + \mathbf{B}_{R_H} \lozenge \mathbf{B}_{T_V} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{VH}.$$

Since, all four possible cases out of (4.61) (H→HV, V→HV, HV→H, and HV→V) have the same structure, we restrict ourselves to the derivation of the equations for the Jacobian matrix and the Fisher information matrix for (4.72) only. Using the definitions (4.68) and (4.69) the component matrices of the Jacobian matrices are given by Table 4-3.

Table 4-3: Component matrices of the Jacobian matrix for polarimetric measurements at one link end.

$\boldsymbol{\theta} \to$	α	τ	φ_T	ϑ_T	φ_R	ϑ_R	$\gamma_{HH,r}$	$\gamma_{HH,i}$	$\gamma_{HV,r}$	$\gamma_{HV,i}$	
$\mathbf{D}_t =$ $[$	$\mathbf{D}_\alpha$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$]$
$\mathbf{D}_f =$ $[$	$\mathbf{B}_f$	$\mathbf{D}_\tau$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$]$
$\mathbf{D}_{T_H} =$ $[$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{D}_{\varphi_T,H}$	$\mathbf{D}_{\vartheta_T,H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$]$
$\mathbf{D}_{R_H} =$ $[$	$\mathbf{B}_H$	$\mathbf{B}_H$	$\mathbf{B}_H$	$\mathbf{B}_H$	$\mathbf{D}_{\varphi_R,H}$	$\mathbf{D}_{\vartheta_R,H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_V}$	$\mathbf{B}_{R_V}$	$]$
$\mathbf{D}_g =$ $[$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T$	$\mathbf{1}^T \mathrm{j}$	$\mathbf{1}^T$	$\mathbf{1}^T \mathrm{j}$	$]$

Using these definitions, the Jacobian matrix for the model (4.72) can be expressed as

■ $$\mathbf{D} = \mathbf{D}_{R_H} \lozenge \mathbf{D}_{T_H} \lozenge \mathbf{D}_g \lozenge \mathbf{D}_f \lozenge \mathbf{D}_t. \tag{4.73}$$

Assuming the covariance matrix of the stochastic process $\mathbf{n}$ is $\mathbf{R}_{nn} = \sigma^2 \mathbf{I}$ the Fisher information matrix of the normalised deterministic parameters is given by

$$\mathcal{J}(\boldsymbol{\theta}, \sigma^2) = \frac{2}{\sigma^2} \Re\{\mathbf{D}_t^H \mathbf{D}_t \circ \mathbf{D}_f^H \mathbf{D}_f \circ \mathbf{D}_g^H \mathbf{D}_g \circ \mathbf{D}_{T_H}^H \mathbf{D}_{T_H} \circ \mathbf{D}_{R_H}^H \mathbf{D}_{R_H}\}, \tag{4.74}$$

and for the structured covariance matrix $\mathbf{R}_{nn} = \mathbf{I} \otimes \mathbf{I} \otimes \mathbf{R}_f \otimes \mathbf{I}$ we yield

■ $$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\{\mathbf{D}_t^H \mathbf{D}_t \circ \mathbf{D}_f^H \mathbf{R}_f^{-1} \mathbf{D}_f \circ \mathbf{D}_g^H \mathbf{D}_g \circ \mathbf{D}_{T_H}^H \mathbf{D}_{T_H} \circ \mathbf{D}_{R_H}^H \mathbf{D}_{R_H}\}. \tag{4.75}$$

4.2.3   Model for Non-polarimetric Measurements

If only one field component has been excited at the transmitter and only one field component can be received, the radio channel model reduces to

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{B}_{R_H} \lozenge \mathbf{B}_{T_H} \lozenge \mathbf{B}_f \lozenge \mathbf{B}_t \cdot \boldsymbol{\gamma}_{HH}. \tag{4.76}$$

Again, using the definitions (4.68) and (4.69) the component matrices of the Jacobian matrices in Table 4-4 are given. Using these definitions, the Jacobian matrix for the (non-polarimetric) model (4.76) can be expressed as follows

■ $$\mathbf{D} = \mathbf{D}_{R_H} \lozenge \mathbf{D}_{T_H} \lozenge \mathbf{D}_g \lozenge \mathbf{D}_f \lozenge \mathbf{D}_t. \tag{4.77}$$

Assuming the covariance matrix of the stochastic process $\mathbf{n}$ is $\mathbf{R}_{nn} = \mathbf{I} \otimes \mathbf{I} \otimes \mathbf{R}_f \otimes \mathbf{I}$ the Fisher information matrix of the normalised deterministic parameters is given by

■ $$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}) = 2 \cdot \Re\{\mathbf{D}_t^H \mathbf{D}_t \circ \mathbf{D}_f^H \mathbf{R}_f^{-1} \mathbf{D}_f \circ \mathbf{D}_g^H \mathbf{D}_g \circ \mathbf{D}_{T_H}^H \mathbf{D}_{T_H} \circ \mathbf{D}_{R_H}^H \mathbf{D}_{R_H}\}. \tag{4.78}$$

Table 4-4: Component matrices of the Jacobian matrix for non-polarimetric measurements.

$\boldsymbol{\theta} \rightarrow$	α	τ	φ_T	ϑ_T	φ_R	ϑ_R	$\gamma_{HH,r}$	$\gamma_{HH,i}$	
$\mathbf{D}_t =$	$[\ \mathbf{D}_\alpha$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$\mathbf{B}_t$	$]$
$\mathbf{D}_f =$	$[\ \mathbf{B}_f$	$\mathbf{D}_\tau$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$\mathbf{B}_f$	$]$
$\mathbf{D}_{T_H} =$	$[\ \mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{D}_{\varphi_T,H}$	$\mathbf{D}_{\vartheta_T,H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$\mathbf{B}_{T_H}$	$]$
$\mathbf{D}_{R_H} =$	$[\ \mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$\mathbf{D}_{\varphi_R,H}$	$\mathbf{D}_{\vartheta_R,H}$	$\mathbf{B}_{R_H}$	$\mathbf{B}_{R_H}$	$]$
$\mathbf{D}_g =$	$[\ \boldsymbol{\gamma}_{HH}^T$	$\boldsymbol{\gamma}_{HH}^T$	$\boldsymbol{\gamma}_{HH}^T$	$\boldsymbol{\gamma}_{HH}^T$	$\boldsymbol{\gamma}_{HH}^T$	$\boldsymbol{\gamma}_{HH}^T$	$\mathbf{1}^T$	$\mathbf{1}^T \mathrm{j}$	$]$

Observe that the elementary blocks in Table 4-2, Table 4-3, and Table 4-4 have a triangular structure. This can be exploited to significantly reduce the computational complexity in implementations. See also the discussion regarding the numerically efficient computation of the Khatri-Rao product $\mathbf{d} = (\mathbf{A} \lozenge \mathbf{B})^H \cdot \mathbf{c}$ outlined in Section 5.2.9 (cf. [51]). Table 4-5 summarises the computation steps for the Fisher information matrix and the Cramér-Rao lower bound of the deterministic model parameters, i.e., the deterministic maximum likelihood problem.

Table 4-5: Calculation steps for the Fisher information matrix and the Cramér-Rao lower bound on the variance of any unbiased estimator for $\boldsymbol{\theta}_{sp}$ (DML)

> Input: Parameter vector $\boldsymbol{\theta}_{sp}$, $\mathbf{R}_{nn}$ ($\mathbf{R}(\boldsymbol{\theta}_{DMC}) + \alpha_0 \mathbf{I}$)
>
> 1) Compute all matrices $\mathbf{B}_r$ $r \in [f, t, T_H, T_V, R_H, R_V, H, V]$ using equations (4.62), (4.63), and (4.66) (Basis functions of the channel model)
>
> 2) Compute all matrices $\mathbf{D}_i$ $i \in [\alpha, \tau, \varphi_T, \vartheta_T, \varphi_R, \vartheta_R]$ using equations (4.64), (4.65), and (4.67)
>
> 3) Compute the FIM (4.70) $\mathcal{J}(\boldsymbol{\theta}_{sp}, \mathbf{R}_{nn})$.
>
> 4) Compute the Cramér-Rao lower bound $\mathbf{CRB}_{\boldsymbol{\theta}_{sp}} = \mathcal{J}^{-1}(\boldsymbol{\theta}_{sp}, \mathbf{R}_{nn})$.
>
> Optional Step:
>
> 5) Compute the Cramér-Rao lower bound of the physical parameters from the Cramér-Rao lower bound of the normalised parameters using the chain rule (4.25).

4.3 Cramér-Rao lower bound for the Stochastic Model Parameters

The derivation of the Cramér-Rao lower bound for the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}_{dmc}$ for the parameters $\boldsymbol{\theta}_{dmc}$ of the covariance-matrix $\mathbf{R}(\boldsymbol{\theta}_{dmc})$ is straightforward. Since this section is solely dedicated to the lower bound on the variance of $\hat{\boldsymbol{\theta}}_{dmc}$, we will for the time being drop the subscript $_{dmc}$ from the parameter vector $\boldsymbol{\theta}_{dmc}$ and write $\boldsymbol{\theta}$.

For simplicity, we assume that the observed channel contains dense multipath components only

$$\mathbf{x} = \mathbf{n}_{dan} \in \mathbb{C}^{M \times 1}.$$

Furthermore, we generalize the model and assume that some independent observations of the stochastic process $\mathbf{n}_{dan}$ are available. We collect all observations in the matrix $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \in \mathbb{C}^{M \times N}$. The log-likelihood function of (2.66) is

$$\mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = -MN \cdot \ln(\pi) - N \ln(\det(\mathbf{R}(\boldsymbol{\theta}))) - \mathrm{tr}\!\left(\mathbf{X}^{\mathrm{H}} \cdot \mathbf{R}^{-1}(\boldsymbol{\theta}) \cdot \mathbf{X}\right). \tag{4.79}$$

Observe that the derivation of explicit expressions for the Cramér-Rao lower bound requires the differentiation of complex-valued matrices. There exist various ways to get closed form expressions for complex matrix derivatives. An elegant way to accomplish that is described in [52]. The author uses the generalised complex derivative (cf. [53]) and provides results for typical expressions.

The score-function, i.e., the first gradient of the log-likelihood function [40] is

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = -N\,\mathrm{tr}\!\left(\frac{1}{\det(\mathbf{R}(\boldsymbol{\theta}))}\,\mathrm{adj}(\mathbf{R}(\boldsymbol{\theta}))\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\right) + \mathrm{tr}\!\left(\mathbf{X}^{\mathrm{H}}\,\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\mathbf{X}\right). \tag{4.80}$$

Since $\det(\mathbf{R}(\boldsymbol{\theta}))^{-1} \cdot \mathrm{adj}(\mathbf{R}(\boldsymbol{\theta}))$ is just the inverse $\mathbf{R}^{-1}(\boldsymbol{\theta})$ of the covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ [53], we can simplify expression (4.80) to

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = \mathrm{tr}\!\left(\mathbf{X}^{\mathrm{H}}\,\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\mathbf{X}\right) - N\,\mathrm{tr}\!\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\right). \tag{4.81}$$

Using the relationship between the inner and outer product of two vectors $\mathbf{x}^{\mathrm{H}}\mathbf{y} = \mathrm{tr}\!\left(\mathbf{y}\,\mathbf{x}^{\mathrm{H}}\right)$ and the property of the trace operator $\mathrm{tr}(\mathbf{ABCD}) = \mathrm{tr}(\mathbf{BCDA})$ [53] we rewrite (4.81) to

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = N\,\mathrm{tr}\!\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}} - \mathbf{R}(\boldsymbol{\theta})\right)\right). \tag{4.82}$$

We now note that the term $1/N \cdot \mathbf{X} \cdot \mathbf{X}^{\mathrm{H}}$ in (4.82) can be understood as an estimate of the covariance matrix $\hat{\mathbf{R}}$ and introduce the non-parametric estimate of the covariance matrix as

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}}. \tag{4.83}$$

Using (4.83) in (4.82) yields for the score function

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = N\,\mathrm{tr}\!\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\cdot\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta})\right)\right) \tag{4.84}$$

or equivalently

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = N\,\mathrm{tr}\!\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\cdot\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\hat{\mathbf{R}} - \mathbf{I}\right)\right). \tag{4.85}$$

The stochastic Fisher matrix is the second gradient of the log-likelihood function [40], hence we take the second partial derivative with respect to the parameter θ_k and obtain

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}) = N\,\mathrm{tr}\Bigg(&-\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_k}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\!\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\hat{\mathbf{R}} - \mathbf{I}\right) + \\
&+\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial^2}{\partial \theta_i \partial \theta_k}\mathbf{R}(\boldsymbol{\theta})\right)\!\left(\mathbf{R}^{-1}(\boldsymbol{\theta})\hat{\mathbf{R}} - \mathbf{I}\right) + \\
&-\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\!\left(\frac{\partial}{\partial \theta_k}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\hat{\mathbf{R}}\Bigg).
\end{aligned} \tag{4.86}$$

Now assuming an unbiased estimator for the covariance matrix

$$\mathrm{E}\big\{\mathbf{R}(\boldsymbol{\theta}) - \mathbf{R}(\hat{\boldsymbol{\theta}})\big\} = \mathbf{0}, \text{ and } \mathrm{E}\big\{\mathbf{R}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{R}(\boldsymbol{\theta})\big\} = \mathbf{I}$$

we yield for the elements of the *Fisher Information Matrix* the equation

$$\mathcal{J}_{ik} = -\mathrm{E}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_k}\mathcal{L}(\mathbf{X}|\boldsymbol{\theta})\right\} = N\,\mathrm{tr}\left\{\mathbf{R}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_k}\mathbf{R}(\boldsymbol{\theta})\right)\right\}. \qquad (4.87)$$

The sum of the main diagonal elements (trace) of the product between a square Hermitian matrix $\mathbf{A} = \mathbf{A}^{\mathrm{H}}$ and an arbitrary square matrix $\mathbf{B}$ can be expressed using the vector operator $\mathrm{vec}\{\bullet\}$ as

$$\mathrm{tr}(\mathbf{A}\cdot\mathbf{B}) = \mathrm{tr}\big(\mathbf{A}^{\mathrm{H}}\cdot\mathbf{B}\big) = \sum_{n=1}^{N}\mathbf{a}_n^{\mathrm{H}}\mathbf{b}_n = \mathrm{vec}\{\mathbf{A}\}^{\mathrm{H}}\,\mathrm{vec}\{\mathbf{B}\}.$$

Consequently, using the decomposition $\mathbf{R}(\boldsymbol{\theta}) = \mathbf{L}(\boldsymbol{\theta})\cdot\mathbf{L}^{\mathrm{H}}(\boldsymbol{\theta})$, a single element of the Fisher information matrix can be expressed as

$$\mathcal{J}_{ik} = -\mathrm{E}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_k}\mathcal{L}(\mathbf{X}|\boldsymbol{\theta})\right\} = N\cdot\mathrm{vec}\left\{\mathbf{L}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{L}^{-\mathrm{H}}(\boldsymbol{\theta})\right\}^{\mathrm{H}}\,\mathrm{vec}\left\{\mathbf{L}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_k}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{L}^{-\mathrm{H}}(\boldsymbol{\theta})\right\}.$$

Employing the vector operator $\mathrm{vec}\{\bullet\}$, the matrix containing the partial derivatives with respect to all L elements of the parameter vector $\boldsymbol{\theta}$ can be defined as

$$\blacksquare \qquad \mathbf{D}(\boldsymbol{\theta}) = \left[\mathrm{vec}\left\{\mathbf{L}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_1}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{L}^{-\mathrm{H}}(\boldsymbol{\theta})\right\} \quad \cdots \quad \mathrm{vec}\left\{\mathbf{L}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_L}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{L}^{-\mathrm{H}}(\boldsymbol{\theta})\right\}\right]. \qquad (4.88)$$

Using the matrix $\mathbf{D}(\boldsymbol{\theta})$ the Fisher Information matrix $\mathcal{J}(\boldsymbol{\theta})$ of the parameters $\boldsymbol{\theta}$ can be expressed in a compact form by

$$\mathcal{J}(\boldsymbol{\theta}) = -\mathrm{E}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_k}\mathcal{L}(\mathbf{X}|\boldsymbol{\theta})\right\} = N\cdot\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta})\cdot\mathbf{D}(\boldsymbol{\theta}). \qquad (4.89)$$

Now we can easily derive the asymptotic Cramér-Rao lower bound [40] on the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ for the parameters $\boldsymbol{\theta}$ as

$$\blacksquare \qquad \mathrm{CRB}_{\boldsymbol{\theta}} = \mathcal{J}^{-1}(\boldsymbol{\theta}) = \frac{1}{N}\big(\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta})\big)^{-1} \le \mathrm{E}\left\{\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big)\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big)^{\mathrm{T}}\right\}. \qquad (4.90)$$

Furthermore, using the matrix $\mathbf{D}(\boldsymbol{\theta})$ in the expression for the score function (4.84) yields the compact form by

$$\blacksquare \qquad \mathbf{q}(\mathbf{X},\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{L}(\mathbf{X},\boldsymbol{\theta}) = N\cdot\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta})\cdot\mathrm{vec}\left\{\mathbf{L}^{-1}(\boldsymbol{\theta})\hat{\mathbf{R}}\,\mathbf{L}^{-\mathrm{H}}(\boldsymbol{\theta}) - \mathbf{I}\right\}. \qquad (4.91)$$

So far, we have derived a general expressions for the score function, the Fisher information matrix, and the Cramér-Rao lower bound of the parameters $\boldsymbol{\theta}$.

Since $\mathrm{toep}(\bullet)$ is a linear operator the first order partial derivatives of (2.67) with respect to the parameters $\boldsymbol{\theta}_{DMC}$ are given by

$$\blacksquare \qquad \frac{\partial}{\partial\theta_i}\mathbf{R}(\boldsymbol{\theta}_{DMC}) = \mathrm{toep}\left(\frac{\partial}{\partial\theta_i}\boldsymbol{\kappa}(\boldsymbol{\theta}_{DMC}), \frac{\partial}{\partial\theta_i}\boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{DMC})\right). \qquad (4.92)$$

Hence, we have only to calculate the first order partial derivatives of (2.68) with respect to the four parameters in the parameter vector $\boldsymbol{\theta}_{DMC} = [\alpha_0 \ \alpha_1 \ \beta \ \tau_0]^T$ to derive the related Fisher Information matrix. The four partial derivatives are

$$\frac{\partial}{\partial \alpha_0} \boldsymbol{\kappa}(\boldsymbol{\theta}) = \mathbf{e}_0, \tag{4.93}$$

$$\frac{\partial}{\partial \alpha_1} \boldsymbol{\kappa}(\boldsymbol{\theta}) = \frac{1}{M} \left[\frac{1}{\beta_d} \quad \frac{e^{-j2\pi\tau_0}}{\beta_d + j2\pi\frac{1}{M}} \quad \cdots \quad \frac{e^{-j2\pi(M-1)\tau_0}}{\beta_d + j2\pi\frac{M-1}{M}} \right]^T, \tag{4.94}$$

$$\frac{\partial}{\partial \beta} \boldsymbol{\kappa}(\boldsymbol{\theta}) = -\frac{\alpha_1}{M} \left[\frac{1}{(\beta_d)^2} \quad \frac{e^{-j2\pi\tau_0}}{(\beta_d + j2\pi\frac{1}{M})^2} \quad \cdots \quad \frac{e^{-j2\pi(M-1)\tau_0}}{(\beta_d + j2\pi\frac{M-1}{M})^2} \right]^T, \tag{4.95}$$

and

$$\frac{\partial}{\partial \tau_0} \boldsymbol{\kappa}(\boldsymbol{\theta}) = \frac{\alpha_1}{M} \left[0 \quad \frac{-j2\pi \cdot e^{-j2\pi\tau_0}}{\beta_d + j2\pi\frac{1}{M}} \quad \cdots \quad \frac{-j2\pi(M-1) \cdot e^{-j2\pi(M-1)\tau_0}}{\beta_d + j2\pi\frac{M-1}{M}} \right]^T. \tag{4.96}$$

Using equations (4.92) - (4.96) in (4.90) one can now easily calculate the Cramér-Rao lower bound of any unbiased estimator $\hat{\boldsymbol{\theta}}_{DMC}$ for the parameters $\boldsymbol{\theta}_{DMC}$ of the covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{DMC})$. Table 4-6 summarises the processing steps for the calculation of the Cramér-Rao lower bound on $\hat{\boldsymbol{\theta}}_{DMC}$.

Table 4-6: Summary of calculation steps for the Cramér-Rao lower bound on any unbiased estimator for $\boldsymbol{\theta}_{DMC}$ (SML)

Input: Parameter vector $\boldsymbol{\theta}_{DMC}$, Number of independent realisations N

1) Compute the first order derivatives (equation (4.92) - (4.96))

2) Compute the matrix $\mathbf{D}(\boldsymbol{\theta})$ using equation (4.88) (section 6.1.12 describes an computationally efficient algorithm for the multiplication with the inverse Toeplitz matrix $\mathbf{R}^{-1}(\boldsymbol{\theta}_{DMC})$).

3) Compute the Cramér-Rao lower bound $\mathbf{CRB}_{\boldsymbol{\theta}} = \frac{1}{N} \left(\mathbf{D}^H(\boldsymbol{\theta}) \cdot \mathbf{D}(\boldsymbol{\theta}) \right)^{-1}$

4.4   Joint DML and SML Problem

So far, we have derived explicit expressions for the DML problem and the SML problem independently. However, for channel parameter estimation from channel sounding measurements all parameters of the model must be estimated jointly. Therefore, we derive in the following the Fisher information matrix and the Cramér-Rao lower bound for the joint estimation problem.

The general structure of the Fisher information matrix of any estimator for the parameters of a multivariate normal distribution

$$p(\mathbf{x}|\boldsymbol{\theta},(\boldsymbol{\theta})) = \frac{1}{\pi^M \det(\mathbf{R}(\boldsymbol{\theta}))} e^{-(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}))^H \cdot \mathbf{R}(\boldsymbol{\theta})^{-1} \cdot (\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}))}$$

is due to [40], [43], [54], and [55] given by

$$\{\mathcal{J}\}_{ik} = 2 \cdot \Re\left\{\left(\frac{\partial}{\partial\theta_i}\mathbf{s}^{\mathrm{H}}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_k}\mathbf{s}(\boldsymbol{\theta})\right)\right\} + \mathrm{tr}\left\{\mathbf{R}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_i}\mathbf{R}(\boldsymbol{\theta})\right)\mathbf{R}^{-1}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_k}\mathbf{R}(\boldsymbol{\theta})\right)\right\}. \quad (4.97)$$

Since the parameters $\boldsymbol{\theta}_{sp}$ of the deterministic components, i.e., of the concentrated paths $\mathbf{s}(\boldsymbol{\theta}_{sp})$ are not parameters of the distribution of the dense multipath components $\mathbf{d}_{dmc}$, i.e., of the covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ and vice versa, the Fisher information matrix is a block diagonal matrix

∎ $$\mathcal{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{J}(\boldsymbol{\theta}_{sp}) & \mathbf{0} \\ \mathbf{0} & \mathcal{J}(\boldsymbol{\theta}_{dmn}) \end{bmatrix}. \quad (4.98)$$

Consequently, the Cramér-Rao lower bound is a block diagonal matrix too with one block for the parameters of the deterministic components and one block for the parameters of the dense multipath components

∎ $$\mathrm{CRB}_{\boldsymbol{\theta}_{chn}} = \mathcal{J}(\boldsymbol{\theta}_{chn})^{-1} = \begin{bmatrix} \mathrm{CRB}_{\boldsymbol{\theta}_{sp}} & \mathbf{0} \\ \mathbf{0} & \mathrm{CRB}_{\boldsymbol{\theta}_{dmn}} \end{bmatrix}. \quad (4.99)$$

This has two consequences. First of all the expressions for the Cramér-Rao lower bounds (4.21), (4.70), (4.75), (4.78), and (4.90) are also valid for the joint estimation problem. Secondly the disjoint parameter subsets $\boldsymbol{\theta}_{SP}$ and $\boldsymbol{\theta}_{DMC}$ are asymptotically uncoupled. The latter statement has important consequences for the structure of the parameter estimator to develop. We will make use of this model property in the subsequent chapters.

4.5 Conclusion on the CRLB of the Deterministic Parameters

For the model developed in Chapter 2 explicit expressions for the Fisher information matrix and the Cramér-Rao lower bound can be given. The equations are structured in such a way, that almost all possible channel-sounding setups are covered.

The Cramér-Rao lower bound provides an effective means to decide, whether a propagation path should be classified as a concentrated propagation path or not. If it is too weak, it should be added to the stochastic process describing the dense multipath components or to the measurement noise. Here the term too weak refers to the lower bound on the relative variance of the propagation path.

Furthermore, it has been shown that the FIM is not only a matrix on the way to the Cramér-Rao lower bound of the model parameters. It provides insight in the properties of the parameter estimation problem. In general, it is a powerful tool for diagnostic purposes. In addition, the FIM plays an important role in the parameter estimators described in the following chapters.

We can split the joint estimation problem for all propagation paths often into decoupled sub-problems. Each sub-problem contains a smaller amount of propagation paths. This can be exploited to reduce the complexity of the parameter estimator. The probability that a given problem can be separated grows with M_i and with the amount of parameter dimensions R_P, available to resolve the propagation paths. Nevertheless, the structure of the FIM shows also that we should not split the entire problem into P sub problems a priori, since some of the P propagation paths might be coupled. As we shall see later, this observation is of crucial importance for the choice of the parameter estimator, e.g., whether the Expectation Maximization algorithm is a good choice or not.

The Cramér-Rao lower bound of closely spaced, i.e., coupled propagation paths depends on the parameters of all propagation paths in a path group.

It is important to recognize that the data model is the crucial point in parameter estimation. The real art is the selection of a model, which is accurate enough. In the same context we have to state that an excessive accurate model is unsuitable as well. This is due to the fact that all radio channel observations contain finite information about the underlying noise-free radio channel. Observe that for the concentrated propagation paths the term "noise-free" implies that the observed channel contains no dense multipath components.

The bound on the relative variance of the weight parameters provides a means to control the model complexity, as stated in Example 4-4. That means for every propagation path the following inequality must be fulfilled

$$\frac{\mathrm{CRB}_{|\gamma|}}{|\gamma|^2} < \varepsilon_{|\gamma|}^2 < 1 \,. \qquad (4.100)$$

Equation (4.100) is an effective way to determine the number of assessable propagation paths in an observation. The estimate of the relative variance, i.e., the individual SNR of a propagation path is used in Section 5.2.7 for model order selection. That means it is used to determine the number of specular propagation paths in a radio channel observation. Furthermore, reasonable values for $\varepsilon_{|\gamma|}^2$ are given therein. As already discussed in Example 4-4, the expressions for the Cramér-Rao bound will not provide reasonable estimates of the variance of the path weight if the SNR of a propagation path is close to 0dB or negative. However, this is a weak restriction since only propagation paths with a positive SNR (dominant paths) are of interest. For further discussion on the validity of the Cramér-Rao bound see, e.g., [49], [56]. In [57] it has been shown that the maximum likelihood estimator developed within this work reaches the Cramér-Rao lower bound for the propagation path parameters derived in this chapter, provided the radio channel sounder including the antenna arrays is well calibrated.

5   Estimation of Path Parameters

In recent years, high-resolution channel parameter estimation from channel sounding measurements has attracted a great deal of interest. Initial research work was focused on the estimation of direction of arrival (DoA) of the propagation paths at the base station and was shortly after, extended to the joint estimation of time-delay and DoA [58], [59]. The pioneering work of Steinbauer et al. [60] on the measurement of the double directional radio channel roused the interest in the joint estimation of direction of departure, time-delay, and direction of arrival [61], [1], [6] from channel sounding measurements.

As already discussed in Chapter 2, the radio channel observation $\mathbf{x}$ is a realisation of a stochastic process, and the parameters $\boldsymbol{\theta}$ are stochastic too. The aim of radio channel measurements in combination with channel parameter estimation is the determination of the distribution $p(\boldsymbol{\theta})$. We assume that no prior distribution $p(\boldsymbol{\theta})$ of the unknown parameters is available. In other words, we choose the uniform distribution as a prior for $\boldsymbol{\theta}$. Hence the maximum likelihood estimate

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left(p(\mathbf{x}, \boldsymbol{\theta})\right)$$

is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left(p(\mathbf{x}|\boldsymbol{\theta})\right),$$

since

$$p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

and $p(\boldsymbol{\theta}) = const$. Nevertheless, one should observe that if reliable probability densities of the parameters $\boldsymbol{\theta}$ are available, the parameter estimator developed within this work, can be improved further by taking them into account. This will lead to better parameter estimates, i.e., to parameter estimates with a lower variance.

For the estimation of propagation path parameters from channel sounding measurements, two classes of parameter estimators have been deployed so far. Namely, various algorithms belonging to the so-called subspace based class of algorithms and algorithms based on the SAGE (space alternating generalised expectation maximization) procedure [62]. Since the authors of [63], [64], and [11] did not assign a name to their SAGE based channel parameter estimation algorithm we will refer to it in the following as SAGECPE1. Subspace based channel parameter estimation algorithms reportedly used to estimated propagation path parameters are Unitary ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [65], [44], [15], MUSIC (MUltiple SIgnal Classification), and RARE (Rank Reduction) [66], [67]

77

All algorithms published so far have widely ignored the contribution of the dense multipath components to the radio-channel observation. The models used to develop the parameter estimators are usually based on the assumption that the channel observation contains only concentrated propagation paths and white receiver noise. The first and obvious drawback is the lack of information about the distribution of the dense multipath components in the parameter estimation results. Omitting the dense multipath components in the channel parameter estimator's data model causes also significant problems while estimating the parameters of the concentrated propagation paths. Mainly two problems arise if we assume that the covariance matrix is a scaled version of the identity matrix. The first problem observed, is that highly sophisticated model order selection criteria based on stochastic complexity arguments such as AIC (Akaike information criterion) or MDL (minimum description length) fail to detect a useable number of assessable propagation paths, since the data model is wrong. Often the estimated model order is completely wrong, leading to unacceptable estimation results of the parameter estimator. Secondly, the estimator approximates the distribution of the dense multipath components with concentrated propagation paths what will fail since the observed information is not sufficient to accomplish that. It is important to note, that both errors described are not caused by any inability of the estimators or the used estimation framework, but rather by the incompleteness of the data model. In general, the robustness of parameter estimators depends heavily on the completeness of the data model used to derive the estimator.

The objective of the subsequent sections is to develop an algorithm framework for high-resolution parameter estimation from measurements. The parameter estimator should compute parameter estimates for every observation and provide reliability information about the parameter estimates. In general, we expect the estimator to process data without user interaction. The derivation of sufficient statistics for radio channel parameters $p(\boldsymbol{\theta})$ requires the processing of large amounts of data, i.e., thousands of observations. The main challenge of the parameter estimation problem is the complexity. If the number of observable propagation paths is 100 and all model parameters can be estimated, the total number of real-valued propagation path parameters is 1400. Consequently, we can classify the problem as a mid- or large-scale problem. Additionally, no algorithm for the joint estimation of propagation path parameters and DMC parameters exists.

5.1   Global Maximization Algorithms

Given a channel observation $\mathbf{x}$, the maximization problem to solve is

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_{sp} \\ \hat{\boldsymbol{\theta}}_{dan} \end{bmatrix} = \arg\max\left(\frac{1}{\pi^{M} \det(\mathbf{R}(\boldsymbol{\theta}_{dan}))} e^{-(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}_{sp}))^{H} \cdot \mathbf{R}^{-1}(\boldsymbol{\theta}_{dan}) \cdot (\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}_{sp}))} \right),$$

or equivalently

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_{sp} \\ \hat{\boldsymbol{\theta}}_{dan} \end{bmatrix} = \arg\max\left(-\ln(\det(\mathbf{R}(\boldsymbol{\theta}_{dac}))) - (\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}_{sp}))^{H} \cdot \mathbf{R}^{-1}(\boldsymbol{\theta}_{dan}) \cdot (\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}_{sp})) \right). \qquad (5.1)$$

The methods to solve (5.1) described in the following are implicit in nature. The parameter estimates are defined as the maximizing or minimizing arguments of some nonlinear objective function. In practice, finding the solution of (5.1) is far from trivial, especially since the number of nonlinear parameters (i.e. structural parameters) is large. In the following, we propose a computationally attractive and reasonably reliable procedure to solve such a problem. The procedure is based on numerical nonlinear local optimisation techniques, such as Gauß-Newton, Levenberg-Marquardt [68], or iterative maximum likelihood [40]. The initial values

for the local optimisation algorithms are estimated with suitable global search algorithms or are handed down from the previous observation.

Usually a long sequence of channel observations, taken over time, has to be processed. Since the structural parameters (time-delay, angles, Doppler-shift) of the radio channel change slowly over time, the parameter estimates of the actual observation can be used as the initial values for the nonlinear parameters of the subsequent observation. This reduces the computational complexity of the global search procedure significantly, because only initial values for propagation paths, emerged in the actual observation, have to be found. Candidate algorithms for the determination of the initial values are based on the Expectation-Maximization approach [69]. In some cases also subspace based algorithms such as multidimensional Unitary ESPRIT, or RARE [66], [67] can be applied to find initial values. The applicability of the subspace-based algorithm depends primarily on the structure of the measurement system matrix $\mathbf{G}$ (cf. Section 5.3.11).

The joint maximization of the log-likelihood function (5.1) with a multidimensional exhaustive search is not feasible in practice. A way to find a solution is to separate the problem into smaller sub-problems. Due to [62] we can solve the joint maximization problem if we choose parameter subsets and maximize the objective function in an alternating manner with respect to the subsets. An obvious choice for two parameter subsets are the parameters sets $\boldsymbol{\theta}_{sp}$ and $\boldsymbol{\theta}_{dan}$. So we suppose for the time being the covariance matrix $\mathbf{R}_{nn}$ of the stochastic process $\mathbf{n}$ is known, i.e., we know the parameters $\boldsymbol{\theta}_{dan}$. Consequently, the maximization problem (5.1) reduces to

∎
$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})). \tag{5.2}$$

The term $(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))$ in equation (5.2) is the so-called Mahalanobis norm or Mahalanobis distance [70], [71].

Because the objective function (5.2) is a nonlinear function in the structural parameters $\boldsymbol{\mu}$ and since we minimize a weighted error (Mahalanobis norm) the maximum likelihood estimation problem can also be classified as a nonlinear weighted least squares problem NWLS. To be precise, the problem to solve is an optimally weighted least squares problem, since we do not use an arbitrary weighting matrix. Optimally weighted refers to the fact, that the inverse noise covariance matrix $\mathbf{R}_{nn}^{-1}$ is used as the weighting matrix.

It is a well-known fact from estimation theory [40] that the maximum likelihood estimation of data corrupted by additive normal distributed noise is equivalent to the solution of a (weighted) least squares problem. Surprisingly, this is rarely discussed in publications dedicated to maximum likelihood radio channel parameter estimation. Nevertheless, it is an important observation since a large amount of numerical optimisation algorithms exist for this class of estimation problems since a long time.

If the stochastic part of the observation $\mathbf{n}$ is a zero-mean circular Gaussian i.i.d. process with covariance $\mathbf{R}_{nn} = \sigma^2 \mathbf{I}$, the minimization problem reduces to a classical nonlinear least squares problem NLS. We search for the value $\hat{\boldsymbol{\theta}}$ minimizing the error $\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})$, i.e., we minimize the Euclidian norm or Frobenius norm

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})\|_F^2.$$

The general structure of the model describing the contribution of the concentrated propagation paths to the channel is due to (3.3) given by

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}, \quad \mathbf{s}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times 1} \tag{5.3}$$

with $\mathbf{B}(\boldsymbol{\mu}) \in \mathbb{C}^{M \times N_{pol}P}$, the structural parameters $\boldsymbol{\mu} \in \mathbb{R}^{L_{\mu} \times 1}$, and the weights $\boldsymbol{\gamma} \in \mathbb{C}^{N_{pol}P \times 1}$. The scalar $N_{pol} \in [1\ 2\ 4]$ denotes the number of path weights per path, i.e., how many weights from the complete set γ_{HH}, γ_{HV}, γ_{VH}, γ_{VV} are contained in the model. Using (5.3) in (5.2) yields

$$\begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \arg \min_{\boldsymbol{\mu}, \boldsymbol{\gamma}} (\mathbf{x} - \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma})^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot (\mathbf{x} - \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}). \tag{5.4}$$

Since the parameters $\boldsymbol{\gamma}$ are linear parameters, the minimization problem (5.4) can be solved directly for $\hat{\boldsymbol{\gamma}}$ given a parameter set $\hat{\boldsymbol{\mu}}$. For any $\hat{\boldsymbol{\mu}}$ the "best linear unbiased estimate" (BLUE)[9] is determined by

■ $$\hat{\boldsymbol{\gamma}} = \left(\mathbf{B}^{\mathrm{H}}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\hat{\boldsymbol{\mu}}) \right)^{-1} \cdot \mathbf{B}^{\mathrm{H}}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{x}. \tag{5.5}$$

The value $\hat{\boldsymbol{\gamma}}$ is the maximum likelihood estimate of $\boldsymbol{\gamma}$. The estimate is consistent and the estimator (5.5) is statistically efficient [22]. Replacing $\boldsymbol{\gamma}$ in equation (5.4) by its estimate $\hat{\boldsymbol{\gamma}}$ yields

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} \left(\left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{x} \right) - \left(\left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right) \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right)^{-1} \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{x} \right) \right) \right). \tag{5.6}$$

Hence, the maximum likelihood estimate $\hat{\boldsymbol{\mu}}_{ML}$ is the value minimizing the function

$$\mathcal{L}'_N(\mathbf{x}|\boldsymbol{\mu}) = \left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{x} \right) - \left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right) \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right)^{-1} \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{x} \right). \tag{5.7}$$

Observe that the cost function $\mathcal{L}'_N(\mathbf{x}|\boldsymbol{\mu})$ is not a function of the linear parameters, i.e., the path weights, any more. Function (5.7) is in fact the negative compressed log-likelihood function [40] for the structural parameters $\boldsymbol{\mu}$. Dropping the constant term in (5.7) yields the general correlation function

■ $$C(\boldsymbol{\mu}, \mathbf{x}) = \left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right) \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right)^{-1} \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}_{nn}^{-1} \cdot \mathbf{x} \right). \tag{5.8}$$

The reduced nonlinear optimisation problem to solve is now

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} \left(\mathcal{L}'_N(\mathbf{x}|\boldsymbol{\mu}) \right) = \arg \max_{\boldsymbol{\mu}} \left(C(\boldsymbol{\mu}, \mathbf{x}) \right).$$

Since the joint optimisation problem with respect to $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\mu}}$ can be separated into a optimisation problem of $\hat{\boldsymbol{\mu}}$ only, the estimation problem constitutes a separable least squares problem (SLS) [72], [73]. Once we have estimated $\hat{\boldsymbol{\mu}}$ we can calculate the best linear unbiased estimate $\hat{\boldsymbol{\gamma}}$ directly using the estimator (5.5).

Without additional knowledge about the structure of $\mathbf{B}(\boldsymbol{\mu})$ the nonlinear maximization problem can only be solved by an exhaustive multidimensional search over all parameters of the parameter vector $\boldsymbol{\mu}$. Such a global search algorithm has exponential complexity in the number of structural parameters $\boldsymbol{\mu}$ and is therefore only applicable if L_{μ} is small. In the context of channel parameter estimation L_{μ} can be considered as small only if $L_{\mu} \in [1\ 2]$.

Let us, for the time being, assume the observation contains a single propagation path. In the subsequent examples, we will discuss some typical cases of (5.8). Depending on the structure of the basis function $\mathbf{B}(\boldsymbol{\mu})$ and the covariance matrix $\mathbf{R}_{nn}$, the expression for the correlation function can be simplified.

---

[9] Observe that the estimate (5.5) is also known as the weighted least squares or Gauß-Markov estimate [135].

Example 5-1: i.i.d. circular complex Gaussian process

First, let us assume the noise covariance matrix has structure $\mathbf{R} = \sigma^2 \cdot \mathbf{I}$

$$C(\boldsymbol{\mu}, \mathbf{x}) = \frac{1}{\sigma^2} \left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{B}(\boldsymbol{\mu}) \right) \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{B}(\boldsymbol{\mu}) \right)^{-1} \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{x} \right). \tag{5.9}$$

Using the QR decomposition to factorize the matrix-valued function $\mathbf{B}(\boldsymbol{\mu})$ yields

$$\blacksquare \qquad\qquad \mathbf{B}(\boldsymbol{\mu}) = \mathbf{Q}(\boldsymbol{\mu}) \cdot \mathbf{T}(\boldsymbol{\mu}) \tag{5.10}$$

with

$$\mathbf{Q}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{Q}(\boldsymbol{\mu}) = \mathbf{I}, \text{ and } \mathbf{T}(\boldsymbol{\mu}) \in \mathbb{C}^{N_{pol} \times N_{pol}}.$$

The matrix valued function $\mathbf{Q}_l(\boldsymbol{\mu})$ can be interpreted as a unitary parametric subspace. Using (5.10) in (5.9) yields

$$C(\boldsymbol{\mu}, \mathbf{x}) = \frac{1}{\sigma^2} \left(\mathbf{x}^{\mathrm{H}} \cdot \mathbf{Q}(\boldsymbol{\mu}) \cdot \mathbf{T}(\boldsymbol{\mu}) \right) \cdot \left(\mathbf{T}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{Q}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{Q}(\boldsymbol{\mu}) \cdot \mathbf{T}(\boldsymbol{\mu}) \right)^{-1} \cdot \left(\mathbf{T}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{Q}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{x} \right). \tag{5.11}$$

This expression for the cost function can be simplified using the definition

$$\blacksquare \qquad\qquad \mathbf{z}(\boldsymbol{\mu}) = \mathbf{Q}^{\mathrm{H}}(\boldsymbol{\mu}) \mathbf{x} \tag{5.12}$$

to

$$C(\boldsymbol{\mu}, \mathbf{x}) = \frac{1}{\sigma^2} \mathbf{x}^{\mathrm{H}} \cdot \mathbf{Q}(\boldsymbol{\mu}) \cdot \mathbf{Q}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{x} = \frac{1}{\sigma^2} \| \mathbf{z}_l(\boldsymbol{\mu}) \|_F^2. \tag{5.13}$$

We can interpret $\mathbf{Q}(\boldsymbol{\mu})$ as a set of matched filters, and the objective is to find the matched filter and the related parameter set $\hat{\boldsymbol{\mu}}$ that maximizes the correlation function (5.13). Having estimated $\hat{\boldsymbol{\mu}}$ we can calculate the best linear unbiased estimate of the linear model parameters using

$$\hat{\boldsymbol{\gamma}} = \mathbf{T}^{-1}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{Q}^{\mathrm{H}}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{x}. \tag{5.14}$$

The estimated data can be reconstructed using one of the following equations

$$\begin{aligned}
\mathbf{s}(\hat{\boldsymbol{\theta}}) &= \mathbf{B}(\hat{\boldsymbol{\mu}}) \cdot \hat{\boldsymbol{\gamma}} \\
&= \mathbf{Q}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{T}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{T}^{-1}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{Q}^{\mathrm{H}}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{x} \\
&= \mathbf{Q}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{Q}^{\mathrm{H}}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{x} \\
&= \mathbf{Q}(\hat{\boldsymbol{\mu}}) \cdot \mathbf{z}(\hat{\boldsymbol{\mu}}).
\end{aligned} \tag{5.15}$$

Note that the estimate $\hat{\boldsymbol{\gamma}}$ is not necessarily required to reconstruct the complete estimated data $\mathbf{s}(\hat{\boldsymbol{\theta}})$.

Example 5-2: i.i.d. circular complex Gaussian process with covariance matrix R

To simplify the expression for the correlation function

$$C(\boldsymbol{\mu}, \mathbf{x}) = \mathbf{x}^{\mathrm{H}} \cdot \mathbf{R}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \right)^{-1} \cdot \mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{R}^{-1} \cdot \mathbf{x} \tag{5.16}$$

we decompose the covariance matrix into

$$\mathbf{R} = \mathbf{L} \mathbf{L}^{\mathrm{H}}. \tag{5.17}$$

This decomposition can be accomplished by using the Cholesky decomposition. Using (5.17) in (5.16) yields

$$C(\boldsymbol{\mu}, \mathbf{x}) = \mathbf{x}^{\mathrm{H}} \cdot \left(\mathbf{LL}^{\mathrm{H}}\right)^{-1} \cdot \mathbf{B}(\boldsymbol{\mu}) \cdot \left(\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \left(\mathbf{LL}^{\mathrm{H}}\right)^{-1} \cdot \mathbf{B}(\boldsymbol{\mu})\right)^{-1} \cdot \mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \left(\mathbf{LL}^{\mathrm{H}}\right)^{-1} \cdot \mathbf{x}$$

what can be rewritten to

$$C(\boldsymbol{\mu}, \mathbf{x}) = \mathbf{x}^{\mathrm{H}} \mathbf{L}^{-\mathrm{H}} \mathbf{L}^{-1} \mathbf{B}(\boldsymbol{\mu}) \cdot \left(\left(\mathbf{L}^{-1}\mathbf{B}(\boldsymbol{\mu})\right)^{\mathrm{H}} \cdot \mathbf{L}^{-1}\mathbf{B}(\boldsymbol{\mu})\right)^{-1} \cdot \left(\mathbf{L}^{-1}\mathbf{B}(\boldsymbol{\mu})\right)^{\mathrm{H}} \cdot \mathbf{L}^{-1}\mathbf{x}. \qquad (5.18)$$

Now we introduce the observation with applied pre-whitening $\mathbf{x}' = \mathbf{L}^{-1}\mathbf{x}$, and the matrix valued function with pre-whitening $\mathbf{B}'(\boldsymbol{\mu}) = \mathbf{L}^{-1}\mathbf{B}(\boldsymbol{\mu})$. Using these definitions, the expression for the cost function can be simplified to

$$C(\boldsymbol{\mu}, \mathbf{x}) = \left(\mathbf{x}'^{\mathrm{H}} \cdot \mathbf{B}'(\boldsymbol{\mu})\right) \cdot \left(\mathbf{B}'^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{B}'(\boldsymbol{\mu})\right)^{-1} \cdot \left(\mathbf{B}'^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{x}'\right). \qquad (5.19)$$

It is instructive to observe, that the new cost function (5.19) is equivalent to (5.9), that is reasonable since the coloured noise in the observation $\mathbf{x}$ has been transformed into i.i.d. normal distributed noise by the pre-whitening matrix $\mathbf{L}^{-1}$. Using, the QR-decomposition of the transformed matrix valued function $\mathbf{B}'(\boldsymbol{\mu}) = \mathbf{Q}'(\boldsymbol{\mu}) \cdot \mathbf{T}'(\boldsymbol{\mu})$ yields $\mathbf{z}'(\boldsymbol{\mu}) = \mathbf{Q}'^{\mathrm{H}}(\boldsymbol{\mu})\mathbf{x}'$. Therefore, the cost function can simply be calculated as

■ $$C(\boldsymbol{\mu}, \mathbf{x}) = \mathbf{x}'^{\mathrm{H}} \cdot \mathbf{Q}'_l(\boldsymbol{\mu}) \cdot \mathbf{Q}'^{\mathrm{H}}(\boldsymbol{\mu}) \cdot \mathbf{x}' = \left\|\mathbf{z}'_l(\boldsymbol{\mu})\right\|_F^2.$$

Example 5-3: i.i.d. circular complex Gaussian process, complex exponential model, one parameter $\mu \in \left[-\pi, +\pi\right)$.

In this example, we investigate the estimation of the scalar parameter μ of the complex exponential model

■ $$\mathbf{B}(\mu) = \mathbf{Q}(\mu) = \mathbf{a}(\mu) = \frac{1}{\sqrt{M}}\left[\mathrm{e}^{-\mathrm{j}\left(-\frac{M-1}{2}\right)\cdot\mu} \quad \cdots \quad \mathrm{e}^{-\mathrm{j}\left(+\frac{M-1}{2}\right)\cdot\mu}\right]^{\mathrm{T}}. \qquad (5.20)$$

Since the basis-function $\mathbf{a}(\mu)$ has unit norm the cost function can be calculated by

$$C(\mu, \mathbf{x}) = \frac{1}{\sigma^2}\mathbf{x}^{\mathrm{H}} \cdot \mathbf{a}(\mu) \cdot \mathbf{a}^{\mathrm{H}}(\mu) \cdot \mathbf{x} = \frac{1}{\sigma^2}\left|\mathbf{a}^{\mathrm{H}}(\mu) \cdot \mathbf{x}\right|^2. \qquad (5.21)$$

Now let us suppose, we sample the parameter domain of μ equidistantly on N_s points yielding the set $\boldsymbol{\mu}_s = 2\pi \cdot \left[-\frac{1}{2}, -\frac{1}{2}+\frac{1}{N}, ..., \frac{1}{2}-\frac{1}{N}\right]^{\mathrm{T}}$. Furthermore, we assume that the number of samples of the parameter domain is equal to the number of samples in the observation $\mathbf{x}$. Consequently, the cost function sampled at the points $\boldsymbol{\mu}_s$ can be efficiently computed using the FFT, i.e., by

■ $$C(\boldsymbol{\mu}_s, \mathbf{x}) = \frac{1}{\sigma^2}\left(\mathbf{F}^{\mathrm{H}}\mathbf{x}\right) \circ \left(\mathbf{F}^{\mathrm{H}}\mathbf{x}\right)^* \in \mathbb{R}^{N_s \times 1}. \qquad (5.22)$$

Here, the matrix $\mathbf{F}$ denotes the DFT-matrix. Furthermore, if the number of samples of the parameter domain is larger than the number of samples in the observation $N_s > M$ the cost function can be computed using the FFT by using zero padding, i.e., by computing

■ $$C(\boldsymbol{\mu}_s, \mathbf{x}) = \frac{N_s^2}{\sigma_w^2 M^2}\left(\mathbf{F}^{\mathrm{H}}\mathbf{J}^{\mathrm{H}}\mathbf{x}\right) \circ \left(\mathbf{F}^{\mathrm{H}}\mathbf{J}^{\mathrm{H}}\mathbf{x}\right)^*$$

using the selection matrix $\mathbf{J} = [\mathbf{I}\ \mathbf{0}] \in \mathbb{R}^{M \times N_k}$. One should note that the FFT is fast if the number of samples N_s can be factorized into a product of small prime numbers.

Furthermore, one should observe that the cost function is a representation of the impulse response if we assume that the observation $\mathbf{x}$ contains samples of the channel transfer function. If the observation contains data measured with a uniform linear array the cost function is effectively a beam-space representation of the data. Finally, if we assume that the observation contains samples of the radio channel taken over time at the same frequency, and between the same Tx-, Rx-antenna pair, than the cost function is simply a representation of the Doppler-spectrum. The estimated data are reconstructed without explicit estimation of $\hat{\gamma}$ by

$$\mathbf{s}(\hat{\boldsymbol{\theta}}) = \mathbf{a}(\hat{\mu}) \cdot \hat{\gamma} = \mathbf{a}(\hat{\mu}) \cdot \mathbf{a}^{\mathrm{H}}(\hat{\mu}) \cdot \mathbf{x} . \tag{5.23}$$

Example 5-4: circular complex Gaussian process with Toeplitz-structured covariance matrix R, complex exponential model, one parameter $\mu \in [-\pi, +\pi)$.

In this example, we investigate the estimation of the scalar parameter μ of the complex exponential model

$$\mathbf{B}(\mu) = \mathbf{Q}(\mu) = \mathbf{a}(\mu) = \frac{1}{\sqrt{M}} \left[e^{-j(-\frac{M-1}{2}) \cdot \mu} \cdots e^{-j(+\frac{M-1}{2}) \cdot \mu} \right]^{\mathrm{T}}$$

in coloured noise. The cost function can be expressed as

$$C(\mu, \mathbf{x}) = \frac{\left| \mathbf{x}^{\mathrm{H}} \mathbf{R}^{-1} \mathbf{a}(\mu) \right|^2}{\mathbf{a}(\mu)^{\mathrm{H}} \mathbf{R}^{-1} \mathbf{a}(\mu)} .$$

At first, we calculate the solution $\mathbf{a}'(\mu) = \mathbf{R}^{-1} \mathbf{a}(\mu)$. Using this solution in the expression for the cost function yields

$$C(\mu, \mathbf{x}) = \frac{\left| \mathbf{x}^{\mathrm{H}} \mathbf{a}'(\mu) \right|^2}{\mathbf{a}(\mu)^{\mathrm{H}} \mathbf{a}'(\mu)} = \left| \frac{\mathbf{x}^{\mathrm{H}} \mathbf{a}'(\mu)}{\sqrt{\mathbf{a}(\mu)^{\mathrm{H}} \mathbf{a}'(\mu)}} \right|^2 .$$

So, we introduce the vector valued function

■
$$\mathbf{q}(\mu) = \frac{\mathbf{a}'(\mu)}{\sqrt{\mathbf{a}(\mu)^{\mathrm{H}} \mathbf{a}'(\mu)}} , \tag{5.24}$$

and compute the cost function using the expression

$$C(\mu, \mathbf{x}) = \left| \mathbf{x}^{\mathrm{H}} \mathbf{q}(\mu) \right|^2 .$$

Here, it is important to note, that the computation of N solutions to a system of equations involving a Toeplitz matrix can be computed in $O(NM \log(M))$ operations. Such an algorithm is described in Section 6.1.12. The application of the pre-whitening approach leads to an algorithm for the computation of the cost function with a higher computational complexity. This is due to the fact, that the decomposition of the covariance matrix used in the pre-whitening approach leads to a matrix $\mathbf{L}$ without Toeplitz structure.

5.1.1   Basis Functions with Kronecker Structure

The radio channel model for a *single* propagation path can be expressed as a weighted sum of the Kronecker product of four basis functions

$$\mathbf{s}\big(\boldsymbol{\theta}_{sp}\big) = \big(\mathbf{B}_T \otimes \mathbf{B}_R \otimes \mathbf{B}_\tau \otimes \mathbf{B}_\alpha\big)\cdot\boldsymbol{\gamma} \in \mathbb{C}^{M\times 1},$$

with $\mathbf{B}_T = \mathbf{B}_T\big(\mu^{(\varphi_T)},\mu^{(\vartheta_T)}\big) \in \mathbb{C}^{M_T\times 2}$, $\mathbf{B}_R = \mathbf{B}_R\big(\mu^{(\varphi_R)},\mu^{(\vartheta_R)}\big)\in \mathbb{C}^{M_R\times 2}$, $\mathbf{B}_f = \mathbf{B}_f(\tau)\in\mathbb{C}^{M_f\times 1}$,
$\mathbf{B}_t(\alpha)\in\mathbb{C}^{M_t\times 1}$, $\boldsymbol{\gamma} = \big[\gamma_{HH}\ \gamma_{HV}\ \gamma_{VH}\ \gamma_{VV}\big]^{\mathrm{T}}$, and the parameter vector describing the single path entirely

$$\boldsymbol{\theta}_{sp,p} = \big[\mu^{(\alpha)}\ \mu^{(\tau)}\ \mu^{(\varphi_T)}\ \mu^{(\vartheta_T)}\ \mu^{(\varphi_R)}\ \mu^{(\vartheta_R)}\ \gamma_{HH,r}\ \gamma_{HH,i}\ \gamma_{HV,r}\ \gamma_{HV,i}\ \gamma_{VH,r}\ \gamma_{VH,i}\ \gamma_{VV,r}\ \gamma_{VV,i}\big]^{\mathrm{T}}.$$

$$(5.25)$$

This follows directly from the data model (2.29). We exploit in the following the Kronecker-structured of basis functions to get more insight into the structure of the correlation function (5.8). Since all matrices $\mathbf{B}_T$, $\mathbf{B}_R$, $\mathbf{B}_f$, $\mathbf{B}_t$ have the same structure, we drop in the following the index (T, R, f, t) and use the general form $\mathbf{B}_i\big(\boldsymbol{\mu}^{(i)}\big)$. First, we decompose the matrix valued function $\mathbf{B}(\boldsymbol{\mu})$ using the matrices $\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big)\in\mathbb{C}^{M_1\times 2}$, $\mathbf{B}_2\big(\boldsymbol{\mu}^{(2)}\big)\in\mathbb{C}^{M_2\times 2}$ into

$$\mathbf{B}(\boldsymbol{\mu}) = \mathbf{B}_2\big(\boldsymbol{\mu}^{(2)}\big)\otimes\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big).\qquad(5.26)$$

And we factorize the covariance matrix $\mathbf{R}_{nn}$ using $\mathbf{R}_1\in\mathbb{C}^{M_1\times M_1}$, $\mathbf{R}_2\in\mathbb{C}^{M_2\times M_2}$ into

$$\mathbf{R}_{nn} = \mathbf{R}_2\otimes\mathbf{R}_1.\qquad(5.27)$$

Using (5.26) and (5.27) in expression (5.8) for the correlation function yields

$$C(\boldsymbol{\mu}) = \mathbf{x}^{\mathrm{H}}\bigg(\Big(\mathbf{R}_2^{-1}\cdot\mathbf{B}_2\big(\boldsymbol{\mu}^{(2)}\big)\cdot\big(\mathbf{B}_2^{\mathrm{H}}\big(\boldsymbol{\mu}^{(2)}\big)\cdot\mathbf{R}_2^{-1}\cdot\mathbf{B}_2\big(\boldsymbol{\mu}^{(2)}\big)\big)^{-1}\cdot\mathbf{B}_2^{\mathrm{H}}\big(\boldsymbol{\mu}^{(2)}\big)\cdot\mathbf{R}_2^{-1}\Big)\otimes$$
$$\otimes\Big(\mathbf{R}_1^{-1}\cdot\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big)\cdot\big(\mathbf{B}_1^{\mathrm{H}}\big(\boldsymbol{\mu}^{(1)}\big)\cdot\mathbf{R}_1^{-1}\cdot\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big)\big)^{-1}\cdot\mathbf{B}_1^{\mathrm{H}}\big(\boldsymbol{\mu}^{(1)}\big)\cdot\mathbf{R}_1^{-1}\Big)\bigg)\cdot\mathbf{x}.$$

$$(5.28)$$

Observe that the cost function retains the same structure. The two-dimensional case can be easily extended to the R_D-dimensional case due to its regular structure. Here, R_D-dimensional refers to the number of data dimensions and not to the number of parameter dimensions R_P (cf. Section 2.4.5). The decomposition of the matrix valued function in the R_D-dimensional case becomes

$$\mathbf{B}(\boldsymbol{\mu}) = \mathbf{B}_{R_D}\big(\boldsymbol{\mu}^{(R_D)}\big)\otimes\ldots\otimes\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big),\qquad(5.29)$$

and the decomposition of the covariance matrix yields

$$\mathbf{R}_{nn} = \mathbf{R}_{R_D}\otimes\ldots\otimes\mathbf{R}_1.\qquad(5.30)$$

Using the expressions (5.26) and (5.27) to express the correlation function leads to

$$C(\boldsymbol{\mu}) = \mathbf{x}^{\mathrm{H}}\bigg(\Big(\mathbf{R}_{R_D}^{-1}\cdot\mathbf{B}_{R_D}\big(\boldsymbol{\mu}^{(R_D)}\big)\cdot\big(\mathbf{B}_{R_D}^{\mathrm{H}}\big(\boldsymbol{\mu}^{(R_D)}\big)\cdot\mathbf{R}_{R_D}^{-1}\cdot\mathbf{B}_{R_D}\big(\boldsymbol{\mu}^{(R_D)}\big)\big)^{-1}\cdot\mathbf{B}_{R_D}^{\mathrm{H}}\big(\boldsymbol{\mu}^{(R_D)}\big)\cdot\mathbf{R}_{R_D}^{-1}\Big)\otimes$$
$$\vdots$$
$$\otimes\Big(\mathbf{R}_1^{-1}\cdot\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big)\cdot\big(\mathbf{B}_1^{\mathrm{H}}\big(\boldsymbol{\mu}^{(1)}\big)\cdot\mathbf{R}_1^{-1}\cdot\mathbf{B}_1\big(\boldsymbol{\mu}^{(1)}\big)\big)^{-1}\cdot\mathbf{B}_1^{\mathrm{H}}\big(\boldsymbol{\mu}^{(1)}\big)\cdot\mathbf{R}_1^{-1}\Big)\bigg)\mathbf{x}.$$

$$(5.31)$$

It is instructive to see that the correlation function keeps its structure in the individual data domains. However, expression (5.31) is somewhat unwieldy. Therefore, we apply the same

pre-whitening approach as in Example 5-2, equations (5.16) - (5.19) . To this end, we define the transformed matrices

$$\mathbf{B}'_r\left(\boldsymbol{\mu}^{(r)}\right) = \mathbf{L}_r^{-1}\mathbf{B}_r\left(\boldsymbol{\mu}^{(r)}\right) \quad \forall r, r = 1,\ldots,R_D ,\tag{5.32}$$

using the decompositions $\mathbf{R}_r = \mathbf{L}_r \mathbf{L}_r^{\mathrm{H}}$. The QR-decomposition of the R_D matrix-valued functions (5.32) is given by

$$\mathbf{B}'_r\left(\boldsymbol{\mu}^{(r)}\right) = \mathbf{Q}'_r\left(\boldsymbol{\mu}^{(r)}\right)\cdot \mathbf{T}'_r\left(\boldsymbol{\mu}^{(r)}\right) \quad \forall r, r = 1,\ldots,R_D .\tag{5.33}$$

The pre-whitening of the observation can be expressed as

■
$$\mathbf{x}' = \left(\mathbf{L}_R^{-1}\otimes\cdots\otimes\mathbf{L}_1^{-1}\right)\cdot\mathbf{x} .\tag{5.34}$$

Using some simple math yields the expression for the correlation function

$$C(\boldsymbol{\mu},\mathbf{x}) = \mathbf{x}'^{\mathrm{H}}\left(\left(\mathbf{Q}'_{R_D}\left(\boldsymbol{\mu}^{(R_D)}\right)\cdot\mathbf{Q}'^{\mathrm{H}}_{R_D}\left(\boldsymbol{\mu}^{(R_D)}\right)\right)\otimes\ldots\otimes\left(\mathbf{Q}'_1\left(\boldsymbol{\mu}^{(1)}\right)\cdot\mathbf{Q}'^{\mathrm{H}}_1\left(\boldsymbol{\mu}^{(1)}\right)\right)\right)\cdot\mathbf{x}'$$

that is equivalent to

$$C(\boldsymbol{\mu},\mathbf{x}) = \mathbf{x}'^{\mathrm{H}}\left(\mathbf{Q}'_{R_D}\left(\boldsymbol{\mu}^{(R_D)}\right)\otimes\ldots\otimes\mathbf{Q}'_1\left(\boldsymbol{\mu}^{(1)}\right)\right)\cdot\left(\mathbf{Q}'^{\mathrm{H}}_{R_D}\left(\boldsymbol{\mu}^{(R_D)}\right)\otimes\ldots\otimes\mathbf{Q}'^{\mathrm{H}}_1\left(\boldsymbol{\mu}^{(1)}\right)\right)\cdot\mathbf{x}' .$$

This expression is useful. We define the vector valued function

■
$$\mathbf{z}'(\boldsymbol{\mu},\mathbf{x}) = \left(\mathbf{Q}'^{\mathrm{H}}_{R_D}\left(\boldsymbol{\mu}^{(R_D)}\right)\otimes\ldots\otimes\mathbf{Q}'^{\mathrm{H}}_1\left(\boldsymbol{\mu}^{(1)}\right)\right)\cdot\mathbf{x}'\tag{5.35}$$

and express the correlation function simply by

$$C(\boldsymbol{\mu},\mathbf{x}) = \left\|\mathbf{z}'(\boldsymbol{\mu},\mathbf{x})\right\|_F^2 .\tag{5.36}$$

The advantage of expression (5.35) is that every data dimension can be processed individually. This is a natural result, if we consider how we would transform channel transfer functions measured with a ULA into the time-delay and the beam-space domain. We apply a 2-D DFT to accomplish that. However, we can also first transform the data into the time-delay domain with a 1-D DFT and in the next step the beam-space using another 1-D DFT. Furthermore, the order of the two transformations can be chosen arbitrarily.

The derived expression for the correlation function (5.35), (5.36) allows the processing of polarimetric channel sounding measurements, as well as the incorporation of the covariance matrix of the dense multipath components. The expression is a prerequisite for a number of parameter estimation algorithms, as we shall see in the next sections. It leads furthermore to a general form of the multidimensional power profile also covering the polarimetric case.

5.1.2 Efficient Computation of the Correlation Function

The straightforward implementation of expression (5.35) leads to an algorithm, which is inefficient especially in terms of memory usage, but also in terms of computational effort. The main drawback of the direct implementation is the computation of the Kronecker product. This leads to large matrices and requires therefore a large amount of memory. Observe that the storage of the Kronecker product of all matrices $\mathbf{Q}_r = \mathbf{Q}_r\left(\boldsymbol{\mu}^{(r)}\right)$ requires more memory in the polarimetric case as the storage of the observation itself. The relationship between the number of samples in the observation $\mathbf{x} \in \mathbb{C}^{M\times1}$ and the column-size of the matrices $\mathbf{Q}_r \in \mathbb{C}^{M_r\times N_r}$ is given by

$$M = \prod_{i=1}^{R_D} M_i .$$

The relation between the length of the vector $\mathbf{z} \in \mathbb{C}^{N \times 1}$ and the row-size of the matrices $\mathbf{Q}_r$ is given by

$$N = \prod_{i=1}^{R_D} N_i .$$

Instead of computing the Kronecker products, one can as well multiply the individual data dimensions r of the observed data with the related matrices $\mathbf{Q}_r$. For the first dimension, we yield the expression

■
$$\mathbf{x}_1 = \mathrm{vec}\left\{ \left(\mathbf{Q}_1^{\mathrm{H}} \cdot \mathrm{mat}\left\{ \mathbf{x}, M_1, \prod_{i=2}^{R_D} M_i \right\} \right)^{\mathrm{T}} \right\} . \tag{5.37}$$

For the penultimate dimension we get

■
$$\mathbf{x}_q = \mathrm{vec}\left\{ \left(\mathbf{Q}_q^{\mathrm{H}} \cdot \mathrm{mat}\left\{ \mathbf{x}_{q-1}, M_q, \prod_{i=1}^{q-1} N_i \cdot \prod_{i=q+1}^{R_D} M_i \right\} \right)^{\mathrm{T}} \right\} , \tag{5.38}$$

and in the last step we compute the result

■
$$\mathbf{z} = \mathrm{vec}\left\{ \left(\mathbf{Q}_{R_D}^{\mathrm{H}} \cdot \mathrm{mat}\left\{ \mathbf{x}_{R_D-1}, M_{R_D}, \prod_{i=1}^{R_D-1} N_i \right\} \right)^{\mathrm{T}} \right\} . \tag{5.39}$$

Observe that the order of the data dimensions can be chosen arbitrarily, since the data dimensions are independent from each other. Observe that the pre-whitening step is not necessary for the data model developed in Chapter 2. Furthermore, the expressions for the matrices $\mathbf{Q}_r$ have been derived in the previous section. Table 5-1 gives an overview of the parameter domains and the related expressions.

Table 5-1: List of parameter domains and related expressions for the construction of the matrices $\mathbf{Q}_r$ required for correlation function computations.

Data domain	Parameter domain	Example and Expression
Antenna ports Tx	Tx azimuth, Tx elevation	Example 5-1, (5.10)
Antenna ports Rx	Rx azimuth, Rx elevation	Example 5-1, (5.10)
Frequency domain	Time delay of arrival	Example 5-4, (5.24)
Time domain	Doppler-shift	Example 5-3, (5.20)

Suppose we want to calculate some samples of the correlation function with respect to one data-, or parameter-domain. For example, we want to calculate the PDP for a specific Doppler-shift $\mu^{(\alpha)}$, Tx-angle $\boldsymbol{\mu}^{(\mathrm{T})} = \left[\mu^{(\varphi_T)} \; \mu^{(\vartheta_T)} \right]$, and Rx- angle $\boldsymbol{\mu}^{(\mathrm{R})} = \left[\mu^{(\varphi_R)} \; \mu^{(\vartheta_R)} \right]$. To this end, we compute two matrices representing two beam formers $\mathbf{Q}(\boldsymbol{\mu}^{(T)})$, $\mathbf{Q}(\boldsymbol{\mu}^{(R)})$. In addition, we compute a vector $\mathbf{q}(\mu^{(\alpha)})$, which plays the role of a Doppler-filter. Applying the three matrices to the respective data dimensions yields a reduced observation containing channel transfer functions $\mathbf{X}'_f \in \mathbb{C}^{M_f \times N_{pol}}$. Using this matrix the PDP can be calculated in an arbitrary resolution by correlation with vectors constructed according to (5.24). Using this approach the computational effort is significantly reduced, if some samples of the correlation function in the same parameter domain have to be computed. Furthermore, the memory requirements for the calculation of a section of the cost function are lower in comparison with the straightforward implementation based on (5.35).

5.1.3   Relation between Correlation function and Multidimensional Power Spectra

Observe also that the correlation function (5.8) has a direct physical interpretation. If we calculate a set of samples of this correlation function in all six parameter domains, we transform the observed data into the time-delay-, Doppler- and angular-domain. One can understand the correlation function as correlation values in a six-dimensional hypercube. Every dimension represents one of the physical parameter dimensions (TDoA, Tx azimuth and elevation, Rx azimuth and elevation, and the Doppler-shift). The large magnitudes of this correlation function correspond to the observable propagation paths. Due to the weighting with the covariance matrix of the stochastic part of the measured channel, the magnitudes of the peaks of this correlation function corresponds directly to the total signal-to-noise-ratio of the individual propagation paths, in terms of channel parameter estimation. The appropriate weighting with the covariance matrix $\mathbf{R}_{nn}$ as well as the handling of the polarisation is the main difference to the usual representation of a radio channel in the joint time-delay- (impulse response), Doppler- (Doppler spectra) and angular-domains (angular power spectra). This is a by-product of the development of the maximum likelihood estimator. Nonetheless, it is a powerful tool providing insight into the characteristics of the radio channel. Furthermore, it is useful for diagnostics purposes, e.g., it can be valuable while implementing a high-resolution parameter estimation algorithm.

Replacing the covariance matrix $\mathbf{R}_{nn}$ with the identity matrix $\mathbf{I}$ yields a representation of the observed radio channel $\mathbf{x}$ as a power profile even for polarimetric channel measurements. Here, the magnitudes of the peaks correspond to the total power transfer factor of the individual propagation paths.

5.1.4   Iterative ML Estimation using Parameter Subset Update Techniques

Suppose the number of observable propagation paths is P. Then the radio channel observation can be expressed as

$$\mathbf{x} = \sqrt{\alpha_0}\,\mathbf{w} + \mathbf{d}_{dmc} + \sum_{p=1}^{P} \mathbf{s}\!\left(\boldsymbol{\theta}_{sp,p}\right).$$

For the clarity we combine the vector $\sqrt{\alpha_0}\,\mathbf{w}$ describing the measurement noise and the contribution of the DMC to the observation in one vector $\mathbf{n}_{dan} = \mathbf{d}_{dmc} + \sqrt{\alpha_0}\,\mathbf{w}$ and express the radio channel observation as

$$\mathbf{x} = \mathbf{n}_{dan} + \sum_{p=1}^{P} \mathbf{s}\!\left(\boldsymbol{\theta}_{sp,p}\right). \tag{5.40}$$

Since the contribution of the concentrated propagation paths can be expressed by the *superposition* of individual paths, we can maximize the correlation function sequentially in respect to the parameters $\boldsymbol{\theta}_{sp,p}$ of theirs. In other words, we minimize the objective function in (5.2) sequentially with respect to parameter subsets.

Let us suppose we have estimates $\hat{\boldsymbol{\theta}}_{sp,p}^{\{i\}}$ for all propagation paths in the ith optimisation step. If we want to improve the estimates of the parameters describing path p, we partially remove the contributions of the other paths by

■ $$\mathbf{x}_p = \mathbf{x} - \sum_{k=1,k\neq p}^{P} \mathbf{s}\!\left(\hat{\boldsymbol{\theta}}_{sp,k}\right). \tag{5.41}$$

This approach reduces the parameter estimation problem to the estimation of the propagation path parameters of path p only, i.e.,

$$\hat{\boldsymbol{\theta}}_{sp,p}^{\{i+1\}} = \arg\min_{\boldsymbol{\theta}_{sp,p}}\left(\mathbf{x}_p^{\{i\}} - \mathbf{s}\left(\boldsymbol{\theta}_{sp,p}\right)\right)^{\mathrm{H}} \cdot \mathbf{R}_{nn}^{-1} \cdot \left(\mathbf{x}_p^{\{i\}} - \mathbf{s}\left(\boldsymbol{\theta}_{sp,p}\right)\right).$$

This iteration can be understood as an application of the expectation maximization algorithm (EM) [69]. Since the parameter estimation problem constitutes a separable least squares problem [72], [73], the non-linear parameters $\boldsymbol{\mu}_p$ and the linear parameters $\boldsymbol{\gamma}_p$ can be improved sequentially. At first, we update the non-linear parameters by using the maximization step

$$\hat{\boldsymbol{\mu}}_p^{\{i+1\}} = \arg\max\left(C\left(\boldsymbol{\mu}, \mathbf{x}_p^{\{i\}}\right)\right). \tag{5.42}$$

In the second step we solve for the linear parameters using a closed from solution, e.g.,

$$\hat{\boldsymbol{\gamma}}_p^{\{i+1\}} = \left(\mathbf{B}^{\mathrm{H}}\left(\boldsymbol{\mu}_p^{\{i+1\}}\right)\cdot \mathbf{R}_{nn}^{-1}\cdot \mathbf{B}\left(\boldsymbol{\mu}_p^{\{i+1\}}\right)\right)^{-1}\cdot \mathbf{B}^{\mathrm{H}}\left(\boldsymbol{\mu}_p^{\{i+1\}}\right)\cdot \mathbf{R}_{nn}^{-1}\cdot \mathbf{x}_p^{\{i\}}.$$

Together, the linear and the non-linear parameters yield the updated parameter vector $\hat{\boldsymbol{\theta}}_{sp,p}^{\{i+1\}}$. Since we update the parameters sequentially for every path, and sequentially for the non-linear and linear parameters, the resulting algorithm can be understood as an application of the SAGE method for maximum-likelihood estimation [62].

Having noticed, that every data-dimension of the path model can be processed separately, when computing the correlation function $C(\boldsymbol{\mu}, \mathbf{x})$, we reduce the size of the jointly updated parameter subspaces further. Since we have four independent data dimensions, the admissible subsets are $\mu_p^{(\varphi_T)}, \mu_p^{(\vartheta_T)}$, $\mu_p^{(\varphi_R)}, \mu_p^{(\vartheta_R)}$, $\mu_p^{(\tau)}$, and $\mu_p^{(\varphi)}$. Observe, that the subsets $\mu_p^{(\varphi_T)}, \mu_p^{(\vartheta_T)}$ and $\mu_p^{(\varphi_R)}, \mu_p^{(\vartheta_R)}$ should not be divided any further. This is due to the fact that, the response of an antenna array cannot be decomposed into a product of two independent array responses, e.g., into an azimuth- and an elevation-response. However, there exist some specially array structures where this decomposition is possible. Namely, the array response of an ULA or an URA can be factorized in this way as discussed in Section 2.4.

The outlined algorithm is very simple and regular in its structure and therefore very attractive. Unfortunately, the estimator has a serious drawback. The fundamental problem lies in the very first simplification, i.e., to optimize the parameter sets for every propagation path independently. This approach will result in a fast converging algorithm only if the modes of the individual paths are almost orthogonal. Otherwise, we have to accept that, due to the coupling between the modes, the convergence speed is significantly reduced. This issue of parameter coupling has already been addressed in Section 4.1.4. The solution to iterative maximum likelihood estimation and further discussions regarding their convergence speed can be found in Section 5.2.3 and the following.

One should note that the outlined maximization strategy is similar to the alternating projection algorithm [74] developed by Ziskind et al. and the RELAX algorithm proposed by Li et al. in [75], [76], [44]. Although they have not been designed for the same data model, the underlying estimator concept is clearly the same. Furthermore, the SAGECPE1 [63], [64], [11] is based on the same estimation strategy, but ignores the contribution of the dense multipath components. The algorithm is based on the assumption that a channel observation contains only concentrated propagation paths and i.i.d. circular Gaussian noise.

5.1.5   Path Parameter Initialization for Iterative Maximum Likelihood Algorithms

A critical part in all iterative optimisation algorithms is the choice of a good initial value. An exhaustive search over all possible parameter combinations is not feasible considering the high computational complexity. In radio channel parameter estimation, often sequences of some hundreds even several thousands of channel observations, taken over time, have to be processed. Since the structural parameters are slow time variant, it is not necessary to deter-

mine initial parameters for all propagation paths. Instead, we use the estimates of the structural parameters of the previous observation as initial values for the actual observation. Having estimates for the structural parameters, we can always calculate initial values for the linear model parameters, i.e., the path weights using a closed form solution such as (5.5). This approach to parameter initialization inherently provides a means to track propagation paths over a long time. We will only loose track of an individual path, if it coincides with a second path, if it is shadowed for some time, or if the estimator jumps to another path. The tracking of propagation paths is out of the focus of this work. A first attempt to improve the tracking performance of the channel parameter estimator developed in this work has been made by Algeier et al., the approach and first results are presented in [48] and [77].

Since some propagation paths become suddenly visible if a time variant radio channel is measured, we are still left with the task to estimate initial parameters of "newborn" propagation paths. To accomplish this we remove at first the contributions of the tracked propagation paths from the actual observation. This can be carried out using

$$\mathbf{x}_r = \mathbf{x} - \sum_{p=1}^{P} \mathbf{s}\!\left(\hat{\boldsymbol{\theta}}_{sp,p}\right). \tag{5.43}$$

In the remainder $\mathbf{x}_r$, we search for new propagation paths and their parameters. If the number of parameter-domains is small, $R_p \leq 3$ as a rule of thumb, we can calculate the correlation function of the remainder $\mathbf{x}_r$ sampled sufficiently densely in all parameter domains. The position of the maximum of the correlation function determines the initial structural parameters $\boldsymbol{\mu}_{P+1}$ for the new propagation path. Using these parameters the linear parameters can be estimated from $\mathbf{x}_r$ by the close form solution (5.5). However, if the number of parameter domains is large the effort for the computation of the correlation function, which is dense sampled in all parameter dimensions, is very high. If we accept an approach being statistically less efficient, we can reduce the computational cost. The basic idea is to exploit the independence of the data domains, while respecting the covariance matrix $\mathbf{R}_{nn}$ of the stochastic part of the observation.

Let us suppose for the beginning that the covariance matrix $\mathbf{R}_{nn}$ has the structure $\mathbf{R}_{nn} = \alpha_0 \mathbf{I}$. Applying expression (5.37) to $\mathbf{x}_r$ yields the matrix

$$\mathbf{X}_{r,1} = \mathrm{mat}\left\{\mathbf{x}_r, M_1, \prod_{i=2}^{R_D} M_i\right\}. \tag{5.44}$$

Now motivated by the Kronecker-structure of the data model we interpret the columns of $\mathbf{X}_{r,1}$ as realisations of the same process. Using a matrix $\mathbf{Q}_1\!\left(\boldsymbol{\mu}^{(1)}\right)$, we calculate a set of $\prod_{i=2}^{R_D} M_i$ correlation values by

$$\mathbf{Z}_1\!\left(\boldsymbol{\mu}^{(1)}\right) = \mathbf{Q}_1^{\mathrm{H}}\!\left(\boldsymbol{\mu}^{(1)}\right) \cdot \mathbf{X}_{r,1}. \tag{5.45}$$

Ignoring the interrelationship between the columns of $\mathbf{Z}_1\!\left(\boldsymbol{\mu}^{(1)}\right)$ we calculate an incoherent averaged correlation function by

$$C\!\left(\boldsymbol{\mu}^{(1)}, \mathbf{X}_{r,1}\right) = \left\|\mathbf{Z}_1\!\left(\boldsymbol{\mu}^{(1)}\right)\right\|_F^2. \tag{5.46}$$

Solving the reduced maximization problem

$$\hat{\boldsymbol{\mu}}^{(1)} = \arg\max_{\boldsymbol{\mu}^{(1)}} C\!\left(\boldsymbol{\mu}^{(1)}, \mathbf{X}_{r,1}\right) \tag{5.47}$$

we yield an estimate of the parameter(s) $\hat{\boldsymbol{\mu}}^{(1)}$ for a new candidate propagation path. One should observe that the oversampling of the correlation function by a factor of 4 is sufficient

in most cases. This is because we search for a raw initial estimate of the parameters only. In the next step, we use the estimated parameter $\hat{\boldsymbol{\mu}}^{(1)}$ to reduce the dimensionality of the estimation problem by computing

$$\mathbf{x}_{r,2} = \text{vec}\left\{ \left(\mathbf{Q}_1^{\text{H}}\left(\hat{\boldsymbol{\mu}}^{(1)}\right) \cdot \text{mat}\left\{ \mathbf{x}_r, M_1, \prod_{i=2}^{R_D} M_i \right\} \right)^{\text{T}} \right\}, \tag{5.48}$$

i.e. we apply the same idea as in Section 5.1.2. In the next step, we determine raw estimates for $\boldsymbol{\mu}^{(2)}$ by replacing $\mathbf{x}_r$ with $\mathbf{x}_{r,2}$ in equation (5.44) and the following. Having estimates $\hat{\boldsymbol{\mu}}$ of all nonlinear parameters the estimates of the path weights can be calculated using the closed form expression (5.5). Together the estimates $\hat{\boldsymbol{\mu}}_{P+1}$ and $\hat{\boldsymbol{\gamma}}_{P+1}$ yield an estimate $\hat{\boldsymbol{\theta}}_{P+1}$ of the parameters of a new propagation path. Since some propagation paths are removed from the entirety of observed concentrated propagation paths after some channel observations, and new propagation paths are added from time to time it is important to assign every new propagation path a unique identifier. This is crucial important for further processing of estimated radio channel parameters, e.g., to track the parameters of individual propagation paths over their lifetime.

To clarify the global search strategy let us discuss an example. Suppose the channel transfer function has been measured using a 10-element ULA at one link end. At first, we calculate the 10 related individual channel impulse responses, treat them as independent realisations of the same process, and maximize the log-likelihood function with respect to the time delay of arrival. Here, a non-coherent combining procedure avoids any assumption on unknown DoA, or DoD in this step. In the next step, we keep the estimated time delay fixed and maximize for the DoA or DoD. This reduces the maximization problem to two concatenated one-dimensional problems. Any arbitrary assumption of the DoA or DoD angles in the example would implicitly realise coherent combining, which potentially disregards paths impinging from other angles by beam forming. This kind of non-coherent handling of data dimensions related to unknown parameters (e.g. DoA, DoD, Doppler-shift) gives us a higher probability to detect the relevant parameters, which is the time delay of arrival in the example.

The drawback of the outlined procedure is the loss of correlation gain, since we process some of the data domains incoherently. The loss of correlation gain is highest in the first step. This is the reason why the data domain with the largest amount of samples M_r should be processed first. The outlined procedure is a trade-off between computational complexity and robustness. A similar procedure is used in the SAGECPE1 [63], [64] and the RELAX [44] algorithms.

The outlined algorithm is not yet applicable to determine raw initial estimates for new propagation paths, since we have ignored the contribution of the DMC to the channel observation. In general, we have to work with the pre-whitened data (5.34). However, we can avoid the pre-whitening step, since the stochastic process n_{dan} is only correlated in the frequency or time delay domain. To this end, we start the global search for an initial solution always in the frequency domain, i.e., we search for a raw estimate of the time delay of arrival $\hat{\tau}_{P+1}$. The respective correlation vectors, taking the colouring of the stochastic process n_{dan} into account, are defined in (5.24). The necessary condition for this approach is that the full covariance matrix allows the factorisation

$$\mathbf{R} = \mathbf{R}_{R_D} \otimes \ldots \otimes \mathbf{R}_2 \otimes \mathbf{R}_1 .$$

The prewhitening step is necessary in general to guarantee maximum ratio combining of the correlation values in (5.46).

Having estimated raw parameters of the new propagation path, the outlined procedure can be repeated to find additional paths by removing the contribution of the new path from $\mathbf{x}_r$. I.e. we increase the number of propagation paths $P = P+1$ and add the new parameters $\hat{\boldsymbol{\theta}}_{P+1}$ to the complete parameter set $\hat{\boldsymbol{\theta}}$. The algorithm for the estimation of raw parameters $\hat{\boldsymbol{\theta}}_{P+1}$ is outlined in Table 5-2. For further discussions on global optimisation techniques, see also Chapter 4 in [78]. Observe that the explicit estimation of the linear parameters $\boldsymbol{\gamma}$ is not necessarily required throughout the algorithm since the estimated data $\mathbf{s}(\hat{\boldsymbol{\theta}}_{P+1})$ can often be reconstructed without computation of the estimate $\hat{\boldsymbol{\gamma}}$, e.g., using equation (5.15) or (5.23).

One should note that the outlined procedure might not determine a valid initial estimate for the parameters of new propagation paths if multiple propagation paths arise at the same time in a measured scenario having the same power. This issue is for example addressed in [29], [30]. The authors show that the antenna array design has a significant influence on the probability of false detection. The probability of false detection depends clearly on the orthogonality between the basis-vectors of the new propagation paths in a scenario see also Cond. 5, Section 4.1. If the antenna arrays used for channel sounding are properly designed, and parameter tracking is applied the probability of "ghost path" detection can be kept to a minimum.

Table 5-2: Computation of initial path parameter estimates for a new propagation path.

Input:

Data vector $\mathbf{x}$, parameters of the known propagation paths $\hat{\boldsymbol{\theta}}_{sp}$.

Preprocessing:

Sample all parameter domains sufficiently dense (oversampling) yielding $\boldsymbol{\mu}_s^{(r)}$. Compute all matrices $\mathbf{Q}_r(\boldsymbol{\mu}^{(r)})$ related to the samples $\boldsymbol{\mu}_s^{(r)}$.

1) Compute the remainder $\mathbf{x}_{rem} = \mathbf{x} - \sum_{p=1}^{P} \mathbf{s}(\hat{\boldsymbol{\theta}}_{sp,p})$.

2) Form the data matrix $\mathbf{X}_{rem,1} = \mathrm{mat}\left\{ \mathbf{x}_{rem}, M_1, \prod_{i=2}^{R_D} M_i \right\}$.

3) Compute all correlation values $C(\boldsymbol{\mu}^{(1)}, \mathbf{X}_{rem,1})$ for the set $\boldsymbol{\mu}_s^{(r)}$ using

$$C(\boldsymbol{\mu}^{(1)}, \mathbf{X}_{rem,1}) = \left\| \mathbf{Z}_1(\boldsymbol{\mu}^{(1)}) \right\|_F^2 \text{ and } \mathbf{Z}_1(\boldsymbol{\mu}^{(1)}) = \mathbf{Q}_1^H(\boldsymbol{\mu}^{(1)}) \cdot \mathbf{X}_{r,1}.$$

4) Determine the raw estimate $\hat{\boldsymbol{\mu}}^{(1)} = \arg\max_{\boldsymbol{\mu}^{(1)}} C(\boldsymbol{\mu}^{(1)}, \mathbf{X}_{rem,1})$.

5) Compute the reduced remainder

$$\mathbf{x}_{r,2} = \mathrm{vec}\left\{ \left(\mathbf{Q}_1^H(\hat{\boldsymbol{\mu}}^{(1)}) \cdot \mathrm{mat}\left\{ \mathbf{x}_r, M_1, \prod_{i=2}^{R_D} M_i \right\} \right)^T \right\}.$$

6) Repeat steps 2 - 5 for the remaining data-dimensions.

5.2   Local Maximization

The algorithms described in the previous section are all based on grid search procedures. There are basically two drawbacks of the estimation procedure outlined so far. At first the grid search strategies are in general inefficient in terms of computational complexity, since we

calculate some function values to improve only one estimate. Secondly, a full search over all parameter dimensions is not feasible. Therefore, we have to break down the full multidimensional search into one- or two-dimensional search steps. Consequently, we will inevitably ignore the coupling between parameters. This leads in turn to a considerable reduction in convergence speed if two parameters of the estimated data model are coupled.

Hence we make a step back to the formula (5.4). Basically we have to find the minimum of the function

$$\left(\mathbf{x}-\mathbf{B}(\boldsymbol{\mu})\cdot\boldsymbol{\gamma}\right)^{\mathrm{H}}\cdot\mathbf{R}_{nn}^{-1}\cdot\left(\mathbf{x}-\mathbf{B}(\boldsymbol{\mu})\cdot\boldsymbol{\gamma}\right). \tag{5.49}$$

Now we recall an important property of the data model. The model has continuous first and second order partial derivatives, which can be calculated algebraically. Namely, the score function (4.10) is the first gradient of (5.49). The calculation of the score function requires only the computation of $\mathbf{s}(\boldsymbol{\theta})$ and of the related Jacobian matrix $\mathbf{D}(\boldsymbol{\theta})=\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\mathbf{s}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}$. In Section (4.2) we have already derived the expressions for $\mathbf{D}(\boldsymbol{\theta})$, see equations (4.68), (4.73), and (4.77). Furthermore, the second gradient at point $\boldsymbol{\theta}$ of the function to minimize (5.49) is the stochastic Fisher matrix as defined in (4.15). Hence, we reformulate the parameter estimation problem as follows

$$\mathbf{q}\!\left(\mathbf{x}\big|\hat{\boldsymbol{\theta}}\right)=\mathbf{0}. \tag{5.50}$$

Since the Mahalanobis norm is not a convex function of the structural parameters $\boldsymbol{\mu}$ multiple solutions to (5.50) exist. No closed form solution is available for this optimisation problem. Therefore, we have to use an iterative procedure to find $\hat{\boldsymbol{\theta}}$.

Candidate algorithms for the iterative minimization of (5.2) are the steepest descend [79], [80], the Newton-Raphson [22], the Gauß-Newton [22], [81], and the Levenberg-Marquardt algorithm [82], [68]. All methods update the actual estimate $\hat{\boldsymbol{\theta}}^{\{i\}}$ along some descend direction $\Delta\hat{\boldsymbol{\theta}}^{\{i\}}$. The general parameter update expression for all gradient-based methods is

$$\hat{\boldsymbol{\theta}}^{\{i+1\}}=\hat{\boldsymbol{\theta}}^{\{i\}}+\Delta\hat{\boldsymbol{\theta}}^{\{i\}}.$$

If the new parameter $\hat{\boldsymbol{\theta}}^{\{i+1\}}$ reduces the Mahalanobis norm

$$\left(\mathbf{x}-\mathbf{s}\!\left(\hat{\boldsymbol{\theta}}^{\{i+1\}}\right)\right)^{\mathrm{H}}\cdot\mathbf{R}_{nn}^{-1}\cdot\left(\mathbf{x}-\mathbf{s}\!\left(\hat{\boldsymbol{\theta}}^{\{i+1\}}\right)\right)<\left(\mathbf{x}-\mathbf{s}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}\right)\right)^{\mathrm{H}}\cdot\mathbf{R}_{nn}^{-1}\cdot\left(\mathbf{x}-\mathbf{s}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}\right)\right) \tag{5.51}$$

it is accepted. If the norm is not reduced, a new trial step $\Delta\hat{\boldsymbol{\theta}}^{\{i\}}$ has to be calculated. In the following, a short overview of the algorithms is given.

5.2.1   Steepest Descent Method

The steepest descend method updates the parameter vector $\hat{\boldsymbol{\theta}}^{\{i\}}$ along the gradient at $\hat{\boldsymbol{\theta}}^{\{i\}}$

$$\Delta\hat{\boldsymbol{\theta}}_{sd}^{\{i\}}=\varsigma\cdot\mathbf{q}\!\left(\mathbf{x}\big|\hat{\boldsymbol{\theta}}^{\{i\}},\mathbf{R}_{nn}\right).$$

Where λ has to be chosen such that (5.51) is satisfied. The update equation of the steepest descend algorithm is

$$\hat{\boldsymbol{\theta}}^{\{i+1\}}=\hat{\boldsymbol{\theta}}^{\{i\}}+\varsigma\cdot\mathbf{q}\!\left(\mathbf{x}\big|\hat{\boldsymbol{\theta}}^{\{i\}},\mathbf{R}_{nn}\right). \tag{5.52}$$

Expression (5.52) can be understood in terms of the EM algorithm as a one-dimensional search over the parameter ς. In infinite precision arithmetic the steepest descend algorithm will converge to the next local minima or saddle point. The advantage of the algorithm in

general is the low computational cost of a single iteration. The disadvantage is the suboptimal search direction [22].

5.2.2   Newton-Raphson Method

The Newton or Newton-Raphson procedure is a means to find zeros of a smooth function. It is based on the assumption that the objective function is quadratic between the actual point $\hat{\boldsymbol{\theta}}^{\{i\}}$ and the solution $\hat{\boldsymbol{\theta}}$.

The Taylor series expansion of the log-likelihood function around $\boldsymbol{\theta}_0$ up to the quadratic term yields

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn}) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn}) \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) .$$

Using the definition of the score function (4.6) and the stochastic Fisher matrix (4.15) in the Taylor series expansion yields

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn}) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \mathbf{q}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn}) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \tilde{\mathcal{J}}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn})(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \quad (5.53)$$

Consequently, the Taylor series for the score function at point $\boldsymbol{\theta}_0$ is

$$\mathbf{q}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R}_{nn}) = \mathbf{q}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn}) - \tilde{\mathcal{J}}(\mathbf{x}|\boldsymbol{\theta}_0, \mathbf{R}_{nn})(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \quad (5.54)$$

Using (5.50) in (5.54) yields the Newton-Raphson step for iterative maximum likelihood estimation as

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \tilde{\mathcal{J}}^{-1}\left(\mathbf{x}|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) \mathbf{q}\left(\mathbf{x}|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right). \quad (5.55)$$

Observe, that the expression for the stochastic Fisher matrix is equivalent to the exact Hessian of the objective function at $\hat{\boldsymbol{\theta}}^{\{i\}}$.

The Newton-Raphson method converges only to a solution of (5.50) if the higher order terms neglected in the Taylor series expansion are small compared to the absolute value of the function approximation. For nonlinear least squares problems this condition is not always satisfied. Therefore, condition (5.51) may not be fulfilled by the provisional step $\hat{\boldsymbol{\theta}}^{\{i+1\}}$. One can reduce the step length to improve global convergence leading to the form

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \varsigma\, \tilde{\mathcal{J}}^{-1}\left(\mathbf{x}|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) \mathbf{q}\left(\mathbf{x}|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right).$$

However, the reduction of the step size does not assure strict convergence of the Newton-Raphson method. The crucial point is the selection of the exact Hessian, i.e., of the stochastic Fisher matrix as a projector for the gradient. The authors of [22] show that the exact Hessian is not necessarily positive definite. Consequently, the sign of the projected gradient may change and the method will not converge to the maximum of the likelihood function or to the minimum of the negative log-likelihood function. Hence the Newton-Raphson method is not globally convergent. Nevertheless, if the method converges to a solution it will need significantly less iterations than the steepest descent algorithm.

5.2.3   Gauß-Newton Method

The Gauß-Newton method can be understood as a derivative of the Newton-Raphson method. The main advantage is the improved global convergence. As stated above the main drawback of the Newton-Raphson method is the choice of the exact Hessian to project the gradient. In

the Gauß-Newton method the Hessian is replaced by the first order approximation of the exact Hessian

$$2 \cdot \Re\left(\mathbf{D}^{\mathrm{H}}\left(\hat{\boldsymbol{\theta}}^{\{i\}}\right)\mathbf{R}_{nn}^{-1}\,\mathbf{D}\left(\hat{\boldsymbol{\theta}}^{\{i\}}\right)\right).$$

Interestingly, this approximation of the Hessian is equivalent to the expression for the Fisher information matrix defined in (4.11). Since the Fisher information matrix is a covariance matrix it is always positive semi definite. The iteration for the Gauß-Newton algorithm is

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \varsigma\,\mathcal{J}^{-1}\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\mathbf{q}\left(\mathbf{x}\big|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right).$$

Again, the step length ς has to be chosen such that (5.51) holds. The global convergence of the Gauß-Newton method is better than the global convergence of the Newton-Raphson method due to the choice of the approximated Hessian [22]. Nevertheless, also the convergence speed of the Gauß-Newton algorithm is not always sufficient. The inverse of $\mathcal{J}\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)$ is subjected to numerical errors if the problem is bad conditioned at $\hat{\boldsymbol{\theta}}^{\{i\}}$. Furthermore, the projected gradient $\mathcal{J}^{-1}\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\mathbf{q}\left(\mathbf{x}\big|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)$ may not point to the solution if the function to maximize or minimize is not quadratic close to $\hat{\boldsymbol{\theta}}^{\{i\}}$.

Close to the solution, the convergence speed of the Gauß-Newton algorithm is significantly higher than the convergence speed of the steepest-descend algorithm. In the early iterations, it cannot be expected that the modified-Newton method will converge quickly to a solution. However, locally a Gauß-Newton algorithm has at least super-linear convergence and quadratic convergence for zero residual problems (see, e.g., [83]).

The application of the Gauß-Newton algorithm for high resolution propagation path parameter estimation has been described in [84], [85].

5.2.4   Levenberg-Marquardt Method

The Levenberg method was introduced first in [82] and 19 years later derived in a different framework by D. Marquardt in [68]. For many nonlinear optimisation tasks, the Levenberg-Marquardt algorithm is the method of choice (cf. [81], [86], and [87]). It belongs to the class of trust-region algorithms. The Gauß-Newton and the Newton-Raphson method are based on the assumption that the objective function is quadratic near $\hat{\boldsymbol{\theta}}^{\{i\}}$. For nonlinear functions, this assumption holds only for a small region around $\hat{\boldsymbol{\theta}}^{\{i\}}$. Consequently, both algorithms do not converge if the step size $\Delta\boldsymbol{\theta}^{\{i\}}$ is too large. The basic idea of the trust region algorithms is to constrain the step size such that the assumption of a quadratic objective function holds. The Taylor series for the residual $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{L}^{-1}(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))$ at $\boldsymbol{\theta}_0$ produces the following linear model

$$\mathbf{r}(\boldsymbol{\theta}) \approx \mathbf{L}^{-1}\left(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}_0) - \mathbf{D}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right),$$

where $\mathbf{L}$ denotes the decomposition of the covariance matrix $\mathbf{R}_{nn} = \mathbf{L}^{\mathrm{H}} \cdot \mathbf{L}$. Hence the predicted Mahalanobis norm of the residual, assuming a quadratic objective function in $\Delta\boldsymbol{\theta}^{\{i\}}$, is

$$\left\|\mathbf{r}\left(\boldsymbol{\theta}^{\{i+1\}}\right)\right\|_F^2 \approx \left\|\mathbf{L}^{-1}\left(\mathbf{x} - \mathbf{s}\left(\boldsymbol{\theta}^{\{i\}}\right) - \mathbf{D}\left(\boldsymbol{\theta}^{\{i\}}\right) \cdot \Delta\boldsymbol{\theta}^{\{i\}}\right)\right\|_F^2.$$

The idea behind the trust-region algorithms is to choose a step size $\Delta\boldsymbol{\theta}^{\{i\}}$ such that

$$\frac{\left(\mathbf{x} - \mathbf{s}\left(\boldsymbol{\theta}^{\{i+1\}}\right)\right)^{\mathrm{H}}\mathbf{R}_{nn}^{-1}\left(\mathbf{x} - \mathbf{s}\left(\boldsymbol{\theta}^{\{i+1\}}\right)\right) - \left(\mathbf{x} - \mathbf{s}\left(\boldsymbol{\theta}^{\{i\}}\right)\right)^{\mathrm{H}}\mathbf{R}_{nn}^{-1}\left(\mathbf{x} - \mathbf{s}\left(\boldsymbol{\theta}^{\{i\}}\right)\right)}{\left(\Delta\boldsymbol{\theta}^{\{i\}}\right)^{\mathrm{T}}\mathcal{J}\left(\boldsymbol{\theta}^{\{i\}}, \mathbf{R}_{nn}\right)\Delta\boldsymbol{\theta}^{\{i\}} - 2 \cdot \Re\left\{\left(\mathbf{x} - \mathbf{s}\left(\boldsymbol{\theta}^{\{i\}}\right)\right)^{\mathrm{H}}\mathbf{R}_{nn}^{-1}\mathbf{D}\left(\boldsymbol{\theta}^{\{i\}}\right)\Delta\boldsymbol{\theta}^{\{i\}}\right\}} = \rho^{\{i\}} \sim 1. \qquad (5.56)$$

The numerator in (5.56) is referred to as the actual change and the denominator is called the predicted change. If the ratio $\rho^{\{i\}}$ is close to one, the actual change of the objective function is

similar to the predicted change of the objective function that means we can trust in the step $\Delta\boldsymbol{\theta}^{\{i\}}$. In other words, the assumption of a quadratic function holds. If the ratio is close to zero or negative, the trust region must be shrunk. The trust region radius $\Delta^{\{i\}}$ determines the trust region where $\left\|\Delta\boldsymbol{\theta}^{\{i\}}\right\| \leq \Delta^{\{i\}}$. Levenberg proposed in [82] to modify the Gauß-Newton method as follows

$$\hat{\boldsymbol{\theta}}_l^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \left(\boldsymbol{\mathcal{J}}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) + \varsigma^{\{i\}} \mathbf{I}\right)^{-1} \mathbf{q}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right). \tag{5.57}$$

J. Nocedal and S. Wright have show in [88] that a $\varsigma^{\{i\}}$ exist such that $\left\|\Delta\boldsymbol{\theta}^{\{i\}}\right\| = \Delta^{\{i\}}$. They proposed the following iterative procedure to determine its value. Given an initial value $\varsigma_0^{\{i\}}$ and a trust region radius $\Delta^{\{i\}}$, compute the Cholesky factorisation in every iteration l

$$\boldsymbol{\mathcal{J}}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) + \varsigma_l^{\{i\}} \mathbf{I} = \left(\mathbf{U}^{\{i\}}\right)^{\mathrm{T}} \mathbf{U}^{\{i\}}. \tag{5.58}$$

Using the decomposition solve $\left(\mathbf{U}^{\{i\}}\right)^{\mathrm{T}} \mathbf{U}^{\{i\}} \cdot \Delta\boldsymbol{\theta}_l^{\{i\}} = \mathbf{q}\!\left(\mathbf{x}\middle|\boldsymbol{\theta}^{\{i\}}, \mathbf{R}_{nn}\right)$ and $\left(\mathbf{U}^{\{i\}}\right)^{\mathrm{T}} \mathbf{z}_l^{\{i\}} = \Delta\boldsymbol{\theta}_l^{\{i\}}$. Finally, update $\varsigma_l^{\{i\}}$ as

$$\varsigma_{l+1}^{\{i\}} = \varsigma_l^{\{i\}} + \frac{\left\|\Delta\boldsymbol{\theta}_l^{\{i\}}\right\|^2}{\left\|\mathbf{z}_l^{\{i\}}\right\|^2} \cdot \frac{\left\|\Delta\boldsymbol{\theta}_l^{\{i\}}\right\| - \Delta^{\{i\}}}{\Delta^{\{i\}}}. \tag{5.59}$$

The iteration is terminated if the ratio $\frac{\varsigma_{l+1}^{\{i\}}}{\varsigma_l^{\{i\}}}$ is close to one, i.e., if the update is below some tolerance. Observe that for positive $\varsigma_l^{\{i\}}$ the Cholesky decomposition (5.58) exist since $\boldsymbol{\mathcal{J}}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) + \varsigma_l^{\{i\}} \mathbf{I}$ is positive semi definite. Using the solution $\varsigma^{\{i\}}$ the update step $\Delta\boldsymbol{\theta}^{\{i\}}$ is computed. Finally, the trust region has to be adjusted based on the ratio between the actual change and the predicted change of the objective function $\rho^{\{i\}}$, e.g., by

$$\Delta^{\{i+1\}} = \begin{cases} \frac{1}{4}\left\|\Delta\boldsymbol{\theta}^{\{i\}}\right\| & \rho^{\{i\}} < \frac{1}{4} \\ \min\!\left(2\Delta^{\{i\}}, \Delta_{\max}\right) & \rho^{\{i\}} > \frac{3}{4} \\ \Delta^{\{i\}} & otherwise \end{cases} \tag{5.60}$$

where $\Delta_{\max}$ is the largest admissible trust region radius. Provided the initial trust region radius $\Delta^{\{0\}}$ is small enough, this solution for $\varsigma^{\{i\}}$ assures that the Levenberg algorithm converges to the closest maximum or minimum of the objective function. The drawback of this approach is the computational complexity of the iteration (5.59). For problems with some hundred unknowns the Cholesky factorisation (5.58) and the computation of the solutions $\Delta\boldsymbol{\theta}_l^{\{i\}}, \mathbf{z}_l^{\{i\}}$ is too expensive.

Hence, a different suboptimal approach is often used for large-scale problems [89]. If the objective function is improved for a given $\varsigma^{\{i\}}$ the weight is decreased by a certain factor, e.g., $\varsigma^{\{i+1\}} = \frac{1}{10}\varsigma^{\{i\}}$. If the objective function is not improved the weight is increased, e.g., $\varsigma^{\{i+1\}} = 10\varsigma^{\{i\}}$. In terms of maximum likelihood estimation improved means that the likelihood of the new parameter vector $\boldsymbol{\theta}^{\{i+1\}}$ is larger than the likelihood of $\boldsymbol{\theta}^{\{i\}}$. For the estimation of the propagation path parameters (DML), this is also equivalent to a decrease of the Mahalanobis norm (5.49).

The resulting algorithm is computationally less expensive than an algorithm based on the approach of Nocedal and Wright. However, in a strict sense the simplified algorithm is not a trust region algorithm anymore, since we do not verify whether the assumption of a quadratic objective function was valid or not. The simplified algorithm is sufficient for path parameter estimation if it is initialized with a parameter vector, which is close to the solution. The global search algorithm described in the previous section using a sufficiently dense search grid can ensure this.

It is instructive to analyze extreme values of the weight $\varsigma^{\{i\}}$. If the weight $\varsigma^{\{i\}}$ is significantly smaller than the smallest eigenvalue λ_i of $\mathcal{J}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)$, i.e., $\varsigma << \lambda_i, \forall i$ the Levenberg step becomes a Gauß-Newton step

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \mathcal{J}^{-1}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\mathbf{q}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right).$$

If the weight is larger than the maximum eigenvalue $\varsigma >> \lambda_i, \forall i$, the Levenberg step becomes a Steepest Descent step

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \frac{1}{\varsigma}\mathbf{q}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right).$$

As already mentioned the method described by D. Marquardt in [68] is similar to the Levenberg method. The main drawback of the Levenberg algorithm is that we ignore the metric of the individual parameters. D. Marquardt observed that the main diagonal elements of $\mathcal{J}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)$ are related to the metric of the parameters. He proposed to scale the main diagonal elements of $\mathcal{J}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)$ by $1+\varsigma$ yielding

$$\hat{\boldsymbol{\theta}}_m^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \left(\mathcal{J}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) + \varsigma\,\mathbf{I} \circ \mathcal{J}\!\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\right)^{-1} \mathbf{q}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right). \qquad (5.61)$$

If the factor $\varsigma^{\{i\}}$ is much smaller than one, the step becomes a Gauß-Newton step, i.e.,

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \mathcal{J}^{-1}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\mathbf{q}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right).$$

Moreover, if the weight $\varsigma^{\{i\}}$ is significantly larger than 1.0 it becomes a Steepest Descent method with matched metric, i.e.,

■ $$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \left(\varsigma\,\mathbf{I} \circ \mathcal{J}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\right)^{-1} \mathbf{q}\!\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right).$$

Therefore, the weight $\varsigma^{\{i\}}$ is a relative value in (5.61) and an absolute value in (5.57). Hence, if we apply the Marquardt method, we can make sure that the algorithm starts with small steps by choosing a value, which is significantly larger than one.

Since the algorithms developed by K. Levenberg and D. Marquardt are similar they are usually called Levenberg-Marquardt methods in the literature. The complete algorithm is outlined in Table 5-3.

Table 5-3: Iterative optimisation of the parameters $\boldsymbol{\theta}_{sp}$ using the Levenberg-Marquardt algorithm.

Input: observation $\mathbf{x}$, initial solution $\hat{\boldsymbol{\theta}}^{\{0\}}$, covariance matrix $\mathbf{R}_{nn}$, ς.

1) Compute the component matrices of the Jacobian matrix Table 4-2 - Table 4-4.

2) Compute the score function
$$\mathbf{q}\left(\mathbf{x}\middle|\boldsymbol{\theta}^{\{i\}},\mathbf{R}_{nn}\right)=2\cdot\mathfrak{R}\left\{\mathbf{D}^{\mathrm{H}}\left(\boldsymbol{\theta}^{\{i\}}\right)\mathbf{R}_{nn}^{-1}\left(\mathbf{x}-\mathbf{s}\left(\boldsymbol{\theta}^{\{i\}}\right)\right)\right\}.$$

3) Compute the approximation of the Hessian using equations (4.70), (4.75), or (4.78)
$$\mathcal{J}\left(\hat{\boldsymbol{\theta}}^{\{i\}},\mathbf{R}_{nn}\right).$$

4) Compute the parameter update
$$\hat{\boldsymbol{\theta}}_{m}^{\{i+1\}}=\hat{\boldsymbol{\theta}}^{\{i\}}+\left(\mathcal{J}\left(\hat{\boldsymbol{\theta}}^{\{i\}},\mathbf{R}_{nn}\right)+\varsigma\,\mathbf{I}\circ\mathcal{J}\left(\hat{\boldsymbol{\theta}}^{\{i\}},\mathbf{R}_{nn}\right)\right)^{-1}\mathbf{q}\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{\{i\}},\mathbf{R}_{nn}\right).$$

5) Check strict maximization $\mathcal{L}\left(\mathbf{x}\middle|\boldsymbol{\theta}^{\{i+1\}}\right)>\mathcal{L}\left(\mathbf{x}\middle|\boldsymbol{\theta}^{\{i\}}\right)$; yes: set $\varsigma=\frac{\varsigma}{4}$ go to 6. , no: set $\varsigma=8\varsigma$ go to 4.

6) Check convergence; not converged: set $i=i+1$ and go to 1.

Observe that in contrast to other large-scale problems the computation of the Hessian is computationally inexpensive. For the computation of the gradient and the Mahalanobis norm the same approach as in (5.37) - (5.39) can be applied to avoid the computation of the full Jacobian matrix and of the full matrix $\mathbf{B}(\boldsymbol{\mu})$.

The following examples demonstrate the performance of the Gauß-Newton based algorithm compared with the SAGECPE1 or AP algorithm. The simulation results in Figure 5-1 and Figure 5-2 compare the convergence behaviour of the gradient-based ML search to the parameter wise search of the SAGECPE1 in a noise free, closely spaced coherent path scenario. In this case, the paths differ only in DoA and are separated by 5 deg. in angle of arrival which is closer than the Rayleigh resolution of the array. The path magnitudes are equal and the phase difference is zero in Figure 5-1 and 180 deg. in Figure 5-2. Although these constellations maybe considered as worst-case situations, they frequently occur in a practical propagation scenario since path length difference has to change only by 2.5 cm to move from one worst-case situation to the other. The antenna array was a 24-element circular patch array. Only matched vertical polarisation was considered. The two figures depict the iteration steps, which are plotted on the cost function surface. Note that both constellations cause completely different cost function surfaces, which are characterized by shaped, narrow valleys. The parameter wise search of the SAGECPE1 forces very small zigzag steps in the direction of the individual parameters, which can be seen most clearly in Figure 5-2. In both cases, final convergence requires more than 2000 iterations of the SAGECPE1 procedure whereas the gradient search needs only 26 and 13 steps, respectively, to reach the solution. Figure 5-2 (right) also indicates the initial SAGE steps before starting the final gradient steps. The example also shows that quantization of the data model would be detrimental since very small steps are required by the SAGECPE1 in order to achieve some progress. The example also shows that the data model quantization is not directly related to the desired parameter quantization. Actually much finer steps are required. Figure 5-3 further compares coordinate wise (alternating) and gradient based optimisation in terms of the number of iteration vs. the angular separation

of two coherent paths. It becomes clear that especially for paths which are closely spaced than Rayleigh resolution the number of the required iterations becomes prohibitively large.
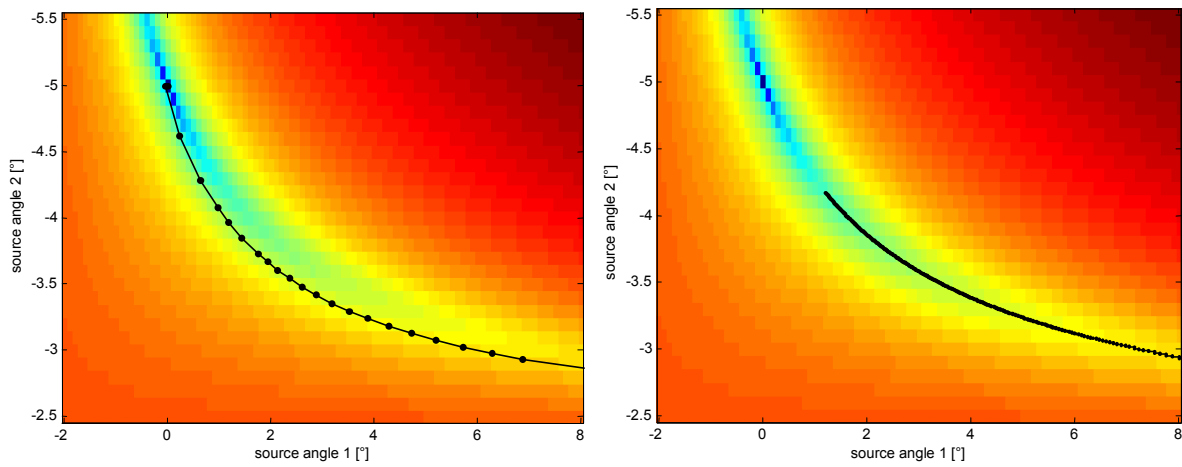
Figure 5-1:   Convergence behaviour of the SAGECPE1 algorithm (right) compared to the gradient based algorithm (left) in case of two strong coupled paths with an angular separation of 5° and a phase difference of 0°. The solution (minimum) is (0°,-5°).
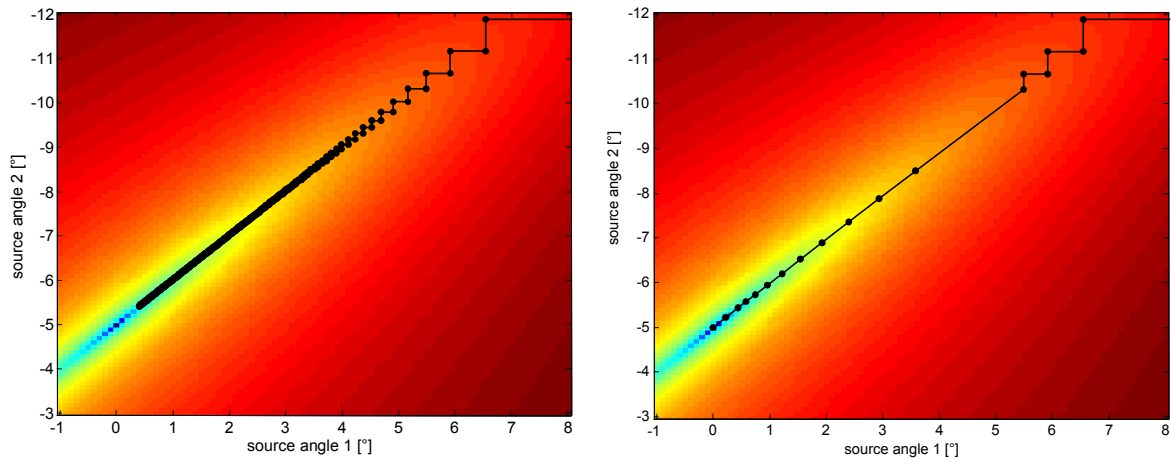
Figure 5-2:   Convergence behaviour of the SAGECPE1 algorithm (right) compared to the gradient based algorithm (left) in case of two strong coupled paths with 5° angular separation and 180° phase difference. The solution (minimum) is (0°,-5°).
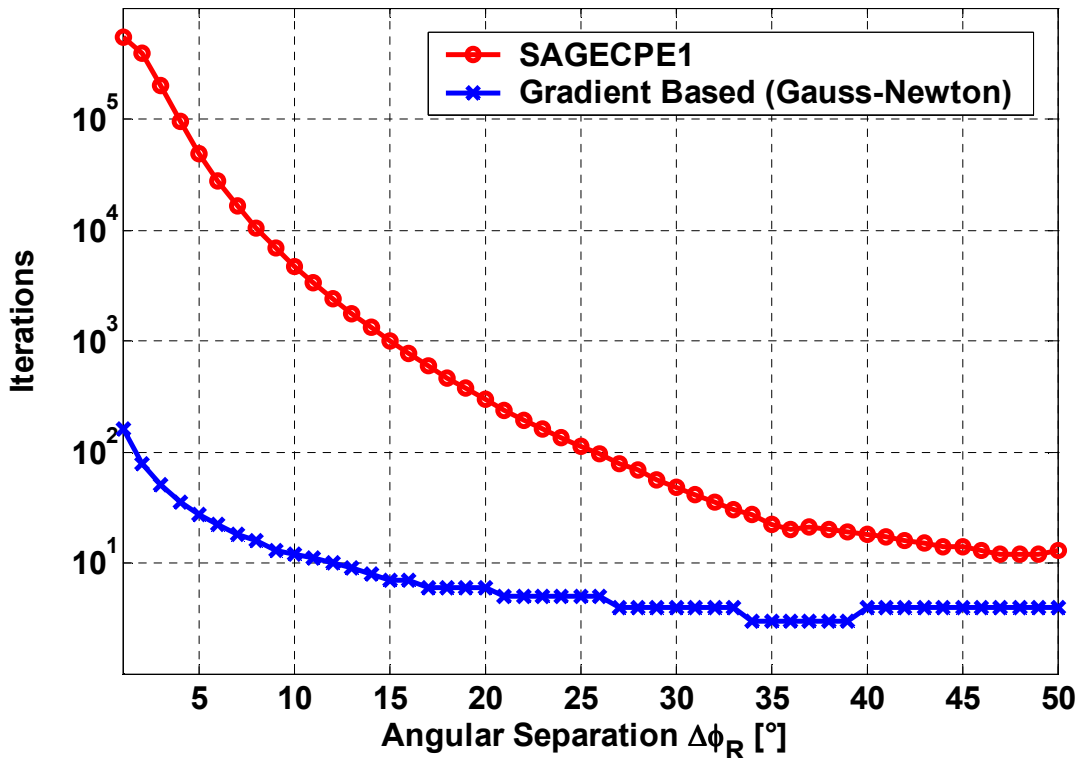
Figure 5-3: Convergence speed of SAGECPE1 and the Gauss-Newton based algorithm as a function of path separation.

5.2.5 Optimisation of Parameter Subsets

Now we recall the discussion about the Fisher information matrix and parameter coupling. We have discussed in Section 4.1.4 that some model parameters are strongly coupled and some are nearly uncoupled. This can be exploited to reduce the computational complexity of the gradient-based algorithms. We divide the complete parameter space into non-overlapping parameter subspaces, and optimize the subspaces individually. This approach can be understood as an application of the SAGE [62] method to the parameter estimation problem (5.1). However, in contrast to other channel parameter estimation implementations we choose the parameter subspaces based on parameter coupling. The issue of parameter coupling has already been discussed in Section 4.1.4. The influence of parameter coupling is also clearly visible in Figure 5-3. If the coupling between the parameters is weak, i.e., the parameters are well separated in the parameter domain, the convergence speed of the SAGECPE1 / AP algorithm is sufficient.

Suppose the second partial derivative of the log-likelihood function with respect to the two parameters θ_i, θ_k is zero. This implies that a change of the parameter estimate $\hat{\theta}_i$ has no direct influence on the parameter $\hat{\theta}_k$ and vice versa. Therefore, we can at first minimize (5.49) with respect to θ_i and than with respect to θ_k. Now consider the special case where all parameters are uncoupled, i.e.,

$$\frac{\partial^2}{\partial \theta_i \partial \theta_k} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = 0 \quad \forall i, \forall k, i \neq k .$$

Consequently, the Hessian is a diagonal matrix and we can use the steepest descent method or the AP algorithm to maximize the log-likelihood function without sacrificing the convergence

speed of the parameter estimator. The parameter update step of the steepest descent algorithm respecting the metric of the individual parameters is

$$\hat{\boldsymbol{\theta}}_m^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \left(\mathbf{I} \circ \boldsymbol{\mathcal{J}}\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)\right)^{-1} \mathbf{q}\left(\mathbf{x} \middle| \hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right) .$$

It is important to observe that the inversion of the diagonal matrix $\mathbf{I} \circ \boldsymbol{\mathcal{J}}\left(\hat{\boldsymbol{\theta}}^{\{i\}}, \mathbf{R}_{nn}\right)$ requires $O(L)$ operations only. Now recall that only some model parameters are typically strongly coupled as already discussed in Section 4.1.4. Consequently, the complete parameter space $\boldsymbol{\Theta}$ can be divided into $1 < K \leq L$ non-overlapping subspaces $\boldsymbol{\Theta}_k$. In other words, we can break the parameter vector $\boldsymbol{\theta}$ into smaller parameter vectors $\boldsymbol{\theta}_k$, and assume that

$$\frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_k^{\mathrm{T}}} \mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \mathbf{0} \quad \forall i, \forall k, i \neq k . \tag{5.62}$$

If we reorder the parameters in the parameter vector $\boldsymbol{\theta}$, keeping the parameter-groups $\boldsymbol{\theta}_k$ together the approximation of the Hessian becomes, due to (5.62), a block-diagonal matrix. As a first result of this approximation, the computational complexity of the Levenberg-Marquardt algorithm is reduced. However, this approximation can be further exploited if we generalize the AP algorithm. In the alternating projection algorithm, all parameters are updated sequentially. Since we do not want to sacrifice the overall convergence speed, we apply this strategy to the parameter groups, i.e., we update the parameter groups sequentially. However, within a parameter group we apply the Levenberg-Marquardt method.

The computational complexity can be reduced further if we assign all parameters of a propagation path to one parameter group. That means we form groups of coupled propagation paths instead of coupled parameters. Using this approach, we are able to optimize the parameters of the propagation path groups individually. As a side effect, the computational effort is often reduced further since the parameters of some path groups converge faster than others. Among other things, the complete algorithm can be understood as an application of the SAGE framework, described by Hero and Fessler [62], to the maximum likelihood estimation problem (5.2), where the parameter subspaces $\boldsymbol{\Theta}_k$ are chosen such that the overall convergence speed is not impaired.

5.2.6   Estimation of the Covariance Matrix of the Parameter Estimates

A by-product of the Gauß-Newton based algorithms is an estimate of the Fisher information matrix $\boldsymbol{\mathcal{J}}\left(\hat{\boldsymbol{\theta}}, \mathbf{R}_{nn}\right)$. Hence, the estimator does not only provide an estimate of the path parameters $\boldsymbol{\theta}_{sp}$, but also an estimate of the covariance matrix of the estimation error. It is well known from estimation theory [40] that the maximum likelihood parameter estimates are asymptotically Normal distributed, if the data model is continuous differentiable in $\boldsymbol{\theta}$ with finite differentials. Since the data model is expressed as a linear projection of complex exponentials, this condition is fulfilled. That means we can interpret the parameter estimate $\hat{\boldsymbol{\theta}}$ as a vector draw from a multivariate normal distribution, i.e.,

$$\hat{\boldsymbol{\theta}} \overset{as.d.}{\sim} \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{\mathcal{J}}^{-1}(\boldsymbol{\theta}, \mathbf{R}_{nn})\right).$$

Here as.d. stands for asymptotic in distribution. Since we have no access to the real parameter vector $\boldsymbol{\theta}$ we replace it by its ML-estimate (maximum likelihood) $\hat{\boldsymbol{\theta}}$ leading to

$$\hat{\boldsymbol{\theta}} \overset{as.d.}{\sim} \mathcal{N}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{J}}^{-1}\left(\hat{\boldsymbol{\theta}}, \mathbf{R}_{nn}\right)\right). \tag{5.63}$$

Since the amount of independent samples is significantly larger than the number of model parameters in channel sounding measurements, this approximation of the distribution of the

estimates fits very well to the real distribution due to the central limit theorem. Further discussions about and examples for the distribution of channel parameter estimates can be found in [90] and [36]. It is important to note that the approximation (5.63) is only valid if the number of measurements is sufficiently large in respect to the complexity (number of parameters) of the DML problem. A discussion about the efficiency of DML estimators for large M can be found in [49].

The knowledge of the approximate distribution (5.63) of the parameter estimates increases their value as a radio channel measurement result significantly. Furthermore, it is the prerequisite for further processing of the parameters. In principle the estimated channel parameters can be treated again as observations having the form

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \mathbf{n}_{\boldsymbol{\theta}} ,$$

where the stochastic process $\mathbf{n}_{\boldsymbol{\theta}}$ is distributed according to $\mathcal{N}\left(\mathbf{0}, \mathcal{J}^{-1}\left(\hat{\boldsymbol{\theta}}, \mathbf{R}_{nn}\right)\right)$. This is of crucial importance if someone wants to use the parameter estimation results to develop a radio channel model or derive the related probability density functions for the channel parameters. If the error distribution of the parameter estimates is ignored the derived probability density functions may partially represent the error of the parameter estimator and not only of the channel parameters $\boldsymbol{\theta}$. A discussion of this problem can be found in [90].

For the parameter estimation problem, the information about the variance of the parameter estimates is the basis for a new approach to model selection in high-resolution channel parameter estimation as shown in the next section.

5.2.7 Model Order Selection for Gauß-Newton based Algorithms

The estimate of the parameter variance given in Section 5.2.6 formula (5.63) together with the discussion about the relative variance of the path weights in Section 4.5 equation (4.100) gives rise to a new approach for model selection. In particular, since we have to choose the number of components contained in the model describing the concentrated propagation paths, the problem to solve is a model order selection problem. The term model order selection refers to the fact that we have to choose the number of propagation paths, but not to pick the type of the model for the individual paths. The fundamental idea is to use the condition (4.100) as a criterion for model order selection, i.e., we introduce the following constraint for the path weights

$$\frac{\mathrm{var}\{|\boldsymbol{\gamma}|\}}{|\boldsymbol{\gamma}|^{2}} < \varepsilon_{|\gamma|}^{2} < 1 . \tag{5.64}$$

That means we require that the certainty of the estimated path magnitude must be larger than its uncertainty. We require that the relative variance of the measured path magnitude is smaller than $\varepsilon_{|\gamma|}^{2}$. Observe, that the constraint (5.64) can be directly interpreted as a rule to decide, whether the estimated propagation path belongs to the circular normal distributed dense multipath components or not. If we choose $\varepsilon_{|\gamma|}^{2} = 0.3695$ there is probability of 10% that we select a propagation path, which belongs to the noise or DMC process. If we choose $\varepsilon_{|\gamma|}^{2} = 0.0924$ there is a probability of 1% that we choose a propagation path which belongs to the noise or DMC process. See also Section 4.5 for a discussion of the bound $\varepsilon_{|\gamma|}^{2}$. This approach is a means to avoid so-called path splitting. A typical sign of path splitting are extremely large path magnitude estimates of two propagation paths having almost the same structural parameters, i.e., similar angles, time delay of arrival, and Doppler shift. Typically the phase difference between the two estimated closely spaced paths is close to $\pm \pi$. That means the superposition of the two estimated paths approximate the observed propagation path very well, however the parameter estimates itself does not represent the physical reality.

This effect is not restricted to maximum likelihood estimation. For example, it can also be observed if the ESPRIT algorithm is applied to channel parameter estimation. This is the main reason why the often proposed strategy for model order selection in channel parameter estimation, to use a "large enough number of propagation paths such that all dominant impinging waves can be estimated" [63], [91], [92] does not work in practice.

Figure 5-5 shows an example of a measured channel impulse response and the related parameter estimates in the time delay domain. For every estimated path a dot is plotted representing the parameter pair $|\hat{\gamma}_p|, \hat{\tau}_p$. Furthermore, the estimated variance of the estimated path weight is also shown as a dot. The dots related to a single propagation path are connected by a line. Since the power-delay-profile PDP is shown in a logarithmic scale, the lines represent directly the signal to noise ratio of the estimated path magnitudes. In this example path splitting has occurred twice. There is a pair of closely spaced paths at 3050ns and another pair at 3370ns. The SNR of the first pair is negative, i.e., the estimation error of the path magnitudes is larger than the power of the paths. The estimated path weights are approximately 10dB larger than the magnitude of the observed propagation path, nevertheless it is important to note that the approximation of the path by the two propagation paths is reasonably good. This is an important observation. The goodness of fit, i.e., the approximation of the impulse response is by no means a measure for the goodness of the parameter estimates. In the same context we have to stress the fact that an overestimation of the model order will promise a good approximation of the channel observation, but may lead to incorrect parameter results. Nevertheless, the overestimation of the model order is a good approach to model order selection if it is combined with the condition (5.64). As already discussed the overestimation of the model order is leading to path splitting. Furthermore, it leads to the estimation of propagation paths with low power. In both cases the relative variance of the estimated path weights will be too large. Consequently, we can estimate the model order in the following way. First, we overestimate the number of concentrated propagation paths. In a subsequent step we remove the estimated paths having a small or negative SNR, starting with the worst estimate. If all estimated propagation paths meet the condition (5.64), we have an estimate of the model order and a reliable set of parameter estimates.
The advantage of the outlined approach to model order selection is that the model order is chosen such that the parameter estimates are reliable. The disadvantage is that we have to overestimate the model order and consequently increase the computational effort. However, this drawback is acceptable if we consider that all existing algorithms for model order selection, such as the Akaike information criterion [44] or the Minimum Description Length [65], [28], [44] requires the computation of the log-likelihood values of the candidate models as well and have therefore the same or even a higher computational complexity.

It is instructive to observe, that the SNR of an estimated propagation path, which is well separated from the others, is determined by the path magnitude and power of the dense multipath components only having the same time delay of arrival, provided of course that the measurement noise is small enough. This is an example for the inherent limitation of the reachable path parameter variance through the radio channel itself (cf. Section 4.1.5). Furthermore, the estimated propagation paths close to the direct path (line of sight) show a significant increase in the variance of the estimated path weights. This is a result of the parameter coupling of the closely spaced paths as discussed in Section 4.1.4. In simple terms a short distance between two propagation paths in the parameter space leads to noise enhancement.

Figure 5-7 shows an example for a reliable parameter set, i.e., all paths have a positive SNR.
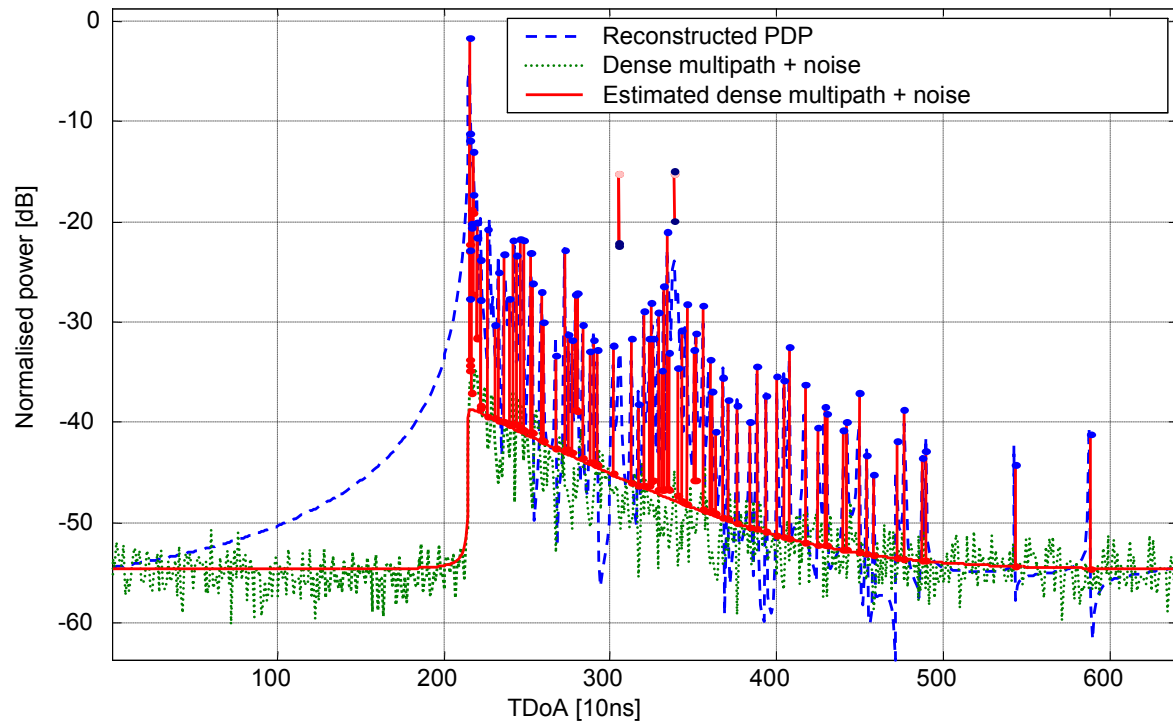
Figure 5-4: An example for a measured CIR and the related parameter estimates (four paths unreliable). The dashed line shows the PDP of the specular propagation paths reconstructed with the measurement bandwidth and the dotted line represents the PDP of the measured DMC and noise. The straight line shows the estimated PDP of the DMC and the measurement noise. The blue and red dots denote the magnitude of the estimated path weights and the estimated variance of the path weights, respectively. The length of the lines connecting the dots represents the SNR of the propagation path.
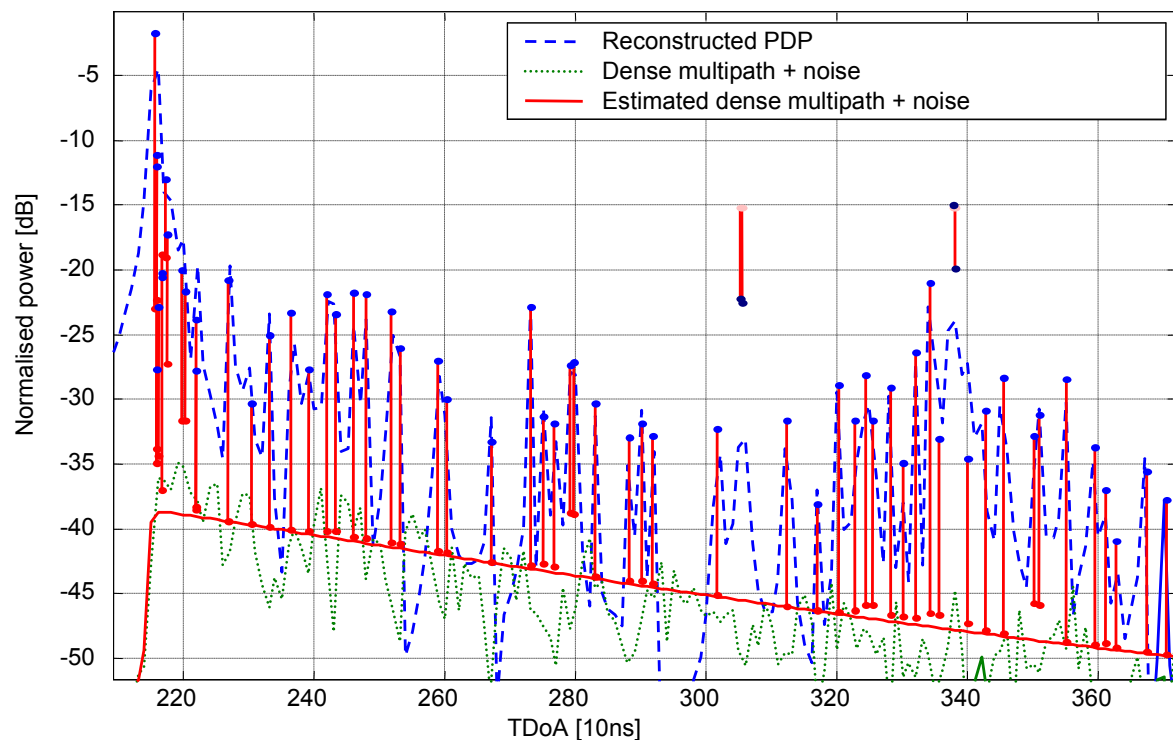
Figure 5-5: Magnified view of Figure 5-4, an example for a measured CIR and the related parameter estimates (four paths unreliable).
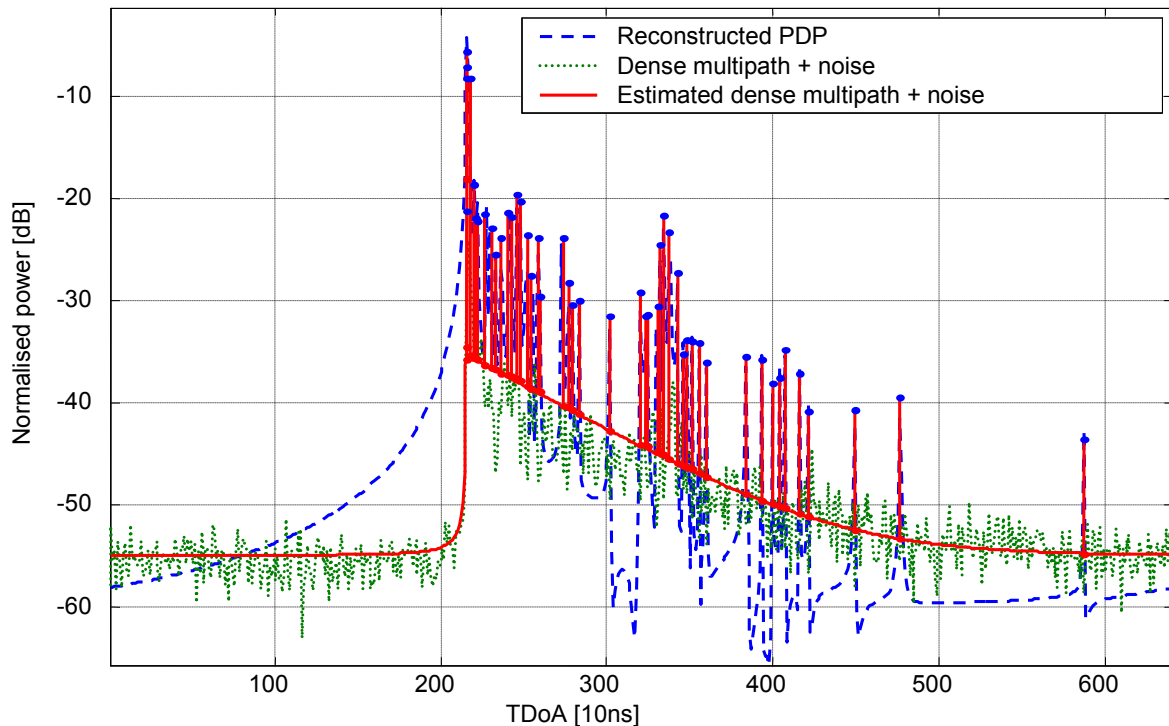
Figure 5-6:   An example for a measured CIR and the related parameter estimates (all parameters reliable). The dashed line shows the PDP of the specular propagation paths reconstructed with the measurement bandwidth and the dotted line represents the PDP of the measured DMC and noise. The straight line shows the estimated PDP of the DMC and the measurement noise. The blue and red dots denote the magnitude of the estimated path weights and the estimated variance of the path weights, respectively. The length of the lines connecting the dots represents the SNR of the propagation path.
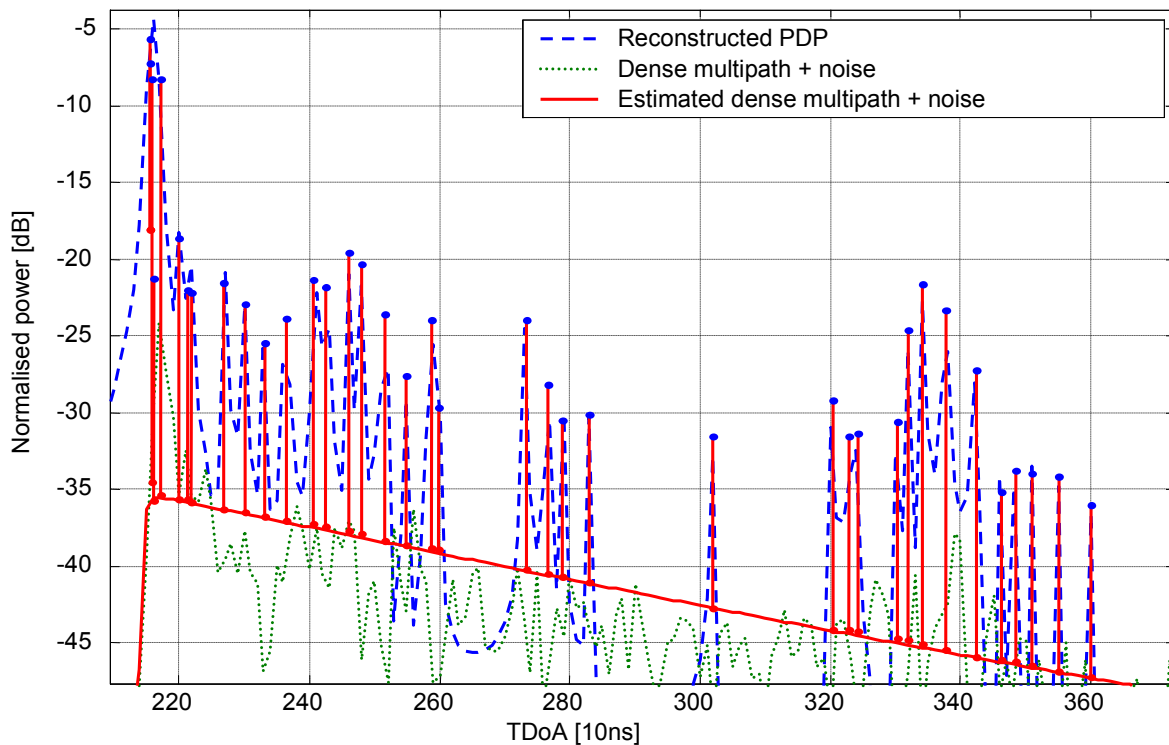
Figure 5-7:   Magnified view of Figure 5-6, an example for a measured CIR and the related parameter estimates (all parameters reliable).

Another way to select an appropriate model for a given observation are so called model selection criteria. A criterion often proposed for model order selection is the *minimum description length* (MDL) criterion [16], [17], [93]. The objective function of the MDL criterion is

$$d_{MDL}(P) = -\ln\left(p\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{(P)}\right)\right) + p(P) = -\mathcal{L}\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{(P)}\right) + \frac{p(P)}{2}, \qquad (5.65)$$

where $p(P)$ is a penalty term describing the complexity of the model P. The minimum of the criterion function $d(P)$ gives an estimate of the model order by

$$\hat{\mathbf{P}} = \arg\min(d(P)).$$

Another criterion also proposed for model selection is the Akaike Information Criterion (AIC) [44], [93]. The related criterion function is

$$d_{AIC}(P) = -\ln\left(p\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{(P)}\right)\right) + p(P) = -\mathcal{L}\left(\mathbf{x}\middle|\hat{\boldsymbol{\theta}}^{(P)}\right) + p(P). \qquad (5.66)$$

The main drawback of both information theoretic criteria is the expensive computation of the first term $\mathcal{L}\left(\mathbf{x}\middle|\boldsymbol{\theta}^{(P)}\right)$. We have to estimate the solutions $\hat{\boldsymbol{\theta}}^{(P)}$ for *all* possible models that are not feasible in general. An exemption are subspace based algorithms such as ESPRIT, MUSIC, and RARE. Here, the estimation of the model order is carried out during the subspace estimation step, namely when signal and noise subspace are separated. However, it should be noted that the model order determined from the singular values or eigenvalues as proposed in [28], [65] is not necessarily the best choice for the parameter estimation algorithm used in the subsequent steps to compute the actual parameter estimation results. The model order selection criteria (5.65) and (5.66) require that the parameter estimator employed is a maximum likelihood estimator for the model parameters. If the parameter estimator is statistically less efficient the assumptions made during the derivation of MDL and AIC are not fulfilled. Nevertheless, the solution described in [28], [65] provides at least a raw estimate of the model order. Therefore, the model order selection criteria outlined above are the methods of choice if the multidimensional Unitary ESPRIT is applied to channel parameter estimation.

Finally, it should be noted that the proposed algorithm for model order selection as well as the AIC and MDL criterion would inevitably fail, if the underlying candidate models do not fit to the observation. That means for the problem at hand, the number of the concentrated propagation paths can only be estimated if the contribution of the dense multipath components to the channel observation is taken into account. The advantage of the proposed criteria to determine the number of concentrated propagation paths is that we evaluate the chosen model order and get information on how the model must be changed to improve the result. The drawback is we have to overestimate the model order at the beginning. However, this is acceptable since the number of propagation paths changes slowly over time. To conclude the outlined strategy for model order selection is a feasible way to determine the number of dominant propagation paths in a radio channel observation.

5.2.8   Problem Conditioning

A crucial point of the gradient-based algorithms is the condition of the Hessian or its approximation. The approximation of the Hessian used in the Gauß-Newton (Section 5.2.3) and the Levenberg-Marquardt (Section 5.2.4) algorithm is given by the expressions for the Fisher information matrix derived in Section 4.2. Observe that the second order derivatives with respect to the nonlinear model parameters are functions of the path weights, whereas the second order derivatives with respect to the linear parameters are not. That means they can improve the condition of the approximated Hessian by scaling the observation and the path weights

appropriately. Let $\mathcal{J}(\boldsymbol{\mu}, \mathbf{R}_{nn})$ be the block matrix of $\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})$ describing the Fisher information of the nonlinear parameters only and $\mathcal{J}(\boldsymbol{\gamma}, \mathbf{R}_{nn})$ the block matrix describing the Fisher information of the linear parameters only. An appropriate scaling factor for the channel observation is given by the maximum and minimum values of the main diagonal elements of the two block matrices

$$\iota_{mima} = \sqrt[4]{\frac{\max(\mathrm{diag}\{\mathcal{J}(\boldsymbol{\gamma}, \mathbf{R}_{nn})\})}{\max(\mathrm{diag}\{\mathcal{J}(\boldsymbol{\mu}, \mathbf{R}_{nn})\})} \cdot \frac{\min(\mathrm{diag}\{\mathcal{J}(\boldsymbol{\gamma}, \mathbf{R}_{nn})\})}{\min(\mathrm{diag}\{\mathcal{J}(\boldsymbol{\mu}, \mathbf{R}_{nn})\})}} \ . \tag{5.67}$$

One should observe, that ι_{mima} is not the optimum solution to the minimization problem

$$\iota_{opt} = \arg \min_{\iota}(\mathrm{cond}(\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn}, \iota))) \ .$$

However, it provides a reasonable trade off between computational complexity and condition improvement. The determination of the optimum value ι_{opt} is computationally much more expensive, than the search for the maximum and minimum values of the main diagonal elements of the two block matrices $\mathcal{J}(\boldsymbol{\mu}, \mathbf{R}_{nn})$ and $\mathcal{J}(\boldsymbol{\gamma}, \mathbf{R}_{nn})$.

If the matrix $\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})$ is badly conditioned and the observation has been scaled by ι_{mima}, then the observed information is not sufficient to determine all parameters of the actual model. Recalling the discussion in Section 5.2.7 about the relative variance of the path weights and model order selection this problem can be solved by determining the path having a path weight with a relative variance close to or larger than 1.0. If the problem is badly conditioned, we have one or more very small eigenvalues in $\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})$. That means some parameters have a very large variance in a transformed parameter space. Consequently, one or more propagation paths have to be removed from the data model since they cannot be estimated reliably.

To summarise, badly conditioned problems can be avoided by appropriate scaling of the channel observations using (5.67) and by application of the proposed model order selection criteria outlined in Section 5.2.7.

5.2.9   Implementation Issues

For the iterative maximum likelihood estimation of the parameters $\boldsymbol{\theta}_{sp}$ using a Gauß-Newton based algorithm, three basic function blocks have to be implemented. Namely the gradient, the approximation of the Hessian, and the negative log-likelihood function, i.e., the Mahalanobis norm have to be computed. Furthermore, for the estimation of $\hat{\boldsymbol{\gamma}}$ given the structural parameters $\hat{\boldsymbol{\mu}}$ the computation of (5.5) (BLUE) is required. To study the general structure of the four function blocks we introduce abstract versions of the data model $\mathbf{s}(\boldsymbol{\theta}_{sp})$ and the Jacobian matrix (4.68), (4.73), and (4.77). The data model has the general structure

$$\mathbf{s}(\boldsymbol{\theta}_{sp}) = (\mathbf{B}_3 \Diamond \mathbf{B}_2 \Diamond \mathbf{B}_1) \boldsymbol{\gamma} \ ,$$

and the Jacobian matrix can be expressed as

$$\mathbf{D} = \mathbf{D}_3 \Diamond \mathbf{D}_2 \Diamond \mathbf{D}_1 \ .$$

The related component matrices have sizes $\mathbf{B}_i \in \mathbb{C}^{M_i \times P}$ and $\mathbf{D}_i \in \mathbb{C}^{M_i \times L}$, where L denotes the number of real valued parameters in $\boldsymbol{\theta}_{sp} \in \mathbb{R}^{L \times 1}$. Furthermore, we define the covariance matrix of the stochastic process $\mathbf{n}$ using the covariance matrices related to the three dimensions $\mathbf{R}_i \in \mathbb{C}^{M_i \times M_i}$ as

$$\mathbf{R}_{nn} = \mathbf{R}_3 \otimes \mathbf{R}_2 \otimes \mathbf{R}_1 \ .$$

The expressions for the BLUE, the Hessian, the gradient, and the Mahalanobis norm are summarised in expressions (5.68) to (5.71).

BLUE:

$$
\begin{aligned}
\hat{\boldsymbol{\gamma}} &= \left(\mathbf{B}^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{x} \\
&= \underbrace{\left(\mathbf{B}_3^{\mathrm{H}}\mathbf{R}_3^{-1}\mathbf{B}_3 \circ \mathbf{B}_2^{\mathrm{H}}\mathbf{R}_2^{-1}\mathbf{B}_2 \circ \mathbf{B}_1^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{B}_1\right)^{-1}}_{\text{Term Type I}} \underbrace{\left(\left(\mathbf{R}_3^{-1}\mathbf{B}_3\right)\lozenge\left(\mathbf{R}_2^{-1}\mathbf{B}_2\right)\lozenge\left(\mathbf{R}_1^{-1}\mathbf{B}_1\right)\right)\mathbf{x}}_{\text{Term Type II}},
\end{aligned}
\tag{5.68}
$$

Approximation of the Hessian (Fisher information matrix):

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{\theta},\mathbf{R}_{nn}) &= 2\cdot\Re\left\{\mathbf{D}^{\mathrm{H}}\mathbf{R}_{nn}^{-1}\mathbf{D}\right\} \\
&= 2\cdot\Re\left\{\underbrace{\mathbf{D}_3^{\mathrm{H}}\mathbf{R}_3^{-1}\mathbf{D}_3 \circ \mathbf{D}_2^{\mathrm{H}}\mathbf{R}_2^{-1}\mathbf{D}_2 \circ \mathbf{D}_1^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{D}_1}_{\text{Term Type I}}\right\},
\end{aligned}
\tag{5.69}
$$

Gradient:

$$
\begin{aligned}
\mathbf{q}(\mathbf{x}|\boldsymbol{\theta},\mathbf{R}_{nn}) &= 2\cdot\Re\left\{\mathbf{D}^{\mathrm{H}}(\boldsymbol{\theta})\mathbf{R}_{nn}^{-1}(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}))\right\} \\
&= 2\cdot\Re\left\{\underbrace{\left(\mathbf{R}_3^{-1}\mathbf{D}_3\lozenge\mathbf{R}_2^{-1}\mathbf{D}_2\lozenge\mathbf{R}_1^{-1}\mathbf{D}_1\right)^{\mathrm{H}}\mathbf{x}}_{\text{Term Type II}} - \underbrace{\left(\mathbf{D}_3^{\mathrm{H}}\mathbf{R}_3^{-1}\mathbf{B}_3 \circ \mathbf{D}_2^{\mathrm{H}}\mathbf{R}_2^{-1}\mathbf{B}_2 \circ \mathbf{D}_1^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{B}_1\right)\boldsymbol{\gamma}}_{\text{Term Type I}}\right\},
\end{aligned}
\tag{5.70}
$$

Mahalanobis norm (negative log-likelihood function):

$$
\begin{aligned}
\mathcal{L}_N'(\mathbf{x}|\boldsymbol{\mu}) &= (\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}))^{\mathrm{H}}\mathbf{R}_{nn}^{-1}(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta})) \\
&= \mathbf{x}^{\mathrm{H}}\mathbf{R}_{nn}^{-1}\mathbf{x} + \\
&\quad -2\Re\left\{\boldsymbol{\gamma}^{\mathrm{H}}\underbrace{\left(\mathbf{R}_3^{-1}\mathbf{B}_3\lozenge\mathbf{R}_2^{-1}\mathbf{B}_2\lozenge\mathbf{R}_1^{-1}\mathbf{B}_1\right)^{\mathrm{H}}\mathbf{x}}_{\text{Term Type II}}\right\} + \boldsymbol{\gamma}^{\mathrm{H}}\underbrace{\left(\mathbf{B}_3^{\mathrm{H}}\mathbf{R}_3^{-1}\mathbf{B}_3 \circ \mathbf{B}_2^{\mathrm{H}}\mathbf{R}_2^{-1}\mathbf{B}_2 \circ \mathbf{B}_1^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{B}_1\right)\boldsymbol{\gamma}}_{\text{Term Type I}}.
\end{aligned}
\tag{5.71}
$$

Observe, that all expressions above requires the computation of a term of Type I having structure

$$
\mathbf{B}_3^{\mathrm{H}}\mathbf{R}_3^{-1}\mathbf{B}_3 \circ \mathbf{B}_2^{\mathrm{H}}\mathbf{R}_2^{-1}\mathbf{B}_2 \circ \mathbf{B}_1^{\mathrm{H}}\mathbf{R}_1^{-1}\mathbf{B}_1 .
$$

The computational complexity of an algorithm implementing this basic block is, depending on the structure of the covariance matrices $\mathbf{R}_i$ (see also Section 6.1.12),

$$
O\!\left(P^2\sum_{i=1}^{3}M_i\right) \le O(\text{Term Type I}) \le O\!\left(P^2\sum_{i=1}^{3}M_i + P\sum_{i=1}^{3}M_i^2\right) \quad \text{for (5.68) and (5.71)},
$$

$$
O\!\left(P^2\sum_{i=1}^{3}M_i\right) \le O(\text{Term Type I}) \le O\!\left(PL\sum_{i=1}^{3}M_i + P\sum_{i=1}^{3}M_i^2\right) \quad \text{for (5.70)},
$$

and

$$O\left(P^2\sum_{i=1}^{3}M_i\right) \le O(\text{Term Type I}) \le O\left(L^2\sum_{i=1}^{3}M_i + L\sum_{i=1}^{3}M_i^2\right) \text{ for (5.69)}.$$

Hence, the computational complexity is roughly proportional to the square of the propagation paths and the sum of the samples taken in the individual data domains R_D $\sum_{i=1}^{R_D}M_i$. In contrast, it is easy to see that the computation of a term of Type II requires an algorithm having computational complexity $P\prod_{i=1}^{R_D}M_i$. That means the computation of terms of Type II is typically by far more expensive than the computation of terms of Type I.

As a first result, we can conclude that the computation of the Hessian is, in contrast to other mid- or large-scale problems, feasible since it contains only a term of Type I. This is especially true if we exploit the redundancies in the Jacobian matrices (4.68), (4.73), or (4.77). Furthermore, the computational complexity of one iteration of any gradient based algorithm for $\hat{\boldsymbol{\theta}}_{sp}$ is proportional to the number of propagation paths and to the number of samples in one radio channel snapshot $\mathbf{x}\in\mathbb{C}^{M\times1}$. This estimate of the computational complexity holds as long as $M \gg L^3$, or the parameter subset update strategy outlined in Section 5.2.5 is employed. Otherwise, the computation of the update step $\Delta\boldsymbol{\theta}^{\{i\}}$ (cf. expressions (5.57) and (5.61)) will determine the computational complexity of the parameter update step since the solution to a system of L equations has to be found.

Since the computation of terms having structure of Type II determine the computation time of the whole algorithm, we discuss in the following a way to implement the product

$$\mathbf{y} = \left(\mathbf{B}_{R_D}\Diamond\ldots\Diamond\mathbf{B}_{R_1}\right)^{\mathrm{H}}\mathbf{x}. \tag{5.72}$$

A straightforward way to compute $\mathbf{y}$ is to form $\mathbf{B} = \mathbf{B}_{R_D}\Diamond\ldots\Diamond\mathbf{B}_{R_1}$ and to compute $\mathbf{y} = \mathbf{B}^{\mathrm{H}}\mathbf{x}$. The drawback of this approach is that we need a large amount of memory to store the full matrix $\mathbf{B}$. An algorithm, which avoids the explicit computation of $\mathbf{B}$ is outlined in Table 5-4. In addition to the reduced memory requirements, the outlined algorithm is significantly faster than the direct approach since the first processing step is formulated as a matrix-matrix multiplication. In today's layered memory architectures the time needed to compute the product of two matrices is mainly determined by the number of floating point operations the FPU is able to carry out in a given time, especially if the matrices involved are large. In contrast, the computation of a matrix–vector product is mainly determined by the memory-bandwidth of the system if the involved matrix is large. The algorithm outlined in Table 5-4 is still suboptimal in terms of the required data transfers between FPU and memory. An optimal implementation of (5.72) is described in [51]. Furthermore, the aforementioned redundancies of the Jacobian matrix can easily be exploited if the outlined algorithm is used. In addition, the computation of the full Jacobian matrix $\mathbf{D}$ is not feasible for example if parameters from broadband MIMO-channel measurements have to be estimated.

Table 5-4: Memory efficient Computation of the Product $\mathbf{y} = \left(\mathbf{B}_{R_D} \Diamond \ldots \Diamond \mathbf{B}_{R_1} \right)^{\mathrm{H}} \mathbf{x}$

---

1) Compute the matrix $\mathbf{Z}_1^{\mathrm{T}} = \mathbf{B}_1^{\mathrm{H}} \, \mathrm{mat}\{\mathbf{x}\}$

2) for $k = 1, \ldots, P$ do

3.1) $\quad \mathbf{z}_1 = \{\mathbf{Z}_1\}_k$

3.2) for $r = 2, \ldots, R_D$

$\quad\quad$ Compute the vector matrix product $\mathbf{z}_r^{\mathrm{T}} = \{\mathbf{B}_r\}_k^{\mathrm{T}} \, \mathrm{mat}\{\mathbf{z}_{r-1}\}$.

$\quad$ end

3.3) $\quad y_k = \mathbf{z}_{R_D}$

To summarize, a parameter update step of the algorithm outlined in Table 5-3 has approximately linear computational complexity in the number of concentrated propagation paths P, and in the number of data samples M. The iterative maximum likelihood estimator can be implemented without forming the full matrices $\mathbf{D}$, and $\mathbf{B}$ explicitly.

5.3 Subspace Based Algorithms

Another class of estimators, which can be applied to channel parameter estimation, is the class of so-called subspace based algorithms. The most prominent representative of this class of parameter estimators is the ESPRIT algorithm. It has been applied to radio channel parameter estimation for 10 years [94], [58], [59], [15], [1], [61], [95], [28], [96], [97], [98], and [99]. However, also the RARE and MUSIC algorithm have been applied to high-resolution channel parameter estimation from radio channel sounding measurements [100], [101], [102] and [103]. The general model [44] used in all subspace-based algorithms models M_o observations $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_{M_o} \end{bmatrix} \in \mathbb{C}^{M \times M_o}$ by

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_{M_o} \end{bmatrix} = \mathbf{B}(\boldsymbol{\mu})\mathbf{S} + \mathbf{N}. \tag{5.73}$$

Here, the matrix $\mathbf{S} \in \mathbb{C}^{P \times M_o}$ represents samples of signals received from P signal sources and the matrix $\mathbf{N} \in \mathbb{C}^{M \times M_o}$ (the measurement noise). In general it is assumed, that the signal covariance matrix

$$\mathbf{R}_{\mathbf{SS}} = \mathrm{E}\{\tfrac{1}{M_o}\mathbf{S}\mathbf{S}^{\mathrm{H}}\}$$

has full rank $\mathrm{rank}(\mathbf{R}_{\mathbf{SS}}) = P$ and that the number of observations is large $M_o \gg 1$. The second assumption guarantees that the estimate of the data correlation matrix

$$\hat{\mathbf{R}}_{\mathbf{XX}} = \tfrac{1}{M_o}\mathbf{X}\mathbf{X}^{\mathrm{H}} \tag{5.74}$$

has at least rank P. Furthermore, the algorithms are constructed on the assumption that the correlation matrix has the structure

$$\mathbf{R}_{\mathbf{XX}} = \mathbf{B}(\boldsymbol{\mu})\mathbf{R}_{\mathbf{SS}}\mathbf{B}^{\mathrm{H}}(\boldsymbol{\mu}) + \mathbf{R}_{\mathbf{NN}},$$

where $\mathbf{R}_{\mathbf{NN}} = \mathrm{E}\{\tfrac{1}{M_q}\mathbf{N}\mathbf{N}^{\mathrm{H}}\}$ is the noise covariance matrix. In the context of high-resolution parameter estimation it is crucially important to realise that only one observation of the radio channel is available as discussed in Chapter 2. Consequently, the data correlation matrix can-

not be estimated using (5.74). A solution to this problem is the so-called subarray-smoothing technique [104], [105].

If an observed radio scenario contains highly correlated or even coherent sources, the signal correlation matrix $\mathbf{R_{SS}}$ will have a rank, which is smaller than the number of sources P. Consequently, a signal subspace having rank P could not be estimated from the estimated data correlation matrix (5.74) as shown for example in [104].

The subarray smoothing techniques has been developed for uniform linear arrays to overcome this problem. The idea is to decompose the uniform linear array into smaller over-lapping subarrays. Using this approach the number of independent measurements can be increased, where the array aperture is decreased to the size of the subarrays. Using these approach subspace-based algorithms can be used for parameter estimation even in scenarios with coherent sources. The subarray smoothing can also be applied to increase the number of independent observations in radio channel parameter estimation.

In radio channel parameter estimation, the generalised form of subarray smoothing the multidimensional smoothing technique [65], [44] has to be applied. It is important to keep in mind that the availability of only one observation is the reason why the application of multi-dimensional smoothing is necessary. Although, subarray smoothing is often discussed in combination with ESPRIT it is a signal processing technique on its own. To be precise, it is a signal processing technique for signal-subspace estimation and does not imply the subsequent application of the ESPRIT [106], [44], [107], the Unitary ESPRIT [65], [44], or the multidimensional Unitary ESPRIT [65], [44]. For example, the multidimensional smoothing technique is also a prerequisite for the application of the RARE algorithm [66], [67], [65] to radio channel parameter estimation.

An implicit result of the discussion about multidimensional smoothing for signal subspace estimation for subspace based channel parameter estimation is that the signal model must correspond directly or by a linear transformation to the model

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}, \tag{5.75}$$

where the signal modes $\mathbf{B}(\boldsymbol{\mu})$ must have the structure

$$\mathbf{B}(\boldsymbol{\mu}) = \mathbf{A}\left(\boldsymbol{\mu}^{(R_P)}\right) \Diamond \ldots \Diamond \mathbf{A}\left(\boldsymbol{\mu}^{(1)}\right). \tag{5.76}$$

If the measured channel sounding data can be transformed into the data model (5.76) by a linear projection, we say the observed data exhibits a hidden rotational invariance structure. The processing of data having hidden rotational invariance structure is discussed in section 5.3.11. Observe that equation (5.76) implies that the number of data-dimensions is equal to the number of parameter dimensions $R_P = R_D$. See also Section 2.4.5 for a discussion of data- and parameter dimensions.

To summarise, the availability of only a single observation (single snapshot case) is the reason why multidimensional smoothing must be applied in channel parameter estimation from channel sounding measurements. Consequently, the applicability of subspace based high-resolution parameter estimators is restricted to a subset of the available antenna array structure. The observed channel sounding data have rotational invariance structure if uniform linear arrays or uniform rectangular arrays are used throughout the measurements. Furthermore, the measured data exhibits a hidden rotational invariance structure if a circular uniform beam array (CUBA) is used at the Tx- and/or the Rx-site to observe the radio channel. The CUBA-ESPRIT algorithm was first published first in [108]. The algorithm is outlined in Section 5.3.12. Finally, it should be noted that uniform circular arrays have approximately hidden rotational invariance structure. Consequently, subarray smoothing is also applicable if a UCA

has been used to measure the radio channel. However, the parameter estimates determined by a signal subspace based parameter estimator will be biased since the UCA data model only approximately exhibits the rotational invariance structure. Additionally, the elevation angles of the propagation paths at the UCA must be the same for all propagation paths and known a priori, otherwise, the projector for the transformation of the UCA output data to data having approximately rotational invariance structure cannot be constructed. Altogether, if a UCA is used to measure the radio channel it is preferable to use the maximum likelihood estimator outlined in the previous sections.

In the next sections, the signal subspace estimation from channel sounding measurements is discussed. We will mainly focus on memory efficient techniques for signal subspace estimation. Especially the naïve implementation of multidimensional smoothing and of the subsequent signal subspace estimation step using the EVD (eigenvalue decomposition) or SVD (singular value decomposition) will lead to very large matrices. This is all the more true if we want to estimate channel parameters from MIMO channel sounding measurements.

5.3.1 Signal Subspace Estimation for Polarimetric Measurements

Subspace based algorithms can also be used for DoA or DoD estimation if polarimetric uniform linear or rectangular arrays (PULA, PURA) has been used while measuring the radio channel. Each antenna element of a PULA/PURA has two output ports, one port for horizontal $(\mathbf{e}_\varphi)$ and one for vertical $(\mathbf{e}_\vartheta)$ polarisation. We can interpret the output signal of such an antenna array as the output signals of two independent uniform linear or rectangular arrays, respectively. In both cases the observed data show the same rotational invariance structure. An efficient direction and polarisation estimation algorithm for a special kind of antenna arrays, so called COLD arrays (co-centred orthogonal loop and dipole), has been proposed in [109]. However, this algorithm is not applicable to channel parameter estimation since different antenna arrays are used. Therefore, we outline in the following an alternative approach to signal subspace based direction of departure and/or direction of arrival estimation for polarimetric antenna arrays.

Let us introduce the common complex gain of the antenna array elements (cf. Section 2.4.2) for horizontal and vertical polarisation at the Tx-site

$$b_{com,TH}\left(\mathbf{\Omega}_p^{(T)}\right), \ b_{com,TV}\left(\mathbf{\Omega}_p^{(T)}\right) \tag{5.77}$$

and at the Rx-site

$$b_{com,RH}\left(\mathbf{\Omega}_p^{(R)}\right), \ b_{com,RV}\left(\mathbf{\Omega}_p^{(R)}\right), \tag{5.78}$$

where $\mathbf{\Omega}_p^{(T)} = \begin{bmatrix} \varphi_{T,p} & \vartheta_{T,p} \end{bmatrix}^\mathrm{T}$ represents the DoD and $\mathbf{\Omega}_p^{(R)} = \begin{bmatrix} \varphi_{R,p} & \vartheta_{R,p} \end{bmatrix}^\mathrm{T}$ the DoA of path p. Now we combine the common complex gain of the antenna array elements with the respective path weights $\gamma_{HH}, \gamma_{HV}, \gamma_{VH}, \gamma_{VV}$ yielding

$$\boldsymbol{\gamma}'_{HH} = \left[b_{com,TH}\left(\Omega_1^{(T)}\right) b_{com,RH}\left(\Omega_1^{(R)}\right) \gamma_{HH,1} \ \cdots \ b_{com,TH}\left(\Omega_P^{(T)}\right) b_{com,RH}\left(\Omega_P^{(R)}\right) \gamma_{HH,P} \right]^\mathrm{T},$$

$$\boldsymbol{\gamma}'_{HV} = \left[b_{com,TH}\left(\Omega_1^{(T)}\right) b_{com,RV}\left(\Omega_1^{(R)}\right) \gamma_{HV,1} \ \cdots \ b_{com,TH}\left(\Omega_P^{(T)}\right) b_{com,RV}\left(\Omega_P^{(R)}\right) \gamma_{HV,P} \right]^\mathrm{T},$$

$$\boldsymbol{\gamma}'_{VH} = \left[b_{com,TV}\left(\Omega_1^{(T)}\right) b_{com,RH}\left(\Omega_1^{(R)}\right) \gamma_{VH,1} \ \cdots \ b_{com,TV}\left(\Omega_P^{(T)}\right) b_{com,RH}\left(\Omega_P^{(R)}\right) \gamma_{VH,P} \right]^\mathrm{T},$$

and,

$$\boldsymbol{\gamma}'_{VV} = \left[b_{com,TV}\left(\Omega_1^{(T)}\right) b_{com,RV}\left(\Omega_1^{(R)}\right) \gamma_{VV,1} \ \cdots \ b_{com,TV}\left(\Omega_P^{(T)}\right) b_{com,RV}\left(\Omega_P^{(R)}\right) \gamma_{VV,P} \right]^\mathrm{T}.$$

If polarimetric uniform linear or rectangular arrays have been applied at both link ends, we get four measurements of the radio channel sharing the same rotational invariance structure $\mathbf{A}(\boldsymbol{\mu}) = \mathbf{A}(\boldsymbol{\mu}^{(R)}) \Diamond \ldots \Diamond \mathbf{A}(\boldsymbol{\mu}^{(1)})$. We combine the measurements in one matrix $\mathbf{X}_{pol}$ having structure

$$\mathbf{X}_{pol} = [\mathbf{x}_{HH} \ \mathbf{x}_{HV} \ \mathbf{x}_{VH} \ \mathbf{x}_{VV}] = [\mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}'_{HH} \ \mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}'_{HV} \ \mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}'_{VH} \ \mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}'_{VV}] + \mathbf{N}$$
$$\mathbf{X}_{pol} = \mathbf{A}(\boldsymbol{\mu}) \cdot [\boldsymbol{\gamma}'_{HH} \ \boldsymbol{\gamma}'_{HV} \ \boldsymbol{\gamma}'_{VH} \ \boldsymbol{\gamma}'_{VV}] + \mathbf{N}. \tag{5.79}$$

The structure of the data matrix $\mathbf{X}_{pol}$ allows multidimensional smoothing as a pre-processing step. All subspace-based algorithms developed for non-polarimetric antenna arrays are as well applicable for parameter estimation from measurements carried out with polarimetric antenna arrays. Furthermore, one should note that the data model (5.79) is general insofar as it also covers measurements with a polarimetric antenna array at one link end or with non-polarimetric antenna arrays at both link ends.

Equation (5.79) shows that polarimetric measurements are also a means to increase the resolvability of concentrated propagation paths, since the data matrix $\mathbf{X}_{pol}$ has already a maximum rank of 4 without forward-backward and subarray smoothing.

5.3.2   On the Choice of Subarray Sizes for Multidimensional Smoothing

An issue not discussed so far is how to choose the subarray sizes for the subarray smoothing technique. For the one-dimensional case the optimum subarray size has been found in [104] to be[10]

$$M_{sub_r} = \left\lceil \tfrac{2}{3} M_r \right\rceil + 1. \tag{5.80}$$

This subarray size is optimal insofar that it allows the estimation of the largest number of sources. To understand this result, consider the following example. Suppose the channel is observed by a 9-element uniform linear array. The data model of the channel observation is

$$\mathbf{x} = (\mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma} + \mathbf{n}) \in \mathbb{C}^{9 \times 1}. \tag{5.81}$$

The optimum subarray size due to (5.80) is $M_{sub} = 7$. Using a subarray size of $M_{sub} = 7$ three overlapping subarrays can be selected out of the 9-element uniform linear array. Using forward-backward smoothing [104], [105] a total number of six independent measurements can be generated from the observation (5.81). The smoothed data matrix is

$$\mathbf{X}_{SS} = \begin{bmatrix} x_1 & x_2 & x_3 & x_7^* & x_8^* & x_9^* \\ x_2 & x_3 & x_4 & x_6^* & x_7^* & x_8^* \\ x_3 & x_4 & x_5 & x_5^* & x_6^* & x_7^* \\ x_4 & x_5 & x_6 & x_4^* & x_5^* & x_6^* \\ x_5 & x_6 & x_7 & x_3^* & x_4^* & x_5^* \\ x_6 & x_7 & x_8 & x_2^* & x_3^* & x_4^* \\ x_7 & x_8 & x_9 & x_1^* & x_2^* & x_3^* \end{bmatrix} \in \mathbb{C}^{7 \times 6}. \tag{5.82}$$

It has been shown in [104] that this matrix has full column rank except for special angle and path weight constellations. Consequently, the maximum rank of the signal subspace estimated from this data matrix is

---

[10] Here $\lceil \bullet \rceil$ means rounding to the next integer value greater or equal to $\bullet$.

$$\max_{\boldsymbol{\theta}}(\text{rank}(\mathbf{E}_S)) = \max_{\boldsymbol{\theta}}(\text{rank}(\mathbf{X}_{SS})) = 6.$$

This is sufficient to estimate the angles of six propagation paths (sources) from the observed data using for example ESPRIT. For the six propagation paths six complex path weights have to be estimated. Therefore, the number of real-valued unknowns in the six-path model is 18 and the number of independent real valued measurements is also 18. That means the maximum number of propagation paths (sources) that can be resolved with a 9-element ULA is $P = 6$. However, one should note that the subarray size given in (5.80) is not necessarily optimal in terms of statistical performance. If the observed scenario contains only a single propagation path, the reduction of the array size from a 9-element ULA to a 7-element ULA will lead to an increase of the estimator variance. The authors of [39] researched the optimal subarray size for the case of two equiv.-powered closely spaced coherent sources (propagation paths) and found a solution, which is only slightly different from (5.80), namely

$$M_{sub_r,opt_2} = 0.6(M_r + 1). \tag{5.83}$$

For a small number of samples M_r the difference between (5.80) and (5.83) is negligible. For large M_r the ratio between the two solutions tends to

$$\frac{M_{sub_r}}{M_{sub_r,opt_2}} \rightarrow \frac{10}{9}.$$

In general, the solution (5.80) is preferable for channel sounding application since it leads to a signal subspace estimate having the highest rank possible. Consequently, it allows the estimation of the largest number of propagation paths.

In Figure 5-8, the multidimensional smoothing technique is visualized for $R_P = R_D = 3$ parameter-dimensions. One can understand the measured data as a three-dimensional cuboid. If we apply multidimensional smoothing to this cuboid, we decompose it into a large number of overlapping smaller cuboids. It is straightforward to interpret the general case of multidimensional smoothing in a similar way. In cases with a higher dimensionality, we interpret the measured data as hyper-cuboids.
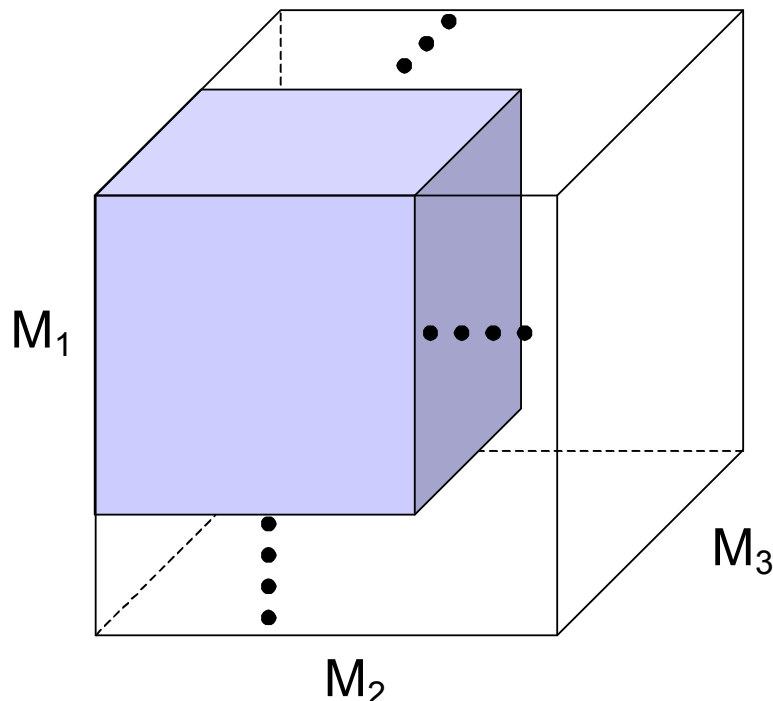
Figure 5-8:   Visualization of the multidimensional smoothing technique for $R_D = 3$

It is important to realise that the smoothed data matrix has a row size of

$$M_{row} = \prod_{r=1}^{R_D} \left\lceil \tfrac{2}{3} M_r \right\rceil + 1$$

and a column size of

$$M_{col} = \prod_{r=1}^{R_D} \left(M_r - \left\lceil \tfrac{2}{3} M_r \right\rceil + 1 \right).$$

The complete data matrix $\mathbf{X}_{SS} \in \mathbb{C}^{M_{row} \times M_{col}}$ may become very large if the number of data dimensions, e.g., parameter dimensions is large since the row- and column size is a product of the individual sizes of the data dimensions. This is the reason why the explicit implementation of the multidimensional smoothing step is not feasible in practice. In this context, it is important to observe that the multidimensional smoothing operation can also be understood as an indexing operation (data addressing) on the data vector $\mathbf{x}$. In Section 5.3.4 it is shown how this indexing approach can be used to carry out multidimensional smoothing implicitly, i.e., without forming the data matrix $\mathbf{X}_{SS}$, during the signal subspace estimation step.

5.3.3 Signal Subspace Estimation for Conjugate Centro-Symmetric Data

In [65] it is shown that, using an appropriate transformation, the ESPRIT algorithm can be implemented using real-valued arithmetic only. The resulting algorithm is the Unitary ES-PRIT. However, a prerequisite for the application of Unitary ESPRIT is that the matrices $\mathbf{A}(\boldsymbol{\mu}^{(r)})$ are conjugate centro-symmetric matrices, i.e.,

$$\mathbf{A}(\boldsymbol{\mu}^{(r)}) = \mathbf{\Pi} \mathbf{A}^*(\boldsymbol{\mu}^{(r)}) \boldsymbol{\Delta} \tag{5.84}$$

must hold, where $\boldsymbol{\Delta}$ is an arbitrary diagonal matrix. Since the same structure is also required for multidimensional smoothing as discussed in the previous sections this requirement is already fulfilled if smoothing is applicable. Consequently, the Unitary ESPRIT algorithm can be applied to estimate the structural parameters $\boldsymbol{\mu}$ from channel sounding measurements if the channel observation can be expressed using (5.76). The transformation of the complex-valued data into real-valued data can be carried out by the sparse unitary matrices

$$\mathbf{Q}_{M_r}^{(s)} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_q & +j\mathbf{I}_q \\ \mathbf{\Pi}_q & -j\mathbf{\Pi}_q \end{bmatrix}, \quad M_r = 2q \tag{5.85}$$

if M_r is even or

$$\mathbf{Q}_{M_r}^{(s)} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_q & \mathbf{0} & +j\mathbf{I}_q \\ \mathbf{0}^T & \sqrt{2} & \mathbf{0}^T \\ \mathbf{\Pi}_q & \mathbf{0} & -j\mathbf{\Pi}_q \end{bmatrix}, \quad M_r = 2q+1 \tag{5.86}$$

if M_r is odd. For multidimensional data, the transformation matrices have to be applied to every data dimensions. For the multidimensional smoothed data, the following transformation matrix may be used

$$\mathbf{Q}_{M_{col}} = \mathbf{Q}_{M_{sub_R}}^{(s)} \otimes \mathbf{Q}_{M_{sub_{R-1}}}^{(s)} \otimes \ldots \otimes \mathbf{Q}_{M_{sub_1}}^{(s)} . \tag{5.87}$$

The transformation from the complex-valued correlation matrix to the real valued correlation matrix is carried out by

$$\Xi_{xx} = \Re\left\{ \mathbf{Q}_{M_{SS}}^H \mathbf{R}_{xx} \mathbf{Q}_{M_{SS}} \right\}. \tag{5.88}$$

Equivalently the complex-valued smoothed data matrix $\mathbf{X}_{SS}$ can be transformed into a real-valued data matrix by

$$\mathbf{X}_{SSR} = \left[\Re\{\mathbf{Q}^H \cdot \mathbf{X}_{SS}\} \quad \Im\{\mathbf{Q}^H \cdot \mathbf{X}_{SS}\}\right].$$

One should note that taking the real- and imaginary-part of the transformed data matrix is equivalent to forward-backward smoothing. Therefore, it is sufficient to create only the forward-smoothed data matrix from the observed data $\mathbf{x}$.

In has been shown in both [1] and [44] that the straight-forward application of multi-dimensional smoothing to MIMO channel sounding measurements may lead to very large data matrices. To overcome that problem an algorithm for economy size signal subspace estimation has been developed (Section 5.3.6). The algorithm calculates only the signal subspace and not the noise subspace of the measured data. Furthermore, the multidimensional smoothing step is carried out implicitly. This avoids the explicit generation of a possibly large matrix containing the smoothed data.

5.3.4   Economy Size Signal Subspace Estimation

In this subsection, we discuss an efficient implementation of the first step of R-D Unitary ESPRIT with smoothing, i.e., the signal subspace estimation step. Throughout the derivation of the algorithm, it is assumed that the left $\mathbf{\Pi}$-real matrices $\mathbf{Q}^{(s)}_{M_{sub_r}}$ as defined in equation (5.85) and (5.86) are used to transform the complex valued data to real valued data. For simplicity, we assume at the beginning that there is only additive white noise.

Note that only the P left singular vectors that correspond to the P largest singular values of

$$\left[\Re\{\mathbf{Q}^H \cdot \mathbf{X}_{SS}\} \quad \Im\{\mathbf{Q}^H \cdot \mathbf{X}_{SS}\}\right] = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \qquad (5.89)$$

have to be computed to estimate the structural parameters of P propagation paths. In the case of channel parameter estimation from multi-dimensional channel sounding measurements, the dimension of the signal subspace may be small compared to the dimension of the noise subspace. Therefore, the standard economy size SVD algorithm based on Householder and modified QR-decompositions, is for the application at hand, not computational efficient as it computes all singular vectors and singular values. Let us discuss an example.

Suppose we have an observation of a 5-dimensional ($R_P = 5$) harmonic retrieval problem with the following dimensions $M_1 = 97$, $M_2 = 8$, $M_3 = 8$, $M_4 = 4$, $M_5 = 5$ and we want to estimate the parameters of $P = 20$ dominant multipath components. Using a subarray size of approximately $M_{sub_r} = \frac{2}{3} \cdot M_r$ for subarray smoothing leads to a data matrix of size $(65 \cdot 6 \cdot 6 \cdot 3 \cdot 4) \times 2 \cdot (33 \cdot 3 \cdot 3 \cdot 2 \cdot 2) = 28080 \times 2376$. The matrix of the left singular vectors $\mathbf{U}$ has a size of 28080×2376. Altogether, we need approximately memory for $140 \cdot 10^6$ values or, in other words, approximately 556 megabytes if we use single precision arithmetic or 1.1 gigabytes of memory if we use double precision arithmetic, only to store $\mathbf{X}_{SS}$ and the results $\mathbf{U}$ and $\mathbf{V}$. But for R-D Unitary ESPRIT, only the $P = 20$ left singular vectors spanning the signal subspace have to be computed. Therefore, we now derive an algorithm to estimate only the signal subspace.

5.3.5   Economy Size Signal Subspace Estimation for Data in White Noise

For notational convenience, let us introduce the smoothing operator

$$\mathbf{X}_{SS} = \mathscr{F}_s\{\mathbf{x}\} \in \mathbb{C}^{M_{sub} \times L}$$

describing the R-dimensional forward smoothing operation introduced in Section 5.3.2. Here the vector $\mathbf{x} \in \mathbb{C}^M$ contains only a single snapshot, i.e., $N = 1$. Note that a product involving a matrix $\mathbf{X}_{SS} = \mathcal{F}_s\{\mathbf{x}\}$ can be computed without forming the matrix $\mathbf{X}_{SS}$ itself. This is because the R-dimensional smoothing operation defines only indexing operations on $\mathbf{x}$. This property has the advantages that no additional memory is needed to store the full matrix $\mathbf{X}_{SS}$. This is also advantageous in layered memory architectures, where memory bandwidth is limited.

A computationally efficient algorithm to calculate some eigenvalues and the related eigenvectors of a matrix is based on Arnoldi-iterations [110], [111]. Let

$$\mathbf{T} = \left[\Re\{\mathbf{Q}^H \mathbf{X}_{SS}\} \quad \Im\{\mathbf{Q}^H \mathbf{X}_{SS}\}\right]^{\mathrm{T}} \cdot \left[\Re\{\mathbf{Q}^H \mathbf{X}_{SS}\} \quad \Im\{\mathbf{Q}^H \mathbf{X}_{SS}\}\right] \tag{5.90}$$

be the correlation matrix of the R-D smoothed data $\mathbf{x}$. The eigenvalue decomposition of $\mathbf{T}$ is equal to

$$\mathbf{T} \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{\Lambda},$$

where the diagonal matrix $\mathbf{\Lambda} = \mathbf{\Sigma}^2$ contains the eigenvalues of $\mathbf{T}$. Expanding the equation for the correlation matrix (5.90) yields

$$\mathbf{T} \cdot \mathbf{V} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{\Lambda}, \tag{5.91}$$

where the block matrices $\mathbf{T}_{11}, \mathbf{T}_{12}, \mathbf{T}_{21}, \mathbf{T}_{22}$ are given by

$$\begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} = \begin{bmatrix} \Re\{\mathbf{X}_{SS}^H \mathbf{Q}\} \cdot \Re\{\mathbf{Q}^H \mathbf{X}_{SS}\} & \Re\{\mathbf{X}_{SS}^H \mathbf{Q}\} \cdot \Im\{\mathbf{Q}^H \mathbf{X}_{SS}\} \\ \Im\{\mathbf{X}_{SS}^H \mathbf{Q}\} \cdot \Re\{\mathbf{Q}^H \mathbf{X}_{SS}\} & \Im\{\mathbf{X}_{SS}^H \mathbf{Q}\} \cdot \Im\{\mathbf{Q}^H \mathbf{X}_{SS}\} \end{bmatrix}.$$

Observe that the products $\Re\{\mathbf{Q}\} \cdot \Re\{\mathbf{Q}^H\}$, $\Im\{\mathbf{Q}\} \cdot \Im\{\mathbf{Q}^H\}$, $\Re\{\mathbf{Q}\} \cdot \Im\{\mathbf{Q}^H\}$, and $\Im\{\mathbf{Q}\} \cdot \Re\{\mathbf{Q}^H\}$ yield the following simple sparse matrices

$$\Re\{\mathbf{Q}\} \cdot \Re\{\mathbf{Q}^H\} = \frac{1}{2} \cdot (\mathbf{I} + \mathbf{\Pi}) = \mathbf{Z}_1,$$

$$\Im\{\mathbf{Q}\} \cdot \Im\{\mathbf{Q}^H\} = \frac{1}{2} \cdot (\mathbf{\Pi} - \mathbf{I}) = \mathbf{Z}_2$$

or even a zero matrix

$$\Re\{\mathbf{Q}\} \cdot \Im\{\mathbf{Q}^H\} = \Im\{\mathbf{Q}\} \cdot \Re\{\mathbf{Q}^H\} = \mathbf{0}.$$

Using these properties, we can express the four block matrices $\mathbf{T}_{11}, \mathbf{T}_{12}, \mathbf{T}_{21}, \mathbf{T}_{22}$ as

$$\mathbf{T}_{11} = \Re\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_1 \cdot \Re\{\mathbf{X}_{SS}\} - \Im\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_2 \cdot \Re\{\mathbf{X}_{SS}\}$$

$$\mathbf{T}_{12} = \Re\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_1 \cdot \Im\{\mathbf{X}_{SS}\} + \Im\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_2 \cdot \Im\{\mathbf{X}_{SS}\}$$

$$\mathbf{T}_{21} = \Im\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_1 \cdot \Re\{\mathbf{X}_{SS}\} + \Re\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_2 \cdot \Re\{\mathbf{X}_{SS}\}$$

$$\mathbf{T}_{22} = \Im\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_1 \cdot \Im\{\mathbf{X}_{SS}\} - \Re\{\mathbf{X}_{SS}^T\} \cdot \mathbf{Z}_2 \cdot \Im\{\mathbf{X}_{SS}\}.$$

Note that the computation of a product that involves $\mathbf{Z}_1$ or $\mathbf{Z}_2$ requires no multiplication (except for the scaling by $\frac{1}{2}$), but only real-valued additions and subtractions are necessary. Moreover, the product of $\mathbf{Z}_1$ and an arbitrary vector $\mathbf{w}$ yields the even part of $\mathbf{w}$ and the product $\mathbf{Z}_2 \cdot \mathbf{w}$ is the odd part.

In one Arnoldi iteration, the product between $\mathbf{T}$ and a vector provided by the Arnoldi core algorithm, e.g., $\mathbf{v}$, has to be calculated.

$$\mathbf{w} = \mathbf{T} \cdot \mathbf{v} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \tag{5.92}$$

Exploiting the redundancies in the expressions for the block matrices $\mathbf{T}_{11}, \mathbf{T}_{12}, \mathbf{T}_{21}, \mathbf{T}_{22}$ and omitting the scaling by $\frac{1}{2}$ in $\mathbf{Z}_1$ and $\mathbf{Z}_2$ leads to the following steps for the calculation of $\mathbf{w} = \mathbf{T} \cdot \mathbf{v}$

$$\mathbf{x}_1 = \Re\{\mathbf{X}_{SS}\} \cdot \mathbf{v}_1$$
$$\mathbf{x}_2 = \Im\{\mathbf{X}_{SS}\} \cdot \mathbf{v}_2$$
$$\mathbf{y}_1 = \mathbf{Z}_1 \cdot (\mathbf{x}_1 + \mathbf{x}_2)$$
$$\mathbf{y}_2 = \mathbf{Z}_2 \cdot (\mathbf{x}_1 - \mathbf{x}_2)$$
$$\mathbf{w}_1 = \Re\{\mathbf{X}_{SS}^{\mathrm{T}}\} \cdot \mathbf{y}_1 - \Im\{\mathbf{X}_{SS}^{\mathrm{T}}\} \cdot \mathbf{y}_2$$
$$\mathbf{w}_2 = \Re\{\mathbf{X}_{SS}^{\mathrm{T}}\} \cdot \mathbf{y}_2 + \Im\{\mathbf{X}_{SS}^{\mathrm{T}}\} \cdot \mathbf{y}_1.$$

By taking into account that the matrix $\mathbf{X}_{SS}$ can be expressed via the R-D smoothing operator that is applied to $\mathbf{x}$, we compute the product $\mathbf{w} = \mathbf{T} \cdot \mathbf{v}$ in the following fashion

$$\mathbf{x}_1 = \mathcal{F}_s\{\Re\{\mathbf{x}\}\} \cdot \mathbf{v}_1$$
$$\mathbf{x}_2 = \mathcal{F}_s\{\Im\{\mathbf{x}\}\} \cdot \mathbf{v}_2$$
$$\mathbf{y}_1 = \mathbf{Z}_1 \cdot (\mathbf{x}_1 + \mathbf{x}_2)$$
$$\mathbf{y}_2 = \mathbf{Z}_2 \cdot (\mathbf{x}_1 - \mathbf{x}_2)$$
$$\mathbf{w}_1 = \mathcal{F}_s\{\Re\{\mathbf{x}\}\}^{\mathrm{T}} \cdot \mathbf{y}_1 - \mathcal{F}_s\{\Im\{\mathbf{x}\}\}^{\mathrm{T}} \cdot \mathbf{y}_2$$
$$\mathbf{w}_2 = \mathcal{F}_s\{\Re\{\mathbf{x}\}\}^{\mathrm{T}} \cdot \mathbf{y}_2 + \mathcal{F}_s\{\Im\{\mathbf{x}\}\}^{\mathrm{T}} \cdot \mathbf{y}_1.$$

Using this scheme in conjunction with an eigenvalue decomposition based on implicitly restarted Arnoldi-iterations (reference netlib ARPACK (Arnoldi Package, functions dsaupd and dseupd)), we get a computationally efficient and also memory efficient algorithm for the calculation of some right singular vectors $\mathbf{V}_S$ of $\mathbf{X}_{SS}$. ARPACK uses a so called reverse communication interfaces. One has to allocate the necessary space for the eigenvectors, eigenvalues and some temporary work memory. The main computation steps are carried out by the core functions, e.g., dsaupd. An iteration step is carried out by calling at first the core function. Depending on the return value one has either to compute the product $\mathbf{w} = \mathbf{T} \cdot \mathbf{v}$, where $\mathbf{v}$ is provided by the core function, or stop the iteration since the algorithm has converged.

Finally, the signal subspace estimate for R-D Unitary ESPRIT can be calculated from $\mathbf{V}_S$ using the following relationship:

$$\mathbf{E}_S = \left[\Re\{\mathbf{Q}^{\mathrm{H}}\mathbf{X}_{SS}\} \quad \Im\{\mathbf{Q}^{\mathrm{H}}\mathbf{X}_{SS}\} \right] \mathbf{V}_S = \mathbf{U}_S \boldsymbol{\Sigma}_S. \tag{5.93}$$

The left singular vectors of $\mathbf{X}_{SS}$ that correspond to the calculated right singular vectors $\mathbf{V}_S$ are obtained from $\mathbf{E}_S$ by taking into account that

$$\mathbf{E}_S^{\mathrm{T}}\mathbf{E}_S = \boldsymbol{\Sigma}_S^{\mathrm{T}}\mathbf{U}_S^{\mathrm{T}}\mathbf{U}_S\boldsymbol{\Sigma}_S = \boldsymbol{\Sigma}_S^{\mathrm{T}}\boldsymbol{\Sigma}_S. \tag{5.94}$$

In other words, the singular values $\mathrm{diag}\{\boldsymbol{\Sigma}_S\}$ are just the vector-norms of the columns of $\mathbf{E}_S$. Nevertheless, one should note that the normalization of the columns of $\mathbf{E}_S$ is not necessary

for R-D Unitary ESPRIT. Therefore, a computationally and memory efficient way to calculate the signal subspace $\mathbf{E}_S$ from $\mathbf{V}_S$ is given by

$$\mathbf{E}_S = \left[\Re\{\mathbf{Q}^H \mathbf{X}_{SS}\} \quad \Im\{\mathbf{Q}^H \mathbf{X}_{SS}\} \right] \mathbf{V}_S . \tag{5.95}$$

The computational complexity of the proposed algorithm for the computation of all singular vectors is approximately twice the effort of the classical SVD using Householder- and modified QR-transformations. However, since the number of concentrated propagation paths is typically significantly smaller than the number of columns of the smoothed data matrix $\mathbf{X}_{SS}$ the outlined algorithm reduces the computational effort for the estimation of the signal subspace. One should note that the accuracy of the described algorithm is similar to the accuracy of classical SVD algorithms.

In view of the memory requirements, the signal subspace estimation algorithm is a prerequisite for the application of the multidimensional Unitary ESPRIT to large scale problems, such as channel parameter estimation from MIMO channel sounding measurements. The ESSSE (economy size signal subspace estimation) algorithm is summarised in Table 5-5.

Table 5-5: Summary of economy size signal subspace estimation in white noise.

---

1) Call the Arnoldi-iteration function (e.g. dsaupd) yielding an estimated eigenvector $\mathbf{v}$.

2) Check convergence: yes - go to step 4; no - continue iteration with step 3.

3) Compute the product $\mathbf{w}^{\{i\}} = \mathbf{T} \cdot \mathbf{v}^{\{i\}} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v}_1^{\{i\}} \\ \mathbf{v}_2^{\{i\}} \end{bmatrix}$, where $\mathbf{v}^{\{i\}}$ has

   been provided by the core function dsaupd via

   $$\mathbf{x}_1 = \mathcal{F}_s\{\Re\{\mathbf{x}\}\} \cdot \mathbf{v}_1^{\{i\}}$$
   $$\mathbf{x}_2 = \mathcal{F}_s\{\Im\{\mathbf{x}\}\} \cdot \mathbf{v}_2^{\{i\}}$$
   $$\mathbf{y}_1 = \mathbf{Z}_1 \cdot (\mathbf{x}_1 + \mathbf{x}_2)$$
   $$\mathbf{y}_2 = \mathbf{Z}_2 \cdot (\mathbf{x}_1 - \mathbf{x}_2)$$
   $$\mathbf{w}_1^{\{i\}} = \mathcal{F}_s\{\Re\{\mathbf{x}\}\}^T \cdot \mathbf{y}_1 - \mathcal{F}_s\{\Im\{\mathbf{x}\}\}^T \cdot \mathbf{y}_2$$
   $$\mathbf{w}_2^{\{i\}} = \mathcal{F}_s\{\Re\{\mathbf{x}\}\}^T \cdot \mathbf{y}_2 + \mathcal{F}_s\{\Im\{\mathbf{x}\}\}^T \cdot \mathbf{y}_1.$$

   Form the result $\mathbf{w}^{\{i\}}$ and return it to the core function dsaupd. Go to step 1.

4) Compute the signal subspace estimate using the computed right singular vectors as

   $$\mathbf{E}_S = \left[\Re\{\mathbf{Q}^H \mathcal{F}_s\{\mathbf{x}\}\} \quad \Im\{\mathbf{Q}^H \mathcal{F}_s\{\mathbf{x}\}\} \right] \mathbf{V}_S .$$

5.3.6   Economy Size Signal Subspace Estimation for Data in Coloured Noise

If we want to use R-D Unitary ESPRIT or RARE for the estimation of the parameters of the specular propagation paths, we have to consider the contribution of the dense multipath components to the observation. In the context of signal subspace estimation, the contributions of the DMC can be understood as contributions from an additive coloured noise process. Consequently, the correlation matrix of the DMC and the measurement noise $\mathbf{R}(\boldsymbol{\theta}_{DMC})$ can be considered as the covariance matrix of a coloured noise process $\mathbf{R}_{ww}$. There exist two general solutions, namely the covariance and the square root approach, for the estimation of signal

subspaces in coloured noise. See [44] (pp. 267-275) for a discussion of the two solutions. Although both solutions can be theoretically applied to the problem at hand, the straightforward implementation may be impractical. The direct implementation of the signal subspace estimation step will lead to a memory inefficient algorithm as already discussed in the previous section. In the following, a memory efficient algorithm for the signal subspace estimation from observations with coloured noise is outlined.

The general structure of the signal correlation matrix is

$$\mathbf{R}_{xx} = \mathbf{A} \cdot \mathbf{R}_{ss} \cdot \mathbf{A}^H + \mathbf{R}_{ww} = \mathbf{E}\left\{ \tfrac{1}{N} \mathbf{X}_{ss} \cdot \mathbf{X}_{ss}^H \right\}, \tag{5.96}$$

and the signal subspace $\mathbf{E}_s$ to estimate spans the same subspace as $\mathbf{A}$. The authors of [44] suggest the decomposition of the noise covariance matrix into

$$\mathbf{R}_{ww} = \mathbf{L}_w \mathbf{L}_w^H.$$

Then pre-whitening of the correlation matrix $\mathbf{R}_{xx}$ can be achieved by calculating

$$\mathbf{R}'_{xx} = \mathbf{L}_w^{-1} \cdot \mathbf{R}_{xx} \cdot \mathbf{L}_w^{-H}. \tag{5.97}$$

Note that inserting (5.97) in (5.96) yields the structure of the data correlation matrix after pre-whitening

$$\mathbf{L}_w^{-1} \mathbf{R}_{xx} \mathbf{L}_w^{-H} = \mathbf{L}_w^{-1} \mathbf{A} \mathbf{R}_{ss} \mathbf{A}^H \mathbf{L}_w^{-H} + \mathbf{I}. \tag{5.98}$$

The eigenvalue decomposition of (5.98) has the form

$$\left[\mathbf{E}'_s \ \ \mathbf{E}'_0 \right] \Sigma^2 \left[\mathbf{E}'_s \ \ \mathbf{E}'_0 \right]^H = \mathbf{L}_w^{-1} \mathbf{A} \mathbf{R}_{ss} \mathbf{A}^H \mathbf{L}_w^{-H} + \mathbf{I},$$

where $\mathbf{E}'_s = \mathbf{L}_w^{-1} \mathbf{A} \mathbf{T}$ is an estimate of the transformed signal subspace and $\mathbf{E}'_0$ is the transformed noise subspace (null space of $\mathbf{E}'_s$). Reversing the pre-whitening with $\mathbf{L}_w$ yields the signal subspace as

$$\mathbf{E}_s = \mathbf{L}_w \mathbf{E}'_s.$$

Applying the pre-whitener $\mathbf{L}_w^{-1}$ directly to the data matrix $\mathbf{X}$ yields the pre-whitened data matrix

$$\mathbf{X}'_{ss} = \mathbf{L}_w^{-1} \mathbf{X}_{ss},$$

which is related to the correlation matrix (5.97) according to

$$\mathbf{R}'_{xx} = \mathbf{E}\left\{ \frac{1}{N} \mathbf{X}'_{ss} \cdot (\mathbf{X}'_{ss})^H \right\}.$$

The singular value decomposition of the pre-whitened data matrix yields

$$\mathbf{U}' \Sigma \mathbf{V}^H = \mathbf{X}'_{ss}. \tag{5.99}$$

If the noise covariance matrix $\mathbf{R}_{ww}$ has full rank the inverse $\mathbf{R}_{ww}^{-1}$ exist, and the following identity holds

$$(\mathbf{X}'_{ss})^H \mathbf{X}'_{ss} = \mathbf{X}_{ss}^H \underbrace{\mathbf{L}_w^{-H} \mathbf{L}_w^{-1}}_{\mathbf{R}_{ww}^{-1}} \mathbf{X}_{ss} = \mathbf{V} \Sigma^2 \mathbf{V}^H.$$

Therefore, we calculate the eigenvalue decomposition of

$$\mathbf{X}_{ss}^{H}\mathbf{L}_{w}^{-H}\mathbf{L}_{w}^{-1}\mathbf{X}_{ss} = \mathbf{V}\,\mathbf{\Sigma}^{2}\mathbf{V}^{H}$$

to get the right singular vectors of (5.99). Using the singular vectors belonging to the right signal subspace $\mathbf{V}_s$ the left singular vectors spanning the estimated signal subspace $\hat{\mathbf{E}}_s$ can be derived from the observed smoothed data $\mathbf{X}_{SS}$ according to

$$\hat{\mathbf{E}}_{s} = \mathbf{U}_{s}\mathbf{\Sigma}_{s} = \mathbf{X}_{ss} \cdot \mathbf{V}_{s} = \hat{\mathbf{A}} \cdot \mathbf{T} \,.$$

As already discussed at the in Section 5.2.9 signal parameter estimation using signal subspace based parameter estimation techniques requires multidimensional smoothing as a pre-processing step. The data matrix $\mathbf{X}_{SS}$ is related to the observed data $\mathbf{x}$ according to

$$\mathbf{X}_{SS} = \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\} = \left[\mathcal{F}_{s}\{\mathbf{x}\} \quad \mathbf{\Pi}\,\mathcal{F}_{s}\{\mathbf{x}\}^{*}\right], \tag{5.100}$$

where $\mathcal{F}_{s}^{\mathrm{fb}}\{\bullet\}$ denotes the forward-backward smoothing operator. Assuming we know the covariance matrix of the DMC $\mathbf{R}(\mathbf{\theta}_{dan})$ the eigenvalue problem to solve is

$$\mathbf{V}\mathbf{\Sigma}^{2} = \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\}^{H} \cdot \mathbf{R}(\mathbf{\theta}_{dan})^{-1} \cdot \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\} \cdot \mathbf{V} \,.$$

Again, it is important to realise that only the right singular vectors $\mathbf{V}_s$ belonging to the P dominant generalised eigenvalues must be computed for R-D Unitary ESPRIT.

A numerically stable algorithm for the calculation of some eigenvalues and eigenvectors, i.e., singular vectors of complex matrices is also provided by ARPACK (see http://www.caam.rice.edu/software/ARPACK, functions znaupd and zneupd), which is based on implicitly restarted Arnoldi iterations [110]. The only function the user has to supply to the algorithm is the product $\mathbf{w} = \mathbf{A} \cdot \mathbf{v}$, i.e.,

$$\mathbf{w} = \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\}^{H} \cdot \mathbf{R}(\mathbf{\theta}_{dan})^{-1} \cdot \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\} \cdot \mathbf{v} \,. \tag{5.101}$$

Hence, an estimate of the signal subspace can be computed without forming the smoothed data matrix $\mathbf{X}_{SS}$ explicitly. To compute the product (5.101) three steps are necessary

$$\mathbf{y}_{1} = \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\} \cdot \mathbf{v} \Rightarrow \mathbf{y}_{2} = \mathbf{R}(\mathbf{\theta}_{dan})^{-1} \cdot \mathbf{y}_{1} \Rightarrow \mathbf{w} = \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\}^{H} \cdot \mathbf{y}_{2}. \tag{5.102}$$

If the covariance matrix can be factorized using the Kronecker product as discussed in Section 2.5.3 the computational effort to compute $\mathbf{y}_{2} = \mathbf{R}(\mathbf{\theta}_{dan})^{-1} \cdot \mathbf{y}_{1}$ and even more important the storage requirements for $\mathbf{R}(\mathbf{\theta}_{dan})^{-1}$ can be reduced since

$$\mathbf{R}(\mathbf{\theta}_{dan})^{-1} = \left(\mathbf{R}_{R}(\mathbf{\theta}_{dan}) \otimes \mathbf{R}_{T}(\mathbf{\theta}_{dan}) \otimes \mathbf{R}_{f}(\mathbf{\theta}_{dan}) \otimes \mathbf{R}_{t}(\mathbf{\theta}_{dan})\right)^{-1}$$
$$= \mathbf{R}_{R}(\mathbf{\theta}_{dan})^{-1} \otimes \mathbf{R}_{T}(\mathbf{\theta}_{dan})^{-1} \otimes \mathbf{R}_{f}(\mathbf{\theta}_{dan})^{-1} \otimes \mathbf{R}_{t}(\mathbf{\theta}_{dan})^{-1}.$$

Using implicitly restarted Arnoldi iterations in combination with (5.102) the computation and the storage of the complete covariance matrix $\mathbf{R}(\mathbf{\theta}_{dan})$ can be avoided. In particular we can even take advantage of the Toeplitz structure of $\mathbf{R}_{f}(\mathbf{\theta}_{dan})$ and use a computationally efficient algorithm as outlined in section 6.1.12 when computing the product $\mathbf{y}_{2} = \mathbf{R}(\mathbf{\theta}_{dan})^{-1} \cdot \mathbf{y}_{1}$.

Finally one should observe that the calculation of the left singular vectors of (5.100) can be computed without forming the data matrix $\mathbf{X}_{SS}$ explicitly too, using the relation

$$\hat{\mathbf{E}}_{s} = \mathcal{F}_{s}^{\mathrm{fb}}\{\mathbf{x}\} \cdot \mathbf{V}_{s} \,. \tag{5.103}$$

From the estimated complex signal subspace $\hat{\mathbf{E}}_s$ the real valued signal subspace for R-D Unitary ESPRIT $\hat{\mathbf{D}}_s$ can be computed by

$$\hat{\mathbf{D}}_s = \mathbf{Q}^H \hat{\mathbf{E}}_s .$$

One should note that the transformation of the data and the noise covariance matrix to the real domain before signal subspace estimation as proposed in [28] is not reasonable in the case under consideration. This is due to the fact, that the transformation to the real domain destroys the Toeplitz structure of the noise covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{dan})$. For example, the transformation of $\text{toep}\{[2 \; j \; -1]^H\}$ leads to

$$\mathbf{Q}_3^H \begin{bmatrix} 2 & j & -1 \\ -j & 2 & j \\ -1 & -j & 2 \end{bmatrix} \mathbf{Q}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & \sqrt{2} \\ 0 & \sqrt{2} & 3 \end{bmatrix} .$$

Consequently, computationally efficient algorithms for the multiplication with inverse Toeplitz matrices (cf. Section 6.1.12) are not applicable anymore if the complex correlation matrices are transformed to the real domain using equation (5.88).

Table 5-6: Truncated signal subspace estimation in coloured noise.

1) Call the Arnoldi-iteration function (e.g. znaupd).

2) Check convergence: yes - go to step 4; no - continue iteration with step 3.

3) Compute the product $\mathbf{w}^{\{i\}} = \mathcal{F}_s^{\text{fb}}\{\mathbf{x}\}^H \cdot \mathbf{R}(\boldsymbol{\theta}_{dan})^{-1} \cdot \mathcal{F}_s^{\text{fb}}\{\mathbf{x}\} \cdot \mathbf{v}^{\{i\}}$, where $\mathbf{v}^{\{i\}}$ has been provided by the core function znaupd via

$$\mathbf{y}_1 = \mathcal{F}_s^{\text{fb}}\{\mathbf{x}\} \cdot \mathbf{v}$$

$$\mathbf{y}_2 = \mathbf{R}_R(\boldsymbol{\theta}_{dan})^{-1} \otimes \mathbf{R}_T(\boldsymbol{\theta}_{dan})^{-1} \otimes \mathbf{R}_f(\boldsymbol{\theta}_{dan})^{-1} \otimes \mathbf{R}_t(\boldsymbol{\theta}_{dan})^{-1} \cdot \mathbf{y}_1$$

$$\mathbf{w}^{\{i\}} = \mathcal{F}_s^{\text{fb}}\{\mathbf{x}\}^H \cdot \mathbf{y}_2 .$$

Return the vector $\mathbf{w}^{\{i\}}$ to the core function znaupd. Go to step 1.

4) Compute the signal subspace estimate using the computed right singular vectors as

$$\hat{\mathbf{E}}_s = \mathcal{F}_s^{\text{fb}}\{\mathbf{x}\} \cdot \mathbf{V}_s .$$

5) Optional: Compute the signal subspace for Unitary ESPRIT by

$$\hat{\mathbf{D}}_s = \mathbf{Q}^H \hat{\mathbf{E}}_s .$$

5.3.7 Subspace Rotation Invariance - ESPRIT

Paulraj, Roy, and Kailath first introduced the ESPRIT algorithm in [107]. ESPRIT is a search free fully parametric estimation procedure for undamped cisoids[11] in noise. The original method often referred to as LS (least squares) – ESPRIT is not statistically efficient, i.e., not minimum variance unbiased (MVUB). Nevertheless, since it is very attractive in terms of computational complexity various improved algorithms have been developed based on the pioneering work of Paulraj, Roy, and Kailath. An overview of the variants is given in [44, pp. 21-26].

In the following a short summary of the LS-ESPRIT, the Unitary ESPRIT and the R-dimensional Unitary ESPRIT is given.

---

[11]Cisoid is an abbreviation for complex-valued sinusoidal signal, e.g., complex exponential

5.3.8   LS-ESPRIT

The basis of the standard ESPRIT algorithm is the so-called invariance equation. It has the form

$$\mathbf{J}_1\mathbf{A}(\boldsymbol{\mu})\cdot\boldsymbol{\Omega}(\boldsymbol{\mu}) = \mathbf{J}_2\mathbf{A}(\boldsymbol{\mu}), \qquad (5.104)$$

where $\mathbf{J}_1$ and $\mathbf{J}_2$ are selection matrices and $\boldsymbol{\Omega}(\boldsymbol{\mu})$ is a diagonal matrix

$$\boldsymbol{\Omega}(\boldsymbol{\mu}) = \begin{bmatrix} e^{-j\mu_1} & 0 & \cdots & 0 \\ 0 & e^{-j\mu_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & e^{-j\mu_P} \end{bmatrix}.$$

The invariance equation relates the subset $\mathbf{J}_1\mathbf{A}(\boldsymbol{\mu})$ via a phase rotation to $\mathbf{J}_2\mathbf{A}(\boldsymbol{\mu})$. For the basis function defined in (3.4) the best choice for the selection matrices, providing maximum overlap, is

$$\mathbf{J}_1 = [\mathbf{I}\ \mathbf{0}] \in \mathbb{R}^{M-1\times M}, \text{ and } \mathbf{J}_2 = [\mathbf{0}\ \mathbf{I}] \in \mathbb{R}^{M-1\times M}.$$

As discussed in section 5.3.1 the signal subspace $\mathbf{E}_S$ spans the same subspace as $\mathbf{A}(\boldsymbol{\mu})$, i.e.,

$$\mathbf{A}(\boldsymbol{\mu})\mathbf{T} = \mathbf{E}_s, \qquad (5.105)$$

where $\mathbf{T}$ is an arbitrary full rank matrix $\mathbf{T} \in \mathbb{C}^{P\times P}$. Replacing the exact signal subspace $\mathbf{E}_S$ with the estimate $\hat{\mathbf{E}}_S$ and inserting equation (5.105) in (5.104) yields

$$\mathbf{J}_1\hat{\mathbf{E}}_S\mathbf{T}^{-1}\boldsymbol{\Omega}(\boldsymbol{\mu}) \approx \mathbf{J}_2\hat{\mathbf{E}}_S\mathbf{T}^{-1}. \qquad (5.106)$$

Consequently, the least squares solution of

$$\hat{\boldsymbol{\Psi}}_{LS} = \arg\min_{\boldsymbol{\Psi}_{LS}}\left\|\mathbf{J}_1\hat{\mathbf{E}}_S\boldsymbol{\Psi}_{LS} - \mathbf{J}_2\hat{\mathbf{E}}_S\right\|_F^2 \in \mathbb{C}^{P\times P}$$

has approximately the structure

$$\hat{\boldsymbol{\Psi}}_{LS} \approx \mathbf{T}^{-1}\cdot\boldsymbol{\Omega}(\boldsymbol{\mu})\cdot\mathbf{T}.$$

The eigenvalues of the solution $\hat{\boldsymbol{\Psi}}_{LS}$ are estimates of the P phase factors $e^{-j\mu_p}$. In the final step of LS-ESPRIT one has to calculate the P nonlinear model parameters $\hat{\mu}_p$ from the estimated phase factors, i.e., from the P eigenvalues ϖ_p of $\hat{\boldsymbol{\Psi}}_{LS}$ as

$$\hat{\mu}_p = \Im\{\ln(\varpi_p)\}, \quad 1 \le p \le P.$$

The difference between LS-ESPRIT and its variations the TLS-, WLS-, and SLS-ESPRIT [112] (Total-, Weighted-, Structured Least Squares) is the handling of the error $\boldsymbol{\Delta}_{E_S} = \hat{\mathbf{E}}_S - \mathbf{E}_S$ in the subspace estimate $\hat{\mathbf{E}}_S$. More precisely, since we use overlapping subsets of $\hat{\mathbf{E}}_S$ both sides of equation (5.106) contain some residual error from the subspace estimate. It may be tempting to use the TLS (total least squares) solution instead of the LS solution to solve for $\hat{\boldsymbol{\Psi}}$ in (5.106). However also the TLS solution is suboptimal since the error on the left hand side and the right hand side is correlated. It has been shown in [65], [112] that the optimum solution in a least squares sense is the structured least squares solution (SLS).

5.3.9 Unitary ESPRIT

Since the standard ESPRIT algorithm operates on complex-valued data, complex computations have to be carried out to compute a parameter estimate. In order to reduce the computational complexity, the Unitary ESPRIT [65], [28] algorithm has been developed. It transforms the input data matrix $\mathbf{X}$ to a real-valued representation. Therefore, all computations are real-valued. This transformation $\mathcal{T}_{CR}$ is applicable to the smoothed data matrix $\mathcal{F}_S(\mathbf{x})$ if the smoothed data can be expressed as $\mathcal{F}_S(\mathbf{x}) = \mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\Gamma} + \mathbf{N}$ as already discussed in Section 5.3.3. The transformation $\mathcal{T}_{CR}$ can be expressed in the form

$$\mathcal{T}_{CR} = \mathbf{Z} = \mathbf{Q}_{M_{\text{sub}}}^{\text{H}} \big[\mathcal{F}_s\{\mathbf{x}\} \quad \boldsymbol{\Pi}_{M_{\text{sub}}} \mathcal{F}_s\{\mathbf{x}^*\} \boldsymbol{\Pi}_{M-M_{\text{sub}}+1} \big] \mathbf{Q}_{2M-2M_{\text{sub}}+2}^{\text{H}} \in \mathbb{R}^{M_{\text{sub}} \times 2(M-M_{\text{sub}}+1)}, \qquad (5.107)$$

where $\boldsymbol{\Pi}_N$ is the reflection matrix, i.e., a matrix with ones on its anti-diagonal and zeros elsewhere. Similar to the standard ESPRIT algorithm the signal subspace $\mathbf{E}_S$ is estimated from the real-valued data matrix $\mathbf{Z}$ using a singular value decomposition or using a eigenvalue decomposition of the Gramian $\mathbf{Z}\mathbf{Z}^{\text{H}}$. The real-valued representation of the invariance equation (5.104) is

$$\mathbf{K}_1 \mathbf{E}_S \boldsymbol{\Psi} \approx \mathbf{K}_2 \mathbf{E}_S,$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are the transformed selection matrices $\mathbf{J}_1$ and $\mathbf{J}_2$ having the structure

$$\mathbf{K}_1 = 2 \cdot \Re\{\mathbf{Q}_{M_{\text{sub}}-1}^{\text{H}} \mathbf{J}_2 \mathbf{Q}_{M_{\text{sub}}}\} \qquad (5.108)$$

and

$$\mathbf{K}_2 = 2 \cdot \Im\{\mathbf{Q}_{M_{\text{sub}}-1}^{\text{H}} \mathbf{J}_2 \mathbf{Q}_{M_{\text{sub}}}\}. \qquad (5.109)$$

Solving the Unitary ESPRIT invariance equation using the least squares approach

$$\boldsymbol{\Psi}_{ULS} = (\mathbf{K}_1 \mathbf{E}_S)^+ \mathbf{K}_2 \mathbf{E}_S$$

yields an estimate for the predictor matrix $\boldsymbol{\Psi}_{ULS}$. The similarity transformation to diagonal structure, i.e., the eigenvalue decomposition yields

$$\boldsymbol{\Psi}_{ULS} = \mathbf{T} \boldsymbol{\Omega}_{ULS} \mathbf{T}^{-1} \in \mathbb{R}^{P \times P}.$$

From the eigenvalues $\varpi = \text{diag}\{\boldsymbol{\Omega}_{ULS}\}$ the estimates of the structural parameters $\boldsymbol{\mu}$ can be computed by

$$\hat{\mu}_i = 2 \arctan(\varpi_i). \qquad (5.110)$$

One should observe, that equation (5.110) reveals a weakness of the Unitary ESPRIT. If the true parameter μ are close to $\pm\pi$ the bilinear transformed parameters $\varpi_i = \tan\left(\frac{\mu_i}{2}\right)$ tend to $\pm\infty$. Hence, the price for the real-valued computation of Unitary ESPRIT is a larger eigenvalue spread of $\boldsymbol{\Psi}_{ULS}$ and numerical problems if $|\mu_p| \sim \pi$.

5.3.10 Multidimensional Unitary ESPRIT

The extension of Unitary ESPRIT to multidimensional Unitary ESPRIT is straightforward. After multidimensional smoothing the data matrix has multiple rotational invariance structure, i.e., it can be expressed in the form

$$\mathbf{X} = \mathcal{F}_s^{\text{fb}}\{\mathbf{x}\} = \big(\mathbf{A}(\boldsymbol{\mu}^{(R_P)}) \Diamond \mathbf{A}(\boldsymbol{\mu}^{(R_P-1)}) \Diamond \ldots \Diamond \mathbf{A}(\boldsymbol{\mu}^{(1)})\big) \cdot \boldsymbol{\Gamma} + \mathbf{N} = \mathbf{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\Gamma} + \mathbf{N}.$$

Observe, that the matrices $\mathbf{A}\left(\boldsymbol{\mu}^{(r)}\right)$ have, due to subarray smoothing, a size of $\mathbf{A}\left(\boldsymbol{\mu}^{(r)}\right) \in \mathbb{C}^{M_{sub_r} \times P}$. For every parameter dimension $1 \le r \le R_P$ the following generalised invariance equation holds

$$\mathbf{J}_{1,r}\mathbf{A}(\boldsymbol{\mu})\cdot\boldsymbol{\Omega}_r(\boldsymbol{\mu}) = \mathbf{J}_{2,r}\mathbf{A}(\boldsymbol{\mu}).$$

The related R_P generalised selection matrices are given by

$$\mathbf{J}_{1,1} = \mathbf{I}_{M_{sub,R_P}} \otimes \ldots \otimes \mathbf{I}_{M_{sub,2}} \otimes \mathbf{J}_1^{(M_{sub,1})}$$

$$\mathbf{J}_{1,2} = \mathbf{I}_{M_{sub,R_P}} \otimes \ldots \otimes \mathbf{J}_1^{(M_{sub,2})} \otimes \mathbf{I}_{M_{sub,1}}$$

$$\vdots \qquad \qquad \qquad \ddots$$

$$\mathbf{J}_{1,R_P} = \mathbf{J}_1^{(M_{sub,R_P})} \otimes \ldots \otimes \mathbf{I}_{M_{sub,2}} \otimes \mathbf{I}_{M_{sub,1}}$$

and

$$\mathbf{J}_{2,1} = \mathbf{I}_{M_{sub_{R_P}}} \otimes \ldots \otimes \mathbf{I}_{M_{sub_2}} \otimes \mathbf{J}_2^{(M_{sub_1})}$$

$$\mathbf{J}_{2,2} = \mathbf{I}_{M_{sub_{R_P}}} \otimes \ldots \otimes \mathbf{J}_2^{(M_{sub_2})} \otimes \mathbf{I}_{M_{sub_1}}$$

$$\vdots \qquad \qquad \qquad \ddots$$

$$\mathbf{J}_{2,R_P} = \mathbf{J}_2^{(M_{sub_{R_P}})} \otimes \ldots \otimes \mathbf{I}_{M_{sub_2}} \otimes \mathbf{I}_{M_{sub_1}},$$

where the identity matrices have size $\mathbf{I}_{M_{sub,r}} \in \mathbb{R}^{M_{sub,r} \times M_{sub,r}}$ and the sub-selection matrices the structure

$$\mathbf{J}_1^{(M_{sub,r})} = \begin{bmatrix} \mathbf{I}_{M_{sub,r}} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(M_{sub,r}-1)\times M_{sub,r}}, \text{ and } \mathbf{J}_2^{(M_{sub,r})} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{M_{sub,r}} \end{bmatrix} \in \mathbb{R}^{(M_{sub,r}-1)\times M_{sub,r}}.$$

Replacing the sub-selection matrices $\mathbf{J}_1^{(M_{sub,r})}$ and $\mathbf{J}_2^{(M_{sub,r})}$ by their real valued counterparts defined in (5.108) and (5.109) and transforming the complex valued data to the real domain using the transformation (5.107) leads to the R_P real valued invariance equations for the signal subspace estimate,

$$\mathbf{K}_{1,r}\mathbf{E}_S\boldsymbol{\Psi}_r \approx \mathbf{K}_{2,r}\mathbf{E}_S.$$

In the noiseless case or with an infinite number of independent observations all predictor matrices admit the following eigenvalue decomposition

$$\boldsymbol{\Psi}_r = \mathbf{T}\boldsymbol{\Omega}_r\mathbf{T}^{-1} \in \mathbb{R}^{P\times P}. \tag{5.111}$$

The eigenvalues are related to the structural parameters $\boldsymbol{\mu}$ according to

$$\boldsymbol{\Omega}_r = \operatorname{diag}\left\{\tan\left(\frac{\mu_p^{(r)}}{2}\right)\right\}_{p=1}^{P}.$$

It is important to note that all R_P predictor matrices share the same set of eigenvectors in the noiseless case. Therefore, a joint eigenvalue decomposition of all predictor matrices leads to a joint estimation procedure for all structural parameters. A solution to this joint eigenvalue decomposition problem is the Simultaneous Schur Decomposition (SSD). A derivation of the algorithm can be found in [65], [28], or [113].

5.3.11   Data with Hidden Rotational Invariance Structure

The general data model for the radio channel developed in Section 2 has the structure

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{GA}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma} = \mathbf{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\gamma}$$

(cf. expression (2.54)). Now let us suppose the matrix valued function $\mathbf{B}(\boldsymbol{\mu})$ can be decomposed into

$$\mathbf{B}(\boldsymbol{\mu}) = \left(\mathbf{G}_R \cdot \mathbf{A}\left(\boldsymbol{\mu}^{(R)}\right)\right) \Diamond \dots \Diamond \left(\mathbf{G}_1 \cdot \mathbf{A}\left(\boldsymbol{\mu}^{(1)}\right)\right).$$

The related data model for the radio channel becomes

$$\mathbf{s}(\boldsymbol{\theta}) = \left(\left(\mathbf{G}_R \cdot \mathbf{A}\left(\boldsymbol{\mu}^{(R)}\right)\right) \Diamond \dots \Diamond \left(\mathbf{G}_1 \cdot \mathbf{A}\left(\boldsymbol{\mu}^{(1)}\right)\right)\right) \cdot \boldsymbol{\gamma},$$

or equivalently

$$\mathbf{s}(\boldsymbol{\theta}) = \left(\left(\mathbf{G}_R \otimes \dots \otimes \mathbf{G}_1\right) \cdot \left(\left(\mathbf{A}\left(\boldsymbol{\mu}^{(R)}\right)\right) \Diamond \dots \Diamond \left(\cdot \mathbf{A}\left(\boldsymbol{\mu}^{(1)}\right)\right)\right)\right) \cdot \boldsymbol{\gamma}.$$

Now it can be shown that the model $\mathbf{s}(\boldsymbol{\theta})$ has hidden rotational invariance structure if all sub-matrices $\mathbf{G}_r \in \mathbb{C}^{M_r \times N_r}$ have an appropriate size and rank.

At first let us suppose the observed system response has been over-sampled, i.e., $M_r \geq N_r, \forall r$ and the rank of $\mathbf{G}_r \in \mathbb{C}^{M_r \times N_r}$ is $\text{rank}(\mathbf{G}_r) = N_r$. Then the observed data can be transformed such that the transformed data have the multidimensional rotational invariance structure required for subarray smoothing and for the application of, e.g., multidimensional Unitary ESPRIT. The transformation matrix for every data dimension has to satisfy

$$\mathbf{P}_{G_r} \mathbf{G}_r = \mathbf{I}. \tag{5.112}$$

The Moore-Penrose pseudo inverse yields a solution in the least-squares sense to (5.112). That means we can calculate the projector $\mathbf{P}_{G_r}$ using the singular value decomposition of the system matrix $\mathbf{G}_r$, i.e., we compute first

$$\mathbf{G}_r = \begin{bmatrix} \mathbf{U}_{G_r} & \mathbf{U}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{G_r} \\ \mathbf{0} \end{bmatrix} \mathbf{V}_{G_r}^{\mathrm{H}}$$

and with this decomposition the linear transformation matrix can be calculated by

$$\mathbf{P}_{G_r} = \mathbf{V}_{G_r} \boldsymbol{\Sigma}_{G_r}^{-1} \mathbf{U}_{G_r}^{\mathrm{H}}. \tag{5.113}$$

It is easy to prove that this projector fulfils equation (5.112). However, if the observed system response is under-sampled, i.e., $M_r < N_r$ for some r, the singular value decomposition of the respective system matrix becomes

$$\mathbf{G}_r = \mathbf{U}_{G_r} \begin{bmatrix} \boldsymbol{\Sigma}_{G_r} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{G_r}^{\mathrm{H}} \\ \mathbf{V}_0^{\mathrm{H}} \end{bmatrix}. \tag{5.114}$$

The Moore-Penrose pseudo inverse is now

$$\mathbf{P}_{G_r} = \mathbf{V}_{G_r} \boldsymbol{\Sigma}_{G_r}^{-1} \mathbf{U}_{G_r}^{\mathrm{H}}. \tag{5.115}$$

Using equation (5.115), and (5.114) we yield for the product between the system matrix $\mathbf{G}_r$ and the related projector $\mathbf{P}_{G_r}$ the expression

$$\mathbf{P}_G \mathbf{G} = \mathbf{V}_G \mathbf{V}_G^{\mathrm{H}} = \mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^{\mathrm{H}} \neq \mathbf{I}$$

since

$$\mathbf{V}_G \mathbf{V}_G^{\mathrm{H}} + \mathbf{V}_0 \mathbf{V}_0^{\mathrm{H}} = \mathbf{I}.$$

Hence, the observed data only have hidden rotational invariance structure if the system response is over-sampled and all system matrices have full column rank. To summarise, if the measurement system matrix $\mathbf{G}$ allows the factorisation

$$\mathbf{G} = \mathbf{G}_{R_D} \otimes \cdots \otimes \mathbf{G}_1, \tag{5.116}$$

where each of the R_D sub matrices have full column rank, i.e.,

$$\mathrm{rank}(\mathbf{G}_r) = N_r \quad \forall r,$$

the channel sounding measurements can be transformed such that the transformed data have a multidimensional rotational invariance structure. Furthermore, the transformation matrix $\mathbf{P}_G$ is given by the Kronecker-product of the individual transformation matrices $\mathbf{P}_{G_r}$ as

$$\mathbf{P}_G = \mathbf{P}_{G_{R_D}} \otimes \cdots \otimes \mathbf{P}_{G_r}.$$

Altogether, the parameter estimation from data having hidden rotational invariance structure requires only the extension of the signal subspace estimation step by

$$\mathbf{x}' = \mathbf{P}_G \mathbf{x}. \tag{5.117}$$

The transformed data $\mathbf{x}'$ have the rotational invariance structure necessary for Unitary ES-PRIT or RARE application. In addition, multidimensional smoothing can be applied to the transformed data. Observe that the multiplication of the observed data with $\mathbf{P}_G$ changes the structure of the covariance matrix $\mathbf{R}_{nn}$ of the stochastic part of the observation. The new covariance matrix is

$$\mathbf{R}'_{nn} = \mathbf{P}_G \mathbf{R}_{nn} \mathbf{P}_G^{\mathrm{H}}.$$

If both, the covariance matrix $\mathbf{R}_{nn}$ and the measurement system matrix $\mathbf{G}$ can be factorized into sub matrices

$$\mathbf{R}_{nn} = \mathbf{R}_{nn,R_D} \otimes \cdots \otimes \mathbf{R}_{nn,1}$$

and

$$\mathbf{G} = \mathbf{G}_{R_D} \otimes \cdots \otimes \mathbf{G}_1,$$

the transformed covariance matrix $\mathbf{R}'_{nn}$ has Kronecker structure as well

$$\mathbf{R}'_{nn} = \mathbf{R}'_{nn,R_D} \otimes \cdots \otimes \mathbf{R}'_{nn,1},$$

with the sub matrices

$$\mathbf{R}'_{nn,r} = \mathbf{P}_{G,r} \mathbf{R}_{nn,r} \mathbf{P}_{G,r}^{\mathrm{H}} \quad \forall r. \tag{5.118}$$

This is an important fact in terms of computational complexity, since the observed data have often only a hidden rotational invariance structure in some but not all data dimensions. Consequently, only the data dimensions having hidden rotational invariance structure and the related covariance matrices must be transformed according to (5.117) and (5.118).

Data having such a hidden rotational invariance structure are for example the measurement data acquired with a circular uniform beam array. The resulting CUBA-ESPRIT algorithm is described in the next section.

5.3.12   Unitary ESPRIT for CUBA Configurations

It is often assumed that ESPRIT in combination with subarray smoothing is only applicable if uniform linear or rectangular arrays have been applied during measurements. As already discussed in the previous section, subarray smoothing and consequently ESPRIT is applicable if the observed data have hidden rotational invariance structure. A class of antenna arrays having hidden rotational invariance structure is the class of circular uniform beam arrays (CUBA) [114]. A circular uniform beam array (CUBA) consists of M antennas with identical beam patterns $b_0(\mu)$. The main beam directions of two adjacent elements are rotated by $\mu_0 = 2\pi/M$ with respect to each other as depicted in Figure 5-9.
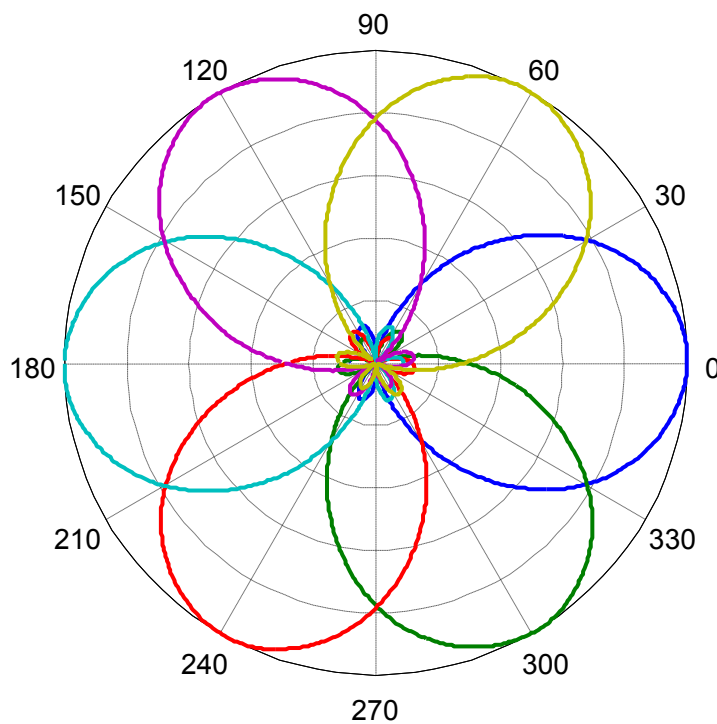
Figure 5-9:   Beam patterns of a 6-element CUBA.

Given a narrowband signal from a single source that impinges on the CUBA from the direction of arrival μ_p, the output signals of the M array elements can be described by means of the sampled beam pattern as

$$b_p(m \cdot \mu_0) = b_0(m \cdot \mu_0 - \mu_p) \qquad (5.119)$$

with $m = 0,1,\ldots,(M-1)$. Since the antenna beam pattern is periodic in 2π, $b_0(\mu) = b_0(\mu + k \cdot 2\pi)$, $k \in \mathbb{Z}$, the equivalent virtual aperture function of a single element can be calculated via the discrete Fourier transform (DFT) as

$$g_a(l \cdot s_0) = \sum_{m=0}^{M-1} b_0(m \cdot \mu_0) e^{-j2\pi\frac{ml}{M}}, \qquad (5.120)$$

where $s_0 = 1/(2\pi)$. Consequently, in the virtual aperture domain the single source response (5.119) equals $g_p(l \cdot s_0) = g_a(l \cdot s_0) e^{-jl\mu_p}$. If $s_p(n)$ denotes the signal received from source p at time n, the virtual aperture function of the array output at time n can be expressed as

$$g_l(n) = \sum_{p=1}^{P} s_p(n) g_a(l \cdot s_0) e^{-jl\mu_p}, \quad 0 \le l \le (M-1).$$

Therefore,

$$g_l(n) = g_a(l \cdot s_0) \sum_{p=1}^{P} s_p(n) e^{-jl\mu_p}, \quad 0 \le l \le (M-1). \tag{5.121}$$

The set of equations (5.121) can be simplified to

$$\mathbf{g}(n) = \mathbf{G}_a \mathbf{A} \mathbf{s}(n), \tag{5.122}$$

where the array steering matrix $\mathbf{A} = [\mathbf{a}(\mu_1) \ \mathbf{a}(\mu_2) \ \cdots \ \mathbf{a}(\mu_P)]$ contains the steering vectors of the P sources

$$\mathbf{a}(\mu_p) = \begin{bmatrix} 1 & e^{-j\mu_p} & \cdots & e^{-j\mu_p(M-1)} \end{bmatrix}^T,$$

the diagonal matrix

$$\mathbf{G}_a = \begin{bmatrix} g_a(0) & 0 & \cdots & 0 \\ 0 & g_a(s_0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & g_a((M-1)s_0) \end{bmatrix} \in \mathbb{C}^{M \times M}$$

represents the virtual aperture function of a single element, the signal vector at time n is formed as

$$\mathbf{s}(n) = \begin{bmatrix} s_1(n) & s_2(n) & \dots & s_P(n) \end{bmatrix}^T \in \mathbb{C}^P,$$

and the virtual aperture output vector at time n is given by

$$\mathbf{g}(n) = \begin{bmatrix} g_0(n) & g_1(n) & \dots & g_{M-1}(n) \end{bmatrix}^T.$$

Since the array output signal depends on samples of the beam pattern $b_0(m \cdot \mu_0)$, the Nyquist sampling criterion has to be considered, i.e., the virtual aperture function $g_a(l \cdot s_0)$ contains the required rotational invariance structure only if the Fourier-transform of $b_0(\mu)$ has finite support in $L \cdot s_0$ and $L \le M$. The virtual aperture output vector $\mathbf{x}(n)$ is calculated from the received data $\mathbf{x}_R(n) = \mathbf{B} \cdot \mathbf{s}(n)$ via a DFT, where

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}(\mu_1) & \mathbf{b}(\mu_2) & \dots & \mathbf{b}(\mu_p) \end{bmatrix}^T$$

is a matrix that contains the P beam vectors

$$\mathbf{b}(\mu_p) = \begin{bmatrix} b_p(0) & b_p(\mu_0) & \dots & b_p((M-1) \cdot \mu_p) \end{bmatrix}^T.$$

The transformation matrix $\mathbf{P}_{G_a}$ that transforms $\mathbf{x}_R(n)$ into the virtual aperture space $\mathbf{x}(n) = \mathbf{P}_{G_a} \cdot \mathbf{x}_R(n)$ is defined as:

$$\mathbf{P}_{G_a} = \mathbf{G}_L^{-1} \mathbf{J}_L \mathbf{F}_M. \tag{5.123}$$

Here,

$$\mathbf{F}_M = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & e^{-j2\pi\frac{(M-1)^2}{M}} \end{bmatrix} \in \mathbb{C}^{M \times M}$$

is the $M \times M$ DFT-matrix and

$$\mathbf{J}_L = \left[\begin{bmatrix} \mathbf{0}_{(\frac{L-1}{2}) \times (\frac{L+1}{2})} \\ \mathbf{I}_{(\frac{L+1}{2}) \times (\frac{L+1}{2})} \end{bmatrix} \quad \mathbf{0}_{L \times (M-L)} \quad \begin{bmatrix} \mathbf{I}_{(\frac{L-1}{2}) \times (\frac{L-1}{2})} \\ \mathbf{0}_{(\frac{L+1}{2}) \times (\frac{L-1}{2})} \end{bmatrix} \right] \in \mathbb{R}^{L \times M}$$

denotes the selection matrix that chooses only the useful part of $g_a(l \cdot s_0)$ and ensures a good condition of $\mathbf{G}_L = \mathbf{J}_L \mathbf{G}_0 \mathbf{J}_L^T$, and thus also of its inverse $\mathbf{G}_L^{-1}$. In this case, the modified virtual aperture space $\mathbf{x}(n)$ shows the required rotational invariance structure. Therefore, the signal subspace can be obtained from $\mathbf{x}(n)$ as in other ESPRIT-type algorithms. Since $\mathbf{A}$ is left $\mathbf{\Pi}$-real, Unitary ESPRIT can also be used. Even the usual subarray smoothing techniques can be applied. Hence, the algorithm can be applied to channel parameter estimation from channel sounding measurements. For notational simplicity, Table 5-7 summarises Unitary ESPRIT for CUBA configurations without the subarray smoothing technique.

Obviously, the ideal beam pattern of a CUBA for the described algorithm is a sinc-function, since a beam pattern $b_0(\mu) = \text{sinc}(L/2 \cdot \mu)$ provides a virtual aperture function $g_a(l \cdot s_0) = \text{rect}(l/L)$ with strictly limited support. In the context of Nyquist sampling, this is the ideal anti-aliasing lowpass filter for an angular sampling rate of $1/(L \cdot s_0)$.

Example 5-5: The following example is based on measured beam patterns of an $M = 6$ element CUBA. The left picture of Figure 5-10 shows the uncalibrated and the calibrated beam pattern $b_0(\mu)$ of a single antenna element. The transformation matrix $\mathbf{P}_{G_a}$ has been estimated using the calibration algorithm described in Chapter 7. Then the calibrated beam pattern has been obtained using the inverse Fourier-transform.

To illustrate the DoA estimation performance of Unitary ESPRIT for CUBA configurations in the noise free case, a simple scenario with $P = 3$ coherent sources was simulated. With 2 sources at fixed angles $(-90°, +90°)$, and simultaneously rotating one source and the array from $-75°$ to $+75°$, 151 scenarios have been created. Figure 5-11 shows the estimated angles. The parameters have been estimated from a single snapshot.

Parameter estimation results using the CUBA-ESPRIT have been published in [1], [61], [96], and [99].
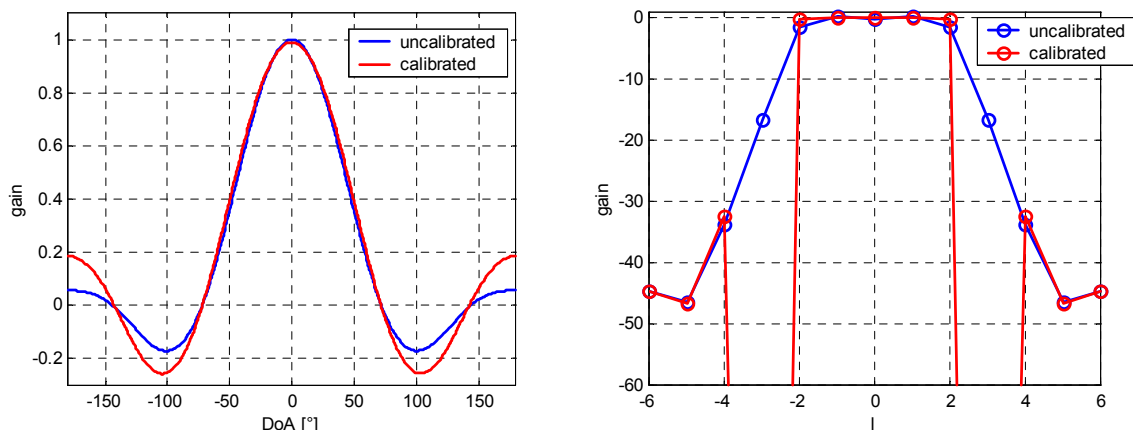
Figure 5-10: CUBA beam pattern (left) and corresponding virtual aperture function (right).
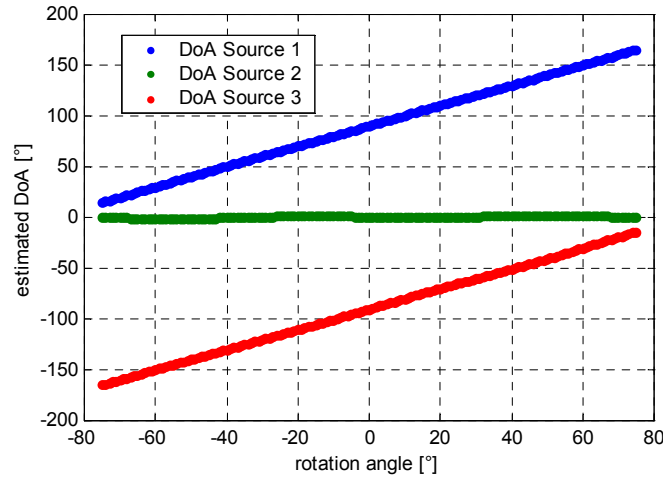
Figure 5-11: DoA estimation results for three coherent sources obtained by Unitary ESPRIT for CUBA configurations.

Table 5-7: Summary of Unitary ESPRIT for CUBA configurations with white noise.

1) Signal Subspace Estimation: Let $\mathbf{x}(n) = \mathbf{P}_{G_a} \mathbf{b}(n)$ and determine the matrix $\mathbf{X} \in \mathbb{C}^{L \times N}$. Then, compute the signal subspace $\mathbf{E}_s \in \mathbb{R}^{L \times P}$ as the P dominant singular vectors of
$$\mathcal{T}_{CR}(\mathbf{X}) \in \mathbb{R}^{L \times 2N}.$$

2) Solution to the Invariance Equations: Then solve
$$\mathbf{K}_1 \mathbf{E}_S \mathbf{\Psi} \approx \mathbf{K}_2 \mathbf{E}_S$$
by means of LS, TLS, or SLS.

3) Frequency Estimation: Calculate the eigenvectors of the resulting real-valued solution
$$\mathbf{\Psi} = \mathbf{T} \mathbf{\Omega} \mathbf{T}^{-1} \text{ with } \mathbf{\Omega} = \mathrm{diag}\{\varpi_p\}_{p=1}^{P}.$$

4) Direction estimation: Calculate the estimates $\hat{\mu}_p = 2\arctan(\varpi_p)$, they are already the estimated directions.

6 Estimation of DMC Parameters

Based on the model (2.63) we derive in this section an estimator $\hat{\boldsymbol{\theta}}_{dan}$ for the parameters of the dense multipath components (DMC) $\boldsymbol{\theta}_{dan}$. As already stated in Section 2.5 we combine the contributions of the measurement noise and the dense multipath components in one model since both are complex circular normal distributed processes. The focus of this chapter is only the estimation of the parameters $\boldsymbol{\theta}_{dan}$. Hence, we assume in the subsequent sections of this chapter that the radio channel observation contains DMC and noise only, i.e.,

$$\mathbf{x} = \mathbf{n}_{dan} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dan})).$$

The joint estimation of all channel parameter $\boldsymbol{\theta}_{chn}$ is treated at the end of this chapter.

6.1 Maximum Likelihood Estimation of DMC Parameters

Since the parametric covariance matrix (2.67), (2.68) is a nonlinear function in some of its parameters, a closed form solution for the maximization problem

$$\hat{\boldsymbol{\theta}}_{dan} = \arg\max_{\boldsymbol{\theta}_{dan}} \mathcal{L}(\boldsymbol{\theta}_{dan}|\mathbf{X}) \tag{6.1}$$

is not available. Hence, we will discuss, in the following section, global and local search strategies to find the parameter vector $\hat{\boldsymbol{\theta}}_{dan}$ maximizing (6.1). Since we usually measure not only one but rather sequences of channel observations over time and since the parameters of the dense multipath components are only slowly time varying we can track the parameters $\boldsymbol{\theta}_{dan}$ from observation to observation. That means that we use the estimated parameters of the previous observation as an initial solution for the parameters of the current observation. Thus, a global search algorithm is needed only once, namely to estimate raw parameters $\boldsymbol{\theta}_{dan}$ for the very first observation of the whole measurement sequence. Consequently, we will focus the discussion on the computational complexity as well as the statistical efficiency of local search strategies. Let us for the time being drop the subscript $_{dan}$ from the parameter vector $\boldsymbol{\theta}_{dan}$ for notational convenience.

6.1.1 Local Search Strategies

Having derived algebraic expressions for the Fisher information matrix and the score function one can say the local search for the maximum of the log-likelihood function has been solved already. Namely, application of the iterative maximum likelihood algorithm requires the knowledge of (4.89) and (4.91) only. The drawback of this approach is its high computational complexity. Therefore, we discuss algorithms with a reduced computational complexity and low storage requirements in the following sections. Nevertheless, we will start with the direct approach to compare the complexity of the other algorithms later on.

131

6.1.2   Direct Approach

Since the score function is the gradient of the log-likelihood function and the Fisher information matrix[12] is an approximation of the related Hessian we can for example use the Gauß-Newton algorithm to iteratively maximize (6.1) using the iteration step

$$\hat{\boldsymbol{\theta}}^{\{i+1\}} = \hat{\boldsymbol{\theta}}^{\{i\}} + \lambda \cdot \mathcal{I}^{-1}\left(\hat{\boldsymbol{\theta}}^{\{i\}}\right) \cdot \mathbf{q}\left(\mathbf{X}, \hat{\boldsymbol{\theta}}^{\{i\}}\right). \tag{6.2}$$

The resulting algorithm is a so-called iterative maximum likelihood algorithm. To find a suitable step size λ, one can use a line search strategy or a trust region algorithm to ensure strict maximization of (6.1), e.g., the Levenberg-Marquardt algorithm. The disadvantage of this local optimisation strategy is the high memory and computation effort for a single iteration. For every iteration we have to invert the Toeplitz matrix $\mathbf{R}\left(\hat{\boldsymbol{\theta}}^{\{i\}}\right)$, if we use the Levinson algorithm this requires $O(M^2)$ operations. To calculate the matrix (4.88) we need the product between this inverse and the Toeplitz matrices (4.93) - (4.96), an efficient algorithm for the product between the inverse of a Toeplitz matrix and an arbitrary vector requires $O(M \log(M))$ operations using fast Fourier-Transforms (FFT). Hence, the calculation of (4.88) requires $O(L \cdot M^2 \log(M))$ operations. The computation of the score function and the *Fisher Information matrix* requires algorithms with a complexity of $O(M^2 \log(M)) + O(M^2 L)$ and $O(M^2 L^2)$, respectively. Therefore, the complexity of one iteration is of order $O(L \cdot M^2 \log(M))$, under the assumption $L < \log(M)$. Additionally, we have to store the complete estimated non-parametric covariance matrix $\hat{\mathbf{R}}$ (4.83). Altogether, the complexity of an iterative maximum likelihood estimator based on the direct implementation of (6.2) is too high. In the next sections we derive an parameter estimator, which is computationally less expensive.

Table 6-1: Iterative optimisation of the parameters $\boldsymbol{\theta}_{dmc}$ using the direct approach
(Gauß-Newton algorithm)

Input data: Data matrix $\mathbf{X}$, initial solution $\boldsymbol{\theta}^{\{0\}}$

Preprocessing: Compute the estimate of the non-parametric covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N} \mathbf{X}\mathbf{X}^{\mathrm{H}}$$

1) Compute the first order derivatives to the parameters $\boldsymbol{\theta}$ of $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using equations (4.93) - (4.96).

2) Compute the Jacobian $\mathbf{D}\left(\boldsymbol{\theta}^{\{i\}}\right)$.

3) Compute $\Delta\boldsymbol{\theta}^{\{i\}} = \arg\min_{\Delta\boldsymbol{\theta}} \left\| \mathbf{D}\left(\boldsymbol{\theta}^{\{i\}}\right) \cdot \Delta\boldsymbol{\theta} - \mathrm{vec}\left\{ \mathbf{R}^{-1}\left(\boldsymbol{\theta}^{\{i\}}\right)\hat{\mathbf{R}} - \mathbf{I} \right\} \right\|_F^2$. Set $\lambda^{\{i\}} = 1$

4) Compute the update $\boldsymbol{\theta}^{\{i+1\}} = \boldsymbol{\theta}^{\{i\}} + \lambda^{\{i\}} \cdot \Delta\boldsymbol{\theta}^{\{i\}}$.

5) Check strict maximization $\mathcal{L}\left(\mathbf{X}|\boldsymbol{\theta}^{\{i+1\}}\right) > \mathcal{L}\left(\mathbf{X}|\boldsymbol{\theta}^{\{i\}}\right)$; yes: go to 1. , no: reduce $\lambda^{\{i\}}$ go to 4.

6) Check convergence; not converged: go to 1.

---

[12] Usage of the stochastic Fisher matrix leads to a Newton-Raphson procedure.

6.1.3 Averaged Covariance Matrix

We can reduce the storage requirements by replacing the non-parametric covariance matrix $\hat{\mathbf{R}}$ with $\text{toep}(\hat{\mathbf{r}}, \hat{\mathbf{r}}^H)$ using the coefficients[13]

∎
$$\hat{r}_l = \frac{1}{M-l+1} \sum_{i=1}^{M-l+1} \{\hat{\mathbf{R}}\}_{i+l-1,i} \quad (l=1,\ldots,M). \tag{6.3}$$

This approach reduces the memory requirements roughly by a factor $\frac{1}{2}M$. The elements of $\hat{\mathbf{r}}$ are the mean values of the diagonals of the matrix $\hat{\mathbf{R}}$, and have an expected value of $E\{\hat{\mathbf{r}}\} = \kappa(\theta)$. We can anticipate that an algorithm using $\text{toep}(\hat{\mathbf{r}}, \hat{\mathbf{r}}^H)$ has the same performance statistically as an algorithm using $\hat{\mathbf{R}}$ directly, since we average in (6.3) only values with the same distribution. Nevertheless, we have not reduced the computational complexity of the iterations (6.2) so far.

Table 6-2: Iterative optimisation of the parameters θ_{dmc} using the direct approach and the averaged covariance matrix (Gauß-Newton algorithm)

Input: data matrix $\mathbf{X}$, initial solution $\theta^{\{0\}}$

Preprocessing: Compute the coefficients $\hat{\mathbf{r}}$ (6.3) of the non-parametric covariance matrix $\text{toep}(\hat{\mathbf{r}}, \hat{\mathbf{r}}^H)$.

1) Compute the first order derivatives to the parameters $\theta^{\{i\}}$ of $\kappa(\theta^{\{i\}})$ using equations (4.93) - (4.96).

2) Compute the Jacobian $\mathbf{D}(\theta^{\{i\}})$.

3) Compute $\Delta\theta^{\{i\}} = \arg\min_{\Delta\theta} \left\| \mathbf{D}(\theta^{\{i\}}) \cdot \Delta\theta - \text{vec}\{\mathbf{R}^{-1}(\theta^{\{i\}})\text{toep}\{\hat{\mathbf{r}}, \hat{\mathbf{r}}^H\} - \mathbf{I}\} \right\|_F^2$. Set $\lambda^{\{i\}} = 1$

4) Compute the update $\theta^{\{i+1\}} = \theta^{\{i\}} + \lambda^{\{i\}} \cdot \Delta\theta^{\{i\}}$.

5) Check strict maximization $\mathcal{L}(\mathbf{X}|\theta^{\{i+1\}}) > \mathcal{L}(\mathbf{X}|\theta^{\{i\}})$; yes: go to 6. , no: reduce $\lambda^{\{i\}}$ go to 4.

6) Check convergence; not converged: go to 1.

6.1.4 Approximation of the Covariance Matrix with a Diagonal Matrix

An approach for the development of an algorithm with a lower computational complexity than the direct solution is based on the observation, that the inverse Fourier-transform of the covariance matrix $\mathbf{R}_f(\theta)$ is a diagonal dominant matrix. Hence, we approximate the covariance matrix in the time-delay domain using its main diagonal elements

$$\text{diag}\{\mathbf{F}^H \cdot \mathbf{R}_f(\theta) \cdot \mathbf{F}\} = \text{diag}\{\mathbf{R}_\tau(\theta)\} = \beta(\theta), \tag{6.4}$$

where $\text{diag}\{\mathbf{A}\}$ denotes an operator selecting the main diagonal elements of $\mathbf{A}$.
The cost function to maximize follows from the log-likelihood function (4.79) and is

---

[13] $\{\mathbf{A}\}_{l,k}$ denotes the element in row l and column k of $\mathbf{A}$.

$$C(\mathbf{X},\boldsymbol{\theta}) = -N\cdot\ln(\det(\mathrm{diag}\{\boldsymbol{\beta}(\boldsymbol{\theta})\})) - \sum_{i=1}^{N}\left(\mathbf{x}_i^{\mathrm{H}}\cdot\mathbf{F}\cdot\mathrm{diag}\{\boldsymbol{\beta}(\boldsymbol{\theta})\}^{-1}\cdot\mathbf{F}^{\mathrm{H}}\cdot\mathbf{x}_i\right). \tag{6.5}$$

For notational convenience we introduce the vector

$$\mathbf{y} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{F}^{\mathrm{H}}\mathbf{x}_i\right)\circ\left(\mathbf{F}^{\mathrm{H}}\mathbf{x}_i\right)^{*}, \tag{6.6}$$

containing the data for the estimator to be constructed. The vector $\mathbf{y}$ is an estimate of the main diagonal elements of the covariance matrix $\mathbf{R}_f(\boldsymbol{\theta})$ in the time-delay domain (the (PDP) of the dense multipath components). Using (6.6) in the cost function (6.5) yields the compact expression

$$C(\mathbf{y},\boldsymbol{\theta}) = -N\cdot\sum_{m=1}^{M}\left(\ln(\beta_m(\boldsymbol{\theta})) + \frac{y_m}{\beta_m(\boldsymbol{\theta})}\right). \tag{6.7}$$

The parameter vector $\hat{\boldsymbol{\theta}}$, maximizing (6.7) is an estimate of the parameter vector describing the covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ and the spectra $\boldsymbol{\kappa}(\boldsymbol{\theta})$. The gradient of the cost function (6.7) is given by the partial derivatives with respect to the parameters θ_i

$$\frac{\partial}{\partial\theta_i}C(\mathbf{y},\boldsymbol{\theta}) = -N\cdot\sum_{m=1}^{M}\left(\frac{1}{\beta_m(\boldsymbol{\theta})}\cdot\left(\frac{\partial}{\partial\theta_i}\beta_m(\boldsymbol{\theta})\right) - \frac{y_m}{\beta_m^2(\boldsymbol{\theta})}\left(\frac{\partial}{\partial\theta_i}\beta_m(\boldsymbol{\theta})\right)\right). \tag{6.8}$$

Equation (6.8) can be rewritten to

$$\frac{\partial}{\partial\theta_i}C(\mathbf{y},\boldsymbol{\theta}) = N\cdot\sum_{m=1}^{M}\left(\left(\frac{y_m}{\beta_m(\boldsymbol{\theta})} - 1\right)\cdot\frac{1}{\beta_m(\boldsymbol{\theta})}\cdot\left(\frac{\partial}{\partial\theta_i}\beta_m(\boldsymbol{\theta})\right)\right). \tag{6.9}$$

Hence the parameters to estimate $\hat{\boldsymbol{\theta}}$ are a solution of

$$\sum_{m=1}^{M}\left(\left(\frac{y_m}{\beta_m(\hat{\boldsymbol{\theta}})} - 1\right)\cdot\frac{1}{\beta_m(\hat{\boldsymbol{\theta}})}\cdot\left(\frac{\partial}{\partial\theta_i}\beta_m(\hat{\boldsymbol{\theta}})\right)\right) = 0 \quad \forall\theta_i.$$

The term

$$\left(\frac{y_m}{\beta_m(\boldsymbol{\theta})} - 1\right) \tag{6.10}$$

can be interpreted as the error function to minimize, and $\beta_m^{-1}(\boldsymbol{\theta})$ as an error weighting. One should observe that for an infinite number of independent observations

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{F}^{\mathrm{H}}\mathbf{x}_i\right)\circ\left(\mathbf{F}^{\mathrm{H}}\mathbf{x}_i\right)^{*} = \mathrm{E}\{\mathbf{y}\} = \boldsymbol{\beta}(\boldsymbol{\theta})$$

holds. Consequently the error function (6.10) tends to zero if $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ and $N\to\infty$. Hence, the estimator is asymptotically unbiased and consistent. For the Gauß-Newton procedure, we need an approximation of the Hessian, i.e., the second gradient of the cost function, which is negative definite. The second gradient of the cost function (6.7), i.e., the second order partial derivatives with respect to the parameter pairs θ_i,θ_k, is

$$\frac{\partial^2}{\partial\theta_i\partial\theta_k}C(\mathbf{y},\boldsymbol{\theta}) = N\cdot\sum_{m=1}^{M_f}\left(\left(1 - 2\frac{y_m}{\beta_m(\boldsymbol{\theta})}\right)\frac{\left(\frac{\partial}{\partial\theta_i}\beta_m(\boldsymbol{\theta})\right)\cdot\left(\frac{\partial}{\partial\theta_k}\beta_m(\boldsymbol{\theta})\right)}{\beta_m^2(\boldsymbol{\theta})} + \left(\frac{y_m}{\beta_m(\boldsymbol{\theta})} - 1\right)\frac{\frac{\partial^2}{\partial\theta_i\partial\theta_k}\beta_m(\boldsymbol{\theta})}{\beta_m(\boldsymbol{\theta})}\right). \tag{6.11}$$

If we use the second gradient itself in the iteration to adjust the metric and the direction of the gradient we end up with a Newton-Raphson algorithm. It is known that the Newton-Raphson algorithm is not always globally convergent since the second gradient, the Hessian, may be positive [22]. In other words the calculated step direction will not point to the maximum. Thus we use an approximation of the Hessian which is at least negative definite. To this end let us analyze the limit of equation (6.11) if $N \rightarrow \infty$. The two terms in the Hessian related to our data $\mathbf{y}$ have the following limit values

$$\lim_{N \rightarrow \infty} \left\{ 1 - 2 \frac{y_m}{\beta_m(\boldsymbol{\theta})} \right\} = -1$$

and

$$\lim_{N \rightarrow \infty} \left\{ \frac{y_m}{\beta_m(\boldsymbol{\theta})} - 1 \right\} = 0 \,.$$

Hence an approximation of the Hessian being always negative definite is given by

$$\frac{\partial^2}{\partial \theta_i \partial \theta_k} C(\mathbf{y}, \boldsymbol{\theta}) \approx -N \cdot \sum_{m=1}^{M_f} \left(\frac{\left(\frac{\partial}{\partial \theta_i} \beta_m(\boldsymbol{\theta}) \right) \cdot \left(\frac{\partial}{\partial \theta_k} \beta_m(\boldsymbol{\theta}) \right)}{\beta_m^2(\boldsymbol{\theta})} \right). \tag{6.12}$$

To calculate the gradient (6.9) and the approximation of the Hessian (6.12) numerically, we need an efficient way to compute $\boldsymbol{\beta}(\boldsymbol{\theta})$ and the related partial derivatives. One can use the direct approach and implement the definitions (6.4) and (2.67). But this approach is numerically unattractive, since we compute many more values than needed, namely all off diagonal elements of $\mathbf{F}^H \cdot \mathbf{R}(\boldsymbol{\theta}) \cdot \mathbf{F}$. A computational efficient way to map $\boldsymbol{\kappa}(\boldsymbol{\theta})$ to $\boldsymbol{\beta}(\boldsymbol{\theta})$ is given by the relation

$$\text{diag}\{ \mathbf{F}^H \cdot \mathbf{R}(\boldsymbol{\theta}) \cdot \mathbf{F} \} = \frac{1}{\sqrt{M}} \cdot \mathbf{F}^H \left(\mathbf{W}_1 \cdot \boldsymbol{\kappa}(\boldsymbol{\theta}) + \mathbf{W}_2 \cdot \boldsymbol{\kappa}^*(\boldsymbol{\theta}) \right), \tag{6.13}$$

using the weighting matrices

$$\mathbf{W}_1 = \text{diag}\{ [M \quad M-1 \quad \cdots \quad 1] \}$$

and

$$\mathbf{W}_2 = \begin{bmatrix} 0 & & \cdots & & 0 \\ & & & \iddots & 1 \\ \vdots & & 0 & 2 & 0 \\ & \iddots & \iddots & \iddots & \vdots \\ 0 & M-1 & 0 & \cdots & 0 \end{bmatrix}.$$

Proof:
Let us suppose $\mathbf{a} \in \mathbb{C}^{M \times 1}$ is an arbitrary vector and $\mathbf{b} = \mathbf{F} \cdot \mathbf{a}, \mathbf{b} \in \mathbb{C}^{M \times 1}$ is its discrete Fourier-transform. Furthermore, we define the matrices $\mathbf{X} = \mathbf{a} \cdot \mathbf{a}^H$ and $\mathbf{Y} = \mathbf{b} \cdot \mathbf{b}^H$. The main diagonal elements of $\mathbf{X}$ can be expressed using the cyclic convolution $\oplus$ between the vectors $\mathbf{b}$ and $\mathbf{\Pi b}^*$ [115]

$$\mathbf{c} = \frac{1}{\sqrt{M}} \mathbf{b} \oplus \left(\mathbf{\Pi} \cdot \mathbf{b}^* \right) = \mathbf{F} \cdot \left(\mathbf{a} \circ \mathbf{a}^* \right) = \mathbf{F} \cdot \text{diag}\{ \mathbf{X} \}, \tag{6.14}$$

where $\mathbf{\Pi}$ is the reflection matrix (see Appendix C). The cyclic convolution between $\mathbf{b}$ and $\mathbf{b}^*$ can be calculated using the vector $\mathbf{z}$ having elements

$$z_l = \sum_{i=1}^{M-l+1} b_{i+l-1} \cdot b_i^*, \quad (l=1,\ldots,M).$$ (6.15)

The elements of $\mathbf{z}$ are also the sums of the diagonals of $\mathbf{Y}$

$$z_l = \sum_{i=1}^{M-l+1} b_{i+l-1} \cdot b_i^* = \sum_{i=1}^{M-l+1} y_{i+l-1,i}, \quad (l=1,\ldots,M),$$

and are related to $\mathbf{c}$ as follows

$$\mathbf{c} = \mathbf{z} + \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi} \end{bmatrix} \mathbf{z}^*.$$

Hence, the main diagonal elements of $\mathbf{X}$ and the total sums along the diagonals of $\mathbf{Y}$ are related in the following way

$$\mathrm{diag}\{\mathbf{X}\} = \frac{1}{\sqrt{M}} \mathbf{F}^H \cdot \left(\mathbf{z} + \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{\Pi} \end{bmatrix} \mathbf{z}^* \right).$$ (6.16)

Since $\mathbf{b} = \mathbf{F} \cdot \mathbf{a}$ is a linear transformation (6.16) is also valid for matrices

$$\mathrm{diag}\{\mathbf{A} \cdot \mathbf{A}^H\} = \mathrm{diag}\{\mathbf{X}\} = \frac{1}{\sqrt{M}} \mathbf{F}^H \cdot \left(\mathbf{z} + \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{\Pi} \end{bmatrix} \cdot \mathbf{z}^* \right),$$

where

$$\mathbf{Y} = \mathbf{B} \cdot \mathbf{B}^H = \mathbf{F} \cdot \mathbf{A} \cdot \mathbf{A}^H \cdot \mathbf{F}^H = \mathbf{F} \cdot \mathbf{X} \cdot \mathbf{F}^H.$$

If $\mathbf{Y}$ has Toeplitz structure $\mathbf{Y} = \mathrm{toep}\{\mathbf{y}, \mathbf{y}^H\}$ the first column of $\mathbf{Y}$, i.e., $\mathbf{y}$, is related to the main diagonal elements of $\mathbf{X} = \mathbf{F}^H \cdot \mathbf{Y} \cdot \mathbf{F}$ as

$$\mathrm{diag}\{\mathbf{F}^H \cdot \mathbf{Y} \cdot \mathbf{F}\} = \frac{1}{\sqrt{M}} \cdot \mathbf{F}^H \cdot \left(\mathbf{W}_1 \cdot \mathbf{y} + \mathbf{W}_2 \cdot \mathbf{y}^* \right)$$ (6.17)

since

$$z_l = \sum_{i=1}^{M-l+1} y_{i+l-1,i} = (M-l+1) \cdot y_i \quad (l=1,\ldots,M).$$

Based on (6.17) we define the transformation $\mathbf{b} = \mathcal{T}_{\mathcal{D}}\{\mathbf{a}\}$

■ $$\mathbf{b} = \frac{1}{\sqrt{M}} \cdot \mathbf{F}^H \cdot \left(\mathbf{W}_1 \cdot \mathbf{a} + \mathbf{W}_2 \cdot \mathbf{a}^* \right) = \mathcal{T}_{\mathcal{D}}\{\mathbf{a}\}.$$ (6.18)

Since $\boldsymbol{\beta}(\boldsymbol{\theta}) = \mathcal{T}_{\mathcal{D}}\{\boldsymbol{\kappa}(\boldsymbol{\theta})\}$ is a linear transformation of $\boldsymbol{\kappa}(\boldsymbol{\theta})$ to $\boldsymbol{\beta}(\boldsymbol{\theta})$, the first order derivatives of $\boldsymbol{\beta}(\boldsymbol{\theta})$ can be calculated using (4.93)-(4.96) in

$$\frac{\partial}{\partial \theta_i} \boldsymbol{\beta}(\boldsymbol{\theta}) = \mathcal{T}_{\mathcal{D}} \left\{ \frac{\partial}{\partial \theta_i} \boldsymbol{\kappa}(\boldsymbol{\theta}) \right\}.$$ (6.19)

Furthermore, the parameter estimates $\hat{\boldsymbol{\theta}}$ are a solution of

$$\sum_{m=1}^{M}\left(\left(\frac{y_m}{\mathcal{T}_{\mathcal{D}}\{\boldsymbol{\kappa}(\hat{\boldsymbol{\theta}})\}_m}-1\right)\cdot\frac{1}{\mathcal{T}_{\mathcal{D}}\{\boldsymbol{\kappa}(\hat{\boldsymbol{\theta}})\}_m}\cdot\mathcal{T}_{\mathcal{D}}\left\{\frac{\partial}{\partial\theta_i}\boldsymbol{\kappa}(\hat{\boldsymbol{\theta}})\right\}_m\right)=0,\quad\forall\theta_i.$$

In terms of numerical complexity an algorithm, using only the main diagonal elements of the covariance matrix $\mathbf{F}^H\cdot\mathbf{R}\cdot\mathbf{F}$, is very attractive. But equation (6.18) reveals also the drawback of the approximation $\mathrm{diag}\{\boldsymbol{\beta}(\boldsymbol{\theta})\}\approx\mathbf{F}^H\cdot\mathbf{R}(\boldsymbol{\theta})\cdot\mathbf{F}$, it leads to a superposition $\mathbf{W}_1\,\mathbf{a}+\mathbf{W}_2\,\mathbf{a}^*$ of values carrying different information about the parameters $\boldsymbol{\theta}$, i.e., with different distributions. The reason is effectively the relation between the Schur-product and the cyclic convolution in (6.14). Hence, in the next section we derive an improved algorithm, avoiding this overlap.

Table 6-3: Iterative optimisation of the parameters $\boldsymbol{\theta}_{dmc}$ using the approximation $\mathrm{diag}\{\mathbf{F}^H\cdot\mathbf{R}(\boldsymbol{\theta})\cdot\mathbf{F}\}=\boldsymbol{\beta}(\boldsymbol{\theta})$ (Gauß-Newton algorithm)

Input: data matrix $\mathbf{X}$, initial solution $\boldsymbol{\theta}^{\{0\}}$

Preprocessing: Compute the PDP $\mathbf{y}$ using equation (6.6) from $\mathbf{X}$.

1) Compute the first order derivatives to the parameters $\boldsymbol{\theta}^{\{i\}}$ of $\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})$ using equations (4.93) - (4.96).

2) Compute $\boldsymbol{\beta}(\boldsymbol{\theta}^{\{i\}})=\mathcal{T}_{\mathcal{D}}(\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}}))$ and

$$\mathbf{D}(\boldsymbol{\theta}^{\{i\}})=\mathrm{diag}\{\boldsymbol{\beta}(\boldsymbol{\theta}^{\{i\}})\}^{-1}\left[\frac{\partial}{\partial\theta_1}\boldsymbol{\beta}(\boldsymbol{\theta})\cdots\frac{\partial}{\partial\theta_4}\boldsymbol{\beta}(\boldsymbol{\theta})\right].$$

3) Compute the error $\boldsymbol{\varepsilon}^{\{i\}}=\mathrm{diag}\{\boldsymbol{\beta}(\boldsymbol{\theta}^{\{i\}})\}^{-1}\mathbf{y}-1$.

4) Compute the solution to $\Delta\boldsymbol{\theta}^{\{i\}}=\arg\min_{\Delta\boldsymbol{\theta}}\|\mathbf{D}(\boldsymbol{\theta}^{\{i\}})\cdot\Delta\boldsymbol{\theta}-\boldsymbol{\varepsilon}^{\{i\}}\|_F^2$ and set $\lambda^{\{i\}}=1$.

5) Compute the update $\boldsymbol{\theta}^{\{i+1\}}=\boldsymbol{\theta}^{\{i\}}+\lambda^{\{i\}}\cdot\Delta\boldsymbol{\theta}^{\{i\}}$.

6) Check strict maximization $\mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i+1\}})>\mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i\}})$; yes: go to 7. , no: reduce $\lambda^{\{i\}}$ go to 5.

7) Check convergence; not converged: go to 1.

6.1.5 A Numerically Efficient Algorithm for the Estimation of DMC Parameters

Let us consider a circulant matrix $\mathbf{C}=\mathrm{circ}\{\mathbf{c}\}$[14], it is completely determined by its first column $\mathbf{c}\in\mathbb{C}^{M\times1}$, each column of $\mathbf{C}$ is equal to the previous column rotated downwards by one element. Hence, the circulant matrix $\mathbf{C}$ contains every element of its first column M times. Let us embed the smoothed covariance matrix $\hat{\mathbf{R}}_s=\mathrm{toep}(\hat{\mathbf{r}},\hat{\mathbf{r}}^H)$ in a circulant matrix $\hat{\mathbf{R}}_c$ in such a way, that all the information in $\hat{\mathbf{r}}$ is as often represented in the smoothed covariance matrix itself

$$\hat{\mathbf{R}}_c=\mathrm{circ}\left(\frac{1}{2M-1}\begin{bmatrix}\mathbf{W}_1\cdot\hat{\mathbf{r}}\\\mathbf{W}_3\cdot\hat{\mathbf{r}}^*\end{bmatrix}\right),\tag{6.20}$$

---

[14] The function $\mathbf{C}=\mathrm{circ}\{\mathbf{c}\}$ maps the vector $\mathbf{c}$ to the circulant matrix $\mathbf{C}$, where the first column of $\mathbf{C}$ equals $\mathbf{c}$.

where $\mathbf{W}_3$ is the weighting matrix

$$\mathbf{W}_3 = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & M-1 & 0 & \cdots & 0 \end{bmatrix}.$$

Now we recall that the eigenvectors of an arbitrary circulant matrix are given by the DFT-matrix. In other words a circulant matrix is diagonalized by the DFT-matrix $\mathbf{F}$

$$\mathrm{diag}\{\hat{\mathbf{y}}_c\} = \mathbf{F}^{\mathrm{H}} \cdot \hat{\mathbf{R}}_c \cdot \mathbf{F}.$$

The eigenvalues $\hat{\mathbf{y}}_c$ are related to the first column of the circulant matrix $\hat{\mathbf{R}}_c$ by

$$\hat{\mathbf{y}}_c = \frac{1}{\sqrt{2M-1}} \cdot \mathbf{F}^{\mathrm{H}} \begin{bmatrix} \mathbf{W}_1 \cdot \hat{\mathbf{r}} \\ \mathbf{W}_3 \cdot \hat{\mathbf{r}}^* \end{bmatrix}.$$

So we define the transformation $\mathbf{b} = \mathcal{T}_c\{\mathbf{a}\}$, mapping vector $\mathbf{a}$ to vector $\mathbf{b}$, as

∎ $$\mathbf{b} = \frac{1}{\sqrt{2M-1}} \cdot \mathbf{F}^{\mathrm{H}} \begin{bmatrix} \mathbf{W}_1 \cdot \mathbf{a} \\ \mathbf{W}_3 \cdot \mathbf{a}^* \end{bmatrix} = \mathcal{T}_c\{\mathbf{a}\}. \tag{6.21}$$

Now, replacing the Toeplitz matrices $\hat{\mathbf{R}}$ and $\mathbf{R}(\boldsymbol{\theta})$ by their circulant counterparts in the log-likelihood function (4.79) yields the new cost function to maximize as

$$C_c(\hat{\mathbf{r}}, \boldsymbol{\theta}) = -N \cdot \sum_{m=1}^{M} \left(\ln\left(\mathcal{T}_c\{\boldsymbol{\kappa}(\boldsymbol{\theta})\}_m\right) + \frac{\mathcal{T}_c\{\hat{\mathbf{r}}\}_m}{\mathcal{T}_c\{\boldsymbol{\kappa}(\boldsymbol{\theta})\}_m} \right). \tag{6.22}$$

The gradient of the new cost function and an approximation of the Hessian can be derived in a similar fashion from (4.85) or (4.87) yielding

$$\mathbf{g}(\hat{\mathbf{r}}, \boldsymbol{\theta})_i = \sum_{m=1}^{M} \left(\left(\frac{\mathcal{T}_c\{\hat{\mathbf{r}}\}_m}{\mathcal{T}_c\{\boldsymbol{\kappa}(\boldsymbol{\theta})\}_m} - 1 \right) \frac{1}{\mathcal{T}_c\{\boldsymbol{\kappa}(\boldsymbol{\theta})\}_m} \mathcal{T}_c\left\{ \frac{\partial}{\partial \theta_i} \boldsymbol{\kappa}(\boldsymbol{\theta}) \right\}_m \right) \tag{6.23}$$

and

$$\tilde{\mathbf{H}}(\boldsymbol{\theta})_{ik} = \sum_{m=1}^{M} \left(\frac{\mathcal{T}_c\left\{ \frac{\partial}{\partial \theta_i} \boldsymbol{\kappa}(\boldsymbol{\theta}) \right\}_m \cdot \mathcal{T}_c\left\{ \frac{\partial}{\partial \theta_k} \boldsymbol{\kappa}(\boldsymbol{\theta}) \right\}_m}{\mathcal{T}_c\{\boldsymbol{\kappa}(\boldsymbol{\theta})\}_m^2} \right). \tag{6.24}$$

In summary the update equation for the Gauß-Newton algorithm becomes

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i + \lambda \cdot \tilde{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}}_i) \cdot \mathbf{g}(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}_i), \tag{6.25}$$

and we have reduced the numerical complexity of one iteration by a factor of M since the complexity for the new iteration is $O(L \cdot M \cdot \log(M))$ only. One should observe that $\hat{\mathbf{y}}_c$ can be calculated directly from the observed channel transfer functions $\mathbf{x}_i$ as follows

$$\hat{\mathbf{y}}_c = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{F}^{\mathrm{H}} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{0}^{(M-1)\times 1} \end{bmatrix} \right) \circ \left(\mathbf{F}^{\mathrm{H}} \begin{bmatrix} \mathbf{x}_i \\ \mathbf{0}^{(M-1)\times 1} \end{bmatrix} \right)^*. \tag{6.26}$$

Equation (6.26) reveals that the embedding of the smoothed covariance matrix in the circulant matrix is effectively a means to avoid the spectral overlap underlying the transformation $\mathcal{T}_{\mathcal{D}}\{\bullet\}$ (6.18), leading to an estimation algorithm with a better statistical efficiency. The computational complexity of an algorithm using $\mathcal{T}_C\{\bullet\}$ is approximately two times higher than the complexity of an algorithm using $\mathcal{T}_{\mathcal{D}}\{\bullet\}$. The parameter estimation algorithm for the parameters $\boldsymbol{\theta}_{DMC}$ of the structured covariance matrix using the circulant matrix approach is summarised in Table 6-4.

Table 6-4: Iterative optimisation of the parameters $\boldsymbol{\theta}_{dmc}$ circulant matrix approach (Gauß-Newton algorithm)

Input: data matrix $\mathbf{X}$, initial solution $\boldsymbol{\theta}^{\{0\}}$

Preprocessing: Compute the PDP $\mathbf{y}_c$ using equation (6.26) from $\mathbf{X}$.

1) Compute the first order derivatives to the parameters $\boldsymbol{\theta}^{\{i\}}$ of $\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})$ using equations (4.93) - (4.96).

2) Compute $\mathbf{D}(\boldsymbol{\theta}^{\{i\}}) = \mathrm{diag}\{\mathcal{T}_C\{\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})\}\}^{-1}\left[\mathcal{T}_C\left\{\dfrac{\partial}{\partial\theta_1}\boldsymbol{\kappa}(\boldsymbol{\theta})\right\} \cdots \mathcal{T}_C\left\{\dfrac{\partial}{\partial\theta_4}\boldsymbol{\kappa}(\boldsymbol{\theta})\right\}\right]$.

3) Compute the error $\boldsymbol{\varepsilon}^{\{i\}} = \mathbf{D}^{\mathrm{T}}(\boldsymbol{\theta}^{\{i\}})\left(\mathrm{diag}\{\mathcal{T}_C\{\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})\}\}^{-1}\mathbf{y}_C - 1\right)$.

4) Compute the solution to $\Delta\boldsymbol{\theta}^{\{i\}} = \arg\min_{\Delta\boldsymbol{\theta}}\|\mathbf{D}(\boldsymbol{\theta}^{\{i\}})\cdot\Delta\boldsymbol{\theta} - \boldsymbol{\varepsilon}^{\{i\}}\|_F^2$ and set $\lambda^{\{i\}} = 1$.

5) Compute the update $\boldsymbol{\theta}^{\{i+1\}} = \boldsymbol{\theta}^{\{i\}} + \lambda^{\{i\}}\cdot\Delta\boldsymbol{\theta}^{\{i\}}$.

6) Check strict maximization $\mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i+1\}}) > \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i\}})$; yes: go to 7. , no: reduce $\lambda^{\{i\}}$ go to 5.

7) Check convergence; not converged: go to 1.

Note, the maximization check $\mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i+1\}}) > \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i\}})$ can be replaced by a gradient check $\|\mathbf{g}(\boldsymbol{\theta}^{\{i+1\}})\| < \|\mathbf{g}(\boldsymbol{\theta}^{\{i\}})\|$, what is computationally less expensive especially since the gradient is required in the next iteration step, anyway. The computation of the log-likelihood function involves the computation of the logarithm $\ln(\bullet)$ at least M times, whereas the calculation of the gradient does not. The robustness of the algorithm outlined in Table 6-4 can be enhanced using the Levenberg-Marquardt optimisation strategy [68], [82], [116] see also Section 5.2.4 for a discussion of the Levenberg-Marquardt method.

Figure 6-1 shows the development of $\Delta\boldsymbol{\theta}^{\{i\}}$ over the iterations i of the algorithm outlined in Table 6-4. The algorithm converges quickly to a solution. Typically, less than 30 iterations are necessary to determine the parameter estimates within double precision accuracy (64 bit IEEE float).

To compare the statistical performance of the proposed estimators for the parameters of the dense multipath components, simulations have been carried out. In Figure 6-2, Figure 6-3, and Figure 6-4 the performance of the estimators outlined in Table 6-4, Table 6-3 and of the ML-estimator are compared to the CRLB. For the DMC parameters, typical values observed in channel sounding measurements have been selected $\boldsymbol{\theta}_{dan} = [\alpha_0\ \alpha_1\ \beta\ \tau] = [0.1\ 1\ 0.07\ 0.1]^{\mathrm{T}}$. The number of frequency samples was the same in all three simulations. In the simulations the number of independent realisations N has been

varied between $N = 1$ and $N = 1024$. For each N, 10000 experiments have been carried out. All estimators attain the CRLB if the number of realisations is $N > 1$, except for the base-delay. The variance of the estimates of the base delay determined by the estimator outlined in Table 6-3 are approx. 2 times the CRLB of the base delay for $N > 100$. The variance of the estimates of the estimator outlined in Table 6-4 for the same parameter is approx. 1.3 times the CRLB for $N > 10$. The ML-estimator attains the CRLB of the base delay for large N. Since the parameter estimator outlined in Section 6.1.5 (Table 6-4) has almost the same performance statistically as the ML-estimator but significantly lower computational complexity, it is the best choice for the estimation of the parameters of the dense multipath components. Figure 6-5 and Figure 6-6 shows the performance of the proposed estimator for a fixed number of realisations $N = 64$ and various numbers of frequency domain or time-delay domain samples M_f. In the simulation for Figure 6-5 the bandwidth was kept constant, i.e., the length of the impulse response has been changed between $M_f = 32$ and $M_f = 1024$ samples. In the simulation for Figure 6-6 the length of the impulse response was constant, i.e., the bandwidth of the observation has been changed between $M_f = 32$ and $M_f = 1024$. The estimator attains the CRLB in both cases except for the base delay estimate. The variance of the base delay estimate is approximately 1.3 times the CRLB in all simulated cases.

Whereas the number of frequency samples cannot be chosen arbitrarily in practice, there are various ways to increase the number of independent observations N. The number of independent observations of the DMC process increases if the number of transmit or receive antennas is increased, provided the DMC process is spatially uncorrelated and white. If the transmitter or the receiver is moving, the number of independent observations can also be increased, by using multiple channel observations over time. However, a necessary condition is that the radio channel is sampled fast enough. Here, fast enough refers to the fact that the estimator treats the parameters of the DMC process as stationary parameters within the observation time. On the other hand, one should note that the sampling interval should not be made arbitrarily small, since only independent observations of the DMC process will enhance the statistical performance of the estimator. Since the radio channel is a correlated process with respect to time, the correlation between the observations is growing if the sampling interval between the channel snapshots decreases. Observe, that a MIMO channel sounding measurement carried out with an 8-element transmit and an 8-element receive array contains 64 realisations of the DMC process. Consequently, MIMO channel sounding measurements are an effective means to acquire a large amount of independent observations for the estimation of the parameters $\boldsymbol{\theta}_{dan}$.
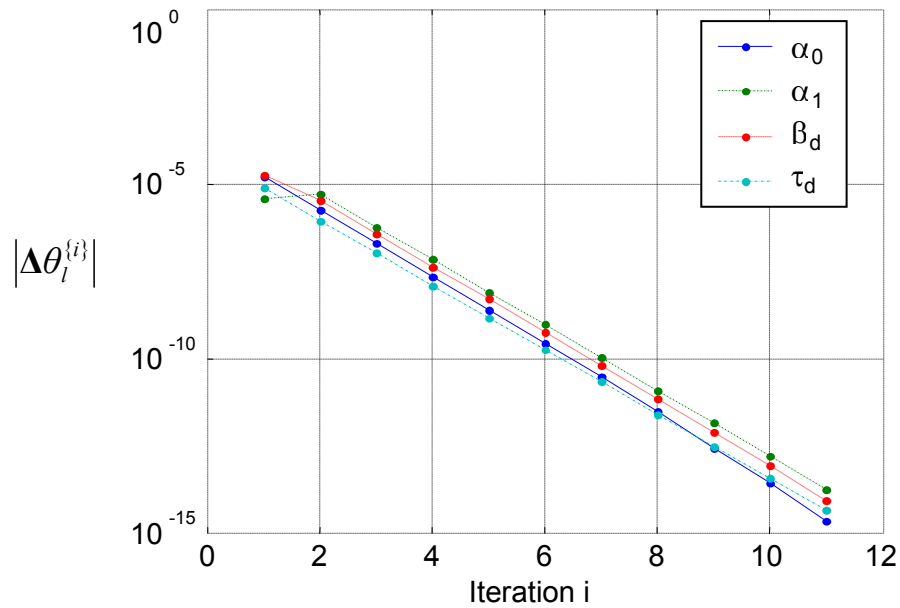
Figure 6-1: Convergence of the parameters versus the number of iterations of the Levenberg-Marquardt algorithm outlined in Table 6-5.
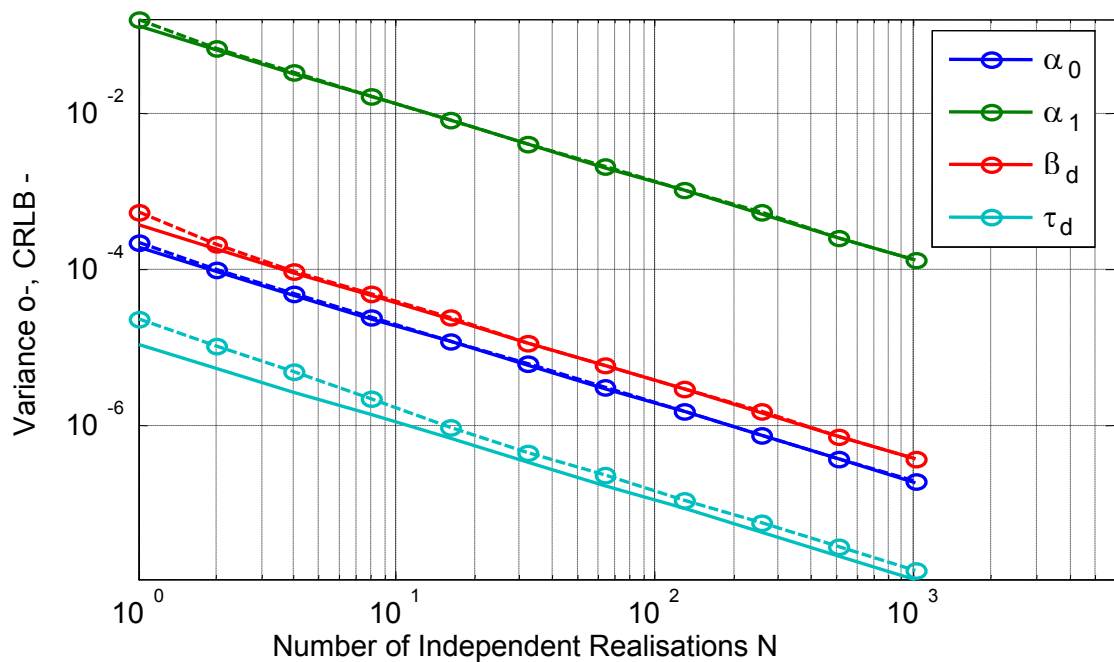
Figure 6-2: Performance of the estimator for the parameters of the dense multipath components (Table 6-4) as a function of realisations N. The number of samples in the frequency domain was $M_f = 128$. The dashed lines denote the variance of the estimates and the straight lines the related CRLB. The Estimator attains the CRLB except for the base delay estimate.
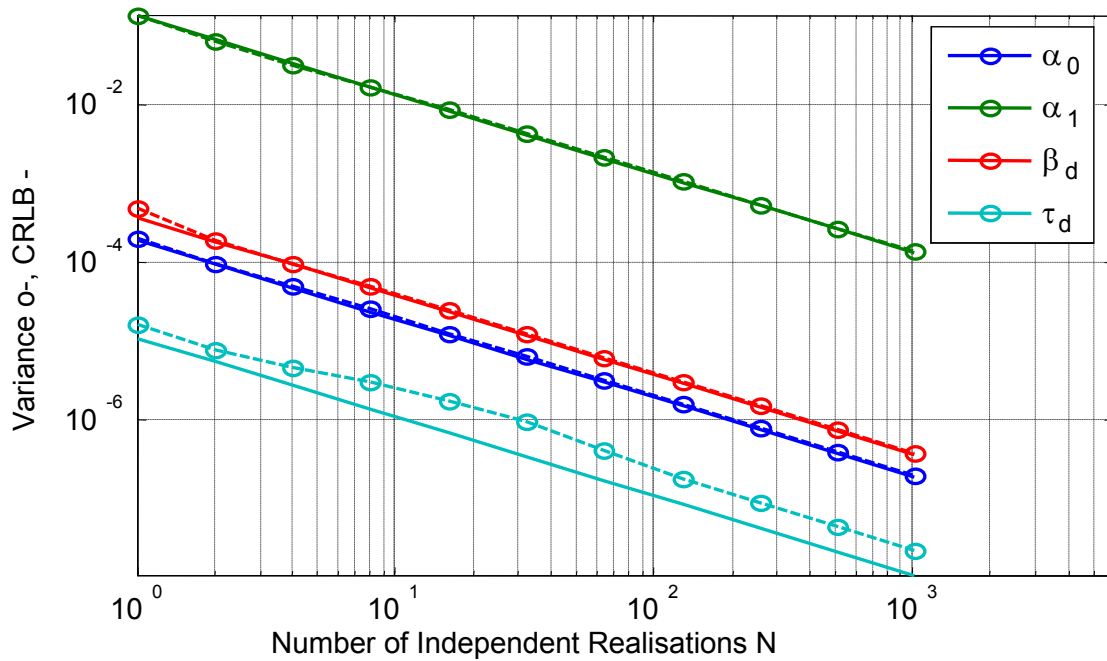
Figure 6-3:   Performance of the estimator for the parameters of the dense multipath components (Table 6-3) as a function of independent realisations N. The number of samples in the frequency domain was $M_f = 128$. The dashed lines denote the variance of the estimates and the straight lines the related CRLB. The Estimator attains the CRLB except for the base delay estimate.
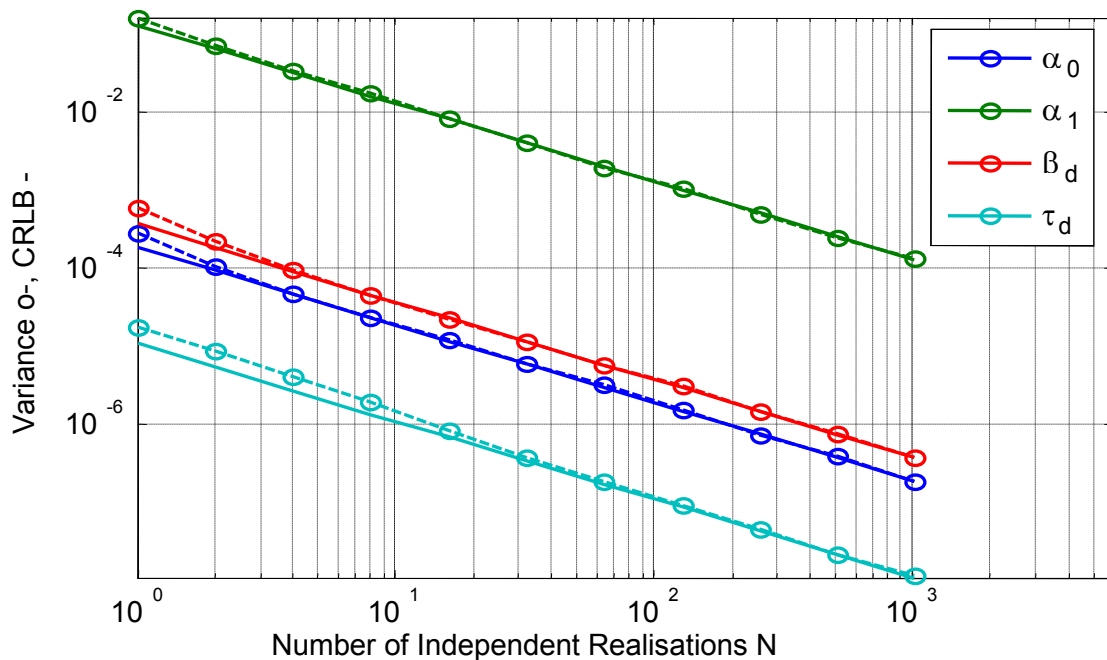
Figure 6-4:   Performance of the ML-estimator for the parameters of the dense multipath components as a function of realisations N. The number of samples in the frequency domain was $M_f = 128$. The dashed lines denote the variance of the estimates and the straight lines the related CRLB. The Estimator attains the CRLB for large N.
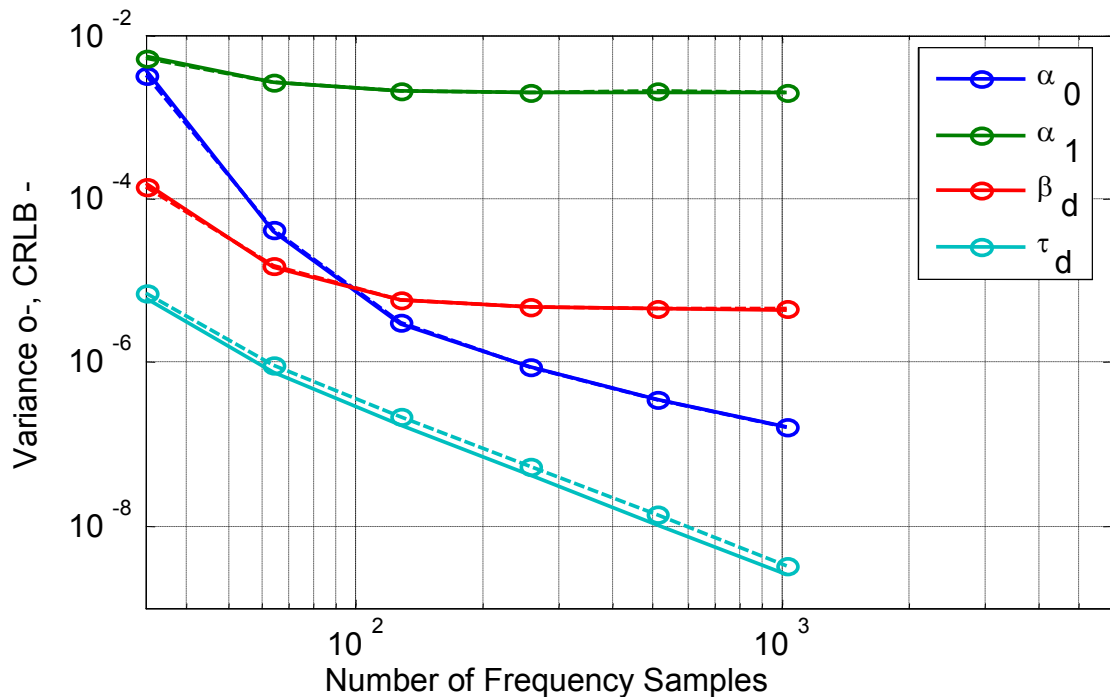
Figure 6-5: Performance of the estimator for the parameters of the dense multipath components (Table 6-4) as a function of number of samples in the frequency domain M_f, i.e., for a fixed bandwidth. The number of observations was $N = 64$. The dashed lines denote the variance of the estimates and the straight lines the related CRLB. The Estimator attains the CRLB except for the base delay estimate.
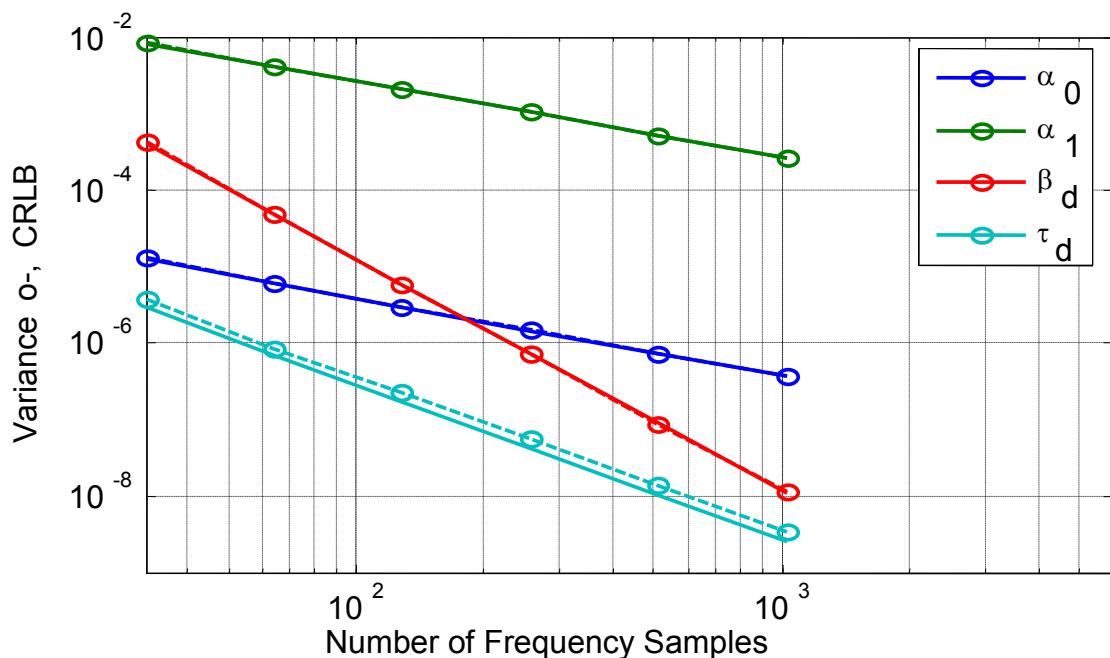
Figure 6-6: Performance of the estimator for the parameters of the dense multipath components (Table 6-4) as a function of number of samples in the frequency domain M_f, i.e., for a fixed impulse response length. The number of observations was $N = 64$. The dashed lines denote the variance of the estimates and the straight lines the related CRLB. The Estimator attains the CRLB except for the base delay estimate.

6.1.6   Model Selection for DMC and Noise

So far, we have always assumed the observed radio channel contains DMC and noise. However, in channel sounding this model is not always valid. Here valid means, the observation does not contain enough information to estimate both components of the model. If the signal to noise ratio is low, the parameters of the DMC cannot be estimated. If the SNR is very high, the power of the noise is not measurable. The three possible models are listed in the following.:

1. The observation contains measurement noise but no DMC, i.e., the observation has a low SNR.
2. The observation contains DMC but no measurement noise, i.e., the observation has a high SNR. No measurement noise means that the measurement noise is hidden by the DMC.
3. The observation contains both model components DMC and measurement noise (best case).

Consequently, we have to solve a model selection problem. It is important to observe that this model selection problem is not a model order selection problem in the strict sense, since all three resulting models are different. However, we can interpret the problem as two independent model order selection tasks. The model orders to choose from are zero (not contained in the model) or one (contained in the model) for both the DMC and the measurement noise. Using this approach, we can use the same reasoning as in Section 5.2.7 for the model selection. That means we estimate the relative variance of the weights of the components noise and DMC, i.e., of α_0 and α_1. Using the estimated relative variance, we can decide the model order individually for the noise model and the DMC model. Observe that an estimate of the variance of the parameter estimates $\boldsymbol{\theta}_{dan}$ can be computed using expression (4.90), replacing the true parameters $\boldsymbol{\theta}_{dan}$ by the estimates $\hat{\boldsymbol{\theta}}_{dan}$.

Figure 6-7 shows estimates of α_0, and α_1. The parameters have been calculated from a MIMO indoor channel sounding measurement, using the algorithm outlined in Table 6-4 and applying the outlined method for model selection. The bound for the relative variance was chosen as 0.3. At the beginning of the measurements, the SNR was very low. Consequently, the data model reduces to a white circular normal distributed process describing the measurement noise, since no radio channel components could be measured. After 300 observations, the SNR was sufficient for the estimation of the DMC parameters. As a reference, the number of propagation paths jointly estimated from the measurements using the RIMAX algorithm (cf. Section 6.2.1) is depicted in Figure 6-8. In addition, the estimated base delay $\hat{\tau}_d$ of the DMC is shown in Table 6-5.
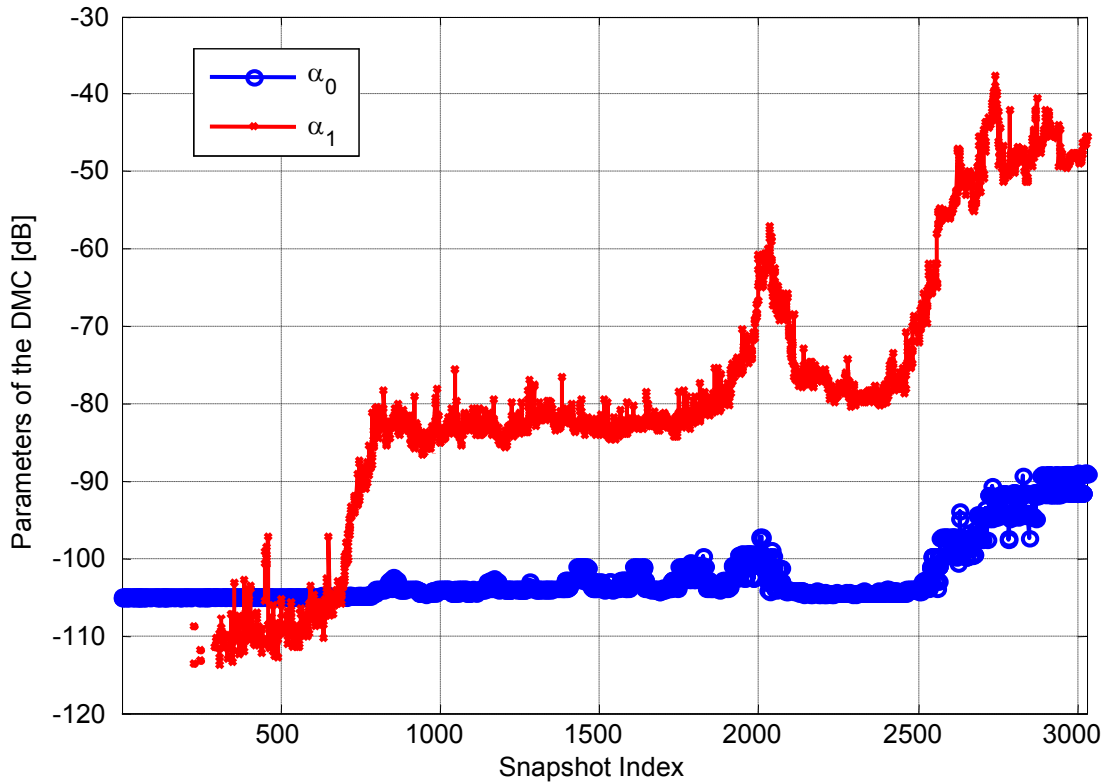
Figure 6-7: Estimation results of the parameters α_0 and α_1 applying the model selection algorithm outlined in Section 6.1.6.
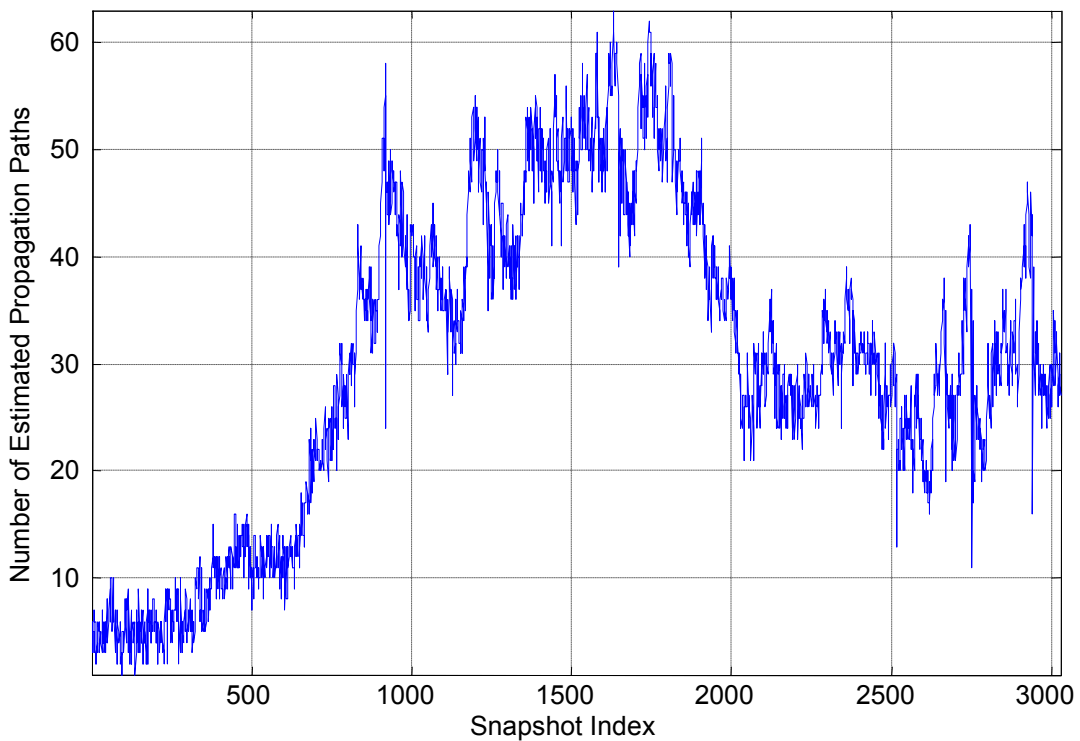
Figure 6-8: Number of propagation paths estimated by the RIMAX algorithm.
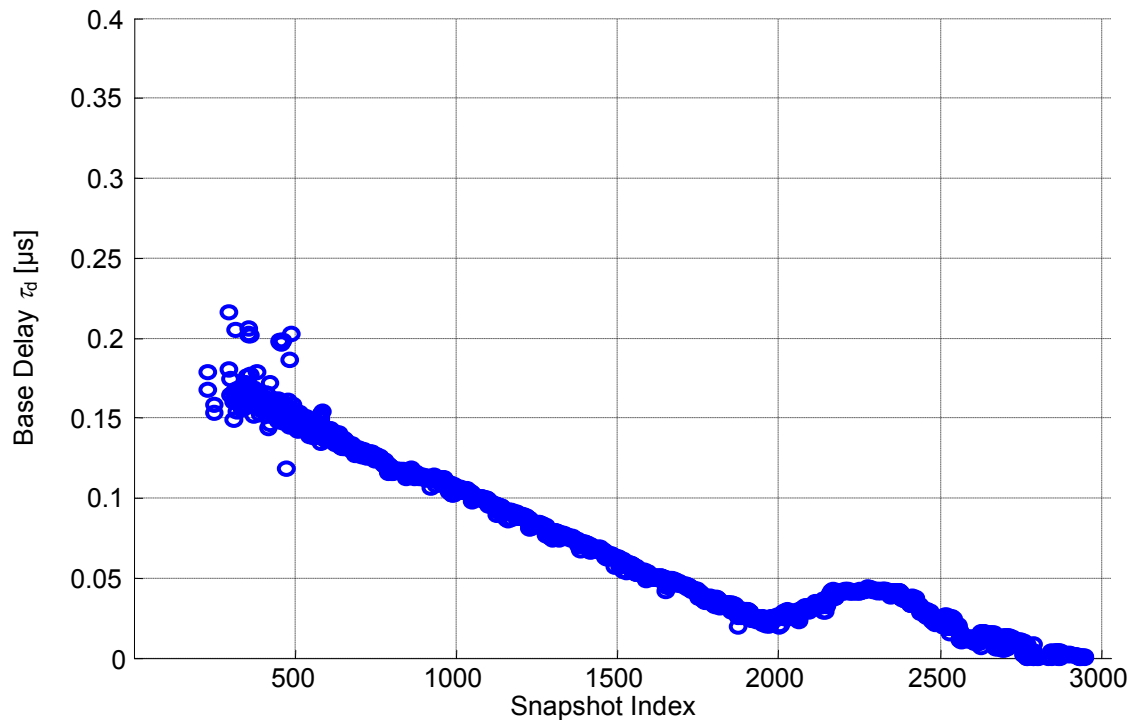
Figure 6-9:   Estimated base delay of dense multipath components.

6.1.7   Estimation of the Parameters of Multiple Independent DMC Processes

Since the developed algorithm exploits the Toeplitz structure of the parametric covariance matrix only, we can apply it too, if the radio channel contains multiple independent versions of the DMC process. Figure 6-10 shows an example of such a radio channel. In this example, a radio channel containing two DMC clusters has been simulated. The parameters of the first process and the second process have been chosen as $\boldsymbol{\theta}_{dmc,1} = \begin{bmatrix} \alpha_1 & \beta_1 & \tau_1 \end{bmatrix}^\mathrm{T} = \begin{bmatrix} 1 & 0.005 & 0.1 \end{bmatrix}^\mathrm{T}$ and $\boldsymbol{\theta}_{dmc,2} = \begin{bmatrix} \alpha_2 & \beta_2 & \tau_2 \end{bmatrix}^\mathrm{T} = \begin{bmatrix} 0.2 & 0.005 & 0.4 \end{bmatrix}^\mathrm{T}$, respectively. The variance of the measurement noise was $\alpha_0 = 0.0316$, i.e., 15dB below the maximum variance of the first DMC process. The figure on the left hand side shows the expected value of the PDP of the simulated radio channel and the right hand side shows the PDP estimated from eight independent realizations.
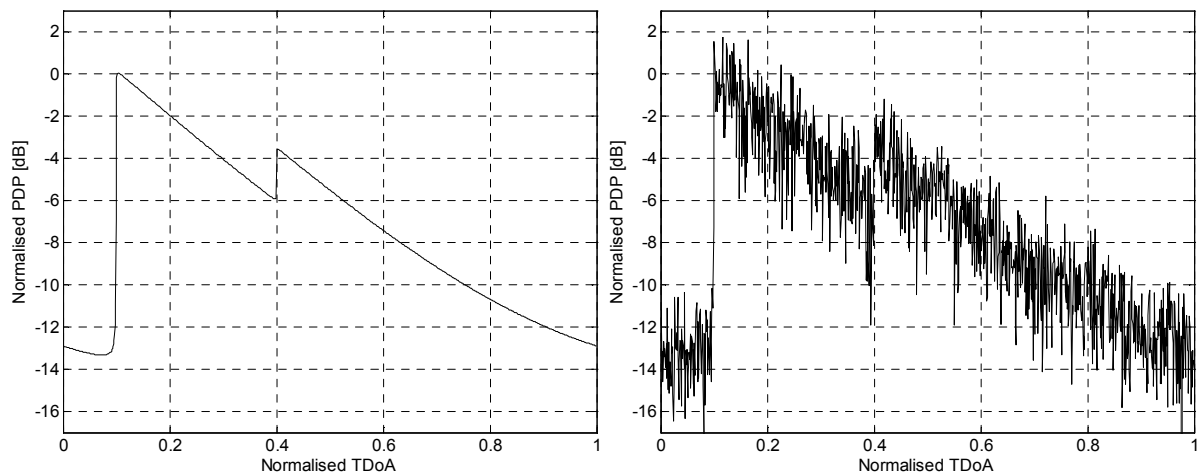
Figure 6-10: Example for a PDP containing two DMC clusters. The left hand side shows the expected value of the PDP and the right hand side shows an example of a PDP estimated from eight realisations of the DMC process.

Let us assume $\boldsymbol{\theta}_{dmc,q} = \begin{bmatrix} \alpha_q & \beta_q & \tau_q \end{bmatrix}^{\mathrm{T}}, 1 \leq q \leq Q$ are independent parameter vectors describing the covariance matrices of Q uncorrelated DMC processes. Consequently, one observation is given by

$$\mathbf{x} = \mathbf{w} + \sum_{q=1}^{Q} \mathbf{d}(\boldsymbol{\theta}_{dmc,q}).$$

Since the processes and the measurement noise are uncorrelated by definition, the covariance matrix of the process $\mathbf{x}$ is given by the sum of the covariance matrices of the Q DMC processes and the covariance matrix of the white measurement noise as

$$\mathbf{R}_{xx} = \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \alpha_0 \mathbf{I} + \sum_{q=1}^{Q} \mathrm{toep}\big(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc,q}), \boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{dmc,q})\big). \tag{6.27}$$

Remembering, that the sum of two Toeplitz matrices has Toeplitz structure too, we represent the covariance matrix (6.27) by

$$\mathbf{R}_{xx} = \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \alpha_0 \mathbf{I} + \mathrm{toep}\big(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}), \boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{dmc})\big) \tag{6.28}$$

using the spectrum

$$\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}) = \sum_{q=1}^{Q} \boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc,q}). \tag{6.29}$$

Since the spectrum $\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc})$ (6.29) is only the sum of the individual spectra of the Q uncorrelated DMC processes, the equations for the first order partial derivatives (4.94)-(4.96) as well as (4.92) are still valid. Consequently, we can apply the algorithm as outlined in Table 6-5 to estimate the set of parameters $\hat{\boldsymbol{\theta}}_{dmc,q}, 1 \leq q \leq Q$ without modifications. Furthermore, one can use the general expression (4.90) to compute the related Cramér-Rao lower bound as outlined in Table 4-6.

What we still need is an initial solution for the first iteration $\hat{\boldsymbol{\theta}}^{\{i\}}$. It should be close enough to the optimum solution to ensure convergence of the iterative algorithms described so far.

6.1.8   Estimation of an Initial Solution

To determine an initial solution $\hat{\boldsymbol{\theta}}_0$ we compute an estimate of the power delay profile from $\hat{\mathbf{R}}$ using equation (6.6) as

$$\hat{\mathbf{y}} = \mathrm{diag}\{\mathbf{F}^{\mathrm{H}} \cdot \hat{\mathbf{R}} \cdot \mathbf{F}\}. \tag{6.30}$$

If the impulse response is observed over a sufficiently long time, an estimate of the noise variance $\hat{\alpha}_0$ is

$$\hat{\alpha}_0 = \min(\hat{\mathbf{y}}). \tag{6.31}$$

Furthermore an estimate of $\hat{\alpha}_1$ can be derived from $\hat{\mathbf{y}}$ using $\hat{\alpha}_0$ by

$$\hat{\alpha}_1 = \max(\hat{\mathbf{y}}) - \hat{\alpha}_0. \tag{6.32}$$

The first element of $\hat{\mathbf{r}}$ defined in equation (6.3), i.e., $\hat{r}_1$ has an expected value of

$$\{\mathrm{E}\{\hat{r}\}\}_1 = \{\boldsymbol{\kappa}(\boldsymbol{\theta})\}_1 = \frac{\alpha_1}{M \cdot \beta_d} + \alpha_0,$$

since it is the mean value of the main diagonal elements of the sample covariance $\hat{\mathbf{R}}$, having expected value

$$\mathrm{E}\{\hat{\mathbf{R}}\} = \mathbf{R}(\boldsymbol{\theta}_{dan}).$$

Table 6-5: Iterative optimisation of the parameters $\boldsymbol{\theta}_{dmc}$ using the circulant matrix approach (Levenberg-Marquardt algorithm)

Input: data matrix $\mathbf{X}$, initial solution $\hat{\boldsymbol{\theta}}^{\{i\}}$, λ.

Pre-processing: Compute the PDP $\mathbf{y}_c$ using equation (6.26) from $\mathbf{X}$.

1) Compute the first order derivatives to the parameters $\boldsymbol{\theta}^{\{i\}}$ of $\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})$ using equations (4.93) - (4.96).

2) Compute $\mathbf{D}(\boldsymbol{\theta}^{\{i\}}) = \mathrm{diag}\{\mathcal{T}_C\{\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})\}\}^{-1}\left[\mathcal{T}_C\left\{\dfrac{\partial}{\partial\theta_1}\boldsymbol{\kappa}(\boldsymbol{\theta})\right\} \cdots \mathcal{T}_C\left\{\dfrac{\partial}{\partial\theta_4}\boldsymbol{\kappa}(\boldsymbol{\theta})\right\}\right]$.

3) Compute the error $\boldsymbol{\varepsilon}^{\{i\}} = \mathrm{diag}\{\mathcal{T}_C\{\boldsymbol{\kappa}(\boldsymbol{\theta}^{\{i\}})\}\}^{-1}\mathbf{y}_C - 1$.

4) Compute the solution to $\Delta\boldsymbol{\theta}^{\{i\}} = (\mathbf{D}^{\mathrm{T}}(\boldsymbol{\theta}^{\{i\}})\mathbf{D}(\boldsymbol{\theta}^{\{i\}}) + \lambda\mathbf{I})^{-1}\mathbf{D}^{\mathrm{T}}(\boldsymbol{\theta}^{\{i\}})\boldsymbol{\varepsilon}^{\{i\}}$.

5) Compute the update $\boldsymbol{\theta}^{\{i+1\}} = \boldsymbol{\theta}^{\{i\}} + \lambda^{\{i\}}\cdot\Delta\boldsymbol{\theta}^{\{i\}}$.

6) Check strict maximization $\mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i+1\}}) > \mathcal{L}(\mathbf{X}|\boldsymbol{\theta}^{\{i\}})$; yes: set $\lambda = \frac{\lambda}{4}$ go to 7. , no: set $\lambda = 8\lambda$ go to 4.

7) Check convergence; not converged: go to 1.

Consequently, an raw estimate of $\hat{\beta}_d$ can be determined from $\hat{\mathbf{r}}$ using the raw estimates $\hat{\alpha}_1$ and $\hat{\alpha}_0$ according to

$$\hat{\beta}_d = \frac{\hat{\alpha}_1}{M(\hat{r}_1 - \hat{\alpha}_0)}. \qquad (6.33)$$

Finally, we have to determine a initial solution for the base time delay of the dense multipath components $\hat{\tau}_d$. At this point it is important to note that the determinant of $\mathbf{R}(\boldsymbol{\theta})$ is independent of the parameter τ_d. If we define the matrix valued function

$$\boldsymbol{\Omega}(\nu) = \mathrm{diag}\{[1\ \mathrm{e}^{-\mathrm{j}2\pi\nu}\ \cdots\ \mathrm{e}^{-\mathrm{j}2\pi(M-1)\nu}]\},$$

we can express $\mathbf{R}(\boldsymbol{\theta})$ as

$$\mathbf{R}(\boldsymbol{\theta}_{dan}) = \mathrm{toep}(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dan}), \boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{dan})) = \boldsymbol{\Omega}(\tau_d)\cdot\mathbf{R}\left(\begin{bmatrix}\alpha_0\\\alpha_1\\\beta_d\\0\end{bmatrix}\right)\cdot\boldsymbol{\Omega}^{\mathrm{H}}(\tau_d).$$

Since $\boldsymbol{\Omega}(\tau_d)$ is unitary the determinant $\det(\mathbf{R}(\boldsymbol{\theta}_{dan}))$, with $\boldsymbol{\theta}_{dan} = [\alpha_0\ \alpha_1\ \beta_d\ \tau_d]^{\mathrm{T}}$, is not a function of τ_d. Consequently, $\det(\mathbf{R}(\boldsymbol{\theta}_{dan})) = \det(\mathbf{R}(\boldsymbol{\theta}_{dan,0}))$, with $\boldsymbol{\theta}_{dan,0} = [\alpha_0\ \alpha_1\ \beta_d\ 0]^{\mathrm{T}}$, holds $\forall\tau_d$. This is important insofar that if the log-likelihood function is maximized over τ_d only,

the term $\ln(\det(\mathbf{R}(\boldsymbol{\theta})))$ in the log-likelihood function (4.79) can be neglected and accordingly the term $\ln(\mathcal{T}_c\{\boldsymbol{\kappa}(\boldsymbol{\theta})\})$ in the cost function (6.22), leading to the simplified maximization problem

$$\hat{\tau}_d = \arg\max_{\tau_d} \sum_{m=1}^{M} \left(\frac{\mathcal{T}_c\{\hat{\mathbf{r}}\}_m}{\mathcal{T}_c\left\{ \boldsymbol{\kappa}\left(\begin{bmatrix} \hat{\alpha}_0 & \hat{\alpha}_1 & \hat{\beta}_d & \tau_d \end{bmatrix}^{\mathrm{T}} \right) \right\}_m} \right).$$

So we compute at first the elements of the vector $\mathbf{z}$ from $\boldsymbol{\kappa}\left(\begin{bmatrix} \hat{\alpha}_0 & \hat{\alpha}_1 & \hat{\beta}_d & 0 \end{bmatrix}^{\mathrm{T}} \right)$

$$z_m = \frac{1}{\mathcal{T}_C\left\{ \boldsymbol{\kappa}\left(\begin{bmatrix} \hat{\alpha}_0 & \hat{\alpha}_1 & \hat{\beta}_d & 0 \end{bmatrix}^{\mathrm{T}} \right) \right\}_m}. \tag{6.34}$$

After embedding this vector in the circulant matrix $\mathbf{Z} = \mathrm{circ}\{\mathbf{z}\}$, an estimate of the base delay τ_d can be computed from the index $l_{\max} = \arg\max_{l}\{\mathbf{c}_\tau\}_l$ belonging to the largest element of $\mathbf{c}_\tau = \mathbf{Z}^T \cdot \hat{\mathbf{y}}_c$ using

$$\hat{\tau}_d = \frac{l_{\max} - 1}{2M_f - 1}. \tag{6.35}$$

One should observe that the cyclic convolution between $\mathbf{z}$ and $\hat{\mathbf{y}}_c$ can be calculated in a computationally efficient way using the FFT by

$$\mathbf{c}_\tau = \mathbf{F}^{\mathrm{H}} \cdot \left((\mathbf{F} \cdot \hat{\mathbf{y}}_c) \circ (\mathbf{F} \cdot \mathbf{z})^* \right). \tag{6.36}$$

Equation (6.36) implies also that the cost function can be calculated with an arbitrary resolution using zero-padding. Altogether to calculate an initial solution $\hat{\boldsymbol{\theta}}^{\{0\}}$ we have first to compute $\hat{\mathbf{y}}$ using equation (6.30) and then to establish an estimate for $\hat{\alpha}_0$, $\hat{\alpha}_1$, and $\hat{\beta}_d$ using (6.31), (6.32), and (6.33). Using these estimates, we compute the vector $\mathbf{z}$ from equation (6.34). Finnally, we determine the initial estimate for $\hat{\tau}_d$ from equation (6.36) and (6.35) using the vector $\mathbf{z}$. All four estimates together yield the initial solution $\hat{\boldsymbol{\theta}}^{\{0\}}$ for the local search algorithm. Table 6-6 gives a summary of the complete algorithm for the estimation of an initial solution.

Table 6-6: Computation of an initial solution for the parameters $\boldsymbol{\theta}_{dmc}$.

Input: data matrix $\mathbf{X}$

Preprocessing: Compute the PDP $\mathbf{y}$ using equation (6.6) from $\mathbf{X}$.

1) Compute an estimate of the measurement noise $\hat{\alpha}_0 = \min(\hat{\mathbf{y}})$.

2) Compute the maximum power of DMC $\hat{\alpha}_1 = \max(\hat{\mathbf{y}}) - \hat{\alpha}_0$.

3) Compute $\hat{r}_1 = \sum_{m=1}^{M} y_m$

4) Compute an estimate of the coherence bandwidth of the DMC as

$$\hat{\beta}_d = \frac{\hat{\alpha}_1}{M(\hat{r}_1 - \hat{\alpha}_0)}.$$

5) Compute the reference vector

$$z_m = \frac{1}{\mathcal{T}_C\left\{\boldsymbol{\kappa}\left(\begin{bmatrix} \hat{\alpha}_0 & \hat{\alpha}_1 & \hat{\beta}_d & 0 \end{bmatrix}^{\mathrm{T}}\right)\right\}_m}, 1 \le m \le 2M - 1$$

and the interpolated PDP

$$\hat{\mathbf{y}}_c = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{F}^{\mathrm{H}}\begin{bmatrix} \mathbf{x}_i \\ \mathbf{0}^{(M-1)\times 1} \end{bmatrix}\right) \circ \left(\mathbf{F}^{\mathrm{H}}\begin{bmatrix} \mathbf{x}_i \\ \mathbf{0}^{(M-1)\times 1} \end{bmatrix}\right)^*.$$

6) Compute $\mathbf{c}_\tau = \mathbf{F}^{\mathrm{H}} \cdot \left((\mathbf{F} \cdot \hat{\mathbf{y}}_c) \circ (\mathbf{F} \cdot \mathbf{z})^*\right)$ using the FFT.

7) Find $l_{\max} = \arg\max_l \{\mathbf{c}_\tau\}_l$ and compute the estimate of the base delay of the

   DMC as $\hat{\tau}_d = \frac{l_{\max} - 1}{2M_f - 1}$.

8) Form the initial estimate $\hat{\boldsymbol{\theta}}^{\{0\}} = \begin{bmatrix} \hat{\alpha}_0 & \hat{\alpha}_1 & \hat{\beta}_d & \hat{\tau}_d \end{bmatrix}^{\mathrm{T}}$.

6.1.9 Frequency Domain Smoothing

The global search algorithm yields unreliable results, if the number of independent observations is small and/or if the observations contain strong discrete (specular) components. A way to improve the robustness of the global search strategy is to decompose the observation into overlapping observations. If we reduce the column size of the data matrix $\mathbf{X}$, i.e., the bandwidth, from M_f to M_{sub_f} we can effectively increase the number of observations by a factor of $L_f = M_f - M_{sub_f} - 1$. Using the L_f selection matrices defined by

$$\mathbf{J}_l = \begin{bmatrix} \overbrace{\mathbf{0}}^{l-1} & \overbrace{\mathbf{I}}^{M_{sub}} & \overbrace{\mathbf{0}}^{L-l} \end{bmatrix} \in \mathbb{R}^{M_{sub} \times M}$$

the new set of observations $\mathbf{X}_{SS}$ can be expressed algebraically by

$$\mathbf{X}_{SS} = \begin{bmatrix} \mathbf{J}_1\mathbf{X} & \mathbf{J}_2\mathbf{X} & \cdots & \mathbf{J}_{L_f}\mathbf{X} \end{bmatrix} \in \mathbb{C}^{M_{sub_f} \times L_f N}. \tag{6.37}$$

Notice that the covariance matrix of an arbitrary segment $\mathbf{J}_l\mathbf{X}$ is given by

$$\mathbf{R}_{\mathbf{X}_l\mathbf{X}_l} = \tfrac{1}{N}\,\mathrm{E}\!\left\{\mathbf{J}_l\mathbf{X}\mathbf{X}\mathbf{J}_l^{\mathrm{H}}\right\} = \tfrac{1}{N}\,\mathbf{J}_l\,\mathrm{E}\{\mathbf{X}\mathbf{X}\}\mathbf{J}_l^{\mathrm{H}} = \mathbf{J}_l\cdot\mathrm{toep}\!\left(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}),\boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{dmc})\right)\cdot\mathbf{J}_l^{\mathrm{H}}.$$

For all Toeplitz matrices, the following holds

$$\mathbf{J}_i\mathbf{T}\mathbf{J}_i^{\mathrm{H}} = \mathbf{J}_k\mathbf{T}\mathbf{J}_k^{\mathrm{H}},\ \forall 1\le i\le L_f, 1\le k\le L_f$$

holds. Consequently the covariance matrix of (6.37)

$$\mathbf{R}_{\mathbf{X}_S\mathbf{X}_S} = \mathbf{J}_1\cdot\mathrm{toep}\!\left(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}),\boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{dmc})\right)\cdot\mathbf{J}_1^{\mathrm{H}} = \mathrm{toep}\!\left(\mathbf{J}_1\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}),(\mathbf{J}_1\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}))^{\mathrm{H}}\right)$$

is simply a reduced version of the covariance matrix $\mathrm{toep}\!\left(\boldsymbol{\kappa}(\boldsymbol{\theta}_{dmc}),\boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}_{dmc})\right)$.

The outlined smoothing technique increases the amount of observations, and reduces the influence of strong specular paths at the same time. The principle is similar to the subspace (subarray) smoothing algorithm used for the signal subspace estimation needed by subspace-based estimators (ESPRIT, RARE, MUSIC) but the intention is clearly different.
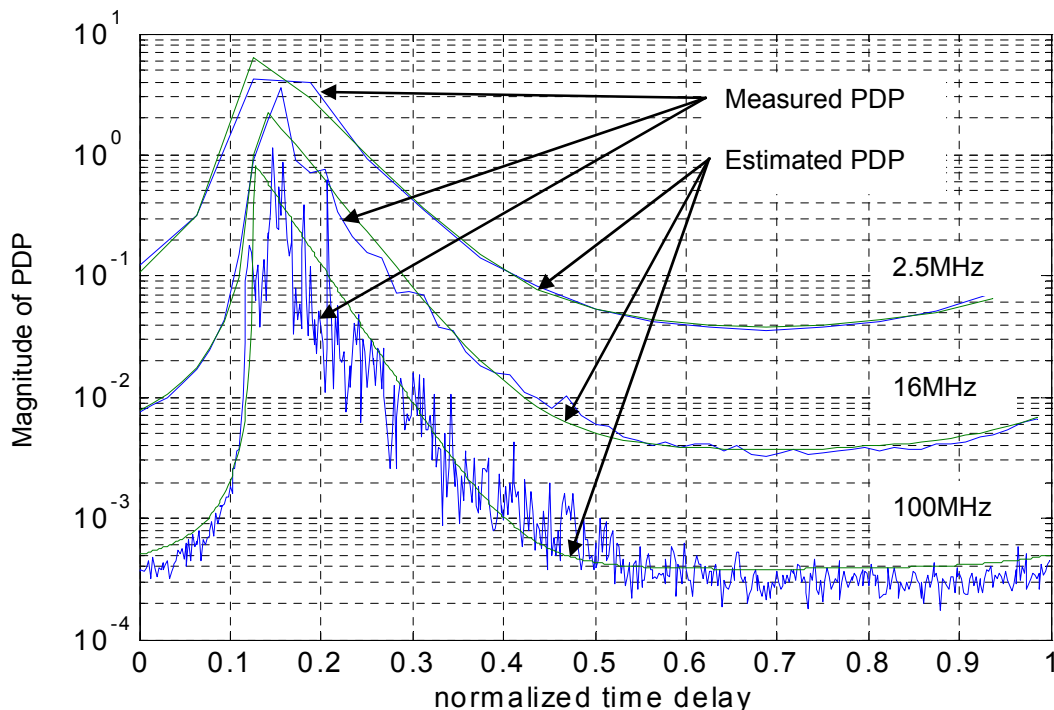
Figure 6-11: Example for the influence of smoothing in the frequency domain to the PDP of a measured SIMO impulse response containing some specular-alike propagation paths and the PDP estimated using the algorithm in Table 6-4.

In Figure 6-11 the influence of smoothing in the frequency domain to the PDP is shown. The PDP has been calculated from a SIMO measurement carried out with an 8-element uniform linear array. The measurement campaign has been carried out in a non-line of sight scenario. Using the smoothed data, at first the parameter of the DMC has been estimated using the global search algorithm outlined in Table 6-6. With these initial estimates, the algorithm summarised in Table 6-4 has been applied to get the final estimate shown in the example. It is instructive to see that the contribution of the specular propagation paths vanishes if the observation bandwidth is reduced. This effect has already been discussed in Section 2.5.4. The reason for this behavior can be found in the nature of the two channel model components. The specular paths contribute to the mean of the channel, i.e., they are the de-

terministic part of the channel. The dense multipath components on the other hand are the stochastic part of the channel.

6.1.10 A Comment on the Least Squares Estimation of Parametric Covariance Matrices

If we have to estimate the parameters $\boldsymbol{\theta}$ of a covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ from an observed covariance matrix $\hat{\mathbf{R}}$ (4.83) we may be tempted to use the simple nonlinear least squares (NLS) approach

$$\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\boldsymbol{\theta}} \left\| \hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2 . \tag{6.38}$$

To understand why this least squares approach is not optimal compared with the maximum likelihood approach (6.1) let us calculate the first order derivatives, i.e., the first gradient of

$$\left\| \hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2 = \mathrm{tr}\left\{ \left(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right)^{\mathrm{H}} \cdot \left(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right) \right\}$$

$$= \mathrm{tr}\left\{ \hat{\mathbf{R}}^{\mathrm{H}}\hat{\mathbf{R}} - \hat{\mathbf{R}}^{\mathrm{H}}\mathbf{R}(\boldsymbol{\theta}) - \mathbf{R}^{\mathrm{H}}(\boldsymbol{\theta})\hat{\mathbf{R}} + \mathbf{R}^{\mathrm{H}}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta}) \right\},$$

to $\boldsymbol{\theta}$. Straight forward calculations yield

$$\frac{\partial}{\partial \theta_i} \left\| \hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2 = 2 \cdot \mathrm{tr}\left\{ \frac{\partial}{\partial \theta_i} \mathbf{R}^{\mathrm{H}}(\boldsymbol{\theta}) \cdot \left(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right) \right\}.$$

Hence, the least squares estimate $\hat{\boldsymbol{\theta}}_{LS}$ is a solution of

$$\mathrm{tr}\left\{ \frac{\partial}{\partial \hat{\theta}_{LS,i}} \mathbf{R}^{\mathrm{H}}\left(\hat{\boldsymbol{\theta}}_{LS}\right) \cdot \left(\hat{\mathbf{R}} - \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{LS}\right) \right) \right\} = 0, \quad \forall \hat{\theta}_{LS,i},$$

whereas the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{ML}$ is according to (4.84) a solution of

$$\mathrm{tr}\left(\mathbf{R}^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right) \left(\frac{\partial}{\partial \theta_i} \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{ML}\right) \right) \cdot \mathbf{R}^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right) \left(\hat{\mathbf{R}} - \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{ML}\right) \right) \right) = 0, \quad \forall \hat{\theta}_{ML,i}.$$

So, the difference between the least squares estimator and the maximum likelihood estimator for the parametric covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ is effectively the weighting with $\mathbf{R}^{-1}\left(\hat{\boldsymbol{\theta}}\right)$ of the difference between the observation and the estimate

$$\left(\hat{\mathbf{R}} - \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{LS}\right) \right) \Leftrightarrow \mathbf{R}^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\left(\hat{\mathbf{R}} - \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{ML}\right) \right),$$

and of the differential

$$\frac{\partial}{\partial \hat{\theta}_{LS,i}} \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{LS,i}\right) \Leftrightarrow \mathbf{R}^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right) \frac{\partial}{\partial \theta_{ML,i}} \mathbf{R}\left(\hat{\boldsymbol{\theta}}_{ML}\right).$$

Finally, note that the maximum likelihood estimator is, although similar, not a weighted nonlinear least squares (WNLS) estimator, since the weighting matrix $\mathbf{R}^{-1}(\boldsymbol{\theta})$ depends on the parameters to estimate $\boldsymbol{\theta}$.

6.1.11 A Generator for the DMC Process

To generate realisations of dense multipath components for numerical simulations we have to generate a circular Gaussian process

$$\mathbf{d}(\boldsymbol{\theta}_{dmc}) \sim \mathcal{N}_C(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dmc})) \in \mathbb{C}^{M_f \times 1},$$

having zero mean and covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{dmc})$. To this end we generate usually a multivariate i.i.d. circular Gaussian process

$$\mathbf{z} \in \mathbb{C}^{M_f \times 1} \sim \mathcal{N}_C(\mathbf{0}, \mathbf{I})$$

and use a transformation matrix $\mathbf{L}(\boldsymbol{\theta}_{dmc})$ satisfying

$$\mathbf{R}(\boldsymbol{\theta}_{dmc}) = \mathbf{L}(\boldsymbol{\theta}_{dmc}) \cdot \mathbf{L}^{\mathrm{H}}(\boldsymbol{\theta}_{dmc}) \tag{6.39}$$

to compute the coloured process

$$\mathbf{d}(\boldsymbol{\theta}_{dmc}) = \mathbf{L}(\boldsymbol{\theta}_{dmc}) \cdot \mathbf{z}. \tag{6.40}$$

The generator (6.40) for the process $\mathbf{d}(\boldsymbol{\theta}_{dmc})$ has a computational complexity of $O(M_f^2 N)$ if we ask for N realisations. And the computational complexity of algorithms for the decomposition (6.39) is of order $O(M_f^3)$, in general. For example, the projector $\mathbf{L}(\boldsymbol{\theta}_{dmc})$ can be calculated using the Cholesky decomposition

$$\mathbf{R} = \mathbf{L} \cdot \mathbf{L}^{\mathrm{H}},$$

or the singular value decomposition (SVD)

$$\mathbf{R} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^{\mathrm{H}} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{U}^{\mathrm{H}} = \left(\mathbf{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\right) \cdot \left(\mathbf{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{H}} = \mathbf{L} \cdot \mathbf{L}^{\mathrm{H}}.$$

The computational complexity can be reduced significantly if we exploit the Toeplitz structure of the covariance matrix (2.67).

The covariance matrix $\mathbf{R}(\boldsymbol{\theta}) = \mathrm{toep}(\boldsymbol{\kappa}(\boldsymbol{\theta}), \boldsymbol{\kappa}^{\mathrm{H}}(\boldsymbol{\theta}))$ can also be expressed using the circulant matrix

$$\mathbf{R}_c(\boldsymbol{\theta}) = \mathrm{circ}\left(\begin{bmatrix} \boldsymbol{\kappa}(\boldsymbol{\theta}) \\ [\,\mathbf{0} \ \boldsymbol{\Pi}_{M_f-1}\,] \cdot \boldsymbol{\kappa}^*(\boldsymbol{\theta}) \end{bmatrix}\right)$$

by

$$\mathbf{R}(\boldsymbol{\theta}) = \mathbf{J} \cdot \mathbf{R}_c(\boldsymbol{\theta}) \cdot \mathbf{J}^{\mathrm{H}}, \tag{6.41}$$

where the selection matrix $\mathbf{J}$ is defined as

$$\mathbf{J} = [\,\mathbf{I}_{M_f} \ \mathbf{0}\,] \in \mathbb{R}^{M_f \times 2M_f - 1}.$$

Now recalling, that any circulant matrix is diagonalized by the DFT-matrix we rewrite expression (6.41) to

$$\mathbf{R}(\boldsymbol{\theta}) = \mathbf{J} \cdot \mathbf{F} \cdot \mathrm{diag}\{\mathbf{w}(\boldsymbol{\theta})\}^2 \cdot \mathbf{F}^{\mathrm{H}} \cdot \mathbf{J}^{\mathrm{H}}.$$

Hence, a computationally efficient procedure to generate the process $\mathbf{d}(\boldsymbol{\theta}_{dmc})$ is given by the relation

$$\mathbf{d}(\boldsymbol{\theta}_{dmc}) = \mathbf{J} \cdot \mathbf{F} \cdot (\mathbf{w}(\boldsymbol{\theta}_{dmc}) \circ \mathbf{z}), \tag{6.42}$$

where $\mathbf{z} \in \mathbb{C}^{(2M_f-1) \times 1}$ is now a realisation of the multivariate i.i.d. circular Gaussian process distributed according to

$$\sim \mathcal{N}_C(\mathbf{0}, \mathbf{I}).$$

The generator (6.42) has a numerical complexity $O(M_f \cdot \log(M_f))$ for one realisation and consequently a complexity of $O(N \cdot M_f \cdot \log(M_f))$ if we ask for N realisations. This is an improvement of $O(M_f / \log(M_f))$ compared with the generator (6.40). Furthermore, one should note that the weighting vector $\mathbf{w}(\boldsymbol{\theta})$ can be calculate in a computationally efficient way using

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \frac{1}{\sqrt{2M_f - 1}} \mathbf{F}^H \begin{bmatrix} \boldsymbol{\kappa}(\boldsymbol{\theta}) \\ [\ \mathbf{0}\ \boldsymbol{\Pi}_{M_f - 1}\ ] \cdot \boldsymbol{\kappa}^*(\boldsymbol{\theta}) \end{bmatrix}, \quad \boldsymbol{\lambda}(\boldsymbol{\theta}) \in \mathbb{R}^{(2M_f - 1) \times 1}$$

and[15]

$$\{\mathbf{w}(\boldsymbol{\theta})\}_i = +\sqrt{\{\boldsymbol{\lambda}(\boldsymbol{\theta})\}_i}\ .$$

Hence, the computation of $\mathbf{w}(\boldsymbol{\theta})$ requires $O(M_f \log(M_f))$ operations only. This is a significant improvement compared with the computational complexity $O(M_f^3)$ necessary for the computation of $\mathbf{L}(\boldsymbol{\theta})$, especially if we have to generate a sequence of realisations having varying parameters $\boldsymbol{\theta}$.

Table 6-7: Generator for the circular zero-mean normal distributed process
$$\mathbf{d}_{dmc} \sim \mathcal{N}_C(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dmc}))$$

---

1) Compute $\boldsymbol{\lambda}(\boldsymbol{\theta}) = \dfrac{1}{\sqrt{2M_f - 1}} \mathbf{F}^H \begin{bmatrix} \boldsymbol{\kappa}(\boldsymbol{\theta}) \\ [\ \mathbf{0}\ \boldsymbol{\Pi}_{M_f - 1}\ ] \cdot \boldsymbol{\kappa}^*(\boldsymbol{\theta}) \end{bmatrix}$ using the FFT.

2) Compute the square roots of the elements of $\boldsymbol{\lambda}(\boldsymbol{\theta})$ as $\{\mathbf{w}(\boldsymbol{\theta})\}_i = +\sqrt{\{\boldsymbol{\lambda}(\boldsymbol{\theta})\}_i}\ $.

3) Compute a realisation of the multivariate i.i.d. circular normal distributed process $\mathbf{z} \sim \mathcal{N}_C(\mathbf{0}, \mathbf{I})$.

4) Computed the coloured process $\mathbf{d}(\boldsymbol{\theta}_{dmc}) = \mathbf{J} \cdot \mathbf{F} \cdot (\mathbf{w}(\boldsymbol{\theta}_{dmc}) \circ \mathbf{z})$.

5) If multiple realisations are required, repeat step 3 and step 4.

6.1.12 Implementation Issues

In this subsection, we investigate computationally efficient ways to solve a set of linear equations

$$\mathbf{T} \cdot \mathbf{B} = \mathbf{A}, \tag{6.43}$$

where $\mathbf{T} \in \mathbb{C}^{M \times M}$ is Hermitian and Toeplitz, $\mathbf{A} \in \mathbb{C}^{M \times N}$ an arbitrary matrix, and $\mathbf{B} \in \mathbb{C}^{M \times N}$ the solution to be calculated. Having covariance matrices in mind, we assume that $\mathbf{T}$ is positive definite, i.e., the inverse $\mathbf{T}^{-1}$ exists. Since (6.43) can be solved separately for every column $\mathbf{a}_n$ of $\mathbf{A}$, we will focus the discussion on the solution of the linear system of equations

$$\mathbf{b} = \mathbf{T}^{-1} \cdot \mathbf{a}. \tag{6.44}$$

In computations with the inverses of Toeplitz matrices, the Gohberg-Semencul formula [117], [118] is useful. This formula represents the inverse of a finite Toeplitz matrix in the form of the sum of products of triangular Toeplitz matrices [119]. The inverse of an arbitrary non-

---

[15] The values of $\boldsymbol{\lambda}(\boldsymbol{\theta})$ are nonnegative since they are the eigenvalues of a covariance matrix.

singular Toeplitz matrix can be expressed using the solutions $\mathbf{x} \in \mathbb{C}^{M \times 1}$ and $\mathbf{y} \in \mathbb{C}^{M \times 1}$ of the equations

$$\mathbf{T} \cdot \mathbf{x} = \mathbf{e}_0 \qquad (6.45)$$

and

$$\mathbf{T} \cdot \mathbf{y} = \mathbf{e}_{M-1} \qquad (6.46)$$

as

$$\mathbf{T}^{-1} = \frac{1}{x_1} \left(\begin{bmatrix} x_1 & 0 & \cdots & \cdots & 0 \\ x_2 & x_1 & \ddots & & \vdots \\ \vdots & x_2 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ x_M & \cdots & \cdots & x_2 & x_1 \end{bmatrix} \cdot \begin{bmatrix} y_M & y_{M-1} & \cdots & \cdots & y_1 \\ 0 & y_M & y_{M-1} & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & y_{M-1} \\ 0 & \cdots & & \cdots & 0 & y_M \end{bmatrix} \right.$$

$$\left. - \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ y_1 & 0 & & & \vdots \\ \vdots & y_1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ y_{M-1} & \cdots & \cdots & y_1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & x_M & \cdots & \cdots & x_2 \\ 0 & 0 & x_M & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & x_M \\ 0 & \cdots & & \cdots & 0 & 0 \end{bmatrix} \right) . \qquad (6.47)$$

A Hermitian Toeplitz matrix satisfies $\boldsymbol{\Pi} \cdot \mathbf{T} \cdot \boldsymbol{\Pi} = \mathbf{T}^*$, hence equation (6.46) can be rewritten to

$$\boldsymbol{\Pi} \cdot \mathbf{T} \cdot \boldsymbol{\Pi} \cdot \boldsymbol{\Pi} \cdot \mathbf{y} = \boldsymbol{\Pi} \cdot \mathbf{e}_{M-1}$$

$$\mathbf{T}^* \boldsymbol{\Pi} \cdot \mathbf{y} = \mathbf{e}_0$$

$$\mathbf{T} \left(\boldsymbol{\Pi} \cdot \mathbf{y}^* \right) = \mathbf{e}_0$$

since $\boldsymbol{\Pi} \cdot \boldsymbol{\Pi} = \mathbf{I}$. Consequently the solution (6.45) and (6.46) are, for a Hermitian Toeplitz matrix, related according to

$$\mathbf{y} = \boldsymbol{\Pi} \cdot \mathbf{x}^* . \qquad (6.48)$$

The solution $\mathbf{x}$ to the equation (6.45) is sufficient to express the inverse of a Hermitian Toeplitz matrix. Using (6.48) in (6.47) yields for the inverse

$$\mathbf{T}^{-1} = \frac{1}{x_1} \left(\begin{bmatrix} x_1 & 0 & \cdots & \cdots & 0 \\ x_2 & x_1 & \ddots & & \vdots \\ \vdots & x_2 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ x_M & \cdots & \cdots & x_2 & x_1 \end{bmatrix} \cdot \begin{bmatrix} x_1^* & x_2^* & \cdots & \cdots & x_M^* \\ 0 & x_1^* & x_2^* & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & x_2^* \\ 0 & \cdots & \cdots & 0 & x_1^* \end{bmatrix} \right.$$

$$\left. - \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ x_M^* & 0 & & & \vdots \\ \vdots & x_M^* & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ x_2^* & \cdots & \cdots & x_M^* & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & x_M & \cdots & \cdots & x_2 \\ 0 & 0 & x_M & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & x_M \\ 0 & \cdots & & \cdots & 0 & 0 \end{bmatrix} \right) . \qquad (6.49)$$

Therefore, we define a lower triangular Toeplitz matrix $\mathbf{T}_1 = \text{toep}(\mathbf{x}, [x_1 \ \mathbf{0}^T]) \in \mathbb{C}^{M \times M}$ and an upper triangular Toeplitz matrix $\mathbf{T}_2 = \text{toep}(\mathbf{0}, [0 \ \ x_M \ \ \cdots \ \ x_2]) \in \mathbb{C}^{M \times M}$ and write the inverse (6.51) in compact form as

$$\mathbf{T}^{-1} = \frac{1}{x_1}\left(\mathbf{T}_1 \mathbf{T}_1^H - \mathbf{T}_2^H \mathbf{T}_2\right).$$

If we embed the matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ as block matrices in the circulant matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 \\ \mathbf{T}_2 & \mathbf{T}_1 \end{bmatrix} = \text{circ}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix}\right) \in \mathbb{C}^{2M \times 2M}, \tag{6.50}$$

we can express the product (6.44) as

$$\mathbf{b} = \mathbf{w}_1 - \mathbf{\Pi} \cdot \mathbf{w}_2^*,$$

where $\mathbf{w}_1 \in \mathbb{C}^{M \times 1}$ and $\mathbf{w}_2 \in \mathbb{C}^{M \times 1}$ are given by

$$\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{C} \cdot \begin{bmatrix} \mathbf{I}^{M \times M} & \mathbf{0}^{M \times M} \\ \mathbf{0}^{M \times M} & \mathbf{0}^{M \times M} \end{bmatrix} \cdot \mathbf{C}^H \cdot \begin{bmatrix} \mathbf{a} \\ \mathbf{\Pi} \cdot \mathbf{a}^* \end{bmatrix}. \tag{6.51}$$

Proof: Using the definition of $\mathbf{C}$ (6.50) in (6.51) yields

$$\mathbf{b} = \frac{1}{x_1}\left(\mathbf{T}_1 \mathbf{T}_1^H \mathbf{a} + \mathbf{T}_1 \mathbf{T}_2^H \mathbf{\Pi} \mathbf{a}^* - \mathbf{\Pi}\left(\mathbf{T}_2 \mathbf{T}_1^H\right)^* \mathbf{a}^* - \mathbf{\Pi}\left(\mathbf{T}_2 \mathbf{T}_2^H\right)^* \mathbf{\Pi} \cdot \mathbf{a}\right)$$

$$\mathbf{b} = \frac{1}{x_1}\left(\mathbf{T}_1 \mathbf{T}_1^H \mathbf{a} + \mathbf{T}_1 \mathbf{T}_2^H \mathbf{\Pi} \mathbf{a}^* - \mathbf{\Pi}\left(\mathbf{T}_2 \mathbf{T}_1^H\right)^* \mathbf{a}^* - \mathbf{\Pi} \mathbf{T}_2^* \mathbf{\Pi} \cdot \mathbf{\Pi} \mathbf{T}_2^T \mathbf{\Pi} \cdot \mathbf{a}\right),$$

which can be simplified to

$$\mathbf{b} = \frac{1}{x_1}\left(\mathbf{T}_1 \mathbf{T}_1^H \mathbf{a} + \mathbf{T}_1 \mathbf{T}_2^H \mathbf{\Pi} \mathbf{a}^* - \mathbf{\Pi}\left(\mathbf{T}_2 \mathbf{T}_1^H\right)^* \mathbf{a}^* - \mathbf{T}_2^H \mathbf{T}_2 \mathbf{a}\right), \tag{6.52}$$

since $\mathbf{\Pi} \cdot \mathbf{T} \cdot \mathbf{\Pi} = \mathbf{T}^T$. The two triangular Toeplitz matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ commute, i.e.,

$$\mathbf{T}_1 \mathbf{T}_2^H = \mathbf{T}_2^H \mathbf{T}_1.$$

Consequently, equation (6.52) can be reduced to

$$\mathbf{b} = \frac{1}{x_1}\left(\mathbf{T}_1 \mathbf{T}_1^H \mathbf{a} - \mathbf{T}_2^H \mathbf{T}_2 \mathbf{a}\right) = \frac{1}{x_1}\left(\mathbf{T}_1 \mathbf{T}_1^H - \mathbf{T}_2^H \mathbf{T}_2\right)\mathbf{a}.$$

Since $\mathbf{C}$ is circulant it can be diagonalized using the DFT-matrix

$$\mathbf{\Lambda} = \mathbf{F}^H \cdot \mathbf{C} \cdot \mathbf{F} = \sqrt{2M} \cdot \text{diag}\left\{\mathbf{F}^H \cdot \begin{bmatrix} \mathbf{x}^{M \times 1} \\ \mathbf{0}^{M \times 1} \end{bmatrix}\right\}. \tag{6.53}$$

Using (6.53) in (6.51) yields the expression

$$\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{F} \cdot \mathbf{\Lambda} \cdot \mathbf{F}^H \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \mathbf{F} \cdot \mathbf{\Lambda}^H \cdot \mathbf{F}^H \cdot \begin{bmatrix} \mathbf{a} \\ \mathbf{\Pi} \cdot \mathbf{a}^* \end{bmatrix}.$$

Hence, the solution $\mathbf{B}$ of equation (6.43) can be computed by

$$\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \mathbf{F} \cdot \mathbf{\Lambda} \cdot \mathbf{F}^{\mathrm{H}} \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \mathbf{F} \cdot \mathbf{\Lambda}^{\mathrm{H}} \cdot \mathbf{F}^{\mathrm{H}} \cdot \begin{bmatrix} \mathbf{A} \\ \mathbf{\Pi} \cdot \mathbf{A}^* \end{bmatrix},$$ (6.54)

and

$$\mathbf{B} = \frac{1}{x_1}\left(\mathbf{W}_1 - \mathbf{\Pi} \cdot \mathbf{W}_2^*\right).$$

Since equation (6.54) contains products with only diagonal- and Fourier-matrices, we can compute the product $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}$ with an algorithm having computational complexity of $O(MN\log(M))$, since the multiplication with a Fourier-matrix $\mathbf{F}$ requires $O(MN\log(M))$ operations, if fast Fourier-transform (FFT) algorithms are used, and the multiplication with a diagonal matrix can be computed using $O(MN)$ operations. Note, that the direct approach not exploiting the Toeplitz structure of $\mathbf{T}$ has a computational complexity of $O(M^2N)$. Furthermore, it is possible to use zero padding to get matrix sizes in (6.54) appropriate for the application of radix-2 or radix-4 FFT algorithms. This is especially important if M is a prime number. Nevertheless, it is important to note, that in today's layered memory architectures mix-radix algorithms may be a better choice than radix-2 algorithms in combination with excessive zero-padding [118].

We have so far ignored the computational complexity of the solution to the set of linear equations

$$\mathbf{T} \cdot \mathbf{x} = \mathbf{e}_0.$$ (6.55)

Algorithms don't exploit the Toeplitz structure of $\mathbf{T}$ have a computational complexity of $O(M^3)$. An Algorithm being computationally more efficient is the Levinson algorithm it requires $O(M^2)$ operations only [120], to compute a solution to the system of equations (6.54). Finally if M is large one can also use a so-called super-fast algorithm to solve (6.54), having computational complexity $O(M\log(M)^2)$ [121]. Table 6-8 summarises the whole algorithm to solve multiple systems of equations of type (6.43) in a computationally efficient way. An computationally efficient implementation of the FFT is available from the fftw-library [122].

6.2   Joint Estimation of Concentrated Propagation Paths and DMC

So far, we have only derived estimators for the independent estimation of $\mathbf{\theta}_{sp}$ and $\mathbf{\theta}_{dan}$. However, the parameter estimation problem to solve is the joint estimation of both parameter vectors, i.e., of

$$\mathbf{\theta}_{chn} = \begin{bmatrix} \mathbf{\theta}_{sp}^{\mathrm{T}} & \mathbf{\theta}_{dan}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}.$$

Therefore, we derive in the next section a joint estimator for all channel parameters. We use the global search algorithm for the raw parameters of new propagation paths outlined in Section 5.1.5, and the global search algorithm for the parameters of the DMC outlined in Section 6.1.8. For local maximization of the likelihood function, we apply the iterative maximum likelihood estimator for the parameters of the propagation paths derived in Section 5.2.4, and the iterative maximum likelihood estimator for the parameters of the DMC derived in Section 6.1.5.

Table 6-8: Fast Matrix Multiplication for Inverse Toeplitz Matrices

Preprocessing: Calculate the diagonal elements $\mathbf{z}$ of the matrix $\mathbf{\Lambda}$ (6.53) using the FFT as $\mathbf{z} = \sqrt{2M} \cdot \mathbf{F}^H \cdot \begin{bmatrix} \mathbf{x}^{M \times 1} \\ \mathbf{0}^{M \times 1} \end{bmatrix}$, in $O(M \log(M))$ operations.

For every column $\mathbf{a}_n$ of the matrix $\mathbf{A}$ do:

1) Compute $\mathbf{z}_1$ using the FFT as

$\mathbf{z}_1 = \mathbf{F}^H \cdot \begin{bmatrix} \mathbf{a}_n \\ \mathbf{\Pi} \cdot \mathbf{a}_n^* \end{bmatrix}$, in $O(M \log(M))$ operations.

2) Compute $\mathbf{z}_2$ using the FFT as

$\mathbf{z}_2 = \mathbf{F} \cdot (\mathbf{z}^* \circ \mathbf{z}_1)$, in $O(M \log(M))$ operations.

3) Compute $\mathbf{z}_3$ using the FFT as

$\mathbf{z}_3 = \mathbf{F}^H \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \mathbf{z}_2$, in $O(M \log(M))$ operations.

4) Compute $\begin{bmatrix} \mathbf{w}_1^T & \mathbf{w}_2^T \end{bmatrix}^T$ using the FFT as

$\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{F} \cdot (\mathbf{z} \circ \mathbf{z}_3)$, in $O(M \log(M))$.

5) Compute the solution $\mathbf{b}_n = \mathbf{T}^{-1} \mathbf{a}_n$

$\mathbf{b} = \dfrac{1}{x_1} (\mathbf{w}_1 - \mathbf{\Pi} \cdot \mathbf{w}_2^*)$, in $O(M)$ operations.

6.2.1   Joint Maximum Likelihood Estimation (RIMAX)

As already discussed in Section 4.4, the Fisher information matrix (4.98) of the complete parameter set $\mathbf{\theta}_{chn}$ is a block diagonal matrix with one block related to $\mathbf{\theta}_{sp}$ and another one related to the parameters $\mathbf{\theta}_{dan}$. Recalling the discussion about parameter coupling and the choice of parameter index sets in the sense of the genuine SAGE algorithm [62] in Section (5.3) we choose two non-overlapping parameter sets $\mathbf{\theta}_{sp}$ and $\mathbf{\theta}_{dan}$. Applying the SAGE idea, we alternate between the maximization of the log-likelihood function

$$\mathcal{L}(\mathbf{x}|\mathbf{\theta}_{sp}, \mathbf{\theta}_{dan}) = -M \ln(\pi) - \ln(\det(\mathbf{R}(\mathbf{\theta}_{dan}))) - (\mathbf{x} - \mathbf{s}(\mathbf{\theta}_{sp}))^H \cdot \mathbf{R}^{-1}(\mathbf{\theta}_{dan}) \cdot (\mathbf{x} - \mathbf{s}(\mathbf{\theta}_{sp})) \qquad (6.56)$$

with respect to the parameters of the concentrated propagation paths and with respect to the parameters of the DMC and measurement noise. Since both parameter sets are not strongly coupled the alternating maximization procedure will converge as fast as a joint iterative maximum likelihood procedure (cf. Section 6.1.2 equation (6.2)). For the maximization of (6.56) with respect to $\mathbf{\theta}_{dan}$, given an estimate $\hat{\mathbf{\theta}}_{sp}$, we apply the estimator outlined in Table 6-5. For the update of the parameters $\hat{\mathbf{\theta}}_{sp}$ given the parameter estimates $\hat{\mathbf{\theta}}_{dan}$, we use the Levenberg-Marquardt algorithm summarised in Table 5-3. The basic concept of the maximum likelihood estimator is outlined in Figure 6-12.

$$\hat{\boldsymbol{\theta}}_{sp} = \arg\min_{\boldsymbol{\theta}_{sp}}\left((\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}_{sp}))^{\mathrm{H}}\cdot\mathbf{R}^{-1}(\hat{\boldsymbol{\theta}}_{dan})\cdot(\mathbf{x}-\mathbf{s}(\boldsymbol{\theta}_{sp}))\right) \longrightarrow \hat{\mathbf{x}}_{dan} = \mathbf{x}-\mathbf{s}(\hat{\boldsymbol{\theta}}_{sp})$$

$$\hat{\boldsymbol{\theta}}_{dan} = \arg\max_{\boldsymbol{\theta}_{dan}}\left(-\ln(\det(\mathbf{R}(\boldsymbol{\theta}_{dan})))-\hat{\mathbf{x}}_{dan}^{\mathrm{H}}\mathbf{R}^{-1}(\boldsymbol{\theta}_{dan})\hat{\mathbf{x}}_{dan}\right)$$

Figure 6-12: Concept for joint maximum likelihood estimation of $\boldsymbol{\theta}_{chn}$.

For the selection of the model order, we use the estimated relative variance of the estimated path weights as described in Section 5.2.7 and drop the propagation paths that are unreliable. The initial estimates are chosen as follows. For the DMC parameters, we use the estimates from the previous observation if available, or employ the global search algorithm outlined in Section 6.1.8. The structural parameters of the propagation paths are initialized using the structural parameters of the previous observation. The related path weights are computed using the closed form solution (5.5) (BLUE). The contribution of the tracked propagation paths is removed from the radio channel observation, and a search for propagation paths that have become visible is carried out using the algorithm outlined in Section 5.1.5. For the channel sounding measurements processed so far, it was sufficient to search for five new propagation paths. One should observe that the global search algorithm for new propagation paths is computationally expensive. Hence, it is important to limit the search for new propagation paths to a reasonable small number of paths. The structure of the RIMAX algorithm is outlined in Figure 6-13.

The number of global iterations required by RIMAX varies strongly. The number of iterations depends heavily on the observation. If the number of assessable propagation paths is small, the SNR of the individual propagation paths is high, and if all paths are well separated, the number of iterations required can be as low as 20. If the scenario is complicated, i.e., the number of propagation paths is high and a lot of closely spaced propagation paths exist in the scenario, the number of iterations required to achieve convergence can be higher than 100.

In Figure 6-14 the PDP of a measured SIMO channel and the related SIMO impulse response reconstructed from the estimated propagation path parameters is shown. The measurement has been carried out in a street micro-cell scenario at 5.2GHz with an 8-element ULA. The next figure shows the PDP after removing the contribution of the concentrated propagation paths, i.e., $\hat{\mathbf{x}}_{dan}$ and the estimated PDP of the DMC. Finally, Figure 6-16 shows the PDP of $\hat{\mathbf{x}}_{dan}$ after whitening with the estimated covariance matrix $\mathbf{R}_{nn}$. This PDP shows that the new data model, describing the radio channel as a superposition of concentrated propagation paths and DMC, is a reasonably good approximation. It is also clearly visible that the data model can be further improved. The variance of the PDP between 800ns and 3000ns is higher then the variance in the other parts of PDP. However, the data model is suitable if the SNR of the measurement is lower. An example is shown in Figure 6-17 to Figure 6-19. Altogether, the derived algorithm (RIMAX) is a significant improvement compared to all existing high-resolution channel parameter estimation algorithms due to the enhanced data model. The RIMAX algorithm has been published in [123], [124], and [136].
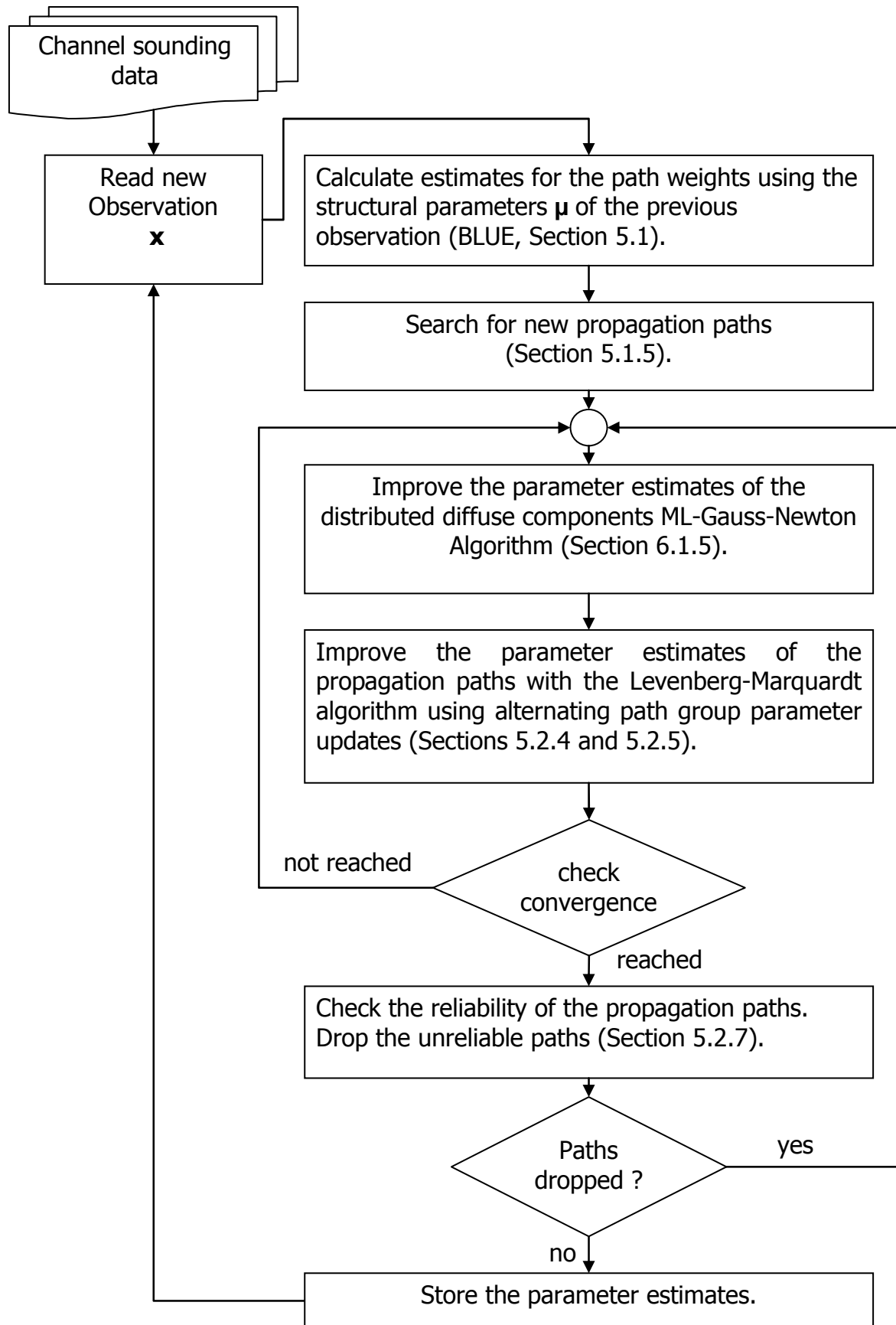
Figure 6-13: Outline of the RIMAX structure
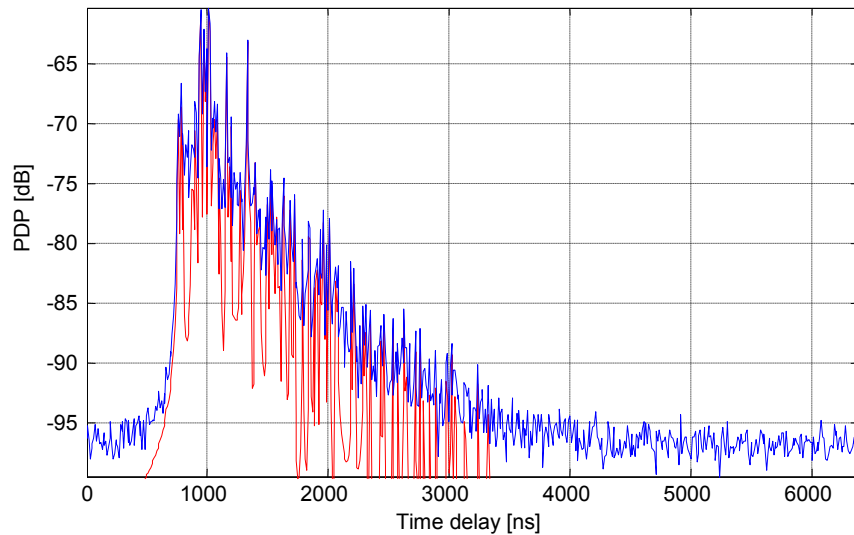
Figure 6-14: Example for the PDP of a measured impulse response and the related estimated concentrated propagation paths.

Figure 6-15: Example for the PDP of the remainder of a measured impulse response after removing the estimated concentrated propagation paths, and the estimated PDP of the DMC.

Figure 6-16: Example for the PDP of the remainder of a measured impulse response after removing the estimated concentrated propagation paths and whitening.

Figure 6-17: Example for the PDP of a measured impulse response and of the estimated concentrated propagation paths.

Figure 6-18: Example for the PDP of the remainder of a measured impulse response after removing the estimated concentrated propagation paths, and the estimated PDP of the DMC.

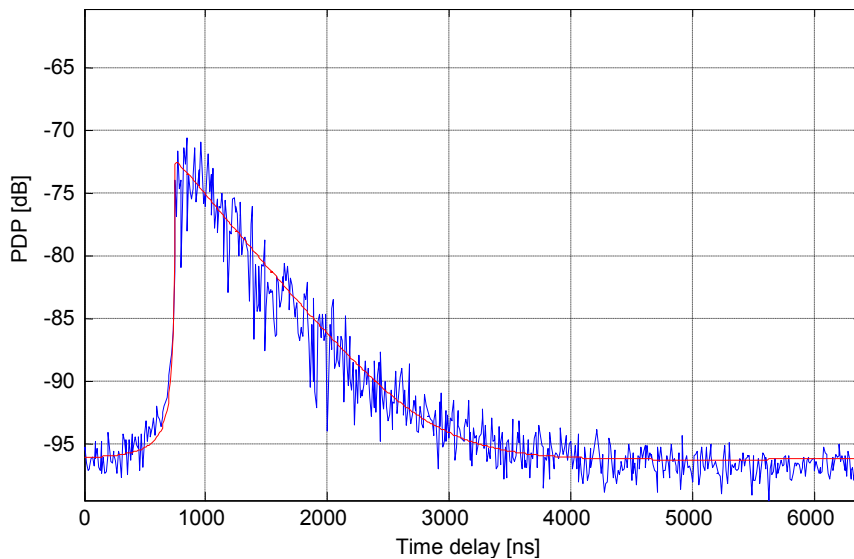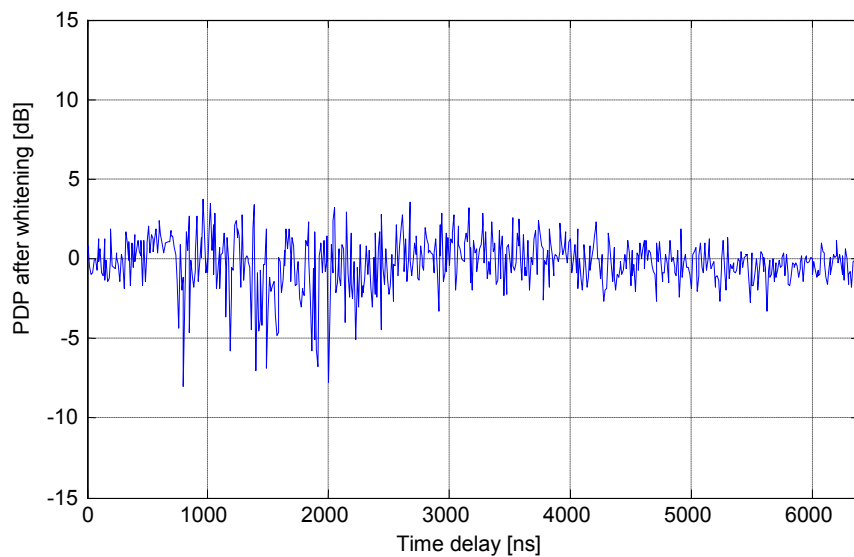

Figure 6-19: Example for the PDP of the remainder of a measured impulse response after removing the estimated concentrated propagation paths and whitening.
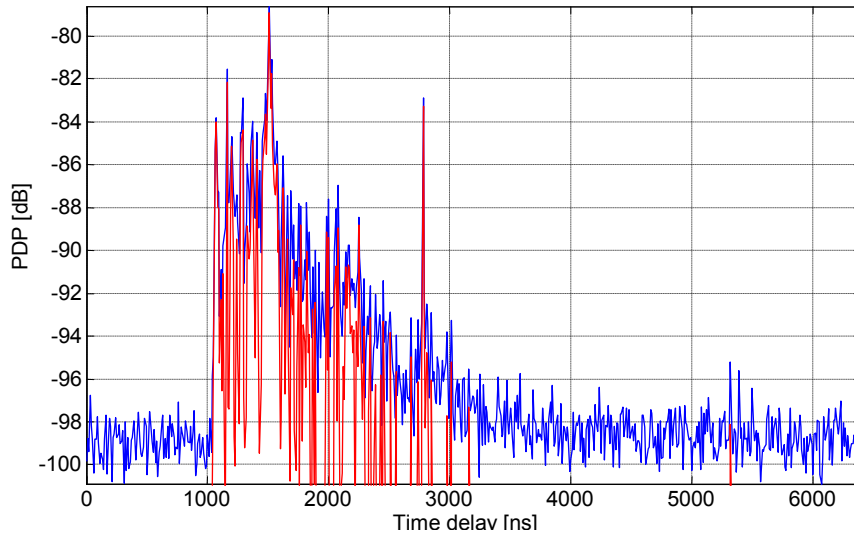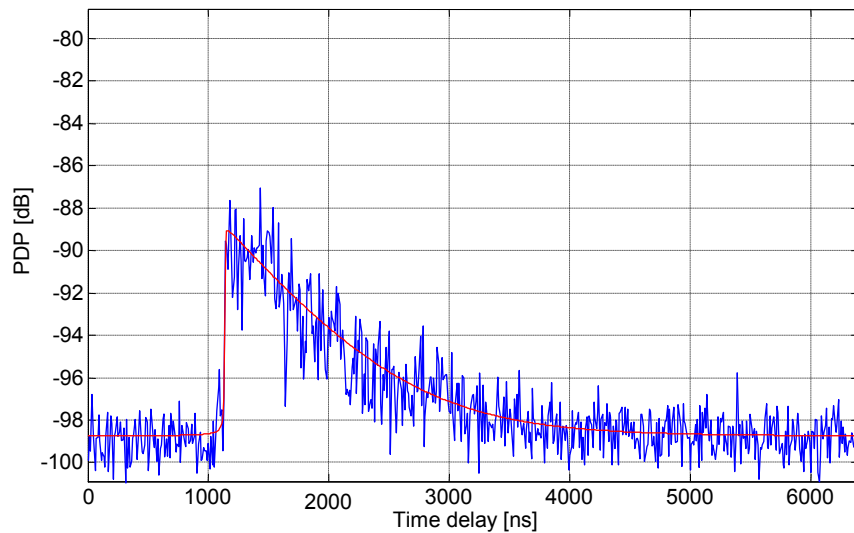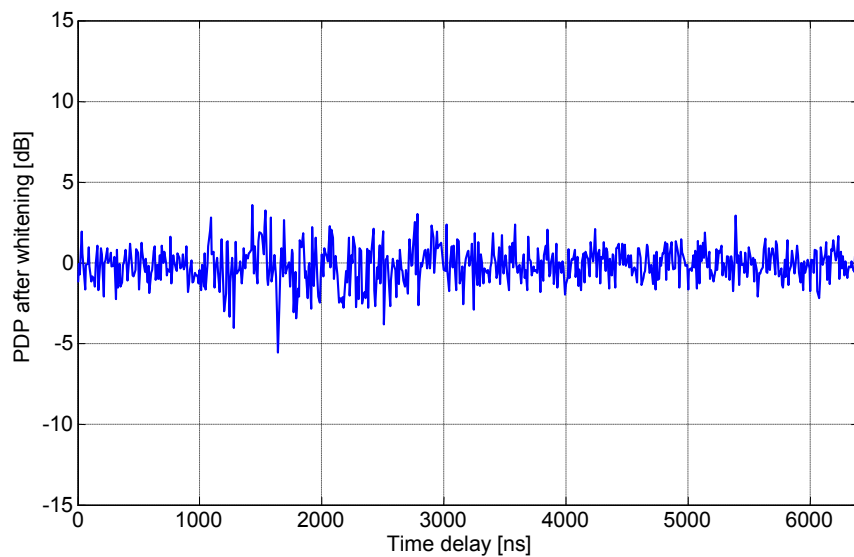
6.2.2 Application of Subspace Based Algorithms for Parameter Estimation in the Presence of DMC

The iterative nature of the estimator for the concentrated propagation paths and the DMC outlined in Section 5.2 and 6.1.5 is a prerequisite for the joint estimation of all channel parameters $\boldsymbol{\theta}_{chn}$. In contrast the joint estimation of the parameters $\boldsymbol{\theta}_{sp}$ and $\boldsymbol{\theta}_{dan}$ is not feasible if a subspace-based algorithm is applied to the estimation of the propagation path parameters $\boldsymbol{\theta}_{sp}$. For the estimation of the signal subspace needed for the estimation of the propagation path parameters, the information about the covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{dan})$ is necessary. However, if the parameters $\boldsymbol{\theta}_{sp}$ are not available, the parameters $\boldsymbol{\theta}_{dan}$ of the covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{dan})$ cannot be estimated. If a sequence of measurements is processed a less than optimal approach can be used to resolve this problem.

One can use the parameter estimates from the previous observation $\boldsymbol{\theta}_{dan}$ and estimate the signal subspace needed for the estimation of $\boldsymbol{\theta}_{sp}$. Having an estimate of the channel parameters $\boldsymbol{\theta}_{sp}$, an estimate of the parameters $\boldsymbol{\theta}_{dan}$ can be determined. However, the estimate $\hat{\boldsymbol{\theta}}_{chn} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{sp}^{T} & \hat{\boldsymbol{\theta}}_{dan}^{T} \end{bmatrix}^{T}$ determined in this way is clearly not a joint estimate of the channel parameters $\boldsymbol{\theta}_{chn}$. In principle, the same approach to joint maximum likelihood estimation as used in the RIMAX algorithm can be applied to compute a joint estimate of the complete channel parameter set. The outlined procedure for the sequential estimation of $\boldsymbol{\theta}_{sp}$, using the subspace based algorithm of choice, and $\boldsymbol{\theta}_{dan}$ has to be carried out in an alternating manner. However, the resulting iterative algorithm is computationally expensive, since we have to perform the complete subspace based parameter estimation algorithm in every iteration. Currently, no subspace based algorithm is known, which can take advantage of an initial solution to speed up the computation substantially. This consideration also applies to parameter tracking from observation to observation. Whereas the RIMAX algorithm can in a natural way, use the parameter estimates from the previous observation as initial values for the actual observation, this a priori information can typically not be used in known subspace-based algorithms. However, since basic algorithms used in, e.g., the multidimensional Unitary ESPRIT are all of iterative nature, the development of an algorithm that takes advantage of previously computed results seems feasible.

6.2.3 Estimation of an initial solution without a priori information

The estimation of an initial solution $\hat{\boldsymbol{\theta}}^{\{0\}}$ for the very first observation (snapshot) of a sequence of channel measurements is critical, since no a-priori information about the channel-parameters is available. The constructed estimator $\hat{\boldsymbol{\theta}}_{dmc}$ (cf. Table 6-5 and Table 6-6) is based on the assumption that the observation of the channel has zero mean

$$\mathbf{x} \sim \mathcal{N}_C\left(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_{dmc})\right) \in \mathbb{C}^{M \times 1},$$

i.e. contains dense multipath components only. The estimator can handle weak concentrated propagation paths especially if frequency domain smoothing is applied during the computation of initial values for $\hat{\boldsymbol{\theta}}_{dmc}$ (cf. Section 6.1.9). However, it will produce unusable estimates if the data fed into the estimator contains strong propagation paths such as the line of sight path. Therefore, one should start with the estimation of some propagation paths at the beginning. Since no a priori information about the parameters of the covariance matrix $\mathbf{R}(\boldsymbol{\theta}_{dmc})$ is available we initialize it with a scaled version of the identity matrix

$$\mathbf{R}(\boldsymbol{\theta}_{dmc}) = \hat{\alpha}_0 \mathbf{I},$$

and choose

$$\hat{\alpha}_0 = \frac{1}{M} \mathbf{x}^H \mathbf{x}.$$

This means that initially we set the estimated of the maximum power of the dense multipath components to $\hat{\alpha}_1 = 0$. The assumption that the stochastic part of the observation is white in all data dimensions has clearly only a small impact on the initial estimates of the strong propagation paths. Having estimated raw parameters $\hat{\boldsymbol{\theta}}_{sp}^{\{0\}}$ of the strongest propagation paths, their contribution can be removed from the observation. The remainder of the observation is then used to estimate initial values for the dense multipath components $\hat{\boldsymbol{\theta}}_{dmc}^{\{0\}}$, using the algorithm outlined in Table 6-6. Both parameter sets $\hat{\boldsymbol{\theta}}_{sp}^{\{0\}}$ and $\hat{\boldsymbol{\theta}}_{dmc}^{\{0\}}$ together yield an initial estimate of the channel parameters $\hat{\boldsymbol{\theta}}_{chn}^{\{0\}}$.

Often it is sufficient to estimate only the strongest propagation path of the observation, before estimating initial parameters of the dense multipath components $\hat{\boldsymbol{\theta}}_{sp}^{\{0\}}$. As a rule of thumb, it is sufficient to search for the $P^{\{0\}} = 5$ strongest propagation paths.

6.2.4 General Limitations

The joint estimation of the parameters of the dense multipath components and of the propagation paths from a SISO channel sounding measurement taken only between two points in space, is not possible without systematic errors. The basic problem here is that only one realisation of the dense multipath components is available. Consequently, we cannot distinguish between the contributions of the propagation paths and of the dense multipath components. Hence, if SISO channel sounding measurements are carried out, it is necessary to move Tx or Rx while measuring the radio channel.

It is generally better to carry out measurements with a moving transmitter and/or receiver. This is due to the stochastic nature of the radio channel. As mentioned in Section 2.5.4, the radio channel is a spatial variant stochastic process. Therefore, taking measurements while moving transmitter and/or receiver over some wavelength, ensures that the complete measurement contains various realisations of the radio channel.

6.3 Conclusions on Parameter Estimation

The joint estimation of the parameters of the dense multipath components and of the propagation paths can be carried out by the RIMAX algorithm outlined in Section 6.2.1. The estimate of the relative variance of the propagation path weights can be used as a reliability measure and for model order selection, i.e., to determine the number of assessable propagation paths. The algorithm has been implemented at the Electronic Measurement Lab at Technische Universität Ilmenau.

The algorithm has been used for statistical analysis of channels sounding campaigns in various scenarios. See for example the results presented in [137] and [138] . Furthermore, the algorithm has been used to evaluate the significance of dense multipath components in real radio channels. In [138] it has been shown that the DMC dominate the radio propagation in some scenarios, whereas the specular propagation paths may dominate in others. Altogether, the outlined parametric channel estimator has proved to be a reliable tool for the analysis of radio propagation channels.

7 Antenna Array Calibration

In practice, the output signals of antenna arrays do not strictly conform to the theoretical antenna array model. There are a variety of reasons for this difference mainly mutual coupling between the antenna elements and geometrical as well as electrical tolerances of the single elements. Often, the mapping of the ideal array manifold vector $\mathbf{a}(\mu)$ to the array manifold vector $\widetilde{\mathbf{a}}(\mu)$ of a real antenna array (ULA, URA, CUBA, UCA) can be, to some extend, expressed via a transformation matrix $\mathbf{K}$ if the mismatch is independent of the spatial frequency μ and the signals are assumed to be narrowband, i.e.,

$$\widetilde{\mathbf{a}}(\mu) = \mathbf{K} \cdot \mathbf{a}(\mu) \tag{7.1}$$

Therefore, the data measured with a real antenna array (ULA, URA, CUBA, UCA) have hidden rotational invariance structure cf. Section 5.3.11. We can reduce the influence of the antenna array imperfections on the parameter estimates if we estimate the transformation matrix $\mathbf{K}$ or even better its inverse $\mathbf{C}$ from reference measurements and use it to calibrate the antenna array. Let us assume that we measure a set of N array steering vectors $\mathbf{a}(\mu_n)$ over the whole angular domain at known spatial frequencies μ_n, $1 \le n \le N$. The noise corrupted measurements will be denoted as $\widetilde{x}_n$. By taking into account that every calibration measurement contains one unknown complex factor γ_n due to the unknown complex beam patterns of the antenna array elements, phase drifts, and so on, the measurement data is modelled in the following way

$$\widetilde{\mathbf{x}}_n = \mathbf{K} \cdot \gamma_n \cdot \mathbf{a}(\mu_n) + \mathbf{n}_n . \tag{7.2}$$

Let us collect all measurements in the matrix

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \widetilde{\mathbf{x}}_1 & \widetilde{\mathbf{x}}_2 & \dots & \widetilde{\mathbf{x}}_N \end{bmatrix},$$

all ideal (sampled) array manifold vectors (or array steering vectors) in

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}(\mu_1) & \mathbf{a}(\mu_2) & \dots & \mathbf{a}(\mu_N) \end{bmatrix},$$

and the related unknown weighting factors γ_n in the diagonal matrix

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \gamma_N \end{bmatrix}.$$

Then we get the following linear data model for all measurements

$$\widetilde{\mathbf{X}} = \mathbf{K} \cdot \mathbf{A} \cdot \mathbf{\Gamma} + \mathbf{W} , \tag{7.3}$$

where $\mathbf{W}$ is assumed to model complex circular i.i.d. measurement noise. The objective is to estimate the calibration matrix $\hat{\mathbf{C}}$ such that

$$\left\| \hat{\mathbf{C}} \cdot \mathbf{K} - \mathbf{I} \right\|_F^2$$

is minimized. Due to the unknown parameters γ_n, this problem cannot be solved directly.

In the following a generalization of the idea proposed in [125] is used to reformulate the optimization problem. For any array steering vector $\mathbf{a}(\boldsymbol{\mu}_n)$ of length M, we can construct a set of $L = M - 1$ linear independent vectors

$$\mathbf{B}_n = \begin{bmatrix} \mathbf{b}_{n,1} & \mathbf{b}_{n,2} & \dots & \mathbf{b}_{n,L} \end{bmatrix}$$

that span the left null space of $\mathbf{a}(\boldsymbol{\mu}_n)$, i.e.,

$$\mathbf{B}_n^H \cdot \mathbf{a}(\boldsymbol{\mu}_n) = \mathbf{0} \, .$$

An orthogonal basis of the left null space is

$$\mathbf{b}_{n,m} = \boldsymbol{\Omega}^m \cdot \mathbf{z}_n \, , \tag{7.4}$$

where the diagonal matrix

$$\boldsymbol{\Omega} = \frac{1}{\sqrt{M}} \begin{bmatrix} e^{-j \cdot 2 \cdot \pi \cdot \frac{1}{M} \frac{-M+1}{2}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & e^{-j \cdot 2 \cdot \pi \cdot \frac{1}{M} \frac{M-1}{2}} \end{bmatrix}$$

and the vectors $\mathbf{z}_n$ are chosen such that

$$\mathbf{z}_n^* \circ \mathbf{a}(\boldsymbol{\mu}_n) = \mathbf{1} \, .$$

Recall that $\circ$ denotes the Hadamard-product (element-wise product) and

$$\mathbf{1}^T = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} .$$

Thus, the calibration matrix $\hat{\mathbf{C}}$ should be chosen such that

$$\mathbf{b}_{n,m}^H \underbrace{\hat{\mathbf{C}} \cdot \mathbf{K}}_{\mathbf{I}_M} \cdot \mathbf{a}(\mu_n) = 0, \quad \forall m, n \, .$$

Using the noise corrupted measurements in (7.2), we, therefore, have to minimize

$$\left\| \mathbf{b}_{n,m}^H \cdot \hat{\mathbf{C}} \cdot \widetilde{\mathbf{x}}_n \right\|_F^2 = \left(\mathbf{b}_{n,m}^H \cdot \hat{\mathbf{C}} \cdot \widetilde{\mathbf{x}}_n \right) \cdot \left(\widetilde{\mathbf{x}}_n^H \cdot \hat{\mathbf{C}}^H \cdot \mathbf{b}_{n,m} \right), \quad \forall m, n \, . \tag{7.5}$$

Hence, the calibration matrix $\hat{\mathbf{C}}$ is obtained as

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \sum_{m=1}^{M-1} \sum_{n=1}^{N} \left(\mathbf{b}_{n,m}^H \cdot \mathbf{C} \cdot \widetilde{\mathbf{x}}_n \right) \cdot \left(\widetilde{\mathbf{x}}_n^H \cdot \mathbf{C}^H \cdot \mathbf{b}_{n,m} \right).$$

Applying the $\text{vec}\{\bullet\}$-operator to $\mathbf{C}^H \mathbf{b}_{n,m}$ and its Hermitian transpose, we obtain

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \sum_{m=1}^{M-1} \sum_{n=1}^{N} \text{vec}(\mathbf{C})^T \cdot \left(\mathbf{I}_M \otimes \mathbf{b}_{n,m}^* \right) \cdot \widetilde{\mathbf{x}}_n \cdot \widetilde{\mathbf{x}}_n^H \cdot \left(\mathbf{I}_M \otimes \mathbf{b}_{n,m}^T \right) \cdot \text{vec}(\mathbf{C})^* \, .$$

Since

$$\text{vec}\{\mathbf{Y}_1 \mathbf{Y}_2 \mathbf{Y}_3\} = \left(\mathbf{Y}_3^T \otimes \mathbf{Y}_1\right) \text{vec}\{\mathbf{Y}_2\}$$

if $\mathbf{Y}_1 \in \mathbb{C}^{y_1 \times y_2}$, $\mathbf{Y}_2 \in \mathbb{C}^{y_2 \times y_3}$, and $\mathbf{Y}_3 \in \mathbb{C}^{y_3 \times y_4}$.

With the definition $\mathbf{c} = \text{vec}\{\mathbf{C}\}$ we get

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \left(\sum_{m=1}^{M-1} \sum_{n=1}^{N} \left(\mathbf{I} \otimes \mathbf{w}_{n,m}\right) \cdot \tilde{\mathbf{x}}_n^* \cdot \tilde{\mathbf{x}}_n^T \cdot \left(\mathbf{I} \otimes \mathbf{w}_{n,m}^H\right) \right) \cdot \mathbf{c}.$$

The choice of $\mathbf{b}_{n,m}$ in (7.4) and some rearrangements lead to

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \left(\sum_{m=1}^{M-1} \left(\mathbf{I}_M \otimes \boldsymbol{\Omega}^m\right) \cdot \left(\sum_{n=1}^{N} \left(\mathbf{I}_M \otimes \mathbf{z}_n\right) \cdot \tilde{\mathbf{x}}_n^* \cdot \tilde{\mathbf{x}}_n^T \cdot \left(\mathbf{I}_M \otimes \mathbf{z}_n^H\right) \right) \cdot \left(\mathbf{I}_M \otimes \boldsymbol{\Omega}^m\right)^H \right) \cdot \mathbf{c}. \qquad (7.6)$$

By utilizing the $\text{vec}\{\bullet\}$-operator again, equation (7.6) can be expressed as

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \left(\sum_{m=1}^{M-1} \left(\mathbf{I} \otimes \boldsymbol{\Omega}^m\right) \cdot \left(\sum_{n=1}^{N} \text{vec}\left(\mathbf{z}_n \cdot \tilde{\mathbf{x}}_n^H\right) \cdot \text{vec}\left(\mathbf{z}_n \cdot \tilde{\mathbf{x}}_n^H\right)^H \right) \cdot \left(\mathbf{I} \otimes \boldsymbol{\Omega}^m\right)^H \right) \cdot \mathbf{c}.$$

Now, let us replace the inner sum by the correlation matrix $\mathbf{R}_{zxzx}$ defined as

$$\mathbf{R}_{zxzx} = \frac{1}{N} \sum_{n=1}^{N} \text{vec}\left\{\mathbf{z}_n \cdot \tilde{\mathbf{x}}_n^H\right\} \cdot \text{vec}\left\{\mathbf{z}_n \cdot \tilde{\mathbf{x}}_n^H\right\}^H$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{z}_n \cdot \mathbf{z}_n^H\right) \otimes \left(\tilde{\mathbf{x}}_n^* \cdot \tilde{\mathbf{x}}_n^T\right).$$

Note that this correlation matrix can directly be computed from the observation using the measured data $\tilde{\mathbf{x}}_n$ and the ideal steering vectors $\mathbf{a}(\mu_n)$ calculated from the known spatial frequencies μ_n. Consequently, we have the following compact form of the optimisation problem

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \left(\sum_{m=1}^{M-1} \left(\mathbf{I}_M \otimes \boldsymbol{\Omega}^m\right) \cdot \mathbf{R}_{zxzx} \cdot \left(\mathbf{I}_M \otimes \boldsymbol{\Omega}^m\right)^H \right) \cdot \mathbf{c}.$$

The multiplications of the $M \times M$ block matrices of $\mathbf{R}_{zxzx}$ by the $M-1$ diagonal matrices $\boldsymbol{\Omega}^m$ can also be expressed as

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \left(\left(\mathbf{1}^{(M \times M)} \otimes \left(\mathbf{F} \cdot \mathbf{F}^H - \frac{1}{M} \mathbf{1}^{(M \times M)} \right) \right) \circ \mathbf{R}_{zxzx} \right) \cdot \mathbf{c}, \qquad (7.7)$$

where $\mathbf{1}^{(M \times M)} = \mathbf{1} \cdot \mathbf{1}^T \in \mathbb{R}^{M \times M}$. Equation (7.7) holds since for an arbitrary matrix $\mathbf{R}$

$$\sum_{m=1}^{M-1} \boldsymbol{\Omega}^m \cdot \mathbf{R} \cdot \left(\boldsymbol{\Omega}^m\right)^H = \left(\mathbf{T} \cdot \mathbf{T}^H\right) \circ \mathbf{R} = \left(\mathbf{F} \cdot \mathbf{F}^H - \tfrac{1}{M} \mathbf{1}^{(M \times M)}\right) \circ \mathbf{R},$$

where the matrix $\mathbf{T}$ is given by

$$\mathbf{T} = \left[\text{diag}\{\boldsymbol{\Omega}^1\} \ldots \text{diag}\{\boldsymbol{\Omega}^{M-1}\}\right] \in \mathbb{C}^{M \times M-1}$$

and the DFT-matrix $\mathbf{F}$ is defined as

$$\mathbf{F} = \left[\text{diag}\{\boldsymbol{\Omega}^0\} \ldots \text{diag}\{\boldsymbol{\Omega}^{M-1}\}\right] \in \mathbb{C}^{M \times M}.$$

Since the DFT matrix $\mathbf{F}$ is unitary, equation (7.7) can be simplified further to

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \underbrace{\left(\left(\mathbf{1}^{(M \times M)} \otimes \mathbf{I}_M \right) \circ \mathbf{R}_{zxzx} - \frac{1}{M} \cdot \mathbf{R}_{zxzx} \right)}_{\mathbf{R}'_{zxzx}} \cdot \mathbf{c} ,$$

so finally we end up with the simple eigenvalue problem

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{c}^H \cdot \mathbf{R}'_{zxzx} \cdot \mathbf{c} ,$$

where we have defined the matrix $\mathbf{R}'_{zxzx}$ as

$$\mathbf{R}'_{zxzx} = \left(\mathbf{1}^{(M \times M)} \otimes \mathbf{I}_M \right) \circ \mathbf{R}_{zxzx} - \frac{1}{M} \cdot \mathbf{R}_{zxzx} . \tag{7.8}$$

The solution of the minimization problem $\hat{\mathbf{c}}$ is given by the eigenvector corresponding to the smallest eigenvalue of

$$\mathbf{R}'_{zxzx} \cdot \hat{\mathbf{c}} - \lambda_{\min} \cdot \mathbf{I}_{M^2} \cdot \hat{\mathbf{c}} = \mathbf{0} . \tag{7.9}$$

Table 7-1 summarises the complete algorithm to estimate the calibration matrix $\hat{\mathbf{C}}$.

Table 7-1: Summary of the Calibration Algorithm

Data Matrix calculation

1) Compute the matrix $\mathbf{R}_{zxzx}$ using the reference measurements $\widetilde{\mathbf{x}}_n$

$$\mathbf{R}_{zxzx} = \frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{z}_n \cdot \mathbf{z}_n^H \right) \otimes \left(\widetilde{\mathbf{x}}_n^* \cdot \widetilde{\mathbf{x}}_n^T \right)$$

2) Compute the data matrix $\mathbf{R}'_{zxzx}$ from $\mathbf{R}_{zxzx}$ as

$$\mathbf{R}'_{zxzx} = \left(\mathbf{1}^{(M \times M)} \otimes \mathbf{I}_M \right) \circ \mathbf{R}_{zxzx} - \frac{1}{M} \cdot \mathbf{R}_{zxzx}$$

3) Solution to the Minimization Problem:
Compute the eigenvectors $\hat{\mathbf{c}}$ corresponding to the smallest eigenvalue of
$$\mathbf{R}'_{zxzx} \cdot \hat{\mathbf{c}} - \lambda_{\min} \cdot \mathbf{I}_{M^2} \cdot \hat{\mathbf{c}} = \mathbf{0}$$

4) Solution Reshaping:
Reshape the vector $\hat{\mathbf{c}}$ to the calibration matrix $\hat{\mathbf{C}}$ using $\hat{\mathbf{C}} = \mathrm{mat}(\hat{\mathbf{c}})$.

If the SNR is low or the estimation problem is ill conditioned, the calibration algorithm can be improved by taking the influence of our vectors $\mathbf{z}_n$ onto the noise correlation into account. To this end, we introduce the noise covariance matrix

$$\mathbf{R}_{znzn} = \mathrm{E}\left\{ \sum_{n=1}^{N} \mathrm{vec}\left\{ \mathbf{z}_n \cdot \mathbf{n}_n^H \right\} \cdot \mathrm{vec}\left\{ \mathbf{z}_n \cdot \mathbf{n}_n^H \right\}^H \right\} .$$

Furthermore we assume that the measurement noise is i.i.d. complex circular Gaussian noise with variance σ_n^2. Consequently, the noise covariance matrix can be expressed as

$$\mathbf{R}_{znzn} = \sigma_n^2 \cdot \left(\mathbf{I} \otimes \mathbf{R}_{zz} \right) \tag{7.10}$$

with

$$\mathbf{R}_{zz} = \frac{1}{N}\left(\sum_{n=1}^{N} \mathbf{z}_n \cdot \mathbf{z}_n^H\right).$$

Replacing $\mathbf{R}_{zxzx}$ in (7.8) by $\mathbf{R}_{znzn}$ from (7.10) yields the following expression for the structure of the noise covariance matrix

$$\mathbf{R}_{znzn}^{'} = \mathbf{I}_{M^2} \circ \mathbf{R}_{zn} - \tfrac{1}{M}\mathbf{R}_{znzn}.$$

With this knowledge we are able to improve our estimate of $\hat{\mathbf{C}}$. We extend the special eigenvalue problem (7.9) to a generalised eigenvalue problem, taking the noise colouring through $\mathbf{z}_n$ into account. The optimum solution to the minimization problem is given by the eigenvector corresponding to the smallest eigenvalue $\lambda_{\min}$ of the following generalised eigenvalue problem

$$\mathbf{R}_{zxzx}^{'} \cdot \hat{\mathbf{c}} - \mathbf{R}_{znzn}^{'} \cdot \hat{\mathbf{c}} \cdot \lambda_{\min} = \mathbf{0}.$$

The knowledge of the noise variance σ^2 itself is not required if we want to find the optimum solution $\hat{\mathbf{C}}$. The complete algorithm to estimate the calibration matrix $\hat{\mathbf{C}}$ considering noise colouring is summarised in Table 7-2.

Table 7-2: Summary of the Calibration Algorithm considering Noise Colouring

Data Matrix calculation

1) Compute the matrix $\mathbf{R}_{zxzx}$ using the reference measurements $\widetilde{\mathbf{x}}_n$

$$\mathbf{R}_{zxzx} = \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{z}_n \cdot \mathbf{z}_n^H\right) \otimes \left(\widetilde{\mathbf{x}}_n^* \cdot \widetilde{\mathbf{x}}_n^T\right)$$

2) Compute the data matrix $\mathbf{R}_{zxzx}^{'}$ from $\mathbf{R}_{zxzx}$ as

$$\mathbf{R}_{zxzx}^{'} = \left(\mathbf{1}^{(M\times M)} \otimes \mathbf{I}_M\right) \circ \mathbf{R}_{zxzx} - \frac{1}{M}\cdot\mathbf{R}_{zxzx}$$

Noise Covariance Matrix calculation:

3) Compute the matrix $\mathbf{R}_{zwzw}$

$$\mathbf{R}_{znzn} = \mathbf{I} \otimes \mathbf{R}_{zz} \text{ with } \mathbf{R}_{zz} = \frac{1}{N}\left(\sum_{n=1}^{N}\mathbf{z}_n \cdot \mathbf{z}_n^H\right)$$

4) Compute the data matrix $\mathbf{R}_{znzn}^{'}$ from $\mathbf{R}_{znzn}$ using

$$\mathbf{R}_{znzn}^{'} = \mathbf{I}_{M^2} \circ \mathbf{R}_{zn} - \tfrac{1}{M}\mathbf{R}_{znzn}$$

5) Solution to the Minimization Problem:
$$\mathbf{R}_{zxzx}^{'} \cdot \hat{\mathbf{c}} - \mathbf{R}_{znzn}^{'} \cdot \hat{\mathbf{c}} \cdot \lambda_{\min} = \mathbf{0}$$

6) Solution Reshaping:
Reshape the vector $\hat{\mathbf{c}}$ to the calibration matrix $\hat{\mathbf{C}}$ using $\hat{\mathbf{C}} = \mathrm{mat}(\hat{\mathbf{c}})$.

Nevertheless, it is important to note that taking the noise colouring by $\mathbf{z}_n$ into account may not necessarily lead to improved estimates. This is because in practical situations expression (7.5) may be dominated by systematic errors such as positioning errors while taking the reference measurements and not by the measurement noise.

Another problem while estimating $\hat{\mathbf{C}}$ may arise if not enough independent information is available from the measurements to determine all of the coefficients in $\hat{\mathbf{C}}$. This will lead to an underdetermined problem. If the matrix $\mathbf{R}'_{zxzx}$ has not only one but several small eigenvalues with a similar magnitude then we have such a case. A good way to solve this issue is to take more measurements. It is, in this context, important to realise that taking measurements over a larger angular range is better than taking more measurements in the same angular range again. One can also solve the problem if a priori information about the structure of $\mathbf{K}$ is available. Usually mutual coupling effects between the antenna array elements dominate $\mathbf{K}$. Since elements with a larger distance in the array often have only small mutual coupling, we can set the corresponding coefficients in the data matrix $\mathbf{R}'_{zxzx}$ to zero. Consequently, the corresponding rows and columns in the data matrix $\hat{\mathbf{C}}$ as well as in the noise covariance matrix $\mathbf{R}'_{znzn}$ can be deleted, leading to a reduced and better-conditioned problem.

The main advantage of the outlined calibration algorithm is its low complexity. This makes it also suitable for the estimation of the calibration matrix of large antenna arrays, such as uniform rectangular arrays.

Figure 7-1 depicts the absolute values of the uncalibrated and calibrated beam patterns of a uniform linear array having eight active elements. On the left hand side only the mean gain of the antenna elements has been calibrated, which corresponds to equalisation with the main diagonal elements of the calibration matrix $\hat{\mathbf{C}}$. On the right hand side, all elements of the calibration matrix have been used to calibrate the antenna array. The resulting beam patterns are much more uniform. This calibration gain will lead to a significant reduction of estimation error in the (angle) estimates obtained with ESPRIT based parameter estimation algorithms. Further discussions about the influence of antenna array imperfections on ESPRIT based parameter estimation algorithms as well as antenna array calibration for channel sounding applications can be found in [28], [126], [127], and [128].
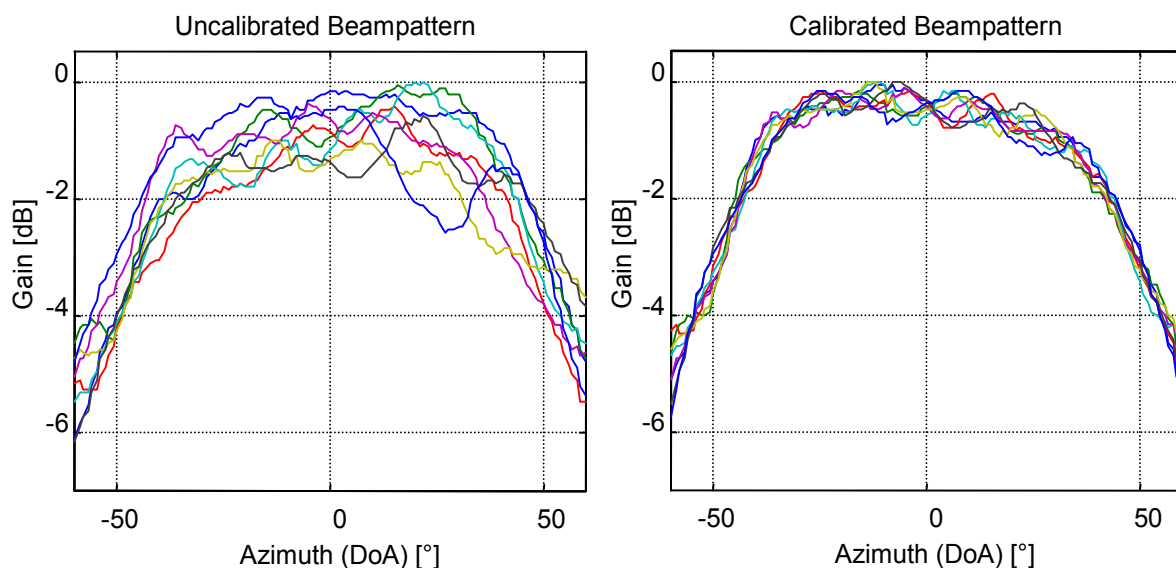
Figure 7-1: Example of an uncalibrated and a calibrated uniform linear antenna array having eight active antenna elements. The figure on the left hand side shows the beam pattern of all eight antenna elements without calibration. On the right hand side the beam pattern of all antenna elements after calibration is depicted.

8   Summary and Conclusions

8.1   Summary and Conclusion

With the RIMAX algorithm, a reliable tool is available to estimate channel parameters from channel sounding measurements. The algorithm is sufficiently robust for the processing of long radio channel measurements containing some thousands of channel observations, almost without user interaction.

To conclude, let me summarise the most important results of this thesis:

- An appropriate data model for parameter estimation from channel sounding measurements describes the radio channel as a superposition of a finite number of concentrated propagation paths and a circular normal distributed process. The circular normal distributed process describes the dense multipath components caused by distributed diffuse scattering.

- The effective aperture distribution function is a mean to express the antenna array response of an arbitrary antenna array. Using the EADF the data model for the concentrated propagation paths can be expressed in an algebraic form. The expression of the data model as a continuous algebraic function is a prerequisite for the application of gradient-based iterative maximum likelihood estimators and for the derivation of closed form expressions for the Cramér-Rao lower bound.

- The covariance matrix describing the zero-mean dense multipath components can be expressed as a continuous matrix-valued function of three channel parameters.

- The Cramér-Rao lower bound on all channel parameters can be expressed in a closed form. The estimation error covariance matrix depends on the coupling between propagation paths, precisely on the coupling of their parameters. The Fisher information matrix is a means to detect groups of coupled propagation paths.

- The estimation of the parameters of concentrated propagation paths is a nonlinear weighted least squares problem. The optimal weighting matrix is the covariance matrix of the dense multipath components and noise. A computationally efficient solution to solve this optimisation problem is provided by the Levenberg-Marquardt algorithm.

- The parameters of the propagation paths and the dense multipath components have to be estimated jointly. The joint maximum likelihood estimation of all radio channel parameters can be carried out using an alternating maximization procedure (SAGE).

- A feasible way to determine the number of propagation paths is given by the relative variance of the propagation path weights. The variance of all parameter estimates can

be estimated using the closed form expressions for the Cramér-Rao lower bound. The estimate of the parameter covariance matrix is an important result of the RIMAX algorithm. The covariance matrix provides important reliability information about the parameter estimates.

- The RIMAX algorithm is a means of estimating radio channel parameters from a long sequence of radio channel observations. The estimator tracks propagation paths over some time. Thus, it is a means to derive parameter statistics for individual propagation paths.

8.2   Further research areas

Since RIMAX provides additional reliability information for the parameters estimated. The parameters can be used to derive probability density functions for the parameters of the deterministic as well the dense multipath components of the radio channel. To this end, a large amount of channel sounding measurements gathered in various radio scenarios has to be processed. The reliability information, provided by the outlined parameter estimator, is important in distinguishing between the parameter variations and the estimation error (variance) of the data processor. Without this information, one may tend to model the statistics of the estimation error and not of the radio channel parameters.

A research area, which has been widely ignored up to now, is the tracking of radio channel parameters, or more precisely the estimation of deterministic changes of channel parameters. The variations of the structural path parameters are closely related to the movement of objects influencing the radio channel. A first promising attempt has been made to estimate the parameter changes of the propagation path parameters using a linear model [48], [77].

A further research topic is the optimal antenna array structure for channel sounding applications. Closely related to this issue is the calibration of antenna arrays, e.g., the estimation of the antenna array model especially the EADF. Since upcoming channel sounding systems will have a larger bandwidth, also the frequency dependence of the array response, e.g., the EADF must be investigated.

A fourth research area is the development of appropriate models to describe the correlation function of the DMC in the spatial domain as well as in the time/Doppler domain. Here, the challenge is the coarse Rayleigh resolution in these domains compared to the resolution in the time delay domain. It is generally a lot easier to increase the measurement bandwidth, than increasing the array aperture, i.e., the angular resolution.

A further research area is related to MIMO communication systems. It is generally accepted that MIMO-systems with knowledge of the radio channel at the transmitter have a higher channel capacity than systems having only knowledge about the radio channel at the receiver. If we need channel information at the transmitter, we have to transmit this information from the receiver back to the transmitter. However, this back transmission will in turn reduce the available channel capacity. Hence, we have to find the shortest code to transmit the channel information back to the transmitter. In other words, we have to measure the stochastic complexity of the radio channel. One way to code the channel information is for example the developed radio channel model, since one has only to transmit the channel parameters with the accuracy (SNR) given by the estimated channel parameter variance. Closely related to this research area are clearly the second and the third research theme mentioned above. Since they will increase the a priori information contained in the channel model and consequently reduce the information necessary to store, e.g. transmit the channel.

Appendix

A   List of Frequently Used Symbols

Symbol	Range	Description	Section
$\mathbf{F}$	$\mathbb{C}^{M \times M}$	unitary Fourier-matrix (DFT-matrix)	**C**
$\mathbf{I}$	$\mathbb{R}^{M \times M}$	identity matrix	**C**
$\mathbf{J}$	$\mathbb{R}^{M \times N}$	selection matrix	**C**
$\boldsymbol{\theta}$	$\mathbb{R}^{L \times 1}$	general model parameters	2
$\mathbf{G}$	$\mathbb{C}^{M \times N}$	General measurement system matrix	2
φ_T	$\mathbb{R}$	Azimuth of departure (Tx)	2.1
φ_R	$\mathbb{R}$	Azimuth of arrival (Rx)	2.1
ϑ_R	$\mathbb{R}$	Elevation of Departure (Tx)	2.1
ϑ_T	$\mathbb{R}$	Elevation of Arrival (Rx)	2.1
τ	$\mathbb{R}$	Time delay of arrival (TDoA)	2.1
$\mathbf{b}_T(\varphi_T, \vartheta_T)$	$\mathbb{C}^{1 \times 2}$	Polarimetric Tx antenna response	2.2
$b_{T_H}(\varphi_T, \vartheta_T)$	$\mathbb{C}$	Tx antenna response for horizontal excitation	2.2
$b_{T_V}(\varphi_T, \vartheta_T)$	$\mathbb{C}$	Tx antenna response for vertical excitation	2.2
$\mathbf{b}_R(\varphi_R, \vartheta_R)$	$\mathbb{C}^{1 \times 2}$	Polarimetric Rx antenna response	2.2
$b_{R_H}(\varphi_R, \vartheta_R)$	$\mathbb{C}$	Rx antenna response for horizontal excitation	2.2
$b_{R_V}(\varphi_R, \vartheta_R)$	$\mathbb{C}$	Rx antenna response for horizontal excitation	2.2
M_T	$\mathbb{R}$	Number of Tx antenna ports	2.2
M_R	$\mathbb{R}$	Number of Rx antenna ports	2.2
P	$\mathbb{R}$	Number of propagation paths	2.2
$\mathbf{G}_{T_f}$	$\mathbb{C}^{M_f \times M_f}$	Frequency response of the transmitter	2.2
$\mathbf{G}_{R_f}$	$\mathbb{C}^{M_f \times M_f}$	Frequency response of the receiver	2.2
$\mathbf{H}(f, t)$	$\mathbb{C}^{M_R \times M_T}$	Frequency and time dependent channel matrix	2.2
α	$\mathbb{R}$	Doppler-shift	2.2
γ, γ_p	$\mathbb{C}$	Scalar path weight	2.2
$\mathbf{s}(\boldsymbol{\theta}_{sp})$	$\mathbb{C}^{M \times 1}$	Parametric model for observed propagation paths (maps propagation paths parameters to measurement system response)	2.3
B_m	$\mathbb{R}$	Observation bandwidth	2.3

M	$\mathbb{R}$	Total number of samples in a channel observation	2.3
M_f	$\mathbb{R}$	Number of frequency samples	2.3
M_t	$\mathbb{R}$	Number of time domain samples	2.3
T_m	$\mathbb{R}$	Observation time	2.3
$\mathbf{A}_\tau$, $\mathbf{A}_\tau(\boldsymbol{\tau})$	$\mathbb{C}^{M_f \times P}$	Complex exponentials describing the TDoA of the propagation paths	2.3
$\mathbf{A}_\alpha$, $\mathbf{A}_\alpha(\boldsymbol{\alpha})$	$\mathbb{C}^{M_t \times P}$	Complex exponentials describing the Doppler-shift of the propagation paths	2.3
$\mathbf{B}(\boldsymbol{\mu})$	$\mathbb{C}^{M \times N_{pol}P}$	General matrix valued function, mapping the structural path parameters $\boldsymbol{\mu}$ to the system response	2.3
$\mathbf{B}_{R_H}$, $\mathbf{B}_{R_H}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R)$	$\mathbb{C}^{M_R \times P}$	Rx-Array response for all propagation paths (horizontal field component)	2.3
$\mathbf{B}_{R_V}$, $\mathbf{B}_{R_V}(\boldsymbol{\varphi}_R, \boldsymbol{\vartheta}_R)$	$\mathbb{C}^{M_R \times P}$	Rx-Array response for all propagation paths (vertical field component)	2.3
$\mathbf{B}_{T_H}$, $\mathbf{B}_{T_H}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T)$	$\mathbb{C}^{M_T \times P}$	Tx-Array response for all propagation paths (horizontal field component)	2.3
$\mathbf{B}_{T_V}$, $\mathbf{B}_{T_V}(\boldsymbol{\varphi}_T, \boldsymbol{\vartheta}_T)$	$\mathbb{C}^{M_T \times P}$	Tx-Array response for all propagation paths (vertical field component)	2.3
$\mathbf{B}_f$	$\mathbb{C}^{M_f \times P}$	Frequency responses of the measurement system for all propagation paths	2.3
$\mathbf{B}_t$	$\mathbb{C}^{M_t \times P}$	Time domain responses of the measurement system for all propagation paths	2.3
$\mathbf{G}_{S_f}$	$\mathbb{C}^{M_f \times M_f}$	Frequency response of the transmission or measurement system	2.3
$\boldsymbol{\gamma}$	$\mathbb{C}^{N_{pol}P \times 1}$	All complex path weights in a model	2.3
$\boldsymbol{\gamma}_{HH}$	$\mathbb{C}^{P \times 1}$	Complex path weights H-H polarisation	2.3
$\boldsymbol{\gamma}_{HV}$	$\mathbb{C}^{P \times 1}$	Complex path weights H-V polarisation	2.3
$\boldsymbol{\gamma}_{VH}$	$\mathbb{C}^{P \times 1}$	Complex path weights V-H polarisation	2.3
$\boldsymbol{\gamma}_{VV}$	$\mathbb{C}^{P \times 1}$	Complex path weights V-V polarisation	2.3
$\boldsymbol{\mu}$	$\mathbb{R}^{L_\mu \times 1}$	Normalised structural parameters (nonlinear parameters)	2.3
$\boldsymbol{\theta}_{sp}$	$\mathbb{R}^{L \times 1}$	Parameters of all propagation paths	2.3
$\mathbf{A}_{\varphi_T}$, $\mathbf{A}_{\varphi_T}(\varphi_T)$	$\mathbb{C}^{N_1 \times P}$	Complex exponentials describing the azimuth of departure of the propagation paths	2.4
$\mathbf{A}_{\vartheta_T}$, $\mathbf{A}_{\vartheta_T}(\vartheta_T)$	$\mathbb{C}^{N_2 \times P}$	Complex exponentials describing the elevation of departure of the propagation paths	2.4
$\mathbf{A}_{\varphi_R}$, $\mathbf{A}_{\varphi_R}(\varphi_R)$	$\mathbb{C}^{N_1 \times P}$	Complex exponentials describing the azimuth of arrival of the propagation paths	2.4
$\mathbf{A}_{\vartheta_R}$, $\mathbf{A}_{\vartheta_R}(\vartheta_R)$	$\mathbb{C}^{N_2 \times P}$	Complex exponentials describing the elevation of arrival of the propagation paths	2.4
$\mathbf{G}_{T_H}$, $\mathbf{G}_{T_V}$	$\mathbb{C}^{M_T \times N_1 N_2}$	Effective aperture distribution function of the Tx-antenna array	2.4
$\mathbf{G}_{R_H}$, $\mathbf{G}_{R_V}$	$\mathbb{C}^{M_R \times N_1 N_2}$	Effective aperture distribution function of the Rx-antenna array	2.4
$\mathbf{H}(\boldsymbol{\theta})$	$\mathbb{C}^{M_R M_f \times M_T M_f}$	Parametric broadband channel matrix	2.4

$\mathbf{K}$	$\mathbb{C}^{M \times N}$	coupling matrix	2.4.2, 7
B_d	$\mathbb{R}$	Coherence bandwidth of DMC	2.5
$\mathbf{R}(\boldsymbol{\theta}_{dmc})$	$\mathbb{C}^{M \times M}$	Covariance matrix of DMC	2.5
$\mathbf{R}_f(\boldsymbol{\theta}_{dmc})$	$\mathbb{C}^{M_f \times M_f}$	Covariance matrix of DMC in the frequency domain	2.5
$\mathbf{R}_\tau(\boldsymbol{\theta}_{dmc})$	$\mathbb{C}^{M_f \times M_f}$	Covariance matrix of DMC in the time delay domain	2.5
$\mathbf{R}_T(\boldsymbol{\theta}_{dmc})$	$\mathbb{C}^{M_T \times M_T}$	Covariance matrix of DMC at Tx	2.5
$\mathbf{R}_R(\boldsymbol{\theta}_{dmc})$	$\mathbb{C}^{M_R \times M_R}$	Covariance matrix of DMC at Rx	2.5
$\mathbf{R}_t(\boldsymbol{\theta}_{dmc})$	$\mathbb{C}^{M_t \times M_t}$	Covariance matrix of DMC in the temporal domain	2.5
α_0	$\mathbb{R}$	Noise power	2.5
α_1	$\mathbb{R}$	Maximum power of the DMC	2.5
β_d	$\mathbb{R}$	Normalised coherence bandwidth	2.5
$\boldsymbol{\kappa}(\boldsymbol{\theta}_{dan})$	$\mathbb{C}^{M_f \times 1}$	Auto-covariance function of the DMC and noise in the frequency domain	2.5
$\boldsymbol{\theta}_{dmc}$	$\mathbb{R}^{L \times 1}$	Parameters of the dense multipath components	2.5
$\boldsymbol{\theta}_{dan}$	$\mathbb{R}^{L \times 1}$	Parameters of the dense multipath components and noise	2.5
τ_d'	$\mathbb{R}$	Base time delay of DMC	2.5
τ_d	$\mathbb{R}$	Base time delay of DMC normalised to measurement bandwidth	2.5
$\mathbf{R}, \mathbf{R}(\boldsymbol{\theta})$	$\mathbb{C}^{M \times M}$	covariance matrix, parametric covariance matrix	2.5, 4
$\mathbf{d}_{dmc}$	$\mathbb{C}^{M \times 1}$	Realisation of the DMC process	2.6
$\boldsymbol{\theta}_{chn}$	$\mathbb{R}^{L \times 1}$	All channel parameters (concentrated propagation paths and DMC)	2.6
$\mathbf{A}(\boldsymbol{\mu})$	$\mathbb{C}^{N \times P}$	Vector valued function, mapping normalised structural parameters $\boldsymbol{\mu}$ to complex exponentials (1-dimensional)	3.6
$\mathbf{A}(\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)})$	$\mathbb{C}^{N \times P}$	Vector valued function, mapping normalised structural parameters $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ to complex exponentials (2-dimensional)	3.6
$\boldsymbol{\mu}^{(\tau)}$	$\mathbb{R}^{P \times 1}$	Normalised time delays of arrival	3.6
$\boldsymbol{\mu}^{(\alpha)}$	$\mathbb{R}^{P \times 1}$	Normalised Doppler-shifts	3.6
$\boldsymbol{\mu}^{(\varphi_T)}$	$\mathbb{R}^{P \times 1}$	Normalised azimuth of departure (Tx)	3.6
$\boldsymbol{\mu}^{(\vartheta_T)}$	$\mathbb{R}^{P \times 1}$	Normalised elevation of departure (Tx)	3.6
$\boldsymbol{\mu}^{(\varphi_R)}$	$\mathbb{R}^{P \times 1}$	Normalised azimuth of arrival (Rx)	3.6
$\boldsymbol{\mu}^{(\vartheta_R)}$	$\mathbb{R}^{P \times 1}$	Normalised elevation of arrival (Rx)	3.6
L	$\mathbb{R}$	Number of model parameters	4
$\mathbf{n}_{dan}$	$\mathbb{C}^{M \times 1}$	Realisation of the DMC process and noise	4
$\mathbf{w}$	$\mathbb{C}^{M \times 1}$	independent circular white Gaussian noise	4
$\mathbf{x}$	$\mathbb{C}^{M \times 1}$	Observation (measured radio channel)	4
$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{R})$	$\mathbb{R}$	Log-likelihood function	4

$\mathbf{D}(\boldsymbol{\mu})$	$\mathbb{C}^{M \times L}$	matrix of first order differentials (Jacobian matrix)	4
$\mathbf{D}_{XX}$		Component matrices of the Jacobian matrix	4
$\mathcal{J}(\boldsymbol{\theta}, \mathbf{R}_{nn})$		Fisher information matrix	4
$\mathbf{X}$	$\mathbb{C}^{M \times N}$	Set of channel observations	4
$c(\mathbf{x}, \boldsymbol{\mu})$	$\mathbb{R}$	Cost-function	5.1
$\mathbf{Q}, \mathbf{Q}(\boldsymbol{\mu})$	$\mathbb{C}^{M \times N}$	orthogonal unitary basis vectors	5.1
$\boldsymbol{\beta}(\boldsymbol{\theta}_{dan})$	$\mathbb{C}^{M_f \times 1}$	Sampled PDP of the DMC in the time-delay domain	6.1.4

B   Abbreviations

3
3GPP   Third Generation Partnership Program

A
AIC   Akaike Information Criterion
APS   Azimuth power spectrum

B
BLUE   Best Linear Unbiased Estimate
BS   Base Station

C
CIR   Complex Impulse Response
COST   European CO-operation in the field of Science and Technical Research
CRLB   Cramér-Rao lower bound
CUBA   Circular uniform beam array

D
DCM   Directional Channel Model
DFT   Discrete Fourier transform
DMC   Dense Multipath Components
DML   Deterministic Maximum Likelihood
DoA   Direction of Arrival
DoD   Direction of Departure

E
EADF   Effective Aperture Distribution Function
EM   Expectation Maximization
ESPRIT   Estimation of Signal Parameters via Rotational Invariance Techniques
ESSSE   Economy Size Signal Subspace Estimation
EVD   Eigenvalue Decomposition

F
FFT   Fast (discrete) Fourier Transform
FIM   Fisher Information Matrix

G
GSCM   Geometry-based Stochastic Channel Model
GSVD   Generalised Singular Value Decomposition

I
i.i.d.   independent identically distributed

L
LS   least squares

M
MDL   Minimum Description Length

MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
ML	Maximum Likelihood
MS	Mobile Station
MUSIC	Multiple Signal Classification
MVUB	minimum variance unbiased
MCSSS	Multi-Carrier Spread Spectrum Signal

O

OFDM	orthogonal frequency division multiplexing

P

PDF	probability density function
PRBS	Pseudo-Random Binary Signal
PULA	polarimetric uniform linear array
PURA	polarimetric uniform rectangular array

Q

QoS	Quality of Service

R

RARE	Rank Reduction
RF	radio frequency
Rx	Receive

S

SAGE	Space Alternating Generalised Expectation Maximization
SCM	spatial channel model (3GPP)
SCUBA	stacked circular uniform beam array
SIMO	single-input multiple-output
SISO	single-input single-output
SLS	Structured Least Squares / Separable Least Squares
SML	Stochastic Maximum Likelihood
SNR	signal to noise ratio
SSE	Signal Subspace Estimation
SVD	singular value decomposition

T

TDoA	Time Delay of Arrival
TLS	Total Least Squares
Tx	Transmit

U

UCA	Uniform Circular Array
ULA	Uniform Linear Array
URA	Uniform Rectangular Array

C   Some Common Definitions and Useful Relations

Matrix definition:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{C}^{m \times n}$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \in \mathbb{C}^{p \times q}$$

Selection of Matrix and Vector elements:

$$\{\mathbf{A}\}_i = \mathbf{a}_i, \ \{\mathbf{A}\}_{i,k} = a_{ik}, \ \{\mathbf{a}\}_i = a_i$$

Schur or Hadamard product ($m = p, n = q$):

$$\mathbf{C} = \mathbf{A} \circ \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{bmatrix} \in \mathbb{C}^{m \times n}$$

Kronecker product:

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \cdot & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \mathbb{C}^{mp \times nq}$$

Identities of Kronecker products:

$$\alpha(\mathbf{A} \otimes \mathbf{B}) = (\alpha\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (\alpha\mathbf{B})$$

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$$

$$(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$$

$$(\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{E} \otimes \mathbf{F}) = (\mathbf{A} \cdot \mathbf{E}) \otimes (\mathbf{B} \cdot \mathbf{F})$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}$$

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \otimes \mathbf{C} = \begin{bmatrix} \mathbf{A} \otimes \mathbf{C} & \mathbf{B} \otimes \mathbf{C} \end{bmatrix}$$

Khatri-Rao product $n = q$ **:**

$$\mathbf{C} = \mathbf{B} \lozenge \mathbf{A} = [(\mathbf{b}_1 \otimes \mathbf{a}_1) \ \dots \ (\mathbf{b}_n \otimes \mathbf{a}_n)] \in \mathbb{C}^{mp \times n}$$

Identities involving the Khatri-Rao product:

$$(\mathbf{D} \lozenge \mathbf{B} \lozenge \mathbf{a}^\mathrm{T}) \cdot \mathbf{c} = (\mathbf{D} \lozenge \mathbf{B}) \cdot (\mathbf{a} \circ \mathbf{c})$$

$$\mathbf{C}^\mathrm{H} \cdot \mathbf{C} = (\mathbf{A} \lozenge \mathbf{B})^\mathrm{H} (\mathbf{A} \lozenge \mathbf{B}) = (\mathbf{B} \lozenge \mathbf{A})^\mathrm{H} (\mathbf{B} \lozenge \mathbf{A}) = (\mathbf{A}^\mathrm{H} \mathbf{A}) \circ (\mathbf{B}^\mathrm{H} \mathbf{B})$$

since

$$(\mathbf{b}_1 \otimes \mathbf{a}_1)^\mathrm{H} \cdot (\mathbf{b}_2 \otimes \mathbf{a}_2) = (\mathbf{a}_1^\mathrm{H} \cdot \mathbf{a}_2) \cdot (\mathbf{b}_1^\mathrm{H} \cdot \mathbf{b}_2)$$

Definition of the vector operator $\mathrm{vec}\{\bullet\}$ **:**

$$\mathrm{vec}\{\mathbf{A}\} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \\ \mathbf{a}_n \end{bmatrix}$$

Identities involving the vector operator:

$$\mathrm{tr}(\mathbf{A}^H \mathbf{B}) = \mathrm{vec}\{\mathbf{A}\}^\mathrm{H} \cdot \mathrm{vec}\{\mathbf{B}\}$$

$$\mathrm{vec}\{\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}\} = (\mathbf{C}^T \otimes \mathbf{A}) \cdot \mathrm{vec}\{\mathbf{B}\}$$

$$\mathbf{a}^T \mathbf{B} = (\mathrm{vec}(\mathbf{B}))^T \cdot (\mathbf{I} \otimes \mathbf{a})$$

$$\mathrm{vec}\{\mathbf{A} \cdot \mathrm{diag}\{\mathbf{c}\} \cdot \mathbf{B}^\mathrm{T}\} = (\mathbf{B} \lozenge \mathbf{A}) \cdot \mathbf{c}$$

$$(\mathbf{D} \lozenge \mathbf{B} \lozenge \mathbf{a}^\mathrm{T}) \cdot \mathbf{c} = (\mathbf{D} \lozenge \mathbf{B}) \cdot (\mathbf{a} \circ \mathbf{c})$$

$$\mathbf{B} \cdot \mathrm{diag}\{\mathbf{a} \circ \mathbf{c}\} \cdot \mathbf{D}^\mathrm{T} = \mathbf{B} \cdot \mathrm{diag}\{\mathbf{c}\} \cdot \mathrm{diag}\{\mathbf{a}\} \cdot \mathbf{D}^\mathrm{T}$$

Matrix operator, reshaping a vector to a $m \times n$ **matrix:**

$$\mathrm{mat}\left\{ \begin{bmatrix} \mathbf{a}_1^{m \times 1} \\ \mathbf{a}_1^{m \times 2} \\ \vdots \\ \mathbf{a}_n^{m \times 1} \end{bmatrix}, m, n \right\} = \mathbf{A}^{m \times n}$$

For $m \equiv n$:

$$\text{mat}\left\{\begin{bmatrix}\mathbf{a}_1^{m\times1}\\\mathbf{a}_2^{m\times1}\\\vdots\\\mathbf{a}_m^{m\times1}\end{bmatrix}\right\} = \mathbf{A}^{m\times m}$$

Definition of a Unitary Matrix:

$$\mathbf{Q}^H\cdot\mathbf{Q} = \mathbf{I}$$

Reflection Matrix, Reverse Permutation Matrix:

$$\mathbf{\Pi} = \begin{bmatrix}0 & \cdots & 0 & 0 & 1\\0 & \cdots & 0 & 1 & 0\\\vdots & \ddots & 1 & 0 & 0\\0 & \ddots & \ddots & \vdots & \vdots\\1 & 0 & \cdots & 0 & 0\end{bmatrix}$$

Identities involving the Reflection Matrix:

$$\begin{bmatrix}a_M\\\vdots\\a_1\end{bmatrix} = \mathbf{\Pi}\cdot\begin{bmatrix}a_1\\\vdots\\a_M\end{bmatrix}$$

Definition of a Left Conjugate Symmetric Matrix:

$$\mathbf{A} = \mathbf{\Pi}\mathbf{A}^*$$

Definition of the Selection Matrix:

$$\mathbf{J}_k^{M\times L} = \begin{bmatrix}\mathbf{0}^{M\times(k-1)} & \mathbf{I}^{M\times M} & \mathbf{0}^{M\times(L-M-k)}\end{bmatrix}$$

Identities involving the Selection Matrix:

$$\begin{bmatrix}a_k\\\vdots\\a_{k+M-1}\end{bmatrix} = \mathbf{J}_k^{M\times L}\cdot\begin{bmatrix}a_1\\\vdots\\a_L\end{bmatrix}$$

Definition of the DFT matrix:

$$\mathbf{F} = \frac{1}{\sqrt{M}} \begin{bmatrix} w^0 & w^0 & w^0 & \cdots & w^0 \\ w^0 & w^1 & w^2 & \cdots & w^{M-1} \\ w^0 & w^2 & w^4 & \cdots & w^{2(M-1)} \\ \vdots & \vdots & & & \vdots \\ w^0 & w^{M-1} & w^{2(M-1)} & \cdots & w^{(M-1)^2} \end{bmatrix}, \quad w = e^{-\frac{j2\pi}{M}}$$

Definition of a Toeplitz matrix:

$$\mathbf{T} = \text{toep}\{\mathbf{c}, \mathbf{r}\} = \text{toep}\left\{\begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix}, \begin{bmatrix} r_1 & \cdots & r_M \end{bmatrix}\right\} = \begin{bmatrix} c_1 & r_2 & \cdots & r_{M-1} & r_M \\ c_2 & \ddots & \ddots & & r_{M-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_{M-1} & & \ddots & \ddots & r_2 \\ c_M & c_{M-1} & \cdots & c_2 & c_1 \end{bmatrix}, c_1 \overset{!}{\equiv} r_1$$

Definition of a Hermitian Toeplitz Matrix:

$$\mathbf{T} = \mathbf{T}^H = \text{toep}\{\mathbf{c}, \mathbf{c}^H\}, \quad c_i \in \left| \begin{matrix} \mathbb{R} & i = 1 \\ \mathbb{C} & otherwise \end{matrix} \right.$$

Definition of a Circulant Matrix:

$$\mathbf{C} = \text{circ}\{\mathbf{c}\} = \text{circ}\left\{\begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix}\right\} = \begin{bmatrix} c_1 & c_M & \cdots & c_3 & c_2 \\ c_2 & c_1 & & c_4 & c_3 \\ \vdots & \vdots & \ddots & & \vdots \\ c_{M-1} & c_{M-2} & & c_1 & c_M \\ c_M & c_{M-1} & \cdots & c_2 & c_1 \end{bmatrix}$$

Identities of Toeplitz Matrices:

$$\mathbf{\Pi} \cdot \mathbf{T} \cdot \mathbf{\Pi} = \mathbf{T}^* = \mathbf{T}^T, \text{ if } \mathbf{T} \text{ is Hermitian}$$

Some Matrix Gradients [53], [40]:

$$\frac{\partial}{\partial \mathbf{A}} \text{tr}(\mathbf{A}) = \mathbf{I}$$

$$\frac{\partial}{\partial \mathbf{A}} \ln(\det(\mathbf{A})) = \left(\mathbf{A}^{-1}\right)^H$$

$$\frac{\partial}{\partial \mathbf{A}} \text{tr}(\mathbf{B}\mathbf{A}^{-1}) = \frac{\partial}{\partial \mathbf{A}} \text{tr}(\mathbf{A}^{-1}\mathbf{B}) = -\left(\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right)^H, \text{ provided } \mathbf{B} \text{ is independent of } \mathbf{A}$$

Calculation of Least Squares Estimates, A and Y known, X unknown:

$$\mathbf{A}\mathbf{X} = \mathbf{Y}$$

solve:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{A}\mathbf{X} - \mathbf{Y}\|_F^2$$

I. direct solution:

$$\hat{\mathbf{X}} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{Y}$$

II. Best solution in terms of diagnostics Moore-Penrose-Pseudo-Inverse $\mathbf{A}^+$

$$\hat{\mathbf{X}} = \mathbf{A}^+\mathbf{Y}.$$

Proof:

$$\mathbf{V}^H\mathbf{\Sigma}^T\mathbf{U}\cdot\mathbf{Y} = \mathbf{V}^H\mathbf{\Sigma}^T\mathbf{U}\cdot\mathbf{U}^H\mathbf{\Sigma}\mathbf{V}\cdot\hat{\mathbf{X}} \quad \left|\mathbf{V}^H\mathbf{\Sigma}^+\left(\mathbf{\Sigma}^T\right)^+\mathbf{V} \rightarrow\right.$$

$$\underbrace{\mathbf{V}^H\left(\mathbf{\Sigma}^T\right)^+\mathbf{U}}_{\mathbf{A}^+}\cdot\mathbf{Y} = \hat{\mathbf{X}}$$

$$\mathbf{A}^+\mathbf{Y} = \hat{\mathbf{X}}$$

III. Computationally efficient solution using the QR decomposition:

$$\mathbf{A} = \begin{bmatrix}\mathbf{Q}_S & \mathbf{Q}_N\end{bmatrix}\begin{bmatrix}\mathbf{R}\\\mathbf{0}\end{bmatrix} = \mathbf{Q}_S\mathbf{R}$$

$$\mathbf{R}^H\mathbf{Q}_S^H\cdot\mathbf{Q}_S\mathbf{R}\cdot\hat{\mathbf{X}} = \mathbf{R}^H\mathbf{Q}_S^H\cdot\mathbf{Y}$$

$$\mathbf{R}\cdot\hat{\mathbf{X}} = \mathbf{Q}_S^H\cdot\mathbf{Y}$$

1) Compute:

$$\mathbf{Z} = \mathbf{Q}_S^H\mathbf{Y}$$

2) Solve

$$\mathbf{R}\cdot\hat{\mathbf{X}} = \mathbf{Z},$$

Observe that $\mathbf{R}$ is an upper triangular matrix!

Bibliography

[1]   A. Richter, D. Hampicke, G. Sommerkorn, and R.S. Thomä, "MIMO Measurement and Joint M-D Parameter Estimation of Mobile Radio Channels," *IEEE Vehicular Technology Conference, VTC 2001-Spring*, Rhodos, May 6-9, 2001.

[2]   G. J. Foschini and M.J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," Wireless Personal Communications, Vol. 6, pp. 311-335, 1998.

[3]   http://www.lx.it.pt/cost259

[4]   http://www.lx.it.pt/cost273

[5]   Andreas F. Molisch, "A Generic Model for MIMO Wireless Propagation Channels in Macro- and Microcells," IEEE Trans. Signal Proc., vol. 52, no. 1, pp. 61--71, Jan. 2004.

[6]   M. Steinbauer, A.F. Molisch, E. Bonek., "The Double-Directional Radio Channel," IEEE Antennas and Propagation Magazine, vol. 45, pp. 51-63, No.4, Aug. 2001.

[7]   R. S. Thomä, D. Hampicke, M. Landmann, A. Richter and G. Sommerkorn, "Measurement-based Channel Modelling (MBPCM)," ICEAA 2003, Torino, Sept. 2003.

[8]   A. Richter, R. S. Thomä," Parametric Modeling and Estimation of Distributed Diffuse Scattering Components of Radio Channels," *COST273*, TD(03)198, Sept. 24-26, 2003, Prague, http://www.lx.it.pt/cost273/.

[9]   A. Richter, R.S. Thomä, T. Taga "Directional Measurement and Analysis of Propagation Path Variations in a Street Micro-Cell Scenario," *IEEE Vehicular Technology Conference, VTC2003-Spring*, Jeju, Korea, April 2003.

[10]   D. Cassioli, M. Z. Win, and A. F. Molisch, "The Ultra-Wide Bandwidth Indoor Channel: From Statistical Model to Simulations," *IEEE Journal on Selected Areas in Communications*, Vol. 20, No. 6. August 2002, pp 1247-1257.

[11]   K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A Stochastic Model of the Temporal and Azimuthal Dispersion Seen at the Base Station in Outdoor Propagation Environments," IEEE Trans. on Vehicular Technology, Vol. 49, No. 2, March 2000.

[12] V. Erceg, D. G. Michelson, S. S. Ghassemzadeh, L. J. Greenstein, A.J. Rustako, P. B. Guerlain, M. K. Dennison, R. S. Roman, D. J. Barnickel, S. C. Wang, R. R. Miller, "A Model for the Multipath Delay Profile of Fixed Wireless Channels," IEEE Journal on Selected Areas in Communications, Vol. 17, No. 3, March 1999.

[13] P. A. Bello, "Characterization of Randomly Time-Variant Linear Channels," IEEE Transactions on Communications, vol. 11, no. 4, 1963.

[14] W. Weichselberger, *Spatial Structure of Multiple Antenna Radio Channels; A Signal Processing Viewpoint*, Ph.D. Thesis, Technische Universität Wien, Vienna, Austria, Dec. 2003.

[15] M. Steinbauer, "The Radio Propagation Channel − A Non-Directional, Directional and Double-Directional Point-of-View," Ph.D. Thesis, Technische Universität Wien, Vienna, Austria, Nov. 2001.

[16] J. Rissanen, "Modeling by shortest data description," Automatica, no. 14, 465-471, 1978.

[17] M. H. Hansen, B. Yu, "Model Selection and the Principle of Minimum Description Length," Journal of the American Statistical Association, vol. 96(454), pp. 746-774, 2001.

[18] T. Jämsä, V. Hovinen, L. Hentilä, " Comparison of Wideband and Ultra-Wideband Channel Measurements," COST 273, Gothenburg, Sweden, June, 2004, TD(04)080.

[19] V. Degli-Esposti, D. Guiducci, A. de´ Marsi, P. Azzi, F. Fuschini, "An Advanced Field Prediction Model Including Diffuse Scattering," IEEE Transactions on Antennas and Propagation, vol. 52, no. 7, pp. 1717-1728, July 2004.

[20] R. Zetik, J. Sachs, R. Thomä, "Ultra-Wideband Real-Time Channel Sounder and Directional Channel Parameter Estimation," 2004 URSI Intl. Symposium On Electromagnetic Theory, Pisa, IT, May 23-27 2004.

[21] J.D. Parsons, *The Mobile Radio Propagation Channel*, John Wiley & Sons, Chichester, 2000.

[22] J. Schoukens, R. Pintelon, *Identification of Linear Systems - A Practical Guideline to Accurate Modeling*. Pergamon Press, 1991.

[23] D. Bückner, H.-J. Kolb, A. Richter, G. Sommerkorn, R. Thomä, U. Trautwein and W. Wirnitzer, "Verfahren zur Bestimmung einer richtungsaufgelösten komplexen Impulsantwort eines Funkkanals und Meßsystems," (in German) (Measurement of Directional Impulse Response), DE 197 41 991.7.

[24] R.S. Thomä, D. Hampicke, A. Richter, A. Schneider, G. Sommerkorn, and U. Trautwein, " Identification of Time-Variant Directional Mobile Radio Channels," In Proc 9th Virginia Tech Symposium on Wireless Personal Communications, Blacksburg, USA, June 2-4, 1999, pp. 11-22.

[25] R. S. Thomä, D. Hampicke, A. Richter, G. Sommerkorn, U. Trautwein, "MIMO Vector Channel Sounder Measurement for Smart Antenna System Evaluation," European Transactions on Telecommunications ETT, Vol. 12, No. 5, Special Issue on Smart Antennas, Sept./Oct. 2001, pp. 427-438.

[26] R.S. Thomä, D. Hampicke, A. Richter, G. Sommerkorn, A. Schneider, U. Traut-wein, "Identification of Time-Variant Directional Mobile Radio Channels," *In Proc* 16th IEEE Instrumentation and Measurement Technology Conference (IMTC'99), Venice, Italy, May 24-26, 1999, vol. 1, pp. 176-181.

[27] U. Trautwein, K. Blau, D. Brückner, A. Richter, G. Sommerkorn, R. Thomä, "Radio Channel Measurement for Realistic Simulation of Adaptive Antenna Arrays," 2nd European Personal Mobile Communications Conference (EPMCC'97), 30.09.-02.10.1997, Bonn, ITG Fachbericht 145, S. 491 – 498.

[28] M. Haardt., R.S. Thomä, A. Richter, "Multidimensional High-Resolution Parameter Estimation with Applications to Channel Sounding," In Yingbo Hua et al. (Ed.): *High-Resolution and Robust Signal Processing*, Marcel Dekker, 2003, p. 253-337.

[29] C. M. Tan, P. Fletcher, M. A. Beach, A. R. Nix, M. Landmann, and R. S. Thomä, "On the Application of Circular Arrays in Direction Finding Part I: Investigation into the estimation algorithms," COST 273, ESPOO, Finland, May 2002.

[30] C. M. Tan, M. Landmann, A. Richter, L. Pesik, M. A. Beach, Ch. Schneider, R.S. Thomä, and A. R. Nix, "On the Application of Circular Arrays in Direction Finding Part II: Experimental evaluation on SAGE with different circular arrays," COST 273, ESPOO, Finland, May 2002.

[31] www.irk-dresden.de

[32] www.channelsounder.de

[33] G. Sommerkorn, D. Hampicke, A. Richter, R. Thomä: "Measurement and Modeling Error Influence to Antenna Array Calibration and its Affect to ESPRIT-based DoA-Estimation," AP 2000 Millennium Conference on Antennas and Propagation, Davos, Switzerland, April 9-14, 2000.

[34] G. Sommerkorn, D. Hampicke, R. Klukas, A. Richter, A. Schneider, R. Thomä, "Reduction of DoA Estimation Errors Caused by Antenna Array Imperfections," *In Proc.* 29th European Microwave Conference, Munich , Oct. 4-8, 1999, Vol. 2, pp. 287-290 (also: Proc. European Wireless'99, Munich, Oct. 6-8, 1999, pp. 21-24)

[35] G. Sommerkorn, D. Hampicke, R. Klukas, A. Richter, A. Schneider, R. Thomä, "Uniform Rectangular Antenna Array Design and Calibration Issues for 2-D ESPRIT Application," *In Proc.* EPMCC Vienna, Austria, February 20th - 22nd, 2001.

[36] M. Landmann, A. Richter and R.S. Thomä," Performance Evaluation of Antenna Arrays for High-Resolution DoA Estimation in Channel Sounding," *In Proc.* IEEE Int. Symp. On Antennas and Propagation, August 17-21, 2004, Sendai, Japan.

[37] R. S. Thomä, M. Landmann, A. Richter, U. Trautwein, "Multidimensional High-Resolution Channel Sounding Measurement," In *Smart Antennas – State of the Art*, EURASIP Book Series, 2005.

[38] M. Landmann, A. Richter, R. S.Thomä, "DoA Resolution Limits in MIMO Channel Sounding," *In Proc.* Int. Symp. on Antennas and Propagation and USNC/URSI National Radio Science Meeting, June 20-26, Monterey, CA, 2004.

[39] A. B. Gershman, V. T. Ermolaev, "Optimal subarray size for spatial smoothing," IEEE Signal Processing Letters, vol. 2, no. 2, Feb. 1995.

[40] L. L. Scharf, *Statistical Signal Processing: detection, estimation, and time series analysis*. Addison-Wesley, New York, 1991, ISBN 0-201-19038-9.

[41] G. Sommerkorn, W. Wirnitzer, R.S. Thomä, A. Richter , "Antenna Multiplexing & Time Alignment for MIMO Channel Sounding", XXVIIth URSI-GA, August 18-24, Maastricht, 2002.

[42] H. Cramér, *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press, 1946.

[43] P. Stoica, A. Nehorai, "MODE, maximum likelihood, and Cramér-Rao bound; Conditional and unconditional results," Center Syst. Sci., Yale University, New Haven, CT, Rep. 8901, Jan. 1989.

[44] Y. Hua, A.B. Gershman, QI. Cheng, "High-Resolution and Robust Signal Processing," Marcel Dekker, New York, ISBN 0-8247-4752-6, 2004.

[45] P. Stoica, A. Nehorai, "Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation," IEEE Trans. On ASSP, Vol. 38, No. 10, October 1990.

[46] P. Stoica, A. Jakobsson, and J. Li, "Cisoid Parameter Estimation in the Colored Noise Case: Asymptotic Cramér-Rao Bound, Maximum Likelihood, and Nonlinear Least-Squares," IEEE Transactions on Signal Processing, vol. 45, no. 8, Aug. 1997.

[47] J. M. Francos and B. Friedlander, "Bounds for Estimation of Complex Exponentials in Unknown Colored Noise," IEEE Transactions on Signal Processing, vol. 43, no. 9, Sep. 1995.

[48] V. Algeier, "Schätzung der Änderung von Mobilfunkkanalparametern aus Funkkanalmessdaten", Diploma-Thesis, Technische Universität Ilmenau, IDN 2112-04D-1, (in german), 2003.

[49] M. Viberg, B. Ottersten, and A. Nehorai, "Performance Analysis of Direction Finding with Large Arrays and Finite Data," IEEE Transactions on Signal Processing, 43(2), pp. 469-477, Feb. 1995.

[50] J. I. Myung and D. J. Navarro, "Information Matrix," *Encyclopaedia of Behavioural Statistics*, Wiley, 2005.

[51] A. Schulz, "Implementierung einer laufzeitoptimierten mehrdimensionalen Korrelationsfunktion auf Basis des x86-SIMD-Befehlssatzes „SSE2", " Technische Universität Ilmenau, May 2004.

[52] A. Hjørungnes, "Complex-Valued Matrix Differentiation with Applications to Optimization of MIMO Systems," to appear in SIAM Journal On Matrix Analysis and Applications.

[53] M. Brookes, "The Matrix Reference Manual" , http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html, Imperial College, London, UK.

[54] S. M. Kay, *Fundamentals of Statistical Signal Processing – Estimation Therory*. Englewood Cliffs, NJ, Prentice Hall, 1993.

[55] B. Porat, *Digital Processing of Random Signals:Theory and Methods*. Englewood Cliffs, NJ, Prentice Hall, 1994.

[56] B. G. Lindsay and B. Li, "On second-order optimality of the observed Fisher information," *The Annals of Statistics*, 1997, Vol. 25, No. 5, 2172-2199.

[57]   M. Landmann, A. Richter, R.S. Thomä, "DoA Resolution Limits in MIMO Channel Sounding," International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting, Monterey, USA, June 2004.

[58]   U. Martin, *Ausbreitung in Mobilfunkkanälen: Beiträge zum Entwurf von Messgeräten und zur Echoschätzung*, Doctor-Thesis, Universität Erlangen-Nürnberg, 1994.

[59]   U. Martin, "Echo estimation – Deriving simulation models for the mobile radio channel," In Proc. IEEE Vehicular Technology Conference, vol. 1, pp. 231-235, Chicago, IL, July 1995.

[60]   M. Steinbauer, "A Comprehensive Transmission and Channel Model for Directional Radio Channels," EURO-COST, TD(98)027, Bern, 2nd - 4th Feb. 1998.

[61]   A. Richter, D. Hampicke, G. Sommerkorn, R.S. Thomä, "Joint Estimation of DoD, Time-delay, and DoA for High-Resolution Channel Sounding," *IEEE Vehicular Technology Conference, VTC2000-Spring*, Tokyo, May 15-18, 2000.

[62]   J. A. Fessler, A. O. Hero, "Space Alternating Generalised Expectation Maximization Algorithm," IEEE Trans. on Signal Processing, Vol. 42, pp. 2664-2677, October 1994.

[63]   B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. Ingeman Pedersen, "Channel Parameter Estimation in Mobile Radio Environments Using the SAGE Algorithm," *IEEE Journal on Selected Areas in Communications*, Vol. 17, No. 3, pp. 434-450, March 1999.

[64]   B. H. Fleury, D. Dahlhaus, R. Heddergott, and M. Tschudin, "Wideband angle of arrival estimation using the SAGE algorithm," *In Proc. IEEE Fourth Int. Symp. spread Spectrum techniques and Applications* (ISSSTA'96), Mainz, Germany, pp. 79-85, Sept. 1996.

[65]   M. Haardt, "Efficient One-, Two-, and Multidimensional High-Resolution Array Signal Processing," Ph.D. Thesis, Shaker Verlag, Aachen, Germany, 1996.

[66]   M. Pesavento, A.B. Gershman, and K.M. Wong, "Direction finding in partly calibrated sensor arrays composed of multiple subarrays," IEEE Trans. Signal Processing, vol. 50, pp. 2103-2115, Sept. 2002.

[67]   C.M.S. See and A.B. Gershman, "Direction-of-arrival estimation in partly calibrated subarray-based sensor arrays," IEEE Trans. Signal Processing, vol. 52, pp. 329-338, Feb. 2004.

[68]   D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," SIAM J. Appl. Math. Vol. 11, pp 431-441, 1963.

[69]   A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, 1977, Series B, 39:1-38.

[70]   C. Therrien, *Decision Estimation and Classification.* John Wiley and Sons, Inc., 1989.

[71]   James E. Gentle, *Elements of Computational Statistics*. Springer-Verlag, 1st edition, 2002, ISBN: 0387954899.

[72]   G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," Inverse Problems, vol. 19, no.2, pp. R1-R26, Apr. 2003.

[73]   K. Mahata and T. Söderström, "Large Sample Properties of Separable Nonlinear Least Squares Estimators," IEEE Transactions on Signal Processing, vol. 52, no. 6, June 2004.

[74]   I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, pp. 1553-1560, Oct. 1988.

[75]   J. Li, D. Zheng and P. Stoica, "Angle and waveform estimation via RELAX." IEEE Trans. on Aerospace and Electronic Systems, vol. 33, 1077-1087, 1997.

[76]   Z.-S. Liu, J. Li and P. Stoica, "RELAX-based estimation of damped sinusoidal signal parameters." Signal Processing, vol. 62, 311-321, 1997.

[77]   V. Algeier, A. Richter, R. Thomä, "A Gradient based Algorithm for Path Parameter Tracking in Channel Sounding", Cost 273, Gothenburg, Sweden, June, 2004, TD(04)124.

[78]   C. Bliek, P. Spellucci, L. N. Vicente, A. Neumaier, L. Granvilliers, E. Monfroy, F. Benhamou, E. Huens, P. V. Hentenryck, D. Sam-Haroud and B. Faltings, "CO-CONUT Deliverable D1 - Algorithms for Solving Nonlinear Constrained and Optimization Problems: The State of The Art," The Coconut Project, June 8, 2001.

[79]   B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm." Proceedings of the IEEE, 63(4):719-720, April 1975.

[80]   B. Widrow and J. McCool, "A comparison of adaptive algorithms based on the method of steepest descent and random search." IEEE Trans. Antennas and Propagation, AP-24:615-636, 1976.

[81]   J. E. Jr. Dennis, "Nonlinear least-squares," State of the Art in Numerical Analysis, ed. D. Jacobs, Academic Press, pp 269-312, 1977.

[82]   K. Levenberg, "A Method for the Solution of Certain Problems in Least Squares," Quart. Appl. Math. 2, 164-168, 1944.

[83]   D. P. Bertsekas, Nonlinear Programming, second edition, Athena Scientific, 1999.

[84]   A. Richter, M. Landmann, and R. S. Thomä, "A Gradient Based Method for Maximum Likelihood Channel Parameter Estimation from Multidimensional Channel Sounding Measurements," XXVIIth URSI-GA, August 18-24 2002, Maastricht.

[85]   A. Richter, M. Landmann, R.S. Thomä, "Maximum Likelihood Channel Parameter Estimation from Multidimensional Channel Sounding Measurements," IEEE Vehicular Technology Conference, VTC2003-Spring, Jeju, Korea, April 2003.

[86]   J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and Theory," Proceedings of the Dundee Conference on Numerical Analysis, G. A. Watson, ed., Springer-Verlag, Berlin, New York, 1978, pp. 105-116.

[87]   J. E. Dennis, Jr., and R. B. Schnabel, "A View of Unconstrained Optimization," Handbooks Oper. Res. Management Sci. 1, North-Holland, Amsterdam, 1989.

[88]   J. Nocedal and S. J. Wright, Numerical Optimization, Springer-Verlag, 1999.

[89] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C*, second edition, Cambridge University Press, October, 1992.

[90] M. Bengtsson and B. Völker, "On the Estimation of Azimuth Distributions and Azimuth Spectra," in Proc. IEEE 54th Vehicular Technology Conference (VTC 2001 Fall), vol. 3, pp. 1612-1625, 2001.

[91] M. Tschudin, R. Heddergott, and P. Truffer, "Validation of a high-resolution measurement technique for estimating the parameters of impinging waves in indoor environments," in Proc. 9 th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC'98) Boston, MA, pp. 1411-1416, Sept. 1998.

[92] M. Tschudin, C. Brunner, T. Kurpjuhn, M. Haardt, and J. A. Nossek, "Comparison between Unitary ESPRIT and SAGE for 3-D Channel Sounding," In Proc. 49th IEEE Vehicular Technology Conference (VTC '99 Spring), vol. 2, pp. 1324-1329, Houston, Texas, USA, May 1999.

[93] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 33, pp. 387-392, Apr. 1985.

[94] J. Fuhl, J. P. Rossi, and E. Bonek, "High-resolution 3D direction-of-arrival determination for urban mobile radio," IEEE Transactions on Antennas and Propagation, vol. 45, no. 4, Apr. 1997.

[95] M. Bronzel, J. Jelitto, N. Lohse, G. Fettweis, R. Thomä, G. Sommerkorn, D. Hampicke, U. Trautwein, A. Richter, "Experimental Verification of Vector Channel Models for Simulation and Design of Adaptive Antenna Array Receivers," *In Proc.* ACTS Mobile Communication Summit ´98, Rhodos, Greece, June 8-11, 1998, vol. 2, pp. 424-429.

[96] T. Zwick, D. Hampicke, A. Richter, G. Sommerkorn, R.Thomä,W. Wiesbeck, "A novel antenna concept for double-directional channel measurements," IEEE Transactions on Vehicular Technology, Volume: 53 , Issue: 2 , 527-537 March 2004.

[97] R. S. Thomä, D. Hampicke, M. Landmann, G. Sommerkorn, A. Richter, "MIMO Measurement for Double-Directional Channel Modeling", *In Proc.* IEE Technical Seminar on MIMO Communication Systems, December 12, 2001, London, UK.

[98] R. Thomä, K. Blau, D. Brückner, D. Hampicke, G. Sommerkorn, A. Richter, U. Trautwein, W. Wirnitzer, "Messung des Übertragungsverhaltens von Mobilfunkkanälen," *In Proc.* 43. Internationales Wissenschaftliches Kolloquium der TU Ilmenau (IWK ´98), 21.-24. September 1998, Bd. 3, S. 345-356.

[99] T. Zwick, D. Hampicke, A. Richter, G. Sommerkorn, R. Thomä, W. Wiesbeck: "Results of Double Directional Channel Sounding Measurements," *In Proc.* IEEE Vehicular Technology Conference, VTC2000-Spring, Tokyo, May 15-18, 2000.

[100] M. Pesavento, C. F. Mecklenbräuker and J. F. Böhme, "Multi-dimensional Rank Reduction Estimator for Parametric MIMO Channel Models," In EURASIP Journal on Applied Signal Processing (JASP), Special Issue on Advances in Smart Antennas, vol. 2004, no. 9, pp. 1354-1363, Aug. 2004.

[101] M. Pesavento, C. F. Mecklenbräuker and J. F. Böhme, "Multi-dimensional harmonic estimation using K-D RARE in application to MIMO channel estimation," *In Proc.* IEEE ICASSP'2003, Int. Conf. on Acoustics, Speech and Signal Processing, vol. 4, pp. IV-644--IV-647, Hongkong, China, Apr. 2003.

[102]   M. Pesavento, C. F. Mecklenbräuker and J. F. Böhme, "Tree-structured multi-dimensional RARE for MIMO channel estimation," COST-273, Temporary Document TD(03)020, Barcelona, Spain, Jan. 2003.

[103]   D. Astély, and B. Ottersten, "The Effects of Local Scattering on Direction of Arrival Estimation with MUSIC," IEEE Transactions on Signal Processing, vol. 47, no. 12, Dec. 1999.

[104]   S.U. Pillai, *Array Signal Processing*. Springer-Verlag, New York, NY, 1989.

[105]   S. U. Pillai and B. H. Kwon, "Forward / backward spatial smoothing techniques for coherent signal identification," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-37, pp. 8–15, Jan. 1989.

[106]   R. Roy, A. Paulraj and T. Kailath, "ESPRIT -- A Subspace Rotation Approach to Estimation of Parameters of Cisoids in Noise," IEEE Trans. on Acoustics, Speech, Signal Processing, 34(5):1340-1344, July 1986.

[107]   A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques – ESPRIT," in Proc. 19th Asimolar Conf. on Signals, Systems, and Computers, pp. 83-89, Pacific Grove, CA, Nov. 1985.

[108]   A. Richter and R. S. Thomä, "CUBA-ESPRIT for angle estimation with circular uniform beam arrays," in Proc. of the Millenium Conference on Antennas and Propagation (AP-2000), Davos, Switzerland, 2000.

[109]   J. Lie, P. Stoica, and D. Zheng, "Efficient Direction and Polarization Estimation with a COLD Array," IEEE Transactions on Antennas and Propagation, vol. 44, no. 4, April 1996.

[110]   R. B. Lehoucq, D. C. Sorensen, and C. Yang, "ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods," Society for Industrial and Applied Mathematics (SIAM), ISBN 0-89871-407-9, 1998.

[111]   W. Arnoldi, "The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem." Quart. Appl. Math. 9, 17-29, 1951.

[112]   M. Haard, "Structured least squares to improve the performance of ESPRIT-type algorithms," IEEE Transactions Signal Processing, vol. 45, pp. 792-799, Mar. 1997.

[113]   M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems", IEEE Trans. Signal Processing, vol. 46, pp. 161–169, Jan. 1998.

[114]   F. Demmerle, W. Wiesbeck, "A biconical multibeam antenna for Space-Division Multiple Access," IEEE Transactions on Antennas and Propagation, vol 46, no.6, June, pp. 782-787, 1998.

[115]   D. Kress, R. Irmer, *Angewandte Systemtheorie*, Verlag Technik 1989.

[116]   P. R. Gill, W. Murray, M.H. Wright, *Practical Optimization*. London: Academic Press, 1981.

[117]   I. Gohberg, I. Feldman, "Convolution equations and projection methods for their solutions," Translations of Mathematical Monographs, 41, Amer. Math. Soc., 1974.

[118] I. Gohberg, A. Semencul, "On the inversion of finite Toeplitz matrices and their continuous analogs," Matem. Issled., Kishinev, 7 (No. 2), pp. 201-233.

[119] I. Gohberg, V. Olshevsky," Fast Algorithms with Preprocessing for Matrix-Vector Multiplication Problems," *Journal of Complexity*, vol. 10, no. 4, pp. 411-427, 1994.

[120] N. Levinson, "The Wiener rms (root-mean-square) error criterion in filter design and prediction," J. Math. Phys., vol. 25, pp. 261-278, 1947.

[121] T. Huckle, "Implementation of superfast algorithms for symmetric positive definite linear equations of displacement rank 2," SCCM Preprint 94-4, in Advanced Signal Processing Algorithms, Proc. of SPIE, 1994.

[122] M. Frigo, S. Johnson, "fftw: an adaptive software architecture for the FFT," *In Proc.* ICASSP, vol. 3, pp. 1381-1384, 1998.

[123] A. Richter, M. Landmann, R.S. Thomä," RIMAX - A Flexible Algorithm for Channel Parameter Estimation from Channel Sounding Measurements," *COST273, TD(04)045*, Jan. 26-28, 2004, Athens, http://www.lx.it.pt/cost273/.

[124] R.S. Thomä, M. Landmann, and A. Richter, "RIMAX – a Maximum Likelihood Framework for Parameter Estimation in Multidimensional Channel Sounding," *In Proc.* Int. Symposium on Antennas and Propagation, Sendai, Japan, August 2004, Best Paper Award.

[125] K. Pensel and J.A. Nossek: "Uplink and Downlink Calibration of Smart Antennas," *In Proc.* Int. Conf. on Telecomm. (ICT98), Porto Carras, Greece, June 1998.

[126] J. X. Zhu and H. Wang, "Effects of sensor positions and pattern perturbations on CRLB for direction finding of multiple narrow-band sources," *In Proc.* 4th ASSP Workshop Spectrum Estimation, Modeling (Minneapolis, MN), Aug. 1988, pp. 98-102.

[127] R. Thomä, K. Blau, D. Hampicke, G. Sommerkorn, A. Richter, U. Trautwein, W. Wirnitzer: Messungen des Übertragungsverhaltens von Mobilfunkkanälen. ITG-Fachtagung Wellenausbreitung bei Mobilfunksystemen und Mikrowellensystemen, 11.-13. Mai 1998, DLR Oberpfaffenhofen, S. 297-312.

[128] R.S. Thomä, D. Hampicke, A. Richter, G. Sommerkorn, A. Schneider, U. Trautwein, W. Wirnitzer, "Identification of Time-Variant Directional Mobile Radio Channels," IMTC/99 Special Issue of IEEE Trans. on Instrumentation and Measurement, April 2000.

[129] D. Hampicke, A. Richter, A. Schneider, G. Sommerkorn, R. Thomä: "Measurement-based characterization of time-variant directional radio channels", COST 259 TD(99) 039, Wien, 22.-23. April 1999 (temporary document).

[130] D. Hampicke, A. Richter, A. Schneider, G. Sommerkorn, R.S. Thomä, U. Trautwein: "Characterization of the Directional Mobile Radio Channel in Industrial Scenarios, Based on Wide-Band Propagation Measurements," *In Proc.* IEEE Vehicular Technology Conference (VTC Fall-1999), Amsterdam, The Netherlands, Sept. 1999, Vol. 4, pp. 2258-2262.

[131] D. Hampicke, A. Richter, A. Schneider, G. Sommerkorn, R.S. Thomä, "Statistical Analysis of Time-Variant Directional Mobile Radio Channels, Based on Wide-Band Measurements," *In Proc.* European Wireless'99, Munich, Oct. 6-8, 1999, pp. 273-278.

[132]   G. H. Golub, C.F. Van Loan, *Matrix Computations*, John Hopkins University Press, ISBN 0-946536-00-7/05-8, 1983.

[133]   R. A. Horn, C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[134]   A. Richter, R. Thomä, D. Hampicke, G. Sommerkorn, A. Schneider, (Ilmenau); R. Klukas, (Dresden), "Anforderungen von Parameterschätzverfahren an die Genauigkeit von Breitband-Vektor-Channel-Soundern," ITG Diskussionssitzung des FA 9.1 "Meßverfahren der Informationstechnik" zum Thema Meßverfahren im Mobilfunk vom 03.03. - 05.03.99 auf Schloß Reisensburg Günzburg / Donau.

[135]   D. G. Luenberger, *Optimization by Vector Space Methods*, New York, NY: John Wiley and Sons, 1969.

[136]   A. Richter, R. Thomä, "Joint Maximum Likelihood Estimation of Specular Paths and Distributed Diffuse Scattering," *In Proc.* IEEE VTC2005-Spring, Stockholm, Sweden, May 2005, (accepted for presentation).

[137]   U. Trautwein, M. Landmann, G. Sommerkorn, R. Thomä, "System-Oriented Measurement and Analysis of MIMO Channels," COST 273 Temporary Document, Bologna, Italy, No. TD(05)063, January 2005.

[138]   A. Richter, M. Landmann, R. S. Thomä, "Parameter Estimation Results of Specular and Dense Multipath Components in Micro- and Macro-Cell Scenarios," *In Proc.* WPMC 2004, Abano Terme, Italy, September 2004.