Fast Algorithms for Multidimensional Harmonic Retrieval

A DISSERTATION
submitted to the Faculty of Electrical Engineering and Information Sciences at the Ruhr-Universität Bochum in a fulfillment of the requirements for the degree of Doctor of Engineering

by
Marius Pesavento

Bochum 2005
Schnelle Algorithmen zur Erkennung
von mehrdimensionalen Harmonischen

DISSEPTION
zur
Erlangung des Grades
eines Doktor-Ingenieurs
der
Fakultät für Elektrotechnik und Informationstechnik
an der Ruhr-Universität Bochum

von
Marius Pesavento

Bochum 2005
Tag der Einreichung: 8. Dezember 2004
Tag der Promotion: 2. Februar 2005
Referent: Prof. Dr.-Ing. Johann F. Böhme
Korreferent: Prof. Dr.-Ing Alex B. Gershman
Abstract

Classic multidimensional harmonic retrieval is the estimation problem in a variety of practical applications, including sensor array processing, radar, mobile communications, multiple-input multiple-output (MIMO) channel estimation and nuclear magnetic resonance spectroscopy. Numerous parametric subspace approaches have been proposed recently to solve this problem, among which the so-called ESPRIT-based algorithms are most popular due to their computational efficiency and comparably simple implementations. In these algorithms certain shift invariances contained in the measurements are exploited to estimate the parameters of interest by solving a joint eigenvalue problem. In many applications the measurements are obtained through uniform sampling along one or multiple dimensions. In these cases, the ESPRIT methods usually fail to exploit all prior information contained in the highly structured measurement data resulting in a significant performance loss in the parameter estimation.

In this work a different approach towards multidimensional harmonic retrieval is taken. A suitable parameterization enables the estimation of the harmonics of interest separately along the various dimensions, thus avoiding the computationally expensive optimization of a multidimensional cost function which would otherwise be required. This procedure makes the estimation problem computationally tractable while retaining much of the benefits inherent in the multidimensional nature of the measurement data such as, for example, relatively mild uniqueness conditions and high resolution capability compared to one dimensional data. Several matrix rank and polynomial rooting criteria are derived to obtain the parameters of interest separately along the various dimensions. New insight is gained from interpreting the proposed rank criteria in diverse contexts: as a relaxation approach in minimizing the classic root-MUSIC criterion, in a Gaussian-elimination framework, and as a rooting-based solution of the multiple invariance equations. The different viewpoints not only yield new stochastic uniqueness conditions for the rank reduction estimators, but also lead to efficient parameter association strategies to correctly group the parameters corresponding to a specific multidimensional harmonic signal. Further, a link between the popular ESPRIT-type methods and the root-MUSIC based approaches is discovered that allows to reformulate the rank reduction idea in terms of a joint generalized eigenproblem. Casting the multidimensional harmonic retrieval problem as an eigenproblem significantly simplifies the parameter estimation and association procedure and makes the algorithm equally applicable to the cases of pure and damped harmonic retrieval.

Simulation results obtained from synthetic data for the single and multiple snapshot case are presented and illustrate that the proposed algorithms are competitive with other existing methods from both a numerical viewpoint and also in terms of estimation performance. Further, in the example of parametric MIMO channel identification, it is demonstrated that the novel algorithms perform well if applied to real measurement data obtained from a channel-sounding campaign.
Kurzfassung


Acknowledgments

I would like to express my deep gratitude to Prof. Johann F. Böhme for his great support and encouragement, which went far beyond purely academic matters. His suggestions and critical advice - in the course of numerous fruitful discussions we had - were and are highly appreciated and were also a significant contribution to the success of this work.

I am also indebted to Prof. Alex B. Gershman for the review of my thesis. He introduced me to the field of signal processing and I began my research activities under his guidance. His enthusiasm, inquiring mind and friendship have been truly inspiring.

I would like to give special thanks to Dr.-Ing Christoph Mecklenbräuker for the comprehensive scientific collaboration, reviewing important parts of this manuscript, as well as for the provision of the experimental data by the Telecommunications Research Center Vienna (ftw).

A critical review of the manuscript was carried out by Dipl. Ing. Markus Bühren and Dipl.-Ing. Rubén Villarino-Villa. The readers of this work owe them as much thanks as I do. I would further like to thank all of my colleagues in the Signal Processing Group, to whom I owe countless valuable discussions and advice, and - last but not least - great fun in the course of my work.

Finally, I would like to express thanks to my parents, Ursula and Modesto Pesavento, for giving me the opportunity to study. Their great expectations in me were expressed through insistent questions concerning my work.
Contents

1 Introduction 1
   1.1 Applications ................................................................. 1
   1.2 Data model ................................................................. 4
       1.2.1 Pure and damped uniform MD HRP ............................ 5
       1.2.2 Partly uniform HRP ................................................. 6
       1.2.3 Partly structured HRP ............................................. 7
       1.2.4 Nonuniform HRP .................................................... 8
       1.2.5 Incomplete data HRP .............................................. 8
   1.3 Outline ................................................................. 8

2 Subspace based 2D harmonic retrieval algorithms 11
   2.1 Covariance approach .................................................. 14
       2.1.1 Subspace relations .............................................. 17
   2.2 Data domain approach ................................................ 18
       2.2.1 Forward-backward averaging .................................. 22
   2.3 ESPRIT algorithms .................................................... 23
   2.4 MUSIC algorithms .................................................... 27
   2.5 Estimation of the linear parameters ............................... 28

3 Rank reduction estimators 29
   3.1 Conventional approach ............................................... 29
   3.2 Relaxed optimization approach ..................................... 32
   3.3 Gaussian-elimination approach and uniqueness .................. 35
3.4 Multiple invariance approach ........................................... 38
3.5 Relations between the approaches .................................... 42

4 Extensions to the remaining array axes 49
4.1 Uniform sampling along all array axis .............................. 49
4.2 Spectral rank reduction estimator .................................. 52

5 Implementation 57
5.1 Polynomial-rooting methods ........................................ 57
  5.1.1 FFT approach .................................................. 57
  5.1.2 Block companion matrix approach ............................. 59
5.2 Noise and finite sample effects .................................... 61

6 Parameter association and MD processing 67
6.1 MD tree-RARE ...................................................... 68
6.2 Eigenvector approach ................................................ 73
6.3 Generalized eigendecomposition approach ...................... 77
  6.3.1 Root-MI-ESPRIT .............................................. 79
  6.3.2 Joint root-MI-ESPRIT ......................................... 81
6.4 Non-uniform sampling case ........................................ 83

7 Simulation results 85
7.1 Synthetic data ........................................................ 85
7.2 Measurement data .................................................... 97

8 Conclusions and Outlook 105

A Useful properties of vector algebra 107
Contents

B Proof of T2 110

C Proof of equivalence between $M_3(a, H \mid \text{“d”})$ and $M_5(a, H \mid \text{“d”})$ 112

D Proof of (3.39) 113

E MPs along remaining dimensions and properties 116

F Finite sample MPs along remaining dimensions 121

G Deterministic CRB for pure and damped HR 122

H Notation and symbols 125

Notation and Symbols 125

Bibliography 133
1 Introduction

The one- and multi-dimensional (MD) harmonic retrieval problem (HRP) is encountered in a variety of classical signal processing applications including sensor array processing, radar and mobile communications and has been studied for many decades. Novel applications in which the MD HRP applies are consistently discovered, recent examples are the parametric Multiple-Input Multiple-Output (MIMO) channel identification and the Nuclear magnetic Resonance (NMR) spectroscopy. Early attempts to solve the HRP were based on non-parametric approaches and merely consisted in Fourier-based spectral analysis. The fundamental drawback in these methods contains in the fact that their performance is limited by the available sample support, regardless the given number of realizations and the Signal-to-Noise Ratio (SNR). In recent years so-called high-resolution methods for parametric MD harmonic retrieval (HR) became very popular due to their ability to yield estimation performance beyond the Fourier-limit [KV96].

On the one hand, the profound understanding gained over the years of intensive research in MD HR and the large variety of methods that are available in signal processing literature motivate the efforts that sometimes have to be made to adapt specific applications to the framework of MD HRP. Formulating the estimation task as a HRP includes suitable design of the experiment and the acquisition system as well as appropriate preprocessing of the measurement data.

One the other hand, the large variety of applications and the demand for new MD algorithms with improved estimation performance for low SNR or small sample support at reduced computational cost make MD HR a challenging problem for ongoing research. The next section briefly describes, based on three examples, namely parametric MIMO channel identification, direction-of arrival (DOA) estimation in array processing, and NMR spectroscopy, how MD HR data is obtained from the measurement systems in these applications.

1.1 Applications

Parametric MIMO channel identification

Stochastic channel models are widely used in MIMO communication systems. Recently, novel parametric channel models have gained increasing attention in MIMO channel sounding. The physical parameters that are considered in these models contain substantial information about the channel characteristics and can provide answers to important questions concerning the scat-
terer distribution of the channel, as well as the existence of a rich multipath environment, dominant propagation paths, and line-of-sight propagation. The model parameters further allow to make statements on the coherence time of the channel, i.e. the time during which the channel can be regarded as stationary. This information can then be used to select the best statistical channel model, to adjust its input parameters, and to develop new realistic channel models. Further the parameter estimates obtained from a channel sounding experiment can be exploited to design site specific wireless networks. Specifically, the knowledge about dominant propagation paths for a given environment allow to optimize the sensor locations of the MIMO system to guarantee high channel capacity.

In the double-directional MIMO channel model the signal is assumed to propagate from the transmitter to the receiver over \( P \) discrete propagation paths. In the three-dimensional (3D) parameter model each path \((p = 1, \ldots, P)\) is characterized by the following parameters: complex path gain \( w_p \), direction-of-departure (DOD) \( \tilde{\gamma}_p \), DOA \( \tilde{\beta}_p \) and propagation delay \( \tilde{\alpha}_p \).

In an idealized data acquisition model for MIMO channel sounders data consists of simultaneous measurements of the individual complex baseband channel impulse responses between all \( M \) transmit antenna elements (Tx) and all \( L' \) receive antenna elements (Rx) after ideal low-pass filtering. These are assembled in a three-way array with dimensions \( K \times L' \times M \). Such a three-way array forms a so-called a “MIMO snapshot” and consists of \( K \) time samples with sampling period \( T_s \).

The MIMO snapshot is modeled as

\[
[Y]_{k,\ell,m} = \sum_{p=1}^{P} w_p \, \text{sinc}(k - \tilde{\alpha}_p / T_s) \, b_p^{\ell} \, c_p^m + \text{noise},
\]

where

\[
b_p = e^{-j \frac{2\pi d_R}{\lambda} \cos \tilde{\beta}_p}, \quad c_p = e^{-j \frac{2\pi d_T}{\lambda} \cos \tilde{\gamma}_p}.
\]

The three indices \( k, \ell, \) and \( m \) represent the time sample, the Rx element number, and the Tx element number, respectively. We have assumed uniform linear receive and transmit arrays, where \( \lambda \) is the wavelength, and \( d_R \) and \( d_T \) denote the elemental spacings of the receive and transmit side, respectively.

The *Discrete Fourier Transform* (DFT) over the time sample index \( k \) yields

\[
[Y]_{k,\ell,m} = \sum_{p=1}^{P} w_p \, a_p^k \, b_p^{\ell} \, c_p^m + \text{noise},
\]

where

\[
a_p = e^{-j \frac{2\pi T_s}{K} \tilde{\alpha}_p}, \quad b_p = e^{-j \frac{2\pi d_R}{\lambda} \cos \tilde{\beta}_p}, \quad c_p = e^{-j \frac{2\pi d_T}{\lambda} \cos \tilde{\gamma}_p}.
\]
1.1 Applications

The MIMO channel estimation problem under the double-directional channel model thus consists of estimating the parameters of interest \( \{a_p, b_p, c_p\}_{p=1}^{P} \), where \( |a_p| = |b_p| = |c_p| = 1 \) and the linear parameter \( w_p \) is considered as an unknown nuisance parameter.

**DOA estimation in array processing**

Direction-of-arrival estimation in sensor arrays appears in a variety of important applications including sonar, radar, and mobile communications. Planar array configurations allow to estimate the azimuth and elevation-angle that the impinging wavefronts form with the \( a \)- and \( c \)-axis, respectively. In a uniform rectangular array as given in figures 1.1 and 1.2.(a) with origin in the sensor element \((1,1)\), the response of the \((k, l)\)th sensor element to a far-field narrow-band signal at azimuth angle \( \bar{\alpha}_p \) and elevation angle \( \bar{\beta}_p \) is represented by the product \( w_p a_p^{k-1} b_p^{l-1} \). Here \( a_p = e^{j\frac{2\pi}{\lambda} d_a \cos \bar{\alpha}_p \sin \bar{\beta}_p} \) and \( b_p = e^{j\frac{2\pi}{\lambda} d_b \sin \bar{\alpha}_p \sin \bar{\beta}_p} \) are the harmonics along the respective axis, \( w_p \) is the signal amplitude, \( \lambda \) denotes the wavelength, \( d_a \) and \( d_b \) are the inter-element separation along the \( a \)- and \( b \)-axis, and the integers \( K \) and \( L \) mark the number of sensors aligned in each row and each column of the array. When \( P \) signals are received by the array in the absence of sensor noise, the measurement obtained at the \((k, l)\)th element can be characterized as the superposition

\[
[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} b_p^{l-1}, \quad (1.5)
\]

for \( k = 1, \ldots, K \) and \( l = 1, \ldots, L \). Similarly, in an array configuration composed of identically oriented uniform linear arrays (ULAs) aligned along the \( a \)-axis with arbitrary inter-subarray displacements as depicted in figure 1.2.(b), the signal received from the \((k, l)\)th sensor element is given by

\[
[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} a_p^{\varepsilon_a,l} b_p^{\varepsilon_b,l} \quad (1.6)
\]

where \( a_p = e^{j\frac{2\pi}{\lambda} d_a \cos \bar{\alpha}_p \sin \bar{\beta}_p} \) and \( b_p = e^{j\frac{2\pi}{\lambda} d_b \sin \bar{\alpha}_p \sin \bar{\beta}_p} \) are the harmonics that contain the DOAs of interest. According to figure 1.2 the parameters \( \varepsilon_{a,l}d_a \) and \( \varepsilon_{b,l}d_b \) denote the displacement of the sensor element indexed by \((1, l)\) with respect to the origin \((\varepsilon_{a,1} = \varepsilon_{b,1} = 0)\) along the \( a \)- and \( b \)-axis, respectively.

**Nuclear magnetic resonance spectroscopy**

Two-dimensional (2D) nuclear magnetic resonance (NMR) data is obtained from exciting a molecular system with a 2D radio-frequency (RF) pulse sequence [BL86] and can be modeled as sum of MD damped harmonics. In the classic 2D nuclear magnetic resonance experiment the two sampling axis contain two time intervals \( t_e \) and \( t_d \). The first time interval \( t_e \) denotes
the so-called evolution time during which the excited nuclei precess freely with its resonance (Larmor) frequency. After applying a sequence of RF pulses in the so-called mixing period, during which the nuclei under investigation are subject to different effects (coupling, chemical exchange ...) the detection phase begins. In this phase the intensity of the resonances at time intervals $t_d$ for $t_d = 0, T_d, \ldots, (L - 1)T_d$ with sampling period $T_d$ is measured. The experiment is repeated from a large number of incremental evolution times $t_e = 0, T_e, \ldots, (K - 1)T_e$ with $T_e$ denoting the sampling period along the second time axis. The measurements are stored in a $K \times L$ matrix which, in the noise-free case, corresponds to the following model

$$[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} b_p^{l-1}. \quad (1.7)$$

Here, $w_p$ denotes the amplitude of the 2D resonances, $a_p = e^{2\pi\frac{\mu}{T_e}(\theta_p + j\beta_p)}$ and $b_p = e^{2\pi\frac{\nu}{T_d}(\theta_p + j\beta_p)}$ are the damped harmonics corresponding the $p$th resonance observed along the first and second time axis, respectively. The damping factors along the two sampling axis are denoted by $\mu_p$ and $\nu_p$ with the corresponding frequencies $\theta_p$ and $\beta_p$. The amplitudes, frequencies and damping factors of the 2D harmonics provide information about the chemical shifts or resonances in a molecule, the couplings between nuclear dipoles, the geometric structure of the molecules and also about chemical exchange between two sites.

### 1.2 Data model

In this section a general description of the data model associated with the HRP is provided. This model marks the general framework under which the different applications given above can be handled. Towards this aim, consider the following 2D mixture

$$[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} f_l(\theta_p, \mu_p, \alpha_p) + \text{noise}, \quad \text{for} \quad \begin{cases} k = 1, \ldots, K \\ l = 1, \ldots, L \end{cases} \quad (1.8)$$

where $K$ and $L$ define the sample support along the first and the second array axis. Here, $w_p$ is the linear parameter denoting the complex signal weight of the $p$th signal. The parameters
1.2 Data model

\( a_p = e^{j\alpha_p} \) for \( p = 1, \ldots, P \) are the harmonics observed along the first array axis and denote the parameters of interest. The \( p \)th harmonic \( a_p \) is fully characterized by its damping factor \( \mu_p \) and its frequency \( \alpha_p \). From applied physical considerations we assume in the following that the damping factors \( \mu_p \leq 0 \). The vectors \( \theta_p \) and \( \psi_p \) are nuisance vectors in the vector spaces \( P \) and \( Q \) that contain all remaining parameters associated with the \( p \)th signal and the measurement setup, respectively. The function \( f_l(\theta_p, \psi_p, \mu_p, \alpha_p) \) describes the dependency of the nuisance parameters on the \( l \)th observation taken along the second array axis. The additive noise term in (1.8) will be specified in chapter 2.

The model formulation in (1.8) represents the MD HR problem in a fairly general form. Specifically, it includes the special cases listed in the following subsections.

### 1.2.1 Pure and damped uniform MD HRP

When dealing with an MD mixture of pure and damped harmonics the data samples form a MD structure. For simplicity of notation, we consider the 3D case for detailed discussion because all features that are of particular importance in the MD HRP can very well be illustrated in the 3D case, and its generalization to the MD case is straightforward. In the pure HRP, the 3D harmonic corresponding to the \( p \)th signal is given by the triplet \((a_p, b_p, c_p)\) were the individual generators observed along the first, second, and third sampling axis read \( a_p = e^{j\alpha_p} \), \( b_p = e^{j\beta_p} \) and \( c_p = e^{j\gamma_p} \), respectively. The parameters in the set \((\alpha_p, \beta_p, \gamma_p)\) denote the 3D frequency that fully characterizes the \( p \)th harmonic. Note that this model applies, for example, to the parametric MIMO channel identification problem as given before.

In the damped HRP the \( p \)th harmonic is described by the generators \( a_p = e^{\mu_p+j\alpha_p} \), \( b_p = e^{\nu_p+j\beta_p} \) and \( c_p = e^{\xi_p+j\gamma_p} \) with damping factors \( \nu_p, \mu_p \) and \( \xi_p \) and frequencies \( \alpha_p, \beta_p \) and \( \gamma_p \) along the first, second, and third array axis, respectively. The integers \( K, L' \) and \( M \) mark the sample support along the three dimensions. Assuming uniform sampling along all array axes, the measurements form a data cube or so-called three-way array [LS02, MSPM04] of dimensions \( K \times L' \times M \) denoted by \( Y \) with entries given as

\[
[Y]_{k,l',m} = \sum_{p=1}^{P} w_p a_p^{k} b_p^{l'} c_p^{m} + \text{noise} .
\] (1.9)

If we concatenate \( M \) consecutive \( K \times L' \) matrices obtained from the three-way array in (1.9) by fixing the sample index along the third axis to successive values \( m = 1, \ldots, M \),\(^1\) we obtain

\(^1\)That is equivalent to introducing the new indices \( l = l' + (m - 1)L' \) for \( l' = 1, \ldots, L' \); \( m = 1, \ldots, M \) and assigning \([X]_{k,l} = [Y]_{k,l',m}\)
1 Introduction

Figure 1.2: (a) Uniform rectangular array. (b) Planar array composed of identical and identically oriented ULAs with arbitrary subarray displacements.

The extended $K \times (L'M)$ matrix with entries

$$[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} \left[ b_p^{l-1} c_p^{m-1} \right] + \text{noise}$$

$$= \sum_{p=1}^{P} w_p a_p^{k-1} f_l \left( [\nu_p, \beta_p, \xi_p, \gamma_p]^T \right) + \text{noise}$$

where according to the general model (1.8) we identify

$$f_l (\theta_p, \varnothing) = f_l (\theta_p) = f_l \left( [\nu_p, \beta_p, \xi_p, \gamma_p]^T \right) = b_p^{l-1} c_p^{m-1}$$

and the new sample index $l = (l' + mL')$ is defined over the sample support $l = 1, \ldots, L$ for $L = L'M$. Note that, for the nuisance parameter vectors we obtain $\theta_p = [\nu_p, \xi_p, \beta_p, \gamma_p]^T$ in the damped harmonic case where $\mathcal{P} = \mathbb{R}^4$. We stress that in this example the nuisance parameter vector $\varnothing$ is the empty vector $\varnothing = \emptyset$. With the given choice of parameters the 3D model (1.10) translates to the general model in (1.8).

1.2.2 Partly uniform HRP

The data model (1.8) further encompasses the 2D HRP with uniform sampling along the first array axis and with (known or unknown) non-uniform sampling pattern along the second array.
1.2 Data model

axis. This case corresponds to the 2D DOA estimation problem in rectangular array geometries where only the sensors along the first array axis are aligned on a uniform grid with common baseline spacing or in partly calibrated subarrays composed of identically oriented ULAs with unknown subarray displacements (see figure 1.2) [PGW02a, PGWB01]. The \((k, l)\)th data sample reads then

\[
[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} \left( a_{p-1} b_p + \text{noise} \right) = \sum_{p=1}^{P} w_p a_p^{k-1} f_l(\nu_p, \beta_p, a_p) + \text{noise} \tag{1.12}
\]

where \(b_p = e^{i\nu_p + j\beta_p}\) is the harmonic of the \(p\)th signal with damping factor \(\nu_p\) and frequency \(\beta_p\) taken along the second array axis over a sample support of \(L\). It is easy to verify that the nuisance parameter vector corresponding to the \(p\)th signal can be written as \(\theta_p = [\nu_p, \beta_p]^T \in \mathcal{P}\) with \(\mathcal{P} = \mathbb{R}^2\). The parameter vector characterizing the non-uniform sampling axis \(\vartheta = [\varepsilon_{a,2}, \ldots, \varepsilon_{a,L}, \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}]^T \in \mathcal{Q}\) with \(\mathcal{Q} = \mathbb{R}^L\) is either assumed to be known perfectly in a calibrated acquisition system or alternatively assumed to be unknown in a partly calibrated system. According to the general model (1.8), we identify \(f_l(\nu_p, \beta_p, a_{p-1}, b_p) = a_p \varepsilon_{a,l} b_p \varepsilon_{b,l}\). Note that the partly uniform sampling case translates to the uniform sampling case for integer \(\varepsilon_{a,l} = 0\) and \(\varepsilon_{b,l} = l\) with \(l = 1, \ldots, L\).

1.2.3 Partly structured HRP

The partly structured HRP is closely related to the previous case of partly uniform HR. Similar to the preceding section, consider now e.g. the problem of 2D DOA estimation in identically oriented subarrays with arbitrary amplitude and phase uncertainties between the individual subarrays. These calibration errors may result from subarray displacements (see figure 1.2), differences in the sensor characteristics or non-identical complex gains in the receiver electronics of different subarrays. In this case, the amplitude and phase relations between samples taken along the second array axis are unknown and the \((k, l)\)th data measurement becomes

\[
[X]_{k,l} = \sum_{p=1}^{P} w_p a_p^{k-1} [B]_{l,p} + \text{noise} = \sum_{p=1}^{P} w_p a_p^{k-1} f_l(\theta_p) + \text{noise} \tag{1.13}
\]

where \(B\) is a complex \(L \times P\) matrix with no particular structure. Taking the first sensor in the first subarray as a reference, the first column of \(B\) contains ones in all entries. The remaining elements \([B]_{l,p}\) for \(l = 2, \ldots, L\) and \(p = 1, \ldots, P\) represent the amplitude and phase of the
pth signal observed in the lth sensor of the kth subarray with respect to the reference element. The parameter vector \( \theta_p \) related to the pth signal is then given by the second to Lth complex entry in the pth column of \( B \). Further it is readily verified that \( f_l(\theta_p) = [B]_{l,p} \) and that the space \( \mathcal{P} = \mathbb{C}^{L-1}\backslash\{0\} \), where we excluded the zero vector to avoid the trivial solution. For sake of completeness, note that here the parameter vector \( \theta \) containing the nuisance parameters associated with the acquisition system is the empty vector. Thus the estimation problem consists of determining the frequencies \( \alpha_1, \ldots, \alpha_P \) and damping factors \( \mu_1, \ldots, \mu_P \) along the a-axis and the unknown entries of the complex signal matrix \( B \) along the b-axis.

1.2.4 Nonuniform HRP

The case where all array axes, including the first axis, are sampled non-uniformly, is not covered by the framework of model (1.8) and is beyond the scope of this work. This estimation problem emerges for example in 2D DOA estimation in sensor arrays composed of identically oriented non-uniform subarrays.

1.2.5 Incomplete data HRP

In the incomplete data HRP some samples in the data matrix \( X \) are missing. This estimation problem is also beyond the scope of model (1.8). If the data matrix becomes sparse, the highly symmetric structure of the measurement setup is lost. Incomplete data sets are obtained for example in 2D DOA estimation with multiple nonidentical but identically oriented subarrays, see [PGWB01, PGW02a]. This includes sparse uniform rectangular array configurations where spatial samples at certain sensor locations on a rectangular grid are not observable due to array design or sensor failure.

1.3 Outline

In this work, the MD estimation problem is formulated and analysed via the compact model (1.8). In the following chapter we briefly review two of the most important subspace algorithms for MD HR. In chapter 3 we consider the problem of estimating only the generators along the first data axis, while the remaining parameters are regarded as nuisance parameters. We shall see that this concept allows a simple separation of the parameters along the first array axis from others along the remaining axes, for any of the model specification of sections 1.2.1-1.2.3. This procedure makes the estimation problem computationally tractable while retaining much of the benefits inherent in the MD nature of the measurement data, such as relatively
mild identifiability conditions and high resolution capability compared to 1D HR data. Chapter 4 provides the means for estimating the harmonics observed along the remaining array axes. Chapter 5 deals with implementation issues in the presence of additive noise. In chapter 6 we treat the problem of how to mutually associate the parameter estimates that are separately obtained along the various dimensions. Simulation result obtained both from synthetic and real measurement data are presented in chapter 7. Finally, in chapter 8 we review and evaluate the main results of this work and provide an outlook on open problems for future research.
1 Introduction
2 Subspace based 2D harmonic retrieval algorithms

Subspace based parameter estimation methods in signal processing and system identification have a tradition of more than 30 years. Starting from the early work by Pisarenko [Pis73] several high resolution algorithms like Multiple Signal Classification (MUSIC) [Sch79, BK80, Sch81, BK83]¹, Estimation of Signal Parameters via Rotation Invariance Techniques (ESPRIT) [RK89], Method Of DOA estimation (MODE) [SS90b, SS90a] and Weighted Subspace Fitting (WSF) [VOK91, VS94] have been proposed in engineering literature. The key idea of subspace methods is to exploit the low-rank structure of the signal components which is shared by many signal processing models. The low-rank structure on the measurement data is efficiently enforced using the singular value decomposition. Originally, subspace based methods, also referred to as high-resolution methods, were developed to increase the resolution of spectral-based DOA and frequency estimation methods beyond the classical Fourier limit [KV96]. Today a large variety of high-resolution techniques have found wide application in radar, sonar and mobile communication systems. Recently, subspace based methods have been successfully applied to estimate the channel parameters of MIMO communication systems [HVU02, SHS+00, THR+99, HMM+02, SHK+01, FRB97, PMB04].

This chapter investigates the low-rank properties associated with the sum-of-harmonic mixtures given in (1.8). Towards this aim, it is convenient to rearrange the entries in the $K \times L$ data matrix $X$ in an appropriate way to form a “long” $KL \times 1$ measurement vector $x$ [PMB04, HN98, JStB01]. Let $\text{vec}\{M\}$ denote the vectorization operator that stacks the individual columns of a matrix $M$ on top of each other so that

$$ x = \text{vec}\{X\}. $$

(2.1)

In vector notation model (1.8) reads

$$ X = \sum_{p=1}^{P} w_p a_p f^T(\theta_p, \vartheta, \mu_p, \alpha_p) + \text{noise}, $$

(2.2)

where

$$ a_p = [1, a_p, a_p^2, \ldots, a_p^{K-1}]^T \in \mathbb{C}^K $$

(2.3)

defines a Vandermonde vector in the generator $a_p$ and

$$ f(\theta_p, \vartheta, \mu_p, \alpha_p) = $$

$$ = [f_1(\theta_p, \vartheta, \mu_p, \alpha_p), f_2(\theta_p, \vartheta, \mu_p, \alpha_p), \ldots, f_L(\theta_p, \vartheta, \mu_p, \alpha_p)]^T \in \mathbb{C}^L. $$

(2.4)

1Even though these are the classic references for the MUSIC algorithm commonly cited in array processing literature, eigenvector based peak estimators with different eigenvalue weighting functions have already been introduced several years before. For a overview on early reference refer to [Böh83] and references therein.
Inserting (2.2) into (2.1) and making use of property (A.6) we obtain

\[ x = \text{vec} \ X = \text{vec} \left\{ \sum_{p=1}^{P} w_p a_p f^T(\theta_p, \vartheta, \mu_p, \alpha_p) \right\} + \text{noise} \]

\[ = \sum_{p=1}^{P} w_p \text{vec} \left\{ (f(\theta_p, \vartheta, \mu_p, \alpha_p) \otimes a_p) \right\} + \text{noise} \]

\[ = (F \circ A) w + \text{noise} \]

\[ = H w + \text{noise} , \quad (2.5) \]

where

\[ w = [w_1, \ldots, w_P]^T \quad (2.6) \]

denotes the complex weight vector, "\( \otimes \)" stands for the Kronecker-product (A.2), and "\( \circ \)" denotes the Khatri-Rao product as specified in (A.3). Further in (2.5) the \( K \times P \) Vandermonde matrix

\[ A = [a_1, a_2, \ldots, a_P], \quad [A]_{k,p} = (a_p)^{k-1} , \quad (2.7) \]

is composed of the generators of interest that are observed along the first array axis. We shall refer to this matrix in the following as the signal matrix along the \( a \)-axis. The \( L \times P \) matrix

\[ F = [f(\theta_1, \vartheta_1, \mu_1, \alpha_1), f(\theta_2, \vartheta_2, \mu_2, \alpha_2), \ldots, f(\theta_P, \vartheta_P, \mu_P, \alpha_P)] \quad (2.8) \]

contains the remaining signal and nuisance parameters along the second array axis. Finally, the \( KL \times P \) signal matrix \( H \) in (2.5) is defined as

\[ H = F \circ A . \quad (2.9) \]

The following assumption establishes a general low-rank model:

**Assumption A1:** The signal matrix \( H \) has full column-rank.

Note that in order to guarantee A1, certain assumptions on the maximum number of harmonics \( P \) that are superimposed in the MD mixture (1.8) need to hold. The number of signals that can uniquely be identified from the observations mainly depends on the number of available samples and the sampling scheme that is used in the data acquisition. The conditions that guarantee a full rank signal matrix are commonly referred to as identifiability conditions of the associated parameter estimation problem. We distinguish between deterministic identifiability conditions [MP98, MPL99, MSD01], which are conditions that only concern the sampling scheme, and stochastic identifiability conditions that also regard the distribution of the generators from which the signal matrix is formed [SLS01, JStB01, SBG00, LS02, MSPM04]. It is clear that the deterministic identifiability of \( P \) signals implies much stronger conditions than
stochastic identifiability, because identifiability needs to be assured for all generator sets including ill-posed cases. Therefore, to prevent overstrict identifiability conditions on the maximum number of signals, it is useful to assume a continuous distribution of the generators and consider the so-called stochastic identifiability of the estimation problem. Stochastic identifiability of \( P \) harmonics in the MD mixture for a given distribution of generators then means that the parameters of \( P \) signals drawn from the indicated distribution are almost-surely, hence with probability one, uniquely resolvable.

In this work the focus lies on uniform sampling along at least one data dimension (1.8). It is well known that uniform sampling schemes often suffer from ambiguities. Deterministic identifiability conditions related with highly regular MD sampling structures are difficult to derive analytically and simulation results show that existing identifiability bounds appear to be overstrict in practically all relevant cases [MP98, MPL99, MSD01]. Therefore, in this work we only consider stochastic identifiability based on generator distributions that seem reasonable in practical applications.

The following result obtained in [JStB01] shall provide further insight in the implication of assumption \( A 1 \) in the uniform sampling case of section 1.2.1.

**Theorem T1:** Given \( N \geq 2 \) Vandermonde matrices \( L_n \in \mathbb{C}^{K_n \times P} \) for \( n = 1, \ldots, N \), with complex generators drawn from a \( NP \)-dimensional complex distribution that is assumed to be continuous with respect to the Lebesgue measure in \( \mathbb{C}^{NP} \), then the following rank result holds almost-surely, i.e. with probability one:

\[
\text{rank}\{L_1 \circ \ldots \circ L_N\} = \min \left\{ P, \prod_{n=1}^{N} K_n \right\} . \tag{2.10}
\]

Theorem T1 reveals that in the uniform sampling case of section 1.2.1 the signal matrix has almost-surely full column rank \( H \) if \( P \leq KL \) and provided that the generators of the \( P \) harmonics along the different axis are drawn from a single MD complex distribution that is assumed to be continuous with respect to the Lebesgue measure (as specified in the Theorem).

For the remaining cases covered by model (1.8) and specified in section 1.2, similar assumptions on the generators can be made to guarantee that the corresponding signal matrix is full rank with probability one. These assumptions can directly be derived following the proof of T1 in [JStB01] and will not be discussed here. Apparently, the case of uniform sampling along all observation axis covers the most restrictive case in terms of existing ambiguities resulting from the sampling scheme.

Provided that \( A 1 \) is satisfied, equation (2.5) reflects the low-rank property of the data model. In other words, if the total number of available samples taken along the first and the second array axis exceeds the number of harmonics \( P \), then the signal matrix \( H \) spans a \( P \)-dimensional
subspace of the $KL$-dimensional complex space. This subspace is in the following referred to as the signal subspace $S$ [Tre02]. Ignoring all noise terms for the time being, equation (2.5) reveals that in the ideal case the observation vector $x$ represents a linear combination of the signal vectors in $H$ and thus lies itself in the signal subspace $S$.

### 2.1 Covariance approach

Additional assumptions are made with respect to the noise term and the signal weights in the data model (2.1). More precisely, specific assumptions on the statistical distribution of the noise contributions must be made. Based on them, this section shows a natural way of separating the signal from the noise subspace using second-order moments and eigendecomposition.

A common approach in HR estimation is to assume that the random noise contributions contained in different samples are independently identically distributed (i.i.d.) complex white Gaussian. This rather ideal noise assumption turns out to be applicable in many applications including radar, DOA estimation in sensor arrays, parametric MIMO channel estimation, and MR spectroscopy, and will be used in the following. Noise models and HR methods applicable under sophisticated noise assumptions are found in [RSB01, BSG91, GBS91, ZA04, PGH00c, PG00, GSPL02].

Let $n \in \mathbb{C}^{KL}$ denote the vector containing the noise contributions of the individual observations in data vector $x$, so that model (2.5) becomes

$$x = \text{vec} \{X\} = Hw + n.$$  \hspace{1cm} (2.11)

The statistical properties of the noise vector are compactly written as

$$\mathbb{E}\{n\} = 0 \hspace{1cm} (2.12)$$

$$\mathbb{E}\{nn^H\} = \sigma^2 I_{KL} \hspace{1cm} (2.13)$$

$$\mathbb{E}\{nn^T\} = 0 \hspace{1cm} (2.14)$$

where $\mathbb{E}\{\cdot\}$ stands for statistical expectation. With the noise properties established here we can now distinguish between two different models of the linear weight vector that are commonly found in signal processing literature [SN89, Tre02]: a) the unconditional or stochastic weight vector model and b) the conditional or deterministic weight vector model.

**Conditional model**

In the conditional weight vector model the weight vector $w$ in (2.11) is assumed to be deterministic and unknown. According to our considerations in context of (2.5) and in absence of
noise the data vector $x$ represents a linear combination of the signal vectors in $H$ with linear coefficients $w_1, \ldots, w_P$. In this case the data vector clearly lies inside the signal subspace $S$. In order to obtain a set of vectors that span the full signal subspace, multiple independent realizations are required. Later on in section 2.2 we shall illustrate how in case of uniform sampling along the array axis forward-backward averaging and smoothing techniques can be applied to acquire multiple data snapshots from a single realization. In case of multiple time-samples the data model (2.11) naturally extends to

$$x(t) = \text{vec} \, X(t) = Hw(t) + n(t)$$

(2.15)

for $t = 1, \ldots, N$. With respect to the noise vector $n(t)$ we assume that the noise is temporally uncorrelated, that is $E\{n(t)n^H(t')\} = \sigma^2 \delta_{t,t'}$ and further uncorrelated along the sampling axis, thus $E\{[n(t)]_k[n^H(t)]_l\} = \sigma^2 \delta_{k,l} \delta_{t,t'}$. Here $\sigma^2$ denotes the noise variance. In the conditional weight vector model, the weight vectors $w(t)$ at the different time instances $t$ are assumed to take arbitrary deterministic values, i.e. no assumptions on the distribution of the weights are made. Hence, disregarding noise, it is simple to observe from (2.15) that at least $N = P$ snapshots corresponding to a linearly independent set of weight vectors $w(1), \ldots, w(N)$ are required to allow full recovery of the complete signal subspace $S$ from the data vectors $x(1), \ldots, x(N)$.

A convenient way of separating the signal and noise subspaces relies on the low-rank property of the data covariance matrix. In fact, the rank of the data matrix is handed over to the rank of the covariance matrix in the noise-free case. Each data vector $x(t)$ in the conditional model contributes a rank one signal component $Hw(t)w^H(t)H^H$ to the data correlation matrix at time instant $t$ that is superimposed to a diagonal noise covariance matrix $\sigma^2 I_{KL}$ yielding

$$R_t = E\{x(t)x^H(t)\} = Hw(t)w^H(t)H^H + E\{n(t)n^H(t)\} = Hw(t)w^H(t)H^H + \sigma^2 I_{KL}.$$

(2.16)

A natural way of creating a correlation matrix with a signal component of rank $P$ is to simply average the covariance matrices $R_t$ over a time interval $t = 1, \ldots, N$ ($N \geq P$). This results in the multiple-snapshot correlation matrix

$$R = \frac{1}{N} \sum_{t=1}^{N} E\{x(t)x^H(t)\} = \frac{1}{N} H \sum_{t=1}^{N} (w(t)w^H(t)) H^H + \sigma^2 I_{KL} = H\hat{P}H^H + \sigma^2 I_{KL}$$

(2.17)

where

$$\hat{P} = \frac{1}{N} \sum_{t=1}^{N} w(t)w^H(t)$$

(2.18)
denotes the so-called sample correlation matrix associated with the weight vector $w(t)$. Clearly, $\hat{P}$ represents a sum of dyadic products that is strictly non-negative definite and which eventually becomes strictly positive definite if the number of time snapshots $N \geq P$, where $P$ is the true number of signals.

Conventionally, signal and noise subspaces separation relies on the singular-value decomposition of the correlation matrix defined in (2.17) as

$$ R = E_S \Lambda_S E_S^H + E_N \Lambda_N E_N^H. \quad (2.19) $$

The diagonal matrices $\Lambda_S \in \mathbb{R}^{(P \times P)}$ and $\Lambda_N \in \mathbb{R}^{(KL-P) \times (KL-P)}$ contain the signal subspace (i.e. the largest $P$) and the noise subspace (i.e. the smallest $(KL-P)$) eigenvalues of $R$ on its main diagonals, respectively. In turn, the columns of the matrices $E_S \in \mathbb{C}^{KL \times P}$ and $E_N \in \mathbb{C}^{KL \times (KL-P)}$ denote the corresponding signal and noise subspace eigenvectors, respectively.

**Unconditional model**

In the unconditional weight vector model we regard the individual signal weights $w_1, \ldots, w_P$ as stochastic quantities with zero mean and non-singular covariance matrix given by

$$ P = E \{ww^H\}. \quad (2.20) $$

In other words, we assume that the weights corresponding to different harmonics are not fully correlated. Under the preceding assumption the data covariance matrix associated with (2.11) reads

$$ R = E \{xx^H\} = HE \{ww^H\} H^H + \sigma^2 I_{KL} = HPH^H + \sigma^2 I_{KL}. \quad (2.21) $$

Apparently the low-rank property of the data model is expressed in the covariance matrix as a rank $P$ contribution of the signal part $HPH^H$ to the overall rank of the positive semi-definite covariance matrix. In fact, in the noise free case we observe a low-rank data covariance matrix $R$ of rank not greater than $P$. The covariance matrix $P$ is per definition non-singular and the signal matrix $H$ is of full column rank per assumption $A1$. Then, Sylvester’s inequality (A.8) yields that the quadratic form $HPH^H$ is positive semi-definite and of rank $P$. Hence, exactly $P$ eigenvalues of the Hermitian matrix $R$ are greater than $\sigma^2$ while $(KL-P)$ eigenvalues are equal to $\sigma^2$. The eigendecomposition of the covariance matrix (2.21) is immediate.

In applications, the true covariance is usually not known. Instead, a finite sample estimate of
the covariance matrix with an eigendecomposition given by

\[
\hat{R} = \frac{1}{N} \sum_{t=1}^{N} x(t)x^H(t) = \hat{E}_S \hat{\Lambda}_S \hat{E}_S^H + \hat{E}_N \hat{\Lambda}_N \hat{E}_N^H
\]

(2.22)
is used. Here, the diagonal matrices \( \hat{\Lambda}_S \in \mathbb{R}^{(P \times P)} \) and \( \hat{\Lambda}_N \in \mathbb{R}^{(KL-P) \times (KL-P)} \) contain, according to (2.19), estimates of the signal subspace and the noise subspace eigenvalues of the sample covariance matrix \( \hat{R} \) on its main diagonals, respectively. In turn, the columns of the matrices \( \hat{E}_S \in \mathbb{C}^{(KL \times P)} \) and \( \hat{E}_N \in \mathbb{C}^{KL \times (KL-P)} \) denote estimates of the corresponding signal subspace and noise subspace eigenvectors.

### 2.1.1 Subspace relations

This section reviews the subspace properties that are fundamental to all subspace-based HR algorithms falling under the framework of the low-rank signal model (2.11). Regardless which of the preceding approaches are followed, the conditional or unconditional covariance data model, in either case there is a well-defined relation between signal and noise subspaces.

When comparing the covariance matrices \( R \) in (2.17) and (2.21) with their corresponding singular value decompositions of the form (2.19) a close relation between the signal matrix \( \mathbf{H} \) and the signal eigenvectors in \( \mathbf{E}_S \) is revealed. This becomes apparent from the covariance matrix in (2.21):

\[
\mathbf{H} \mathbf{P} \mathbf{H}^H + \sigma^2 \mathbf{I}_{KL} = \mathbf{E}_S \mathbf{\Lambda}_S \mathbf{E}_S^H + \mathbf{E}_N \mathbf{\Lambda}_N \mathbf{E}_N^H.
\]

(2.23)

We emphasize the following features of the decomposition in (2.23):

a) the positive definiteness of \( \mathbf{P} \), (see (2.18) ),

b) the full column-rank of \( \mathbf{H} \), (see A1),

c) the positive (semi)-definiteness of \( \mathbf{H} \mathbf{P} \mathbf{H}^H \), (follows from a), b) and Sylvester’s inequality),

d) the “spatial” whiteness of the noise vector (i.e. the diagonal structure of noise covariance matrix \( \sigma^2 \mathbf{I}_{KL} \) (2.13)),

e) the assumption of equal noise power \( (\sigma^2 \geq 0) \) for all data samples,

f) the separation of the signal and noise eigenvectors according to the magnitudes of their corresponding eigenvalues.
These reveal that
\[
\Lambda_S > \sigma^2 I_P \quad (2.24)
\]
\[
\Lambda_N = \sigma^2 I_{KL-P} \quad (2.25)
\]
The signal matrix \( \mathbf{H} \) and the signal eigenvectors in \( \mathbf{E}_S \) span the same signal subspace denoted by \( S \). Hence,
\[
\mathcal{R}\{ \mathbf{H} \} = \mathcal{R}\{ \mathbf{E}_S \} = S. \quad (2.26)
\]
Here, \( \mathcal{R}\{ \mathbf{U} \} \) defines the range-space of a matrix \( \mathbf{U} \), hence the space spanned by the columns of \( \mathbf{U} \). Further, the noise subspace \( \mathcal{N} \), spanned by the columns of \( \mathbf{E}_N \), is a \((KL-P)\)-dimensional space that is orthogonal to \( S \) and that contains all the remaining contributions in the measurements. Thus, \( \mathcal{R}\{ \mathbf{H} \} \perp \mathcal{R}\{ \mathbf{E}_N \} \).

Equation (2.26) implies that there exists a non-singular \( P \times P \) matrix \( \mathbf{K} \) such that
\[
\mathbf{E}_S \mathbf{K} = \mathbf{H}. \quad (2.27)
\]
The full rank matrix \( \mathbf{K} \) relates the unknown signal matrix \( \mathbf{H} \), in which each column vector contains only contributions from one specific signal, with the unitary signal eigenvectors in \( \mathbf{E}_S \) through linear transformation. This matrix plays a substantial role for the derivations given in the following chapters and we shall refer to it as the mixing matrix.

### 2.2 Data domain approach

In many applications, the experimental setup is subject to a rapidly changing environment so that the observation time over which the measurements can be regarded as stationary, the so-called coherence time, is severely limited. This effect is typically observed in MIMO channel sounding experiments, where rapid movements of Tx and Rx positions lead to significant changes of both the scattering environment and the model parameters (DOA, DOD, propagation delay, ...) associated with a specific propagation path between consecutive snapshot. The problem becomes even more severe if time-multiplexing is used to measure the individual transfer functions for all pairs of transmit and receive antennas due to an increase in the acquisition time required for each MIMO snapshot.

Also in MD NMR spectroscopy a lack of stationarity in the experiments often leads to measurements with a coherence time of just a few snapshots. Equation (2.22) shows that in the covariance approach at least \( P \) independent (and stationary) snapshots need to be available to form the required rank \( P \) signal subspace. In this subsection, we shall illustrate how to deduce a low-rank model from the measurements in the single snapshot case. The technique
2.2 Data domain approach

presented here is closely related to spatial smoothing procedures and forward-backward (FB) averaging techniques (in the case of pure exponentials) and requires uniform sampling along all array axes. In specific, we consider the pure and damped MD HRP by means of the 3D HR model in (1.9) and (1.10). Interestingly, the subspace extraction technique addressed here is well known from a variety of different contributions. Similar approaches were developed in [SLS01, LS02, MSPM04, LRL98, HN98].

Consider the data model in (1.9). Adding the $K \times L' \times M$ three-way noise array $N$ to the ideal data array $Y$, the data model becomes

$$
[Y]_{k,l',m} = \sum_{p=1}^{P} w_p a_k^p b_{l'}^p c_m^p + [N]_{k,l',m}, \quad (2.28)
$$

where we assume that the noise contributions are uncorrelated along all sampling axis, hence we suppose that $E\{[N]_{k,l,m} [N^*]_{k',l',m'}\} = \sigma^2 \delta_{k,k'} \delta_{l,l'} \delta_{m,m'}$.

To obtain a low-rank data model of a sufficiently large dimension from the three-way array in (1.9), we rearrange the data samples taken along the first, second and third axis to form a $(K_1 L_1 M_1) \times (K_2 L_2 M_2)$ matrix [SLS01, LRL98]

$$
\tilde{Y} = \begin{bmatrix}
Y_1 & Y_2 & \cdots & Y_{M_2} \\
Y_2 & Y_3 & \cdots & Y_{M_2+1} \\
& & \ddots & \ddots \\
Y_{M_1} & Y_{M_1+1} & \cdots & Y_{M_2+M_1-1}
\end{bmatrix} \quad (2.29)
$$

where

$$
Y_m = \begin{bmatrix}
Y_{1,m} & Y_{2,m} & \cdots & Y_{L_2,m} \\
Y_{2,m} & Y_{3,m} & \cdots & Y_{L_2+1,m} \\
& & \ddots & \ddots \\
Y_{L_1,m} & Y_{L_1+1,m} & \cdots & Y_{L_2+L_1-1,m}
\end{bmatrix} \quad (2.30)
$$

and the integers $K_1, K_2, L_1, L_2, M_1,$ and $M_2$ satisfy

$$
K = K_1 + K_2 + 1, \quad (2.32)
$$

$$
L' = L_1 + L_2 + 1, \quad (2.33)
$$

$$
M = M_1 + M_2 + 1. \quad (2.34)
$$

The integers in (2.32)-(2.34) are chosen such that the reassembled data matrix $\tilde{Y}$ becomes as “large” or as “extended” along both dimensions as possible. It is simple to see that, if we maximize the minimum of $K_1 L_1 M_1$ and $K_2 L_2 M_2$, then the achievable rank of $\tilde{Y}$ and consequently
the maximum number of identifiable signals, is maximized. The reassembled data matrix in (2.28) allows a simple representation [SLS01, LRL98]

$$\tilde{Y} = H_1 W H_2^T + \bar{N}$$  \hspace{1cm} (2.35)

where

$$H_i = C_i \circ B_i \circ A_i, \quad H_i \in \mathbb{C}^{(K_i, L_i, M_i) \times P}$$  \hspace{1cm} (2.36)

$$[A_i]_{k,p} = a_p^{(k-1)}, \quad A_i \in \mathbb{C}^{K_i \times P}$$  \hspace{1cm} (2.37)

$$[B_i]_{l,p} = b_p^{(l-1)}, \quad B_i \in \mathbb{C}^{L_i \times P}$$  \hspace{1cm} (2.38)

$$[C_i]_{l,p} = c_p^{(l-1)}, \quad C_i \in \mathbb{C}^{M_i \times P}$$  \hspace{1cm} (2.39)

$$W = \text{diag}\{w_1, \ldots, w_P\}$$  \hspace{1cm} (2.40)

for \(i = 1, 2\). For reasons of completeness, let us give the structure of the additive \((K_1 L_1 M_1) \times (K_2 L_2 M_2)\) noise matrix, which after reassembling of the data reads

$$\bar{N} = \begin{bmatrix}
N_1 & N_2 & \ldots & N_{M_2} \\
N_2 & N_3 & \ldots & N_{M_2+1} \\
\vdots & \vdots & \ddots & \vdots \\
N_{M_1} & N_{M_1+1} & \ldots & N_{M_2+M_1-1}
\end{bmatrix}$$  \hspace{1cm} (2.41)

where

$$N_m = \begin{bmatrix}
N_{1,m} & N_{2,m} & \ldots & N_{L_2,m} \\
N_{2,m} & N_{3,m} & \ldots & N_{L_2+1,m} \\
\vdots & \vdots & \ddots & \vdots \\
N_{L_1,m} & N_{L_1+1,m} & \ldots & N_{L_2+L_1-1,m}
\end{bmatrix}$$  \hspace{1cm} (2.42)

$$N_{l,m} = \begin{bmatrix}
[N]_{1,l,m} & [N]_{2,l,m} & \ldots & [N]_{K_2,l,m} \\
[N]_{2,l,m} & [N]_{3,l,m} & \ldots & [N]_{K_2+1,l,m} \\
\vdots & \vdots & \ddots & \vdots \\
[N]_{K_1,l,m} & [N]_{K_1+1,l,m} & \ldots & [N]_{K_2+K_1-1,l,m}
\end{bmatrix}. \hspace{1cm} (2.43)$$

Disregarding the noise term \(\bar{N}\) in equation (2.35) for the time being, the singular value decomposition of the reassembled data matrix can be written as

$$\tilde{Y} = U_1 D U_2^T,$$  \hspace{1cm} (2.44)

where \(U_1 \in \mathbb{C}^{(K_1 L_1 M_1) \times P}\) and \(U_2 \in \mathbb{C}^{(K_2 L_2 M_2) \times P}\) denote the matrices composed of the left and right singular vectors, respectively, and the \(P \times P\) diagonal matrix \(D\) contains the corresponding singular values on its main diagonal. In the ideal case it is clear that for low-rank \(\tilde{Y}\) the left signal matrix \(H_1\) and the left singular vectors in \(U_1\) span the same signal subspace. Hence, in allusion to (2.27), there exists a non-singular \(P \times P\) matrix \(K_1\) such that

$$U_1 K_1 = H_1.$$  \hspace{1cm} (2.45)
2.2 Data domain approach

Similarly, it is simple to see that the right signal matrix $H_2$ and the right singular vectors in $U_2$ span the same right signal subspace and there exists a non-singular $P \times P$ matrix $K_2$ such that

$$U_2K_2 = H_2.$$ (2.46)

Note that if additive noise is present in the data samples then the reassembled data matrix is no longer of rank $P$. In this case (2.44) translates to

$$\tilde{Y} = \hat{U}_1 \hat{D} \hat{U}_2^T + N_{\text{residual}},$$ (2.47)

where $\hat{U}_1 \in \mathbb{C}^{(K_1 L_1 M_1) \times P}$ and $\hat{U}_2 \in \mathbb{C}^{(K_2 L_2 M_2) \times P}$ denote the matrices composed of the estimated left and right singular vectors associated with the largest singular values that are arranged on the main diagonal of the $P \times P$ diagonal matrix $\hat{D}$, and the residual term $N_{\text{residual}}$ absorbs all remaining components.

The singular value decomposition in (2.47) is the best rank $P$ approximation of the data matrix $\tilde{Y}$ in a least squares (LS) sense [GvL96]. In other words the singular value decomposition minimizes the Frobenius norm of the residual approximation error matrix, given by $N_{\text{residual}} = \tilde{Y} - \hat{U}_1 \hat{D} \hat{U}_2^T$. However, this only holds in the case that the noise matrix $\tilde{N}$ contains i.i.d. entries [GvL96]. In our case, the noise matrix has the specific redundant block matrix structure displayed in (2.41). Therefore to obtain more reliable estimates of the subspace, it is recommended to design a more sophisticated subspace estimation procedure that incorporates the specific noise structure of the reassembled data matrix in (2.35). This however exceeds the scope of the present work and shall be subject of future research.

It is clear that the block Vandermonde matrices $H_1$ and $H_2$ have in general different sample support along the various array axis according to the integers $K_1$, $K_2$, $L_1$, $L_2$, $M_1$, and $M_2$. However, both signal matrices contain full information about all signal parameters, hence the 3D generators $(a_p, b_p, c_p)$ for $p = 1, \ldots, P$. In the following, we focus on estimating the parameters of interest from the left signal matrix $H_1$. The estimation of the right signal matrix $H_2$ from the right singular vectors in $U_2$ is a dual problem. To simplify notation, and in order to make it consistent with the notation used in the covariance approach presented in the previous section, we introduce the following substitution of identifiers: $H = H_1$, $A = A_1$, $B = B_1$, $C = C_1$, $L' = L_1$, $K = K_1$, and $M = M_1$. (Alternatively we can set $H = H_2$, $A = A_2$, $B = B_2$, $C = C_2$, $L' = L_2$, $K = K_2$, and $M = M_2$.) In either case the new matrices are then defined as

$$H = C \circ B \circ A, \quad H \in \mathbb{C}^{(KL'M) \times P}$$ (2.48)

$$[A]_{k,p} = a_p^{(k-1)}, \quad A \in \mathbb{C}^{K \times P}$$ (2.49)

$$[B]_{l,p} = b_p^{(l-1)}, \quad B \in \mathbb{C}^{L' \times P}$$ (2.50)

$$[C]_{m,p} = c_p^{(m-1)}, \quad C \in \mathbb{C}^{M \times P}$$ (2.51)
Further, with a slight abuse of notation, we substitute the matrix of left singular vectors $U_1$ by $E_S$ and in the following refer to its columns as the signal eigenvectors. This shall simplify the reference on the signal subspace that is estimated either from the covariance in section 2.1.1 or from the data domain approach presented here. Correspondingly, we assign $\hat{E}_S = \hat{U}_1$ for the estimated signal eigenvectors in (2.47) and $K = K_1$ for the mixing matrix in (2.45).

\subsection{Forward-backward averaging}

A popular approach to virtually double the number of samples in the case of pure harmonics and uniform sampling along all array dimensions, is commonly referred to as \textit{forward-backward} (FB) \textit{averaging}. Here the forward part consists of the conventional data processing described above. The backwards part, in turn, stems back from the observation that if taking the complex conjugate of the sum-of-harmonic mixture in the original uniform MD HRP and if also reversing the indices of the samples along all axes then we arrive at a signal subspace formulation in which the same signal vectors $H$ apply as in the \textit{forward-only} approach. To illustrate this in case of 3D pure uniform HRP consider again equation (2.28). The conjugate-reversed version of the three-way array is given by

\begin{equation}
[Y_B]_{k,l',m} = [Y^*]_{K-k+1,L-L'+1,M-m+1} \\
= \sum_{p=1}^{P} w_p^* a_p^{-K+k+1} b_p^{-L-L'+1} c_p^{-M-m+1} + [N^*]_{K-k+1,L-L'+1,M-m+1} \\
= \sum_{p=1}^{P} (w_p^* a_p^{-K+1} b_p^{-L+1} c_p^{-M+1}) a_p^{k} b_p^{l'} c_p^{m} + [N^*]_{K-k+1,L-L'+1,M-m+1} \\
= \sum_{p=1}^{P} w_{B,p} a_p^{k} b_p^{l'} c_p^{m} + [N_B]_{k,l',m} \tag{2.52}
\end{equation}

with the new weights $w_{B,p}, p = 1, \ldots, P$ defined as

\begin{equation}
w_{B,p} = w_p^* a_p^{-K+k+1} b_p^{-L+1} c_p^{-M+1} \tag{2.53}
\end{equation}

and the corresponding noise matrix obtained as $[N_B]_{k,l',m} = [N^*]_{K-k+1,L-L'+1,M-m+1}$. Comparing the backwards data matrix $Y_B$ in (2.52) with the original data matrix in (2.28) and following the same procedure that led to (2.35), it is immediate to show that

\begin{equation}
\tilde{Y}_B = H_1 W_B H_2^T + \tilde{N}_B, \tag{2.54}
\end{equation}

where $Y_B$ and $N_B$ are obtained according to (2.29)-(2.31) and (2.41)-(2.43), replacing $Y$ by $Y_B$ and $N$ by $N_B$, respectively. The diagonal matrix $W_B$ is defined according to (2.40) as

\begin{equation}
W_B = \text{diag} \{ w_{B,1}, \ldots, w_{B,P} \}. \tag{2.55}
\end{equation}
We see from (2.54) that the same subspace relation as in (2.45) and (2.46) can also be formulated for the singular vectors of $\tilde{Y}_B$.

FB averaging, in this context also referred to as MD-folding, was successfully used in [LS02, MSPM04] for the construction of fast estimation procedures and to derive new identifiability results for MD HR in the single snapshot case.

For reasons of completeness we note that also in the covariance approach of sections 2.1 FB averaging is applicable when constructing a backwards covariance matrix $R_B$ from vectorizing the conjugate-reversed data matrix $Y_B$ in lieu of the forward data matrix $Y$. The FB covariance matrix is then defined as

$$R_{FB} = (R + R_B)/2.$$  

(2.56)

It is well known, that in the realistic case estimating the signal subspace from the FB covariance matrix often yields better parameter estimates especially in the case of correlated signals [PGH00c].

In practice, the HR problem consists of estimating the signal matrix $H$ from the signal subspace matrix $E_S$ that itself is obtained from the observations. A large variety of subspace-based HR algorithms can be found in recent literature. The next sections briefly review the two classes of estimators that are most relevant for this work. We intend to classify existing subspace methods in these two types of HR algorithms. The overview shall mark the basis on which new approaches are established. It shall help in putting the novel concepts proposed in the following chapters into context and, without claim of completeness, outline the current state-of-the-art.

### 2.3 ESPRIT algorithms

There exist several subspace algorithms related to the popular ESPRIT technique. This method was derived by Roy [RK89] in the context of 1D DOA estimation and is described in several other publications. It has been generalized to the 2D and MD case and also to multiple-invariance (MI) in numerous different approaches including the unitary ESPRIT approach by Haardt et al. [HN98], the 2D unitary ESPRIT approach by Zoltowski [ZHM96], the MI approach by Swindlehurst et al. [SORK92], the joint diagonalization approach by Van der Veen [vdVVP97, vdVVA98, VvdVP98] and many other related contributions [SLS01, FRB97].

ESPRIT exploits certain invariance structures contained in the measurement setup. Generally speaking, so-called shift or translational invariances emerge when one or more regions of the signal matrix (2.9) translate into another part of the signal matrix by a simple scaling of the individual columns. The highly regular “Khatri-Rao structure” of the signal matrix (2.9) comprises multiple shift-invariances. To illustrate this let us extract specific rows of the signal matrix
to obtain sub-matrices of appropriate structure. The goal is to represent sub-matrices of (2.9) in terms of shifted structures that translate into one another through right-multiplication with appropriate diagonal shifting matrices. Let $\mathbf{J}_{K,k}$ denote the upper $K \times K$ selection matrix

$$
\mathbf{J}_{K,k} = \begin{bmatrix}
I_{K-k} & 0 \\
0 & 0_k
\end{bmatrix} \in \mathbb{R}^{K \times K}.
$$

(2.57)

Then we obtain from (2.9) the $k$th row-reduced upper signal matrix $\mathbf{H}_{a,k}$ defined as

$$
\mathbf{H}_{a,k} = \mathbf{F} \circ \mathbf{A}_k = \mathbf{F} \circ (\mathbf{J}_{K,k} \mathbf{A}) = (\mathbf{I}_L \otimes \mathbf{J}_{K,k}) \mathbf{H}
$$

(2.58)

where the $k$th row-reduced upper Vandermonde matrix

$$
\mathbf{A}_k = \mathbf{J}_{K,k} \mathbf{A}
$$

(2.59)

contains only the elements in the first $K - k$ rows of the original Vandermonde matrix $\mathbf{A}$ while the remaining elements are filled with zero entries. Note that according to the chosen notation, matrix $\mathbf{A}_k$ is not precisely reduced by $k$ rows. In fact $\mathbf{A}_k$ represents a copy of the original Vandermonde matrix $\mathbf{A}$. The original size is left unchanged and only the entries in specific rows (in this case the $(k+1)$th to $K$th row) are set to zero. The same statement holds true for the row-reduced upper signal matrix $\mathbf{H}_{a,k}$ for $k = 1, \ldots, K - 1$.

In the same fashion, we introduce a lower $K \times K$ selection matrix defined as

$$
\mathbf{J}_{K,k} = \begin{bmatrix}
0 & I_{K-k} \\
0_k & 0
\end{bmatrix} \in \mathbb{R}^{K \times K}
$$

(2.60)

such that we obtain the $k$th row-reduced lower signal matrix $\mathbf{H}_{a,k}$ defined as

$$
\mathbf{H}_{a,k} = \mathbf{F} \circ \mathbf{A}_k = \mathbf{F} \circ (\mathbf{J}_{K,k} \mathbf{A}) = (\mathbf{I}_L \otimes \mathbf{J}_{K,k}) \mathbf{H}
$$

(2.61)

where the $k$th row-reduced lower Vandermonde matrix

$$
\mathbf{A}_k = \mathbf{J}_{K,k} \mathbf{A}
$$

(2.62)

is formed from the last $K - k$ rows of the original Vandermonde matrix $\mathbf{A}$. It is worth mentioning that here the lower $K \times K$ selection matrix $\mathbf{J}_{K,k}$ extracts the $(k + 1)$th to last row of $\mathbf{A}$ and restores them in the first to $(K - k)$th rows of the lower rows-reduced signal matrix $\mathbf{A}_k$. The remaining rows are filled with zero elements and appended at the bottom of the matrix, so that
the size of $A_k$ corresponds to the size of the original signal matrix. From representations (2.59) and (2.62) the MI property of Vandermonde matrices is easily identified as
\[
\overline{A}_k \Delta_a^k = J_{K,k} A \Delta_a^k = J_{K,k} A = A_k,
\]
for $k = 1, \ldots, K - 1$, where the diagonal matrix
\[
\Delta_a = \text{diag}\{a_1, a_2, \ldots, a_P\}
\]
contains the $P$ harmonics observed along the first array axis on its main diagonal. Note that in (2.63) shift invariance is represented through right-multiplication with a diagonal matrix that contains the $k$th row of the original Vandermonde matrix $A$ on its main diagonal. This well-known property of Vandermonde structures marks one of the earliest findings from which the original ESPRIT has been developed [RK89] and has further been exploited for example in [SORK92, HN98, ZHM96, SORK92]. Next, we consider the row-reduced Khatri-Rao products of nuisance matrix $F$ and Vandermonde matrices $\overline{A}_k$ and $A_k$ in (2.58) and (2.61), respectively. Clearly, the MI property (2.63) is directly handed to the corresponding row-reduced signal matrices, that is
\[
\overline{H}_{a,k} \Delta_a^k = \overline{H}_{a,k},
\]
for $k = 1, \ldots, K - 1$.

With identity (2.27), property (2.65) can also be represented in terms of row-reduced versions of the signal subspace eigenvectors. Defining the $k$th row-reduced upper signal eigenvector matrix as
\[
\overline{E}_{S,a,k} = (I_L \otimes J_{K,k}) E_S
\]
and analogously the $k$th row-reduced lower signal matrices read
\[
\overline{H}_{a,k} = \overline{E}_{S,a,k} K = (I_L \otimes J_{K,k}) E_S K.
\]
Inserting (2.68) and (2.69) into (2.65) we obtain the identities
\[
\overline{E}_{S,a,k} K \Delta_a^k = \overline{E}_{S,a,k} K
\]
for $k = 1, \ldots, K - 1$. Equation (2.70) forms a set of related eigenproblems. For $k = 1$ identity (2.70) yields the classic ESPRIT algorithm in which the solutions are obtained from solving the single eigenproblem [RK89]. In the literature, different LS or total least squares (TLS)
approaches for solving (2.70) are known [OVK92]. If \( k > 1 \) the set of equations in (2.70) establishes a joint or simultaneous eigenproblem. Note here that the various equations evaluated for distinct values of \( k \) all contain the same matrices \( K \) while the associated eigenvalues on the main diagonal of \( \Delta^k \) differ from one another according to the values in the exponent. It is simple to verify that the columns of the mixing matrix \( K \), defined in (2.27) are up to a complex scaling the eigenvectors of the simultaneous-eigenproblem in (2.70). For \( E^\dagger_{S,a,k} \) representing a generalized inverse of \( E_{S,a,k} \), (2.70) translates in

\[
E^\dagger_{S,a,k} E_{S,a,k} = K \Delta^k K^{-1},
\]

for \( k = 1, \ldots, K - 1 \). Various methods have been proposed in recent literature that provide solutions to the eigenproblems in (2.71) under a framework of joint diagonalization or simultaneous Schur decomposition (see [vdVVA98, VvdVP98] and also [HN98] in a slightly different context). These algorithms are based on iterative optimization schemes that in each step search for an update of the current estimate of a transformation matrix which further reduces the value of the cost function.

A simultaneous Schur decomposition procedure developed in [HN98] relies on a different set of invariance equations obtained from (single) invariances determined along various dimensions. The underlying estimation problem, however, is very similar and the same principles are also applicable here. A real-valued version of equation (2.71) is obtained from unitary transformations of the data covariance matrix. Successive Jacobi transformations are performed to ensure minimization of the cost function on the manifold of unitary matrices (the Grassman manifold). The algorithm consists of a joint Schur approximation. That is, a set of upper triangular Schur matrices and a unitary transformation matrix are computed that approximately solve a real-valued version of the Eigenvalue problem in (2.71). The cost function is a LS-measure of how “upper-triangular” the set of resulting Schur matrices is made. In other words, the strictly lower-triangular part of the Schur matrices are jointly minimized in a LS sense in each iteration step.

In [vdVVA98] a different approach towards solving (2.71) is taken. Here a joint diagonalization algorithm is proposed that uses a Newton iterations scheme. A major drawback in joint diagonalization or simultaneous Schur decomposition approaches lies in their slow convergence rate and their sensitivity to the chosen initial estimates. This may lead to prohibitively high computational complexity associated with the minimization procedure. Sufficiently accurate starting points are usually difficult to obtain and therefore global convergence of these algorithms is not guaranteed. An approach which is free from numerical difficulties exists only in the case of \( k = 2 \). In [ZHM96] an ESPRIT algorithm is presented that is based on the same unitary transformation given in [HN98] and that only requires simple eigendecomposition.

In section 3.4 we will return to the problem of jointly solving the eigenproblem in (2.70). We shall develop a novel algorithm that exploits the specific relation between the eigenvalues in \( \Delta^k \).
for different values of $k$ rather than the fact that all matrices $E_{S,a,k}^f E_{S,a,k}^s$ are posing identical eigenvectors. This relation is completely ignored in existing approaches.

\section*{2.4 MUSIC algorithms}

The MUSIC algorithm was first developed in the context of sensor array processing for DOA estimation [Sch79, BK80, Sch81, BK83]. This section reviews the spectral MUSIC algorithm in the general MD case and the root-MUSIC algorithm that is applicable in the uniform sampling case [Bar83, RH89].

Consider the general HR problem formulated in model (2.5). The signal matrix reads $H = F \circ A$ (2.9). The MD spectral MUSIC algorithm estimates the parameters of interest corresponding to the $P$ harmonics from the deepest minima of the inverse MUSIC spectrum given by

$$f_M(\theta, \vartheta, \mu, \alpha) = h^H(\theta, \vartheta, \mu, \alpha) E_N E_N^H h(\theta, \vartheta, \mu, \alpha)$$

where the signal vector

$$h^H(\theta, \vartheta, \mu, \alpha) = f(\theta, \vartheta, \mu, \alpha) \otimes a$$

for $f(\theta, \vartheta, \mu, \alpha)$ defined according to (2.4) is varied over the MD parameter space, i.e. $-\infty < \mu < 0, 0 \leq \alpha < 2\pi, \theta \in \mathcal{P}$, and $\vartheta \in \mathcal{Q}$.

For uniform sampling, MD root-MUSIC is applicable [WCF01, DMD93, SSJ01, TH92, YLC89, vdVOD92]. Choosing the sampling scheme according to (1.7) in the 3D uniform HR case the inverse 3D MUSIC spectrum along the $a$-, $b$-, and $c$-axis is given by

$$f_M(a, b, c) = h^H(a, b, c) E_N E_N^H h(a, b, c)$$

The deepest nulls of the function in (2.74) yield the true parameters of interest. In the pure HR the generators along all sampling axes are located on the unit circle. Hence we can exploit the conjugate-reciprocity properties which applies in this case. That is, with $a^* = a^{-1}, b^* = b^{-1}$, and $c^* = c^{-1}$, we arrive at the 3D root-MUSIC function

$$f_{r-M}(a, b, c) = h^T(a^{-1}, b^{-1}, c^{-1}) E_N E_N^H h(a, b, c) .$$
28

2 Subspace based 2D harmonic retrieval algorithms

It is important to note that \( f_{-M}(a, b, c) \) represents a three-variate polynomial of degree \( 2K - 1 \), \( 2L' - 1 \), and \( 2M - 1 \) in the parameters \( a, b, \) and \( c \), respectively. From the subspace relation in (2.26) we know that in the ideal case (for exactly known eigenvectors \( E_N \)) the inverse MUSIC spectrum yields zero function values for the true parameters, hence for the 3D root triplets \( (a, b, c) \) equal to one of the true generator triplets \( (a_1, b_1, c_1), \ldots, (a_P, b_P, c_P) \). In other words, in this case, the true generators are obtained from those root triplets \( (a, b, c) \) of the three-variate polynomial in (2.75) that satisfy the unit-norm constraint \( |a| = |b| = |c| = 1 \). The difficulty arising in this context is that, unless in the 1D case, no reliable method for rooting multivariate polynomials is available in literature. Existing methods require good initial estimates, do not guarantee convergence, and suffer from large computational complexity [WCF01, Tre02]. Several interesting estimation procedures have recently been proposed based on the spectral and root-MUSIC algorithm that were especially designed to reduce its large computational cost in the MD case [WCF01, DMD93, HF96, Tre02].

2.5 Estimation of the linear parameters

In this section we briefly address the problem of estimating the linear parameters in the data model (1.8) provided that the nonlinear parameters are previously estimated using one of the estimators proposed in the next chapter. Consider the conditional signal model according to 2.1 (and in the single snapshot case also 2.2). For a given signal matrix \( H \) the standard least squares (LS) estimator that minimizes the norm of the estimation error

\[
n(t) = x(t) - H \hat{w}(t)
\]

is given by

\[
\hat{w}_{LS}(t) = H^\dagger x(t),
\]

where \((\cdot)^\dagger\) denotes Moore-Penrose-Pseudo inverse

\[
H^\dagger = (H^H H)^{-1} H^H,
\]

provided the inverse in (2.78) exists. Further, in this case the LS solution coincides with the Maximum-Likelihood (ML) estimator for the given estimation problem [Böh91]. Exactly these desirable properties of the LS estimator in (2.77) for known signal matrix motivate its use also in the case that only finite sample estimates of the nonlinear signal parameters are available. Hence given an estimate \( \hat{H} \) of the signal matrix we substitute the true signal matrix in (2.77) by its finite sample estimate to obtain

\[
\hat{w}_{LS,\hat{H}}(t) = \hat{H}^\dagger x(t).
\]
3 Rank reduction estimators

The principle of the Rank Reduction Estimator (RARE) was first introduced in [PGWB01, PGW02b, PGW02a]. Even before, in [WZ99, ZW00, SSJ00] rank reduction methods for DOA estimation have been considered in a slightly different context. The rank reduction technique is indeed a very powerful concept and applies to a large variety of problems in MD parameter estimation, calibration and system identification. Interestingly, the RARE concept can be treated from very different perspectives. In appropriating the various viewpoints on the RARE algorithm we not only gain understanding of important properties of the estimator like uniqueness and computational complexity, but we also learn about the affiliations between MUSIC and ESPRIT-based algorithms.

In section 3.1 we start our considerations from the MD MUSIC algorithm. A convenient parameterization of the signal parameter vector is introduced that allows the separation of the generators observed along the various dimensions and leads to a rank reduction estimation criterion. The RARE algorithm is derived and sufficient conditions for unique estimation of the true harmonics along the $\alpha$-axis are proven. In section 3.2 this criterion is interpreted as the optimization of the MUSIC function over a relaxed manifold. Section 3.3 yields first uniqueness results for RARE. In section 3.4 we approach the rank reduction concept from a completely different perspective. From the MI equations in 2.3, a rooting-based rank reduction method referred to as the root MI-ESPRIT algorithm is introduced and its uniqueness conditions are investigated. The relation between the different criteria is then discussed in section 3.5.

3.1 Conventional approach

In this section we follow the first approach towards the RARE algorithm that was taken in [PGWB01, PGW02a, WZ99, ZW00, SSJ00]. We start our considerations from the MUSIC estimator as introduced in chapter 2.4 for the pure HRP case. Recall that the conventional MUSIC algorithm estimates the signal parameters from the deepest minima of the inverse MUSIC spectrum (2.72) for the general model in (2.5). In the ideal case of exactly known covariance matrix $R$, the pure harmonics of interest $\{a_1, \ldots, a_P\}$ can be found from the 3D inverse MUSIC spectrum (2.72) [PGW02a]

$$f_M(\theta, \vartheta, \alpha) = (f(\theta, \vartheta, \alpha) \otimes a)^H E_N^H E_N^H (f(\theta, \vartheta, \alpha) \otimes a) = 0. \quad (3.1)$$

Since the parameter vectors $\theta_1, \ldots, \theta_P$ and $\vartheta$ are considered as unknown (nuisance) parameters, the minimization of (3.1) requires an exhaustive MD search that becomes totally impractical if
the dimensions of vectors \( \theta_1, \ldots, \theta_P \) and \( \vartheta \) are large. Making use of identity (A.6) we represent
the signal vector \( h \) as
\[
h(a, \theta, \vartheta) = f(\theta, \vartheta, \alpha) \otimes a = (I_L \otimes a) f(\theta, \vartheta, \alpha) = T_a(a) f(\theta, \vartheta, \alpha),
\]
where according to (2.3)
\[
a = [1, a, a^2, \ldots, a^{K-1}]^T, \quad a \in \mathbb{C},
\]
and
\[
T_a(a) = (I_L \otimes a)
\]
defines a sparse \( KL \times L \) matrix polynomial (MP) of degree \( K-1 \) in the generator \( a \). Inserting
(3.2) in (3.1) yields
\[
f_M(\theta, \vartheta, \alpha) =
\]
\[
= (f(\theta, \vartheta, \alpha) \otimes a)^H E_N E_N^H (f(\theta, \vartheta, \alpha) \otimes a)
\]
\[
= f^H(\theta, \vartheta, \alpha) T_a^H(a) E_N E_N^H T_a(a) f(\theta, \vartheta, \alpha)
\]
\[
= f^H(\theta, \vartheta, \alpha) T_a^H(a) (I_P - E_S E_S^H) T_a(a) f(\theta, \vartheta, \alpha)
\]
\[
= f^H(\theta, \vartheta, \alpha) M_1(a, E_S | \text{“p”}) f(\theta, \vartheta, \alpha) = 0.
\]
where
\[
M_1(a, E_S | \text{“p”}) = T_a^T(a^{-1}) (I_P - E_S E_S^H) T_a(a)
\]
is the \( L \times L \) Hermitian MP of degree \( 2K-1 \) that is in the following referred to as the RARE MP
of first kind in the generator \( a \). The argument \( E_S \) indicates that in (3.6) the signal subspace is
expressed in terms of signal eigenvectors rather than the signal vectors contained in \( H \). Further
the letter “p” in the argument of the MP specifies that the pure (or undamped) HR case is considered.

Note that (3.6) exploits the conjugate-reciprocity of \( T_a(a) \), that is \( T_a^H(a) = T_a^T(a^{-1}) \) for \( a \) on
the unit circle. A very important observation here is that the parameter vectors \( \theta \) and \( \vartheta \) are
contained in \( f(\theta, \vartheta, \alpha) \) only. Therefore, the polynomial \( M_1(a, E_S | \text{“p”}) \) is not dependent on
the nuisance parameters in \( \theta \) and \( \vartheta \). The following assumption is necessary for unique recovery
of the model parameters.

**Assumption A2:** The number of signals does not exceed the overall number of samples minus
the number of samples taken along the second array axis,
\[
P \leq (K - 1)L.
\]
Note that, if A2 is satisfied, then \( E_N \) and also \( M_1(a, E_S | \text{“p”}) \) are in general full rank. This is
because, according to (3.7), the column rank of \( E_N \) is not less than \( L \). It is clear that equation
(3.5) holds only if the MP \( M_1(a, E_S | \text{“p”}) \) drops rank so that \( \text{rank}\{M_1(a, E_S | \text{“p”})\} < L \)
3.1 Conventional approach

with vector $f(\theta, \vartheta, \alpha)$ located in the nullspace $N\{M_1(a, E_S \mid \text{“p”})\}$. Therefore, the key idea 
of the RARE algorithm is to find the generators of interest for which the MP $M_1(a, E_S \mid \text{“p”})$ 
drops rank, that is

$$\text{rank}\{M_1(a, E_S \mid \text{“p”})\} < L \quad (3.8)$$

or, equivalently, to find the roots of the scalar polynomial

$$P(a) = \det \{M_1(a, E_S \mid \text{“p”})\} = 0 \quad (3.9)$$

From the considerations above it is clear that (3.9) is a necessary condition for the true 
generators along the $a$-axis. However, there are two principal questions concerning the uniqueness of 
the solution.

- Which conditions regarding the number of harmonics, the sample support, the number 
of time samples and the distributions of the generators should be satisfied in order to 
guarantee that the generators can be uniquely identified from (3.1)?

- In the latter case, is the MUSIC solution for the harmonics of interest identical to the 
RARE solution? In other words, can the matrix $M_1(a, E_S \mid \text{“p”})$ become singular 
for some values $a$ that lie on the unit circle but do not nullify the MUSIC polynomial 
$f_M(\theta, \vartheta, \alpha)$, and vice versa, can $f_M(\theta, \vartheta, \alpha)$ become zero for some values $a$ that lie on 
the unit circle but do not nullify the RARE matrix $M_1(a, E_S \mid \text{“p”})$?

The following sections of this chapter provide detailed answers to these important questions, but 
before addressing them we shall derive an alternative formulation of the RARE matrix criterion 
in (3.9) which turns out to be particularly useful if the number of signals $P$ is less than the 
sample support $L$ along the second array axis. Making use of the block determinant lemma in 
(A.7), the RARE polynomial equation given in (3.9) can be rewritten as

$$\det\{M_1(a, E_S \mid \text{“p”})\} =$$

$$= \det\{T_a^T(a^{-1})E_NT_a(a)\}$$

$$= \det\{T_a^T(a^{-1}) (I_{KL} - E_SE_S^H) T_a(a)\}$$

$$= \det\{T_a^T(a^{-1})T_a(a) - T_a^T(a^{-1})E_SE_S^H T_a(a)\}$$

$$= \det\left\{\begin{bmatrix} T_a^T(a^{-1})T_a(a) & T_a^T(a^{-1})E_S \\ E_S^HT_a(a) & I_P \end{bmatrix}\right\}$$

$$= \det\{T_a^T(a^{-1})T_a(a)\} \det\{I_P - E_S^HT_a(a) (T_a^T(a^{-1})T_a(a))^{-1} T_a^T(a^{-1})E_S\}$$

$$= \Omega \det\{I_P - E_S^HT_a(a)\Omega^{-1}T_a^T(a^{-1})E_S\}$$

$$= \Omega \det\{M_2(a, E_S \mid \text{“p”})\} \quad (3.10)$$

where

$$\Omega = T_a^T(a^{-1})T_a(a) = K I_{L \times L} \quad (3.11)$$
is a constant $L \times L$ diagonal matrix that is independent of the generator $a$ and

$$\Omega = \det\{\Omega\}.$$  \hfill (3.12)

The $P \times P$ matrix

$$M_2(a, E_S \mid \text{“p”}) = I_P - E_S^H T_a(a) \Omega^{-1} T_a^T (a^{-1}) E_S$$  \hfill (3.13)

is in the following referred to as the RARE MP of second kind in the generator $a$. Equation (3.13) reveals that the same statements valid for the rank of the RARE MP of first kind, are also valid for the new RARE MP of second kind. Specifically, $M_2(a, E_S \mid \text{“p”})$ is generally (for arbitrary $a$) full rank and becomes singular if $a$ corresponds to one of the true harmonics along the first axis. The RARE MPs of first and second kind therefore show exactly the same singularities, however both matrices differ in their dimension. While $M_1(a, E_S \mid \text{“p”})$ is of dimension $L \times L$, the matrix $M_2(a, E_S \mid \text{“p”})$ is of dimension $P \times P$. This makes one formulation favorable over the other when it comes to evaluating determinants or singularities of the MP as the computational cost associated with this operations grows with the size of the matrix. For details on how to efficiently evaluate the determinants and singularities of MPs, refer to chapter 5.

### 3.2 Relaxed optimization approach

This section delivers new insight in the RARE algorithm and its relation to the conventional MUSIC criterion by taking a closer look at the signal manifolds that are associated with the criteria in (3.9) and (3.1). Towards this aim it appears to be particularly useful to treat both algorithms under a formal optimization-theoretic framework.

In conventional MUSIC the parameters of interest are obtained from minimizing the inverse MUSIC function (2.72) on the parameter spaces, i.e. for $a \in \mathbb{C}, |a| = 1$ and $f(\theta, \vartheta, \alpha) \in \mathbb{C}^L$. The MUSIC algorithm searches for manifold vectors $h(a, \theta, \vartheta)$ that have the smallest distance (in a LS sense) to the signal subspace spanned by the columns of $E_S$. The manifold vector $h(a, \theta, \vartheta)$ describes a surface in the $KL$-dimensional complex space $\mathbb{C}^{KL}$ that in the following we refer to as the original manifold $M_{\text{org}}$. According to the definition of the manifold vector in (3.2) and the actual specification of the nuisance vector $f(\theta, \vartheta, \alpha)$ that depends on which of the cases in section 1.2 are in effect for a specific application, the manifold takes a very characteristic structure. Indeed, as described in section 1.2, the nuisance vectors can either be highly structured as in 1.2.1, moderately structured as in 1.2.2 or unstructured as in case 1.2.3. The structure of the $f(\theta, \vartheta, \alpha)$ and the Vandermonde nature of signal vector $a$ defined in (3.3) are handed to the manifold vector through relation (3.2). In case that the manifold $M_{\text{org}}$ is unambiguous and $AI$ is satisfied, a unique set of exactly $P$ signal vectors
exists that are all located perpendicular to the noise subspace $\mathcal{N}$. However, depending on the number of parameters that describe the manifold $\mathcal{M}_{\text{org}}$, the exact minimization of the inverse MUSIC function (2.72) of the complete manifold becomes a difficult task due to the multi-modal nature of the MD cost function on the original manifold. The RARE criterion can be interpreted as replacing the minimization of the inverse MUSIC function on the original manifold $\mathcal{M}_{\text{org}}$ by the minimization of the inverse MUSIC function over a larger manifold, the RARE manifold $\mathcal{M}_{\text{RARE}}$. This procedure stems back from a technique in optimization theory that is widely known as relaxation of the manifold. Instead of searching for solutions to the optimization criterion on the original manifold, the idea of this optimization technique is to find an appropriately extended manifold, i.e. a larger manifold that fully contains the original one, such that the optimization problem formulated on the new manifold becomes feasible and easier to handle. The solutions on the original manifold are then traced back from the solutions previously obtained on the relaxed manifold. Clearly, since the original manifold is fully contained in the new manifold and the latter is usually larger, not all solutions existing on the relaxed manifold must necessarily correspond to solutions on the original one. However, if for a given relaxed manifold a simple relation between the new and the old manifold exists, and if the number of solutions is finite, then a simple criterion can be found on how to distinguish between the true solutions (i.e. the solutions on the original manifold) and the spurious solutions (i.e. the additional solutions that only exist on the new manifold), and relaxation can truly simplify a complex optimization problem.

In the context of minimizing the inverse MUSIC function (2.72), relaxation consists of replacing the original manifold $\mathcal{M}_{\text{org}}$ defined as

$$\mathcal{M}_{\text{org}} := \{ h(a, \theta, \vartheta) \mid a \in \mathbb{C}, |a| = 1, \theta \in \mathcal{P}, \vartheta \in \mathcal{Q} \}$$

with $h(a, \theta, \vartheta)$ given in (3.2) by a “less-structured” RARE manifold defined as

$$\mathcal{M}_{\text{RARE}} := \{ g(a, k) = (k \otimes a) \mid a \in \mathbb{C}, |a| = 1, k \in \mathbb{C}^L \setminus \{0\} \}.$$  (3.15)

A comparison of the manifolds before and after relaxation reveals that in both cases the manifold vectors can be represented as the Kronecker-product of a $L \times 1$ vector ($f(\theta, \vartheta, \alpha)$ and $k$, respectively) and the Vandermonde vector $a$ (3.3). Hence, both $h(a, \theta, \vartheta)$ and $g(a, k)$ have some degree of structure. However, it is simple to see that depending on the specification of the HRP and the definition of the vector function $f(\theta, \vartheta, \alpha)$ (see section 1.2), the original manifold vector $h(a, \theta, \vartheta)$ is generally restricted to a much specific structure than the counterpart $g(a, k)$. For example, in section 1.2.1 the entries of $f(\theta, \vartheta, \alpha)$ are all restricted to a block Vandermonde structure, while in the RARE manifold the entries of the corresponding vector $k$ can take arbitrary complex values. The considerations above imply, that independently of the specification made on the vector function $f(\theta, \vartheta, \alpha)$, the original manifold $\mathcal{M}_{\text{org}}$ is always fully contained in the relaxed manifold $\mathcal{M}_{\text{RARE}}$, since any non-zero $L \times 1$ vector $f(\theta, \vartheta, \alpha)$ with $\theta \in \mathcal{P}$ and
3 Rank reduction estimators

$\vartheta \in Q$ can be represented by a vector $g \in \mathbb{C}^L \setminus \{0\}$. In other words, the original manifold defines a subset of the RARE manifold. For the sake of completeness, we note that solely in the case described in section 1.2.3 both manifolds cover the same surface of the $KL$-dimensional complex space as both $f(\vartheta, \vartheta, \alpha)$ and $k$ describe arbitrary non-zero complex vectors in $\mathbb{C}^L$.

Minimizing the inverse MUSIC function (2.72) on the new manifold $M_{\text{RARE}}$ instead of $M_{\text{org}}$ results in searching for the nulls of the quadratic form

$$g^H(a, k)E_N E_N^H g(a, k) = (k \otimes a)^H E_N E_N^H (k \otimes a) = 0,$$

for $a \in \mathbb{C}$, $|a| = 1$ and $k \in \mathbb{C}^K \setminus \{0\}$. As illustrated in equation (3.2), the idea of the RARE algorithm consists of introducing a convenient parameterization that allows the separation of the parameters of interest (the generators that are observed along the first array axis) from the remaining parameters. The fact that the RARE manifold vectors and the original manifold vector both consist of a Kronecker-product of a $L \times 1$ vector and a $K \times 1$ Vandermonde vector allows us to represent the new manifold vector in terms of a vector product between a “tall” MP $T_a(a)$ and an $L \times 1$ nuisance vector similar to (3.2):

$$g(a, k) = (k \otimes a) = T_a(a)k.$$

(3.17)

Inserting (3.17) in (3.16) yields the polynomial equation

$$k^H T^T(a^{-1}) E_N E_N^H T(a) k = 0$$

(3.18)

in the parameters $a \in \mathbb{C}$, $|a| = 1$ and with unknown vector $k \in \mathbb{C}^K \setminus \{0\}$. Clearly the roots of (3.18) located on the unit circle are equivalent to the solutions of the RARE polynomial equation (3.9)

$$P(a) = \det \{M_1(a, E_S | \text{“p”})\} = \Omega \det \{M_2(a, E_S | \text{“p”})\} = 0.$$

(3.19)

with $M_1(a, E_S | \text{“p”})$ and $M_2(a, E_S | \text{“p”})$ defined according to (3.6) and (3.10), respectively. Note that (3.19) only depends on the parameter $a$ and not on the complex nonzero vector $k$, and is therefore much easier to solve.

A natural question arising in this context is whether the extension of the manifold on which the MUSIC criterion is minimized affects the number of solutions. We need to find conditions under which no additional solutions, so-called spurious solutions, emerge, that only exist on the RARE manifold but not on the original manifold. Closely related to this problem is the question on whether the relaxation of the manifold effects the uniqueness of the estimation. That is, provided that the solutions of the MUSIC criterion on the original manifold are unique, under which conditions are the roots of the RARE criterion unique solutions to the HRP in (1.8)?

A first attempt to answer this question was undertaken in [PGW02a]. In this contribution, conditions under which the RARE manifold is free of first order ambiguities are derived rather
than uniqueness conditions of the RARE algorithm to avoid higher order ambiguities as claimed in the proof. First order ambiguities emerge when the parameterization of the manifold is not unique, i.e. when the same manifold vector corresponds to two different parameter sets. Higher order or $k$th order ambiguities exist when a manifold vector can be represented as a linear combination of $k$ distinct manifold vectors [MSD01, ASG99]. In order to make a general statement on the uniqueness of the RARE algorithm in a multiple harmonic scenario, further investigations are required.

3.3 Gaussian-elimination approach and uniqueness

To answer the open questions concerning the uniqueness of the solutions on the relaxed manifold, it is not necessary to fully determine whether the RARE manifold $\mathcal{M}_{\text{RARE}}$ is higher order unambiguous. This is because of the following two reasons. First, the data was “generated” by the true manifold and not by the relaxed RARE manifold. For identifiability of the model parameters we need to assume that no first or higher order ambiguities exist, otherwise neither MUSIC nor RARE can guarantee unique parameter estimates. Second, both the MUSIC criterion and the “relaxed” MUSIC criterion are one-dimensional. In other words, we search for single manifold vectors located in the signal subspace $\mathcal{M}$ rather than for a linear combination of manifold vectors. Hence only first order ambiguities of the RARE-manifold are of importance and not higher order ambiguities. It is sufficient to show that for a given model order $P$ and for full rank signal matrix $H$ no linear combination of columns of $H$ exists that can be represented by a manifold vector $g(a, k) \in \mathcal{M}_{\text{RARE}}$ with $a$ not contained in the set of true generators $\mathcal{H}_a = \{a_1, \ldots, a_P\}$. In (3.7) we already found a necessary condition for the uniqueness of the RARE estimator. This section shows that under comparably “mild” conditions on the signal matrix $H$ the inequality $P \leq (K - 1)L$ is also sufficient.

Assume that there exists such a (non-trivial) linear combination of columns of the signal matrix. Then there exists a non-zero vector of linear coefficients $l = [l_1, \ldots, l_P]^T \in \mathbb{C}^P$ such that

$$\sum_{p=1}^{P} l_p h(a_p, \theta_p, \vartheta) = g(a, k)$$

(3.20)

where $h(a_p, \theta_p, \vartheta)$ is the $p$th column of the signal matrix $H$ resulting in

$$Hl - T_a(a)k = 0.$$  

(3.21)

It is simple to verify that (3.21) is satisfied if and only if the augmented matrix

$$M_3(a, H \mid \text{“d”}) = [T_a(a) \mid H] = [(I_L \otimes a) \mid (B \circ A)] \in \mathbb{C}^{KL \times (P+L)}$$

(3.22)
becomes rank deficient. Here the argument $\mathbf{H}$ indicates that the signal subspace is expressed in terms of signal vectors $\mathbf{H}$ rather than the signal eigenvectors $\mathbf{E}_S$. The letter “d” in the argument of the MP specifies that the “damped” HR case is considered.

From the considerations above it is clear that, provided the MP in (3.22) is full rank for all $a$ not contained in the set of true generators $\mathcal{H}_a$, no other relaxed manifold vectors $g(a, k)$ than the ones corresponding to the true harmonics $a_1, \ldots, a_P$ solve the “relaxed” MUSIC criterion. In this case we can state that, with respect to the harmonic along the $a$-axis, the solutions to the MUSIC criterion obtained on the original manifold and the solutions obtained on the relaxed manifold are identical. In order to make general statements on the rank of the MP in (3.22) we need to make the following assumption:

**Assumption A3**: The upper row-reduced signal matrix matrix $\mathbf{H}_{a,1}$ (or equivalently the lower row-reduced signal matrix $\mathbf{H}_{a,1}$) has full column rank.

Note that similar to the discussion after A1 in chapter 2 it is simple to show from theorem T1 that A3 is satisfied with probability one in all practically relevant cases. Equipped with A3 we formulate the following theorem. Further, it is simple to check that A3 implies that assumption A2, i.e. $P \leq L(K - 1)$, is satisfied.

**Theorem T2**: Provided that A3 is satisfied and if all generators are located inside or on the unit-circle ($|a_p| \leq 1$, for $p = 1, \ldots, P$) then

$$\text{rank}\{\mathbf{M}_3(a, \mathbf{H} | \text{“d”})\} = \begin{cases} P + L - \text{mult}\{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P + L, & \text{otherwise}. \end{cases}$$

(3.23)

Here $\text{mult}\{a|\mathcal{H}_a\}$ denotes the multiplicity of the root $a$ in the true generator set $\mathcal{H}_a = \{a_1, \ldots, a_P\}$. This statement holds true for damped and undamped exponential mixtures.

**Proof of T2**: See Appendix B.

As the $\mathbf{H}$ and $\mathbf{E}_S$ span the same signal subspace (2.27) theorem T2 also holds true if the signal matrices $\mathbf{H}$ in (B.3) is replaced by its pendant, the signal subspace matrix $\mathbf{E}_S$ (2.22).

The augmented matrix

$$\mathbf{M}_3(a, \mathbf{E}_S | \text{“d”}) = [\mathbf{T}_a(a)|\mathbf{E}_S] = [\mathbf{I}_L \otimes a]|\mathbf{E}_S] \in \mathbb{C}^{KL \times (P+L)}$$

(3.24)

possesses the same rank properties formulated in theorem T2 concerning the parameter $a$ as the augmented data matrix $\mathbf{M}_3(a, \mathbf{H} | \text{“d”})$, hence:

**Corollary C1**: Provided that A3 is satisfied then

$$\text{rank}\{\mathbf{M}_3(a, \mathbf{E}_S | \text{“d”})\} = \begin{cases} P + L, & \text{for } a \notin \{a_1, \ldots, a_P\}; \\ P + L - \text{mult}\{a|a_1, \ldots, a_P\}, & \text{otherwise}. \end{cases}$$

(3.25)
In practical applications the augmented matrices $M_3(a, H \mid \text{“d”})$ and $M_3(a, E_S \mid \text{“d”})$ are usually non-square and the number of harmonics $P < (K - 1)L$. The difficulty arising in this context is to obtain reliable rank and root estimates of non-square matrices if only perturbed versions of $M_3(a, H \mid \text{“d”})$ and $M_3(a, E_S \mid \text{“d”})$ are available. A natural approach for generators on the unit circle (“p”) that yields a square MP of degree $2K - 1$ from the MP $M_3(a, E_S \mid \text{“d”}) = M_3(a, E_S \mid \text{“d”})|_{a=1}$ of degree $K - 1$ is to simply take the quadratic form

$$M_4(a, E_S \mid \text{“p”}) = M_3^H(a, E_S \mid \text{“p”})M_3(a, E_S \mid \text{“p”})$$

$$= \begin{bmatrix} T_d^T(a^{-1}) & E_S^H \end{bmatrix} \begin{bmatrix} T_d(a) | E_S \\ E_S^H T(a) & I_p \end{bmatrix}$$

(3.26)

With corollary C1 we can now formulate the following corollary for the $L \times L$ MP $M_1(a, E_S \mid \text{“p”})$, the $P \times P$ MP $M_2(a, E_S \mid \text{“p”})$ and the $(L + P) \times (L + P)$ MP $M_4(a, E_S \mid \text{“p”})$.

**Corollary C2:** Provided that A3 is satisfied and all generators are located on the unit circle ($|a_p| = 1$ for $p = 1, \ldots, P$), the MPs $M_1(a, E_S \mid \text{“p”})$, $M_2(a, E_S \mid \text{“p”})$, and $M_4(a, E_S \mid \text{“p”})$ evaluated on the unit-circle ($|a| = 1$) are all non-singular if $a$ is not contained in the set of true generators $\mathcal{H}_a$ and all singular otherwise. Moreover, the order by which $M_1(a, E_S \mid \text{“p”})$, $M_2(a, E_S \mid \text{“p”})$, and $M_4(a, E_S \mid \text{“p”})$ drop rank for a true generator $a = a_p$, i.e. the dimension of the corresponding nullspaces, equals the multiplicity of the harmonic $a$ in the generator set $\mathcal{H}_a$.

**Proof of C2:** The corollary follows immediately from theorem T2 and the fact that on the unit circle (i.e. for $|a| = 1$) the matrix $M_4(a, E_S \mid \text{“p”}) = M_3^H(a, E_S \mid \text{“p”})M_3(a, E_S \mid \text{“p”})$ represents a quadratic form. Hence using (A.9).

$$\text{rank}\{M_4(a, E_S \mid \text{“p”})|_{a=1}\} =$$

$$= \text{rank}\{M_3^H(a, E_S \mid \text{“p”})|_{a=1}\} = \text{rank}\{M_3(a, E_S \mid \text{“p”})|_{a=1}\}$$

$$= P + L - \text{mult}\{a|a_1, \ldots, a_P\}.$$  (3.27)

Further note that with (3.10) we have

$$\det\{M_4(a, E_S \mid \text{“p”})\} = \det\{M_4(a, E_S \mid \text{“p”})\} = \Omega \det\{M_2(a, E_S \mid \text{“p”})\},$$

(3.28)

so that all three MP have identical singularities with identical multiplicity. ■
3.4 Multiple invariance approach

This section comprises several of the most important statements provided in this thesis. The methodology and revised viewpoint from which we contemplate the rank reduction concept is accompanied by a three-fold benefit. First of all, the new approach shall provide us with a MP formulation of significantly reduced degree. In fact the square MP derived in this section has only half the degree of the square polynomials that were previously considered. Second, this new approach shall be equally applicable to undamped and damped HR and yield unique solutions inside and on the unit-circle. Last but not least, this section discovers a close relationship between rooting-based HR algorithms [PGWB01, PGW02a, WZ99, ZW00, SSJ00] and ESPRIT-type methods that exploit (multiple) shift-invariance(s) [RK89, HN98, ZHM96, SORK92, vdVVP97, vdVVA98, VvdVP98, SLS01, FRB97]. Thus a link between these two popular approaches is provided.

Once again, our considerations start from the general model in (1.8). Section 2.3 has shown that the harmonics \( \{a_1, \ldots, a_P\} \) observed along the first array axis can be obtained from the eigenvalues of the joint eigenproblem in (2.70). There we already gave a brief overview on how solutions to the HR problem are obtained in literature based on joint diagonalization approaches. The main advantage of using joint diagonalization of the matrices on the left hand side of (2.71) is that automatically associated estimates along the various array axes can be obtained, an issue on which the following chapter focuses. On the other hand, a major drawback in this approach is that the computational cost related with the use of joined diagonalization algorithms [HN98, vdVVA98, VvdVP98] is considerably high, good starting points need to be available and global convergence is usually not guaranteed especially for closely separated eigenvalues. Further, the relatively poor performance that was for example reported in [PMB04] compared to rooting based HR algorithms like mD-RARE can be explained by the fact that joint diagonalization approaches ignore essential part of the information contained in the MI equations in (2.70). In joint diagonalization the idea is to search for a single eigenvector matrix \( K \) that approximately diagonalizes the matrices on the left side of (2.71) for all values of \( k = 1, \ldots, K - 1 \). In the ideal case the resulting diagonal matrices \( \Delta_k^a \) contain the corresponding eigenvalues on their main diagonals. However, in obtaining a eigenvector matrix \( K \) that is common to all MI equations, the specific relations between the diagonal eigenvector matrices for different values of \( k = 1, \ldots, K \) are ignored. That is, the diagonal eigenvalue matrices \( \Delta_k^a \) represent integer powers of a common basis diagonal matrix \( \Delta_a \) with the true generator on its main diagonal.

The new approach presented in this section overcomes this drawback. It is exactly this relation between the eigenvalue matrices for different values of \( k \) that shall be exploited here. Towards this aim let us write the characteristic equation corresponding to (2.70) as

\[
(\bar{E}_{S,a,k} - \bar{E}_{S,a,k} a_p^k) k_p = 0
\]  

(3.29)
for \( k = 1, \ldots, K \). Here \( k_p \) denotes the \( k \)th generalized eigenvector (GEV) of \( E_{S,a,k} \) and \( \overline{E}_{S,a,k} \), thus the \( k \)th column of \( K \) (2.70), and \( a^k_p \) denotes the corresponding generalized eigenvalue, where \( a_p \) is the true generator along the \( a \)-axis of the \( p \)th harmonic. According to the rank reduction method formulated in the preceding sections let us form a single MP equation from the set of equations in (3.29). By stacking the individual characteristic equations obtained for different values of \( k = 1, \ldots, K - 1 \) on top of each other we obtain a “tall” matrix equation

\[
M_5(a, E_S \mid \text{“d”}) k_p = \begin{bmatrix}
E_{S,a,1} - E_{S,a,1}a^1 \\
E_{S,a,2} - E_{S,a,2}a^2 \\
\vdots \\
E_{S,a,K-1} - E_{S,a,K-1}a^{(K-1)}
\end{bmatrix} k_p = 0. \tag{3.30}
\]

for the MP \( M_5(a, E_S \mid \text{“d”}) \) defined as

\[
M_5(a, E_S \mid \text{“d”}) = \begin{bmatrix}
E_{S,a,1} - E_{S,a,1}a^1 \\
E_{S,a,2} - E_{S,a,2}a^2 \\
\vdots \\
E_{S,a,K-1} - E_{S,a,K-1}a^{(K-1)}
\end{bmatrix} \quad (3.31)
\]

Obviously, in the nontrivial case \( (k \neq 0) \) the harmonic \( a \) that solves (3.30) must necessarily correspond to a matrix \( M_5(a, E_S \mid \text{“d”}) \) of reduced rank.

The MP \( M_5(a, E_S \mid \text{“d”}) \in \mathbb{C}^{K(K-1) \times P} \) of degree \( K - 1 \) possesses similar rank properties as defined in corollary \( C1 \) for the augmented signal matrix \( M_3(a, E_S \mid \text{“d”}) \): the MP in (3.31) drops rank for the true generators and is full rank otherwise. Further in section 3.5 we shall prove that there exists a close interrelation between both MPs. However, before we specify the rank properties of \( M_5(a, E_S \mid \text{“d”}) \), let us illustrate the difficulties in finding the values of \( a \) for which the MP becomes rank deficient.

Section 5.1 provides the means to find the singularities of a square matrix polynomial via determinant evaluation or alternatively via a direct generalized eigenvalue approach. However, these methods are not applicable here as the matrix \( M_5(a, E_S \mid \text{“d”}) \) is, similarly to the MP \( M_3(a, E_S \mid \text{“d”}) \), in general non-square. Hence, the exact evaluation of the roots of the MP becomes difficult if the coefficients of \( M_5(a, E_S \mid \text{“d”}) \) are perturbed due to noise or finite sample effects. Precise greatest right matrix divisor (GRD) estimation or greatest common matrix factor (GCF) extraction is required to determine the harmonics \( a_1, \ldots, a_P \) on the unit circle that cause a drop of the rank in \( M_5(a, E_S \mid \text{“d”}) \). Algorithms which accomplish this task are known from control theory [Kai80, GLR82]. However, existing GRD algorithms are numerically unstable and computationally complex, especially for closely separated harmonics and significant perturbations in the polynomial coefficients. An attempt to adopt these algorithms to the specifications of the HRP can be found in [PGB03]. This algorithm suffers from comparably large
computational complexity and numerical instability in the case of closely spaced generators and therefore will not receive further attention here.

In equation (3.26) we have circumvented the difficulty of determining the nulls of the “tall” MP $M_5(a, E_S | \text{“p”})$ for undamped harmonics. Recall that this was accomplished by transferring rank properties of the “tall” MP to its quadratic form $M_4(a, E_S | \text{“p”})$ and by evaluating the equivalent square RARE polynomials $M_1(a, E_S | \text{“p”})$ or $M_2(a, E_S | \text{“p”})$ on the unit circle instead. The drawback in using the quadratic forms lies in the doubling of the polynomial degree and the associated numerical difficulties in the rooting procedure. Here, a promising approach that avoids quadratic forms shall be promoted. The idea is to multiply the polynomial $M_5(a, E_S | \text{“d”})$ from the left with the MP $M_H^5(a, E_S | \text{“d”})$ evaluated at $a = 0$, i.e. $M_H^5(a, E_S | \text{“d”})|_{a=0}$. Thus, we obtain as square $P \times P$ MP of degree $K - 1$ that is defined as

$$
M_6(a, E_S | \text{“d”}) = M_5^H(a, E_S | \text{“d”})|_{a=0} M_5(a, E_S | \text{“d”})
$$

$$
= \sum_{k=1}^{K-1} (E_{S,a,k}^H E_{S,a,k} - E_{S,a,k}^H E_{S,a,k} a_k). 
$$

Formulation (3.32) is a convenient representation that allows simple interpretation of its underlying rank properties. In allusion to theorem $T2$ the following theorem can be established:

**Theorem T3**: Provided that $A3$ holds true and that all generators are located on or inside the unit-circle, the MPs $M_6(a, E_S | \text{“d”})$ and $M_5(a, E_S | \text{“d”})$ evaluated inside and on the unit-circle ($|a| \leq 1$) are non-singular if $a$ is not contained in the set of true generators $\mathcal{H}_a$, and singular otherwise. Moreover, the order by which $M_6(a, E_S | \text{“d”})$ and $M_5(a, E_S | \text{“d”})$ drops rank for a true generator $a = a_p$, i.e. the dimension of the corresponding nullspaces, equals the multiplicity of the harmonic $a$ in the generator set $\mathcal{H}_a$.

**Proof** of $T3$: To prove this rank properties multiply $M_6(a, E_S | \text{“d”})$ from the left and the right with the full rank matrices $K_H$ and $K$, respectively, where $K$ denotes the mixing matrix defined in (2.27). Clearly, this operation does not change the rank properties of $M_6(a, E_S | \text{“d”})$.

We obtain

$$
M_6(a, H | \text{“d”}) =
$$

$$
= K^H M_6(a, E_S | \text{“d”}) K = \sum_{k=1}^{K-1} (K^H E_{S,a,k}^H E_{S,a,k} K - K^H E_{S,a,k}^H E_{S,a,k} K a_k) =
$$
3.4 Multiple invariance approach

\[
K - 1 \sum_{k=1}^{K-1} \left( H_{a,k}^H H_{a,k} - H_{a,k}^H H_{a,k} \right) = K - 1 \sum_{k=1}^{K-1} H_{a,k}^H H_{a,k} \left( I - \Delta_{a}^{-k} a^k \right)
\]

\[= \left[ \sum_{k=1}^{K-1} \left( H_{a,k}^H H_{a,k} \sum_{m=0}^{k-1} \Delta_{a}^{-m} a^m \right) \right] \left( I - \Delta_{a}^{-1} a \right) \]

\[\approx W_{\text{res}}(a) \]

(3.33)

In other words, the MPs \( M_6(a, E_S \mid \text{“d”}) \) and \( M_6(a, H \mid \text{“d”}) \) are equivalent. In order to show that \( M_6(a, H \mid \text{“d”}) \) becomes singular only at the true generators, it is sufficient to show that the residual MP

\[ W_{\text{res}}(a) = \sum_{k=1}^{K-1} \left( H_{a,k}^H H_{a,k} \sum_{m=0}^{k-1} \Delta_{a}^{-m} a^m \right) \]

(3.34)

is non-singular for any \( a \) inside or on the unit circle. If \( W_{\text{res}}(a) \) is non-singular then it holds that

\[ g^H W_{\text{res}}(a) g \neq 0 \]

(3.35)

for all nonzero \( g \in \mathbb{C}^P \). In our proof we shall actually show that

\[ \text{Re} \{ g^H W_{\text{res}}(a) g \} > 0 \]

(3.36)

which implies (3.35) and hence \( W_{\text{res}}(a) \) is non-singular for any \( a \) inside or on the unit circle. To prove (3.36) we can equivalently show that the Hermitian part of \( W_{\text{res}}(a) \) denoted by

\[ W_{\text{res},h}(a) = \frac{1}{2} \left( W_{\text{res}}(a) + W_{\text{res}}^*(a^*) \right) \]

\[= \frac{1}{2} \sum_{k=1}^{K-1} \sum_{m=1}^{k-1} H_{a,k}^H H_{a,k} \Delta_{a}^{-m} a^m \]

\[+ \frac{1}{2} \sum_{k=1}^{K-1} \sum_{m=1}^{k-1} \Delta_{a}^{-m} a^m H_{a,k}^H H_{a,k} \]

(3.37)

is positive definite, since it satisfies

\[ \text{Re} \{ g^H W_{\text{res}}(a) g \} = g^H W_{\text{res},h}(a) g > 0 \]

(3.38)
In appendix D we show that

\[ 2W_{\text{res},h}(a) = \]

\[ = \sum_{k=1}^{K-2} \sum_{k=1}^{k-1} \sum_{l=0}^{k-2} \sum_{n=0}^{k-2} (\Delta_a^{s-1} a^*)^l \Delta_a^{K-1} F H \Delta_a^{K-1} (\Delta_a^{-1} a)^n \]

\[ + \sum_{l=0}^{K-2} \sum_{k=1}^{k-1} \sum_{l=0}^{k-2} \sum_{n=0}^{k-2} (\Delta_a^{s-1} a^*)^l \Delta_a^{K-1} F H \Delta_a^{K-1} (\Delta_a^{-1} a)^n \]

\[ + \sum_{k=1}^{K-1} \sum_{m=k}^{K-1} \sum_{l=0}^{m} \sum_{n=0}^{m} (1 - |a|^2) (\Delta_a^{s-1} a^*)^l \Delta_a^m F H \Delta_a^m (\Delta_a^{-1} a)^n \]

\[ + \sum_{k=1}^{K-1} \sum_{m=k}^{K-1} \Delta_a^m F H \Delta_a^m. \]

Since \( 1 - |a|^2 \geq 0 \) for \( |a| \leq 1 \), \( W_{\text{res},h}(a) \) is non-negative definite inside and on the unit circle.

Because of (D.1)

\[ \sum_{m=1}^{K-1} \Delta_a^m F H \Delta_a^m = H_{a,1}^H H_{a,1} \]

and with \( H_{a,1} \) assumed to be full column rank \( W_{\text{res},h}(a) \) is positive definite and (3.38) always holds true. This completes the proof.

A direct consequence of theorem T3 is that spurious or noise roots of \( M_6(a, E_S | \text{"d"}) \), the singularities of \( M_6(a, E_S | \text{"d"}) \) which do not correspond to true generators, are located strictly outside the unit circle. It is exactly this property which provides a simple mechanism for separating signal from spurious solutions as we shall see in the section 5.2, where the implementation of the rank reduction methods in the presence of noise is addressed.

### 3.5 Relations between the approaches

Additional links can be derived between the rank properties of the various RARE polynomial matrices \( M_1(a, E_S | \text{"p"}) \) (3.6), \( M_2(a, E_S | \text{"p"}) \) (3.13), \( M_3(a, E_S | \text{"d"}) \) (3.24), \( M_4(a, E_S | \text{"p"}) \) (3.26), \( M_5(a, E_S | \text{"d"}) \) (3.31), and \( M_6(a, E_S | \text{"d"}) \) (3.32).

Performing elementary matrix operations on the rows of \( M_3(a, E_S | \text{"d"}) \), or equivalently on the rows of \( M_3(a, H | \text{"d"}) \) in (3.22), eventually yields that the polynomials \( M_3(a, H | \text{"d"}) \) and \( M_5(a, H | \text{"d"}) \) are equivalent (for a proof see appendix C). It is proven that there exists a square unimodular MP \( U(a) \), i.e. a MP with constant non-zero determinant \( \det\{U(a)\} \neq 0 \), such that

\[ \begin{bmatrix} M_3(a, H | \text{"p"}) \\ 0 \end{bmatrix} = U(a) \begin{bmatrix} I_L & 0 \\ 0 & M_5(a, H | \text{"p"}) \end{bmatrix}. \]
Both MPs posses a common GRD given by \((I - \Delta_a^{-1})a\) (see proof of T3 for details). This property is already clear from corollary C1 and theorem T3, where we have proven that all roots for which \(M_3(a, E_S | "d")\) and \(M_3(a, E_S | "d")\) drop rank are equivalent to the true generators located inside the unit-circle while no additional spurious roots exist.

In contrast to \(M_0(a, E_S | "d")\) and \(M_0(a, E_S | "d")\) in the square MPs \(M_1(a, E_S | "p")\), \(M_2(a, E_S | "p")\), and \(M_4(a, E_S | "p")\) we assume all true generators to represent pure harmonics that are located on the unit circle as indicated by the letter “p”. Further the square MP has noise or spurious roots that do not correspond to the true generators. Corollary C1 proved that the spurious roots are not located on the unit circle. From (3.10) it is immediate that \(M_1(a, E_S | "p")\), \(M_2(a, E_S | "p")\), and \(M_4(a, E_S | "p")\) yield identical scalar polynomial equations (up to scaling by the constant \(\Omega\)). Hence the same statements can be made about signal and noise roots of \(M_1(a, E_S | "p")\) directly apply to the matrices \(M_1(a, E_S | "p")\) and \(M_2(a, E_S | "p")\). Therefore it is sufficient to investigate \(M_4(a, E_S | "p")\).

According to (A.13) the definition of \(M_4(a, E_S | "p")\) in (3.26) as the quadratic form of \(M_3(a, E_S | "p")\) implies that the coefficients of \(M_4(a, E_S | "p")\) are Hermitian-symmetric with respect to the center coefficient (i.e. the coefficient corresponding to \(d^0\)). The Hermitian-symmetry of polynomial coefficients yields the conjugate-reciprocity property of the roots in \(M_4(a, E_S | "p")\) [RH89] (see also the comments on conjugate-reciprocity property in section A). That is, if \(a^{-1}\) is a root of \(M_4(a, E_S | "p")\) then \(a^*\) is also a root of \(M_4(a, E_S | "p")\). Further it follows from Sylvester’s inequality (A.8) and equation (3.26) that each root of \(M_3(a, E_S | "p")\) yields a corresponding root in \(M_4(a, E_S | "p")\) and, according to the remarks above, also a conjugate-reciprocal counterpart for which \(M_4(a, E_S | "p")\) becomes singular. Specifically, each signal root of \(M_3(a, E_S | "p")\) represents a signal root of \(M_4(a, E_S | "p")\) of doubled multiplicity.

It remains to develop a relation between the square MP \(M_2(a, E_S | "p")\) (or equivalently the MPs \(M_1(a, E_S | "p")\) and \(M_4(a, E_S | "p")\)) and the square MP \(M_5(a, E_S | "p")\). From the definitions of \(M_2(a, E_S | "p")\) in (3.13) and \(M_6(a, E_S | "d")\) in (3.32) we know that
\[
K M_2(a, E_S | "p") = K E_S^H E_S - E_S^H T_a(a) T_a(a^{-1}) E_S
\]
\[
= K E_S^H E_S - \left( \sum_{k=1}^{K} E_{S,a,k}^H a^k \right) \left( \sum_{l=1}^{K} E_{S,a,l} a^{-l} \right)
\]  \hspace{1cm} (3.42)

where \(E_{S,a,k}\) is the \(L \times P\) matrix with the \(k\)th, \(k + K\)th, \(k + 2K\)th, \ldots, \((L - 1)K\)th row identical to the corresponding rows in \(E_S\) while the entries in the remaining rows are equal to zero, that is
\[
E_{S,a,k} = (I_L \otimes L_{K,k}) E_S
\]  \hspace{1cm} (3.43)

for \(L_{K,k}\) being the \(K \times K\) selection matrix with the \(k\)th diagonal matrix element equal to 1 and all remaining entries equal to zero.
According to the definition of $\mathbf{E}_{S,a,k}$ and $\mathbf{E}_{S,a,k}$ in (2.66) and (2.67) we have

\[
\mathbf{E}_{S,a,k} = \mathbf{E}_S - \sum_{l=1}^{K-k} \mathbf{E}_{S,a,K-l+1} = \sum_{l=1}^{K-k} \mathbf{E}_{S,a,l}
\]

(3.44)

\[
\mathbf{E}_{S,a,k} = \mathbf{E}_S - \sum_{l=1}^{K-k} \mathbf{E}_{S,a,l} = \sum_{l=k+1}^{K} \mathbf{E}_{S,a,l}
\]

(3.45)

It is simple to check that

\[
\left( \sum_{k=1}^{K} \mathbf{E}_{S,a,k}^H a^k \right) \left( \sum_{l=1}^{K} \mathbf{E}_{S,a,l} a^{-l} \right) = \mathbf{E}_S^H \mathbf{E}_S
\]

\[
+ \sum_{k=1}^{K-1} \left( \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^{-k} + \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^k \right)
\]

(3.46)

Hence inserting (3.46) into (3.42) reveals that

\[
KM_2(a, \mathbf{E}_S | \text{“p”}) = (K - 1) \mathbf{E}_S^H \mathbf{E}_S - \sum_{k=1}^{K-1} \left( \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^{-k} + \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^k \right)
\]

\[
+ \sum_{k=1}^{K-1} \left( \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^{-k} + \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^k \right)
\]

\[
- \sum_{k=1}^{K-1} \left( \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^{-k} + \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} a^k \right)
\]

\[
= \sum_{k=1}^{K-1} \left( \mathbf{E}_{S,a,k}^H - \mathbf{E}_{S,a,k} a^{-k} \right) \left( \mathbf{E}_{S,a,k} - \mathbf{E}_{S,a,k} a^k \right)
\]

(3.47)

where we made use of the identity

\[
(K - 1) \mathbf{I}_p = (K - 1) \mathbf{E}_S^H \mathbf{E}_S = \sum_{k=1}^{K-1} \left( \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} + \mathbf{E}_{S,a,k}^H \mathbf{E}_{S,a,k} \right)
\]

(3.48)

The exact equivalence of the RARE polynomial criterion $\det \{ M_2(a, \mathbf{E}_S | \text{“p”}) \} = 0$ with the polynomial criterion $\det \{ M_5^H(a^{-1}, \mathbf{E}_S | \text{“p”}) \} \det \{ M_5(a, \mathbf{E}_S | \text{“p”}) \} = 0$ that is deduced from equation (3.47) is a surprising result that actually exhibits the close relation between the original RARE approach [PGWBO1, PGW02b, PGW02a, WZ99, ZW00, SSJ00] and the concept of (MI) ESPRIT [SORK92, HN98, ZHM96, vdVVP97, vdVVA98, VvdVP98, SLS01, FRB97].
Moreover the RARE polynomial \( M_2(a, E_S \mid \text{“p”}) \) can be represented as

\[
KM_2(a, E_S \mid \text{“p”}) = M_5^H(a^{-1}, E_S \mid \text{“p”})M_5(a, E_S \mid \text{“p”})
\]

\[
= \sum_{k=1}^{K-1} \left( E_{S,a,k}^H - E_{S,a,k}^H a^{-k} \right) \left( E_{S,a,k} - E_{S,a,k} a^k \right)
\]

\[
= \sum_{k=1}^{K-1} \left( E_{S,a,k}^H E_{S,a,k} - E_{S,a,k}^H E_{S,a,k} a^k \right)
+ \sum_{k=1}^{K-1} \left( E_{S,a,k} E_{S,a,k}^H - E_{S,a,k} E_{S,a,k}^H a^{-k} \right)
= M_6(a, E_S \mid \text{“d”}) + M_7(a, E_S \mid \text{“d”})
\] (3.49)

where

\[
M_7(a, E_S \mid \text{“d”}) = \sum_{k=1}^{K-1} \left( E_{S,a,k}^H E_{S,a,k} - E_{S,a,k}^H E_{S,a,k} a^{-k} \right)
\] (3.50)

can be viewed as the backwards version of the MP \( M_6(a, E_S \mid \text{“d”}) \). That is, if we reverse the samples taken along the \( a \)-axis, \( X_B = J_K X \) and replace \( a \) by \( a^{-1} \), then it is simple to check that equation (3.32) applied on the transformed or so-called backwards data yields the MP in (3.50).

Note that in accordance to definition (3.32) we explicitly define the MP \( M_7(a, E_S \mid \text{“d”}) \) for the damped HRP as indicated by the letter “d”.

We shall in the following prove that the MP \( M_7(a, E_S \mid \text{“d”}) \) is equivalent to \( M_6^*(a^{-1}, E_S \mid \text{“d”}) \). That is, if \( a^{-1}_k \) is a root of \( M_7(a, E_S \mid \text{“d”}) \) (or \( M_7(a, H \mid \text{“d”}) \)) then \( a^*_k \) is a root of \( M_6(a, E_S \mid \text{“d”}) \) (or \( M_6(a, H \mid \text{“d”}) \)). To this end we define

\[
\Delta_{a,b} = \Delta_{b}^{-(L-1)/2} \Delta_{a}^{-(K-1)/2}.
\] (3.51)

Multiplying the signal matrix \( H \) from the left with \( \Delta_{a,b} \) is equivalent to choosing the center of the \( a-b \) plane as the origin of the sampling scheme. If we multiply the MP \( M_7(a, H \mid \text{“d”}) \) from the left with \( \Delta_{a,b}^* \) and from the right with \( \Delta_{a,b} \) we obtain

\[
\Delta_{a,b}^*M_7(a, H \mid \text{“d”})\Delta_{a,b} = \Delta_{a,b}^* \sum_{k=1}^{K-1} \left( H_{a,k}^T H_{a,k} - H_{a,k}^T H_{a,k} a^{-k} \right) \Delta_{a,b}
\]

\[
= \sum_{k=1}^{K-1} \Delta_{a,b}^* H_{a,k}^T H_{a,k} \Delta_{a,b} - \sum_{k=1}^{K-1} \Delta_{a,b}^* H_{a,k}^T H_{a,k} \Delta_{a,b} a^{-k}
\]

\[
= \sum_{k=1}^{K-1} \Delta_{a,b} H_{a,k}^T H_{a,k} \Delta_{a,b}^* - \sum_{k=1}^{K-1} \Delta_{a,b} H_{a,k}^T H_{a,k} a^k \Delta_{a,b}^*
\]

\[
= \Delta_{a,b} \sum_{k=1}^{K-1} \left( H_{a,k}^T H_{a,k}^* - H_{a,k}^T H_{a,k}^* a^k \right) \Delta_{a,b}^*
\]

\[
= \Delta_{a,b} M_6^*(a^{-1}, H \mid \text{“d”}) \Delta_{a,b}^*.
\] (3.52)
where we made use of property

\[ H_{a,k} \Delta_{a,b} = \Pi_{KL} H^*_{a,k} \Delta^*_{a,b} \]  \hspace{1cm} (3.53)

with \( \Pi_{KL} \) denoting the \( KL \times KL \) exchange matrix and \( \Pi_{KL}^H \Pi_{KL} = I_{KL} \). From relation (3.52) it is clear that if \( a_k^* \) is a root of \( M_7(a, H \mid "d") \) then \( a_k^{-1} \) is a root of \( M^*_6(a, H \mid "d") \). Thus it immediately follows that the roots of \( M_7(a, E_S \mid "d") \) and \( M_6(a, E_S \mid "d") \) are conjugate-reciprocal, so that if \( a_k^* \) is a root of \( M_7(a, E_S \mid "d") \) then \( a_k^{-1} \) is a root of \( M_6(a, E_S \mid "d") \).

Relation (3.52) allows us to deduce the following corollary from theorem T3:

**Corollary C3:** Provided that \( A3 \) holds true and if all generators are located on or inside the unit-circle, then the MP \( M_7(a, E_S \mid "d") \) evaluated outside and on the unit-circle (\(|a| \geq 1\)) is non-singular if \( a \) is not contained in the set of conjugate reciprocal true generators given by \( \{1/a_1^*, 1/a_2^*, \ldots, 1/a_p^*\} \), and singular otherwise. The order by which \( M_7(a, E_S \mid "d") \) drops rank for \( a = 1/a_p^* \), i.e. the dimension of the corresponding nullspaces, equals the multiplicity of the harmonic \( a \) in the conjugate reciprocal generator set \( \{1/a_1^*, 1/a_2^*, \ldots, 1/a_p^*\} \). Further, all spurious or noise-solutions are located strictly inside the unit circle.

In the proof of T3 we have shown that the real part of \( M_6(a, E_S \mid "d") \) is positive definite for all roots inside or on the unit-circle that do not correspond the true generators. Hence for generators on the unit circle ("p"), the Hermitian part of \( M_6(a, E_S \mid "d") \) is positive definite for all roots inside the unit circle. Similarly, with (3.52) we have that the Hermitian part of \( M_7(a, E_S \mid "d") \) is positive definite for all roots outside the unit circle. Hence in this context relation (3.49) allows the following intuitive interpretation. While the first summand in the equation (3.49), i.e. the polynomial \( M_6(a, E_S \mid "p") \), is "responsible" for the spurious roots of \( M_2(a, E_S \mid "p") \) outside the unit circle, the conjugate-reciprocal roots inside the unit circle are "due" to
the second summand, i.e. the polynomial \( M_7(a, E_S | \text{“p”}) \). Interestingly, simulation results shown in Fig. 3.1 reveal that the spurious roots of \( M_2(a, E_S | \text{“p”}) \) located in the unit circle and the spurious roots of \( M_6(a, E_S | \text{“p”}) \) are located closely in terms of the corresponding angles in the complex plane. Here, the signal and noise roots of the MP of kind 6 and the MP of kind 2 along the \( a \) axis are displayed, respectively, for the ideal case of exactly known covariance matrix and for a representative uniform 2D pure HRP with 3 harmonics and sample support \( 8 \times 8 \). The generators of the 3 harmonics were chosen as \((a_1, b_1) = (e^{-j0.01\pi}, e^{j0.05\pi})\), \((a_2, b_2) = (e^{j0.1\pi}, e^{j0.12\pi})\), and \((a_3, b_3) = (e^{-j0.07\pi}, e^{-j0.1\pi})\). In contrast, the radii of the spurious roots corresponding to \( M_2(a, E_S | \text{“p”}) \) inside the unit circle are smaller than the corresponding radii of \( M_6(a, E_S | \text{“p”}) \). For an overview, the main rank properties and interrelations between all MPs introduced in this chapter are summarized in table E.1.
3 Rank reduction estimators
4 Extensions to the remaining array axes

In the previous sections we have developed a variety of MPs and formulated rank-reduction criteria that allow unique estimation of the true generators, $a_1, \ldots, a_P$, from the subspace spanned by the columns of $E_S$. The novel rank reduction algorithms, that can be deduced from the rank properties formulated in the previous chapters, efficiently exploit the highly regular structure of the sampling scheme along the $a$-axis that is inherent by the MD HRP under model (1.8). Essentially these algorithms account for the MI or block-Vandermonde structure of the ideal manifold matrix. The parameter estimation of the generators along the $a$-axis, hence the Vandermonde matrix $A$, is separated from the estimation of the remaining signal parameters contained in matrix $F$. This chapter considers the problem of recovering all residual signal parameters and the generators observed along the remaining array axes. In chapter 6 we will then learn about efficient and reliable parameter association techniques that allow to assign the individual parameter estimates, obtained separately along the various dimensions, to a specific MD harmonic.

4.1 Uniform sampling along all array axis

We start our considerations with the special case of pure and damped uniform MD HR described in section 1.2.1. Due to the high symmetry obtained from uniform sampling along all observation axes this case is comparably simple to develop from the results obtained in the previous sections. Consider once again the data model in (1.9). From (1.10), (2.4) and (2.9) and similarly from the considerations on the data domain approach carried out in chapter 2.1.2, the uniform sampling along the three array axes with sample support $K, L'$ and $M$ amounts to a $KL'M \times P$ signal matrix that can be represented as a Khatri-Rao product of three Vandermonde matrices according to

$$
H = F \circ A = C \circ B \circ A
$$

where

$$
F = C \circ B
$$

represents the matrix containing the parameters along the remaining array axes and the Vandermonde matrices $A$, $B$, and $C$ are defined according to (2.49)-(2.51), respectively. It is clear from the definition of the Khatri-Rao product (A.3) and from identity (A.5) that commuting the matrix factors in the product (4.1) results in specific permutation of the rows of the resulting
matrix. Hence it holds that

\[ H_c = B \circ A \circ C = Q_c (C \circ B \circ A) = Q_c H \]  
\[ H_b = A \circ C \circ B = Q_b (C \circ B \circ A) = Q_b H \]  

where \( Q_b \) and \( Q_c \) denote the \( KL'M \times KL'M \) permutation matrices defined as

\[ Q_c = [I_{KL'} \otimes i_{M,1}, I_{KL'} \otimes i_{M,2}, \ldots, I_{KL'} \otimes i_{M,M}] \]  
\[ Q_b = [I_{KM} \otimes i_{L',1}, I_{KM} \otimes i_{L',2}, \ldots, I_{KM} \otimes i_{L',L'}] Q_c \]

and \( i_{K,k} \) denotes the \( k \)th column of a \( K \times K \) identity matrix \( I_K \). Note that with definition

\[ Q_a = [I_{LM} \otimes i_{K,1}, I_{LM} \otimes i_{K,2}, \ldots, I_{LM} \otimes i_{K,K}] Q_c \]

the permutation matrix becomes \( Q_a = I_{KLM} \) such that \( Q_a H = H \) which is intuitive since threefold cyclic commutation of the factors in the product \( C \circ B \circ A \) shall yield the original product. The ordering of the rows in \( H_b \) and \( H_c \) facilitates the formulation of similar matrix identities for the harmonics \( b \) and \( c \), respectively, as the ones previously formulated for the harmonic \( a \) and the signal matrix \( H \). This procedure allows to use the same framework previously used to design MPs in the parameter \( a \) to now set up MPs in \( b \) and \( c \) with corresponding rank properties. In (4.3) and (4.4) the cyclic commutation of the factors in \( H \) (2.48) preserves the structure of the underlying estimation problem. The permutation of the rows of the signal matrix amounts to a cyclic change of variables. Hence the permuted signal matrix \( H_c \) has the same structure as the block Vandermonde signal matrix \( H \) with the difference that in the Khatri-Rao product \( A \) is replaced by \( C \), \( B \) is replaced by \( A \) and \( C \) becomes \( B \). Similarly, comparing the permuted signal matrix \( H_b \) with the original signal matrix \( H \) we note that \( A \) becomes \( B \), \( B \) becomes \( C \) and \( C \) becomes \( A \). Since in the permuted signal matrices \( H_c \) the matrix \( C \) with generators \( c_1, \ldots, c_p \) along \( c \)-axis plays the role of \( A \) with generators \( a_1, \ldots, a_p \) along \( a \)-axis in the original signal matrix \( H \), it is immediate to set up MPs in parameter \( c \) as previously obtained for MPs in the generator \( a \) by consistent replacement of variables. The same rank properties and relations between MPs of kinds 1-7 are obtained as previously derived for MPs in \( a \). Concerning the \( b \)-axis parameters we observe that in \( H_b \) the matrix \( B \) with generators \( b_1, \ldots, b_p \) along \( b \)-axis plays the role of \( A \) with generators \( a_1, \ldots, a_p \) along \( a \)-axis in \( H \), hence straightforward replacement of variables amounts in MPs in parameter \( b \) with the same rank properties as previously obtained for MPs in the generator \( a \).

In summary, the main difference in the use of the new MPs, besides the change of variable \( a \) to \( b \) (and \( c \)) is that now the permuted versions \( H_b \) (and \( H_c \)) are used instead of signal matrix \( H \) as input arguments of the MPs. Also the integers \( K, L, \) and \( M \) indicating the sample support along the various axes are replaced in a cyclic fashion in the new MP formulations. That is to say, for the MPs in parameter \( c \), we replace \( H, K, L' \), and \( M \) by \( H_c, M, K \) and \( L' \). Similarly, for the MPs in \( b \), the parameters \( H, K, L' \), and \( M \) are replaced by \( H_b, L', M \) and \( K \), respectively.
4.1 Uniform sampling along all array axis

From the preceding discussion it is apparent that the permuted signal matrices in (4.3) and (4.4) require corresponding permutations in the rows of the signal eigenvector matrix in $E_S$. We define the signal eigenvector matrix for

$$E_{S,c} = Q_c E_S$$

(4.8)

$$E_{S,b} = Q_b E_S.$$ 

(4.9)

Thus the permuted signal eigenvector matrix $E_{S,c}$ used in the MPs in parameter $c$ and $E_{S,b}$ used in the MPs in parameter $b$. In appendix E we list the exact expression for the MPs in parameter $b$ and $c$ that are based on the results obtained in the previous sections for harmonic $a$ and the permutation introduced above. Following the replacement procedure described in this subsection we obtain the MPs $M_1(b, E_{S,b} \mid "p")$ to $M_7(b, E_{S,b} \mid "d")$ in the parameters $b$ defined in (E.1)-(E.13) and, similarly, the MPs $M_1(c, E_{S,c} \mid "p")$ to $M_7(c, E_{S,c} \mid "d")$ in the parameters $c$ according to (E.2)-(E.14). The formal analogies between MPs of the same kind allow to reformulate and extend the MP characteristics, the rank properties and mutual relations between the MPs in $a$ summarized in table E.1 for MPs in parameters $b$ and $c$. The results can be found in the tables E.2 and E.3 for the $b$-axis and $c$-axis, respectively.

Note finally that in the definitions of the MPs we only considered the practically relevant case of known signal eigenvector matrix $E_S$ (or its permuted versions $E_{S,b}$ and $E_{S,c}$) as this quantity is directly obtained from the measurement data (see e.g. section 2.1 and 2.2). We have mentioned previously that for the MPs of kind 3 and 5-7 equivalent MPs are obtained if we use the signal eigenvector matrix $E_S$ or signal matrix $H$ as input argument (see e.g. equation (3.33) in the proof of $T3$). This is because in those MP formulations merely the signal subspace spanned by the columns of the eigenvector matrix and not the unitarity of its columns is of importance. The same statement holds true for the MPs of kind 3, and 5-7 in the parameters $b$ and $c$. 
4 Extensions to the remaining array axes

4.2 Spectral rank reduction estimator

The preceding section considered the highly symmetrical case that is obtained from uniform sampling along all array axes. Now we consider the general case reported in section 1.2.2 as the partly uniform 2D HRP. This model assumes that uniform sampling is given only along the first dimension, i.e. the $a$-axis, while the second array axis is non-uniformly sampled. The difficulty arising in this context is that, even though the sampling scheme along the second array axis is assumed to be known (e.g. in a calibrated measurement system), an important part of the rich invariance structure along the second array axis is lost in this measurement setup compared to the uniform sampling case.

We start our considerations from the definition of the signal matrix in (2.9) as $H = F \circ A$. The simple exchange of variables $A$ and $F$ and the applications of the previous results to estimate the parameters of $F$ is not feasible since there exists a fundamental difference in the structure of $A$ and $F$. While $A$ is a Vandermonde matrix due to uniform sampling along the $a$-axis, this is not the case for $F$. This makes the formulation of shift invariances along the second array axis more complicated than for the $a$-axis (2.65). We define the block matrices

$$ H_{f,l} = A \Delta_{f,l} = (i_{L,l} \otimes I_K) H $$  \hspace{1cm} (4.10)

for $l = 1, \ldots, L$. Here, the diagonal matrix

$$ \Delta_{f,l} = \text{diag}\{[F]_{l,1}, [F]_{l,2}, \ldots, [F]_{l,P}\} $$  \hspace{1cm} (4.11)

contains the elements in the $l$th row of $F$ on its main diagonal and $i_{L,l} \otimes I_K$ represents a $KL \times L$ matrix that selects only the $(l - 1)K + 1$th to $lK$th row of $H$. For simplicity we assume in the following that $[F]_{l,k} \neq 0$ for $k = 1, \ldots, K$ and $l = 1, \ldots, L$. With the definitions given above, the following invariances concerning the $f$-axis are immediate

$$ H_{f,l} \Delta_{f,l}^{-1} = A = H_{f,n} \Delta_{f,n}^{-1} $$  \hspace{1cm} (4.12)

or equivalently

$$ H_{f,l} \Delta_{f,l}^{-1} \Delta_{f,n} = H_{f,n} $$  \hspace{1cm} (4.13)

for $l, n = 1, \ldots, L$ and $n < l$ to avoid identical invariance equations. The example of section 1.2.2 considers the case that $[F]_{l,p} = a_{p-b}^{e_{b,l}} b_{p-b}^{e_{b,l}}$ for $l = 1, \ldots, L$. For simplicity we assume that $\varepsilon_{a,l} = 0$ for $l = 1, \ldots, L$. Then the MI equations in (4.13) become

$$ H_{f,l} \Delta_{b}^{e_{b,n}-e_{b,l}} \Delta_{b}^{e_{b,n}-e_{b,l}} = H_{f,n} $$  \hspace{1cm} (4.14)

for $l, n = 1, \ldots, L$ and $n < l$. Hence, in terms of signal eigenvector matrices, the MI equations in (4.14) read

$$ E_{S,f,l} K \Delta_{b}^{e_{b,n}-e_{b,l}} = E_{S,f,n} K $$  \hspace{1cm} (4.15)
where in accordance to (4.10)
\[ E_{S,f,l} = (i_l \otimes I_K) E_S \] (4.16)
is obtained from signal eigenvectors in $E_S$ through appropriate row selection for $l = 1, \ldots, L-1$.

Before discussing the means to solve the MI equations in the non-uniform sampling case, let us address the question under which conditions there exists a unique pair of full rank matrix $K$ and diagonal matrix $\Delta_b$ that solves the MI equations in (4.15). From (4.12) we note that in setting up the invariance equations along the second array axis information about the shifted structure of $A$ is lost. The same invariance equations are obtained irrespectively the structure of $A$, hence it is not necessary to know the manifold corresponding to $A$ or the sampling scheme along the $a$-axis in order to set up the system of MI equations. This property, that has been observed in virtually all ESPRIT-type methods, is well-known in array and signal processing literature, where it is for example well established that ESPRIT does not require calibration of the shifted subarrays (as long as all subarrays are identical) but only knowledge about the subarray displacements.

From this perspective and adopting the framework introduced in section 3.2, we can state that in the MI equations (4.14) part of the manifold structure of the original signal vectors

\[ h(a, b \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \begin{bmatrix} 1 \\ b^{b,2} \\ \vdots \\ b^{b,L} \end{bmatrix} \circ a \] (4.17)
is “relaxed”. Instead of searching for manifold vectors $h(a, b \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})$ that are located in the signal subspace, here the estimation problem consists of searching for manifold vectors on a relaxed manifold given by

\[ g(p, b \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \begin{bmatrix} 1 \\ b^{b,2} \\ \vdots \\ b^{b,L} \end{bmatrix} \circ p \] (4.18)

that are located in the signal subspace. Here the original Vandermonde vector $a$ is replaced by an arbitrary non-zero vector $p \in \mathbb{C}^L$. On the original manifold, when searching for manifold vectors $h(a, b \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})$ in the signal subspace, the uniqueness of parameter estimates $a$ and $b$ is guaranteed when there exists no signal vector $h(a, b \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})$ with $a \notin \{a_1, \ldots, a_P\}$ or $b \notin \{b_1, \ldots, b_P\}$ that can be represented as a linear combination of the columns of the true signal matrix $H$. Similarly, on the relaxed manifold the uniqueness of the parameter estimate $b$ requires that no vector $g(p, b \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})$ defined in (4.18) with $b \notin \{b_1, \ldots, b_P\}$ can be represented as a linear combination of the columns of $H$. In the following we assume that this condition is always satisfied. According to the considerations in section 3.3 this is equivalent to assuming the following:
**Assumption A4:** The augmented matrix

\[ M_3(b, H \mid \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}) = \begin{bmatrix}
    I_K & H_{f,1} \\
    I_Kb_{b,2} & H_{f,2} \\
    \vdots & \vdots \\
    I_Kb_{b,L} & H_{f,L}
\end{bmatrix} \]  

is full column rank for \( b \notin \{b_1, \ldots, b_P\} \).

The subscript “3” in the notation of the matrix function \( M_3(b, H \mid \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}) \) points out the analogy to the MP given in section 3.3. It is simple to see that for \( b \) equal to one of the true generators, the matrix \( M_3(b, H \mid \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}) \) becomes low-column-rank with the dimension of the corresponding nullspace equal to the multiplicity of \( b \) in the true generator set. This is because there always exists a linear combination of the first \( K \) columns of \( M_3(b, H \mid \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}) \) that can be represented as one of the last \( P \) columns in \( M_3(b, H \mid \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}) \), e.g., for the true Vandermonde vector \( \alpha_p \) with generator \( a_p \), denoting the vector of linear coefficients, hence

\[ \begin{bmatrix}
    I_K \\
    I_Kb_{b,2} \\
    \vdots \\
    I_Kb_{b,L}
\end{bmatrix} \alpha_p = h(a_p, b_p \mid \varepsilon_{b,2}, \ldots, \varepsilon_{b,L}). \]  

(4.20)

Let us return to the MI shift invariance equations in (4.15). In sections 2.3 and 3.4 we already stressed that the MI equations can be solved by means of joint diagonalization techniques. This approach seems particularly useful in case of non-uniform sampling because the MI equations in (4.15) do not yield a rooting based solution. Hence, in joint diagonalization techniques the set of \( L - 1 \) equations is solved by searching for a common eigenvector matrix \( K \) that approximately diagonalizes all matrices on the left side of

\[ E_{S,f,n}^\dagger E_{S,f,l} = K \Delta_b^{\varepsilon_{b,l}-\varepsilon_{b,n}} K^{-1} \]  

(4.21)

for \( l, n = 1, \ldots, L \) with \( n \leq l \).

The diagonal matrices \( \Delta_b^{\varepsilon_{b,l}-\varepsilon_{b,n}} \) obtained from joint diagonalization contain the corresponding eigenvalues on its main diagonal. The generators of interest, the parameters \( b_1, \ldots, b_P \), are easily obtained from these eigenvalues. This diagonalization approach appears to be straightforward and also yields certain important advantages, as for example automatical pairing of the parameter estimates along the different array axes. However, major drawbacks are the convergence difficulties and the limited performance. This is due to the fact that important information about the relations between the diagonal eigenvector matrices \( \Delta_b^{\varepsilon_{b,l}-\varepsilon_{b,n}} \) for different values of \( l \) and \( n \) are not accounted for (see also section 3.4). In the following, we shall develop a technique that fully incorporates this relations.
In section 3.4, where the MI equations for uniform sampling along the $a$-axis were considered we provided several means to express the MI equations in terms of rank properties of associated MPs in parameter $a$. For non-uniform sampling along the $b$-axis, the characteristic equations in (4.15) have generalized eigenvalues of the form $\epsilon_{p,q}^{b_0-b,n}$ ($p = 1, \ldots, P; l, n = 1, \ldots, L$ and $n < l$) that are not necessarily integer powers of the true generators as it is the case in uniform sampling. In other words, the phase shifts $\epsilon_{b,l} - \epsilon_{b,n}$ are generally arbitrary real numbers. Nevertheless, following the steps that led from the characteristic equations in (3.29) to the MPs in (3.31), i.e. stacking the MI equations for different values of $l$ and $n$ on top of each other to form a “tall” matrix function in $b$, we obtain the $KL(L-1)/2 \times P$ matrix

$$M_5(b, E_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \begin{bmatrix} E_{s,f,1} - E_{s,f,2} b^{f_{b,2}} \\ \vdots \\ E_{s,f,1} - E_{s,f,L} b^{f_{b,L}} \\ E_{s,f,2} - E_{s,f,3} b^{f_{b,3} - \epsilon_{b,2}} \\ \vdots \\ E_{s,f,2} - E_{s,f,L} b^{f_{b,L} - \epsilon_{b,2}} \\ \vdots \\ E_{s,f,L-1} - E_{s,f,L} b^{f_{b,L} - \epsilon_{b,L-1}} \end{bmatrix}$$ (4.22)

which, provided that $A4$ is satisfied and the system of MI equations has a unique solution, drops rank only for $b$ equal to one of the true generators $b_1, \ldots, b_P$. The dimension of the corresponding nullspace is given by the multiplicity of $b$ in the true generator set. In contrast to the MPs in $a$ previously obtained for uniform sampling along the $a$-axis, here we have a general matrix function (MF) in the parameter $b$ which is not necessarily a MP. Hence, instead of efficient rooting procedures a spectral search needs to be performed in order to find the true generators for which the MF become rank deficient [SSJ01, SG04]. The 2D SPEC-RARE function and the 2D SPEC-MI-ESPRIT function formulated in the harmonic along the $b$-axis can for example be defined as

$$f_{2D \text{ SPEC-RARE}}(b, E_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \sigma_{\min}^{-1}\{M_3(b, E_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})\}$$ (4.23)

$$f_{2D \text{ SPEC-MI-ESPRIT}}(b, E_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \sigma_{\min}^{-1}\{M_5(b, E_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})\}$$ (4.24)

where $\sigma_{\min}\{M\}$ denotes the inverse of the minimum singular value of a matrix $M$. In the 2D SPEC-RARE algorithm the parameters of interest along the $b$-axis are obtained from the $P$ highest maxima of the cost function in (4.23). Correspondingly, in 2D SPEC-MI-ESPRIT, (4.24) serves as the cost function whose highest maxima yield the parameters $b_1, \ldots, b_P$. In the finite sample case when only estimates of the signal eigenvectors are available the true signal eigenvector matrix $E_S$ is replaced by its finite sample estimates $\hat{E}_S$ given in (2.22) and (2.47). The finite sample versions of the functions in equations (4.23) and (4.24) then read

$$\hat{f}_{2D \text{ SPEC-RARE}}(b, \hat{E}_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \sigma_{\min}^{-1}\{M_3(b, \hat{E}_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})\}$$ (4.25)

$$\hat{f}_{2D \text{ SPEC-MI-ESPRIT}}(b, \hat{E}_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L}) = \sigma_{\min}^{-1}\{M_5(b, \hat{E}_S \mid \epsilon_{b,2}, \ldots, \epsilon_{b,L})\}$$ (4.26)
In section 6.4 we shall discuss a more efficient way to handle the parameter estimation problem along the non-uniformly sampled axis. There, a suboptimal method is presented in which all parameters of interest are directly obtained from rooting along the uniform sample dimensions. This method can for example be used to initialize the spectral search proposed in this section.
5 Implementation

5.1 Polynomial rooting methods

This section provides the most important tools for rooting square MPs. Efficient MP-rooting methods have been developed and are widely used in control theory literature [Kai80]. Here, we will review the two most popular approaches. The first approach consists of direct application of the determinant expansion rule. In a first step, the coefficients of a scalar matrix polynomial representing the determinant of the square MP are evaluated. In a second step, the roots of this polynomial are determined using standard rooting techniques for scalar polynomials [PGWB01]. We shall see that this method can efficiently be implemented using FFT [Pol00]. The second approach operates directly on the polynomial coefficients of the MP. A so-called block companion matrix (BCM), similar to the well-known companion matrix in the scalar case, is formed from the matrix coefficients [Kai80, GLR82]. The roots of the MP are obtained from the eigenvalues of the BCM. While the first method is reported to be attractive from a numerical point of view [Pol00], section 6.3 shows that the latter method provides some specific advantages for solving the parameter association problem.

5.1.1 FFT approach

Before we start our considerations, we shall recall some of the most important polynomial operations. Let \( P_1(a) = \sum_{n=0}^{K} p_1(n)a^n \) and \( P_2(a) = \sum_{n=0}^{L} p_2(n)a^n \) denote two scalar polynomials of degree \( K \) and \( L \) with polynomial coefficients \( p_1(0), \ldots, p_1(K) \) and \( p_2(0), \ldots, p_2(L) \), respectively. Here, \( p_1(n-1) \) denotes the \( n \)th polynomial coefficient of \( P_1(a) \), hence the polynomial coefficient corresponding to \( a^{n-1} \). It is clear that the sequence of polynomial coefficients \( p_1(0), \ldots, p_1(K) \) and the polynomial \( P_1(a) \) itself forms the following z-transform pair:

\[
P_1(a) = \mathcal{Z}\{p_1(n)\}(a).
\]

Using the convolution property of z-transform we obtain that the product of \( P_1(a) \) and \( P_2(a) \) results in a polynomial \( P_3(a) = \sum_{n=0}^{K+L} p_3(n)a^n \) of degree \( K + L \) which can expressed as

\[
P_3(a) = P_1(a)P_2(a) = \mathcal{Z}\{p_1(n)\}(a)\mathcal{Z}\{p_2(n)\}(a) = \mathcal{Z}\{(p_1 * p_2)(n)\}(a) \quad (5.1)
\]

Here "*" denotes the convolution operator. Hence, multiplication of two scalar polynomials (in z-transform domain) results in the convolution of the corresponding sequences of polynomial coefficients (in data- or “polynomial-coefficient”-domain). In practice, for polynomials of large degrees, a natural approach is to exploit the relation between the DFT and the z-transform to compute the polynomial coefficients of the resulting polynomial \( P_3(a) \). Instead of directly con-
volving the sequences of polynomial coefficients \( p_1(n) \) and \( p_2(n) \), both sequences are first transformed using the DFT or its efficient implementation, the fast Fourier transformation (FFT). The resulting sequences in discrete Fourier domain can be regarded as sampled versions of the polynomials \( P_1(a) \) and \( P_2(a) \) evaluated at \( a = e^{\frac{2\pi i}{N}} \) for \( k = 0, \ldots, N - 1 \), where \( N \) denotes the number of chosen frequency bins. Note that in order to avoid aliasing, \( N \) needs to be larger than the degree of the resulting polynomial \( P_3(a) \), i.e. \( N \geq K + L \). With \( \text{FFT}\{p_1(n)\}(k) \) denoting the FFT of a sequence \( p_1(n) \) and \( \text{IFFT}\{P_1(k)\}(n) \) denoting the inverse FFT (IFFT) of a discrete sequence \( P_1(k) \), we obtain

\[
p_3(n) = (p_1 * p_2)(n) = \text{IFFT}\{ \text{FFT}\{p_1(n)\}(k) \ \text{FFT}\{p_2(n)\}(k) \}(n) \quad (5.2)
\]

Consider next the general MP \( \mathbf{M}(a) \) of dimensions \( P \times P \) and degree \( K \) given by

\[
\mathbf{M}(a) = \sum_{n=0}^{K} \mathbf{C}(n)a^n \quad (5.3)
\]

where the \( P \times P \) matrices \( \mathbf{C}(0), \ldots, \mathbf{C}(K) \) denote the sequence of matrix polynomial coefficients. Let \( \mathbf{M}_{k,l}(a) \) denote the entry in the \( k \)th row and \( l \)th column of \( \mathbf{M}(a) \), thus a scalar polynomial of degree \( K \) with coefficients \( \{\mathbf{C}_{k,l}(0), \ldots, \mathbf{C}_{k,l}(K)\} \). Further let \( \mathbf{M}_{k,l}(a) \) be the \((P-1) \times (P-1)\) matrix polynomial of degree \( K \) that is obtained from \( \mathbf{M}(a) \) by deleting its \( k \)th row and \( l \)th column. Making use of the recursive formula for computing determinants and developing the determinant according to the first column we can write

\[
D_{\mathbf{M}}(a) = \det\{\mathbf{M}(a)\} = \sum_{n=0}^{PK} d_{\mathbf{M}}(n)a^n = \sum_{p=1}^{P} (-1)^{p-1}[\mathbf{M}]_{p,1}(a) \det \mathbf{M}_{p,1}(a). \quad (5.4)
\]

Note that in a \( P \times P \) polynomial of degree \( K \) the maximum degree of the resulting determinant polynomial is given by \( PK \). It is clear from (5.1) and (5.4) that if \( d_{\mathbf{M}}(n) \) denotes the sequence of polynomial coefficients corresponding to the determinant of \( \mathbf{M}(a) \) then this sequence is obtained as

\[
d_{\mathbf{M}}(n) = \sum_{p=1}^{P} (-1)^{p-1}[\mathbf{C}]_{p,1}(n) * d_{\mathbf{M}_{p,1}}(n) \quad (5.5)
\]

The recursive determinant evaluation scheme in (5.4) and (5.5) appears to be useful only for matrix polynomials of moderate degree and small dimension. For larger rooting problems the computational cost and the memory requirements associated with this procedure are prohibitive. In this case we exploit relation (5.2) to circumvent the expensive convolution operation. Hence, the polynomial coefficients of the determinant \( \det\{\mathbf{M}(a)\} \) are computed as

\[
d_{\mathbf{M}}(n) = \text{IFFT}\left\{ \sum_{p=1}^{P} (-1)^{p-1} \text{FFT}\{[\mathbf{C}]_{p,l}(n)\}(k) \ \text{FFT}\{d_{\mathbf{M}_{p,l}}(n)\}(k) \right\}(n) \quad (5.6)
\]
From a computational (and also from a numerical) point of view it is more efficient to perform the recursive evaluation of the determinant in the DFT domain using FFT and transform the obtained sequences back into the “polynomial-coefficient” domain afterwards using the IFFT. Recall that when performing multiplication in FFT domain it is of primary importance to always keep track of the expected resulting polynomial degree in order to prevent aliasing. After evaluating the coefficients of the polynomial \( \det \{ \mathbf{M}(a) \} \), any standard rooting technique for determining the roots of a scalar polynomial can be applied [Pol00].

### 5.1.2 Block companion matrix approach

In this subsection we describe a different way to determine the roots of a MP following the derivation in [Kai80]. Similar to the scalar polynomial case where the roots are obtained from the eigenvalues of the so-called companion matrix, we show that it is possible to design a block matrix version of this procedure to compute the roots of a MP from the solutions of a sparse eigenproblem. Later in section 6.3 we shall find that this rooting procedure is particularly suitable for solving the MD-HR problem based on the singularities of the MPs of chapter 3. This is because in the eigendecomposition based rooting approach the computational complexity of the HR algorithm is significantly reduced since this approach computes the true roots without the overhead of determining also the spurious solutions. Further, the BCM properties yield efficient solutions to the parameter association problem.

Consider again the \( P \times P \) MP \( \mathbf{M}(a) \) of degree \( K \) defined in (5.3) with matrix coefficients \( \mathbf{C}(0), \ldots, \mathbf{C}(K) \). We can use elementary row and column operations to transform the augmented MP of the form

\[
\begin{bmatrix}
\mathbf{M}(a) & 0 \\
0 & \mathbf{I}_{(K-1)P,(K-1)P}
\end{bmatrix}
\]

(5.7)

to the linear MP given as

\[
\mathcal{L}\{\mathbf{M}(a)\} = \mathcal{V}\{\mathbf{M}(a)\} - a \mathcal{T}\{\mathbf{M}(a)\},
\]

(5.8)

where

\[
\mathcal{V}\{\mathbf{M}(a)\} =
\begin{bmatrix}
0 & \mathbf{I}_P & 0 & \cdots & 0 \\
0 & 0 & \mathbf{I}_P & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \mathbf{I}_P \\
-\mathbf{C}(0) & -\mathbf{C}(1) & \cdots & \cdots & -\mathbf{C}(K-1)
\end{bmatrix}
\]

(5.9)
and

\[
T\{M(a)\} = \begin{bmatrix}
I_P & 0 & \cdots & 0 & 0 \\
0 & I_P & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & I_P & 0 \\
0 & 0 & \cdots & 0 & C(K)
\end{bmatrix}
\]  

(5.10)

are sparse constant \(KP \times KP\) matrices formed from the matrix coefficients of \(M(a)\). In other words, there exist two unimodular\(^1\) MPs \(U(a)\) and \(V(a)\), whose exact form can be found e.g. in [Kai80], such that

\[
M(a) = U(a)(V\{M(a)\} - a \ T\{M(a)\})V(a).
\]  

(5.11)

Both MPs, the augmented MP \(\overline{M}(a)\) of degree \(K\) and the linear MP given in (5.8) have exactly the same singularities. Further, it is simple to prove that \(\det\{\overline{M}(a)\} = \det\{M(a)\}\). Hence, instead of rooting the original MP \(M(a)\) we can equivalently determine the roots of the linear MP in (5.8). The procedure of forming the sparse structures according to (5.8)-(5.10) from the MP coefficients in (5.3) is commonly known as linearization of a MP. The matrices in (5.9) and (5.10) represent a BCM pair. It is clear that determining the roots of the linear MP in (5.8) consists of finding the generalized eigenvalues of the BCM pair.

To illustrate that the roots of the MP \(L\{M(a)\}\) indeed coincide with the roots of the original MP \(M(a)\), or its augmented version in (5.7), we assume for example that \(a_i\) is a root of \(M(a)\) of multiplicity \(M_i\) and that \(K_i \in \mathbb{C}^{K \times M_i}\) is the matrix whose columns span the corresponding nullspace. In this case, the associated characteristic equation reads \(M(a_i)K_i = 0\). Then the \(KP \times M_i\) matrix

\[
V_{a,i} = \begin{bmatrix}
K_i \\
a_iK_i \\
\vdots \\
a_i^{K-1}K_i
\end{bmatrix}
\]  

(5.12)

contains the \(M_i\) generalized eigenvectors corresponding to the generalized eigenvalue \(a_i\) that solve the characteristic equation in (5.8). In other words

\[
V\{M(a_i)\}V_{a,i} - a_i \ T\{M(a_i)\}V_{a,i} = 0.
\]  

(5.13)

In order to show that the reverse statement holds equivalently, i.e. that a root \(a_i\) of \(L\{M(a)\}\) with multiplicity \(M_i\) is also roots of the MP \(M(a)\) with multiplicity \(M_i\), we first note that all

\(^1\)i.e. \(\det\{U\}\) and \(\det\{V\}\) are both equal to a nonzero constant and thus independent of \(a\).
generalized eigenvectors of the matrices in (5.10) and (5.10) have the block-structure provided in (5.12). Considering again the characteristic equation it is readily verified that the last $P$ rows (5.13) yield

\[ \sum_{k=0}^{K} C(k) K_i^k a_i^k = M(a_i) K_i = 0. \]  

(5.14)

In other words, $a_i$ is a root of $M(a)$ with $K_i$ spanning the corresponding nullspace. Thus we showed that $M(a)$ and $\mathcal{L}\{M(a)\}$ are equivalent in the sense that both MPs have identical roots. Here, we only intended to provide some intuition on the linearization method. A detailed proof of the statements made in this section can e.g. be found in [Kai80].

### 5.2 Noise and finite sample effects

Throughout the course of the last chapters we have only considered the case where all model assumptions in the measurement data are exactly satisfied and full knowledge about the signal subspace in the form of the true signal eigenvectors in $E_S$ (2.19) are available. This over-idealistic assumption enabled us to exploit the structural prior information about the measurements and to formulate MPs with specific rank properties. Provided that the true signal subspace is known, the harmonics of interest can uniquely be determined from the roots of these MPs. Considering the noise-free case as the starting point for the design of new estimation methods is a common approach in literature. This idealistic approach stems back from the intuition that the asymptotic subspace properties approximately hold true in presence of moderate noise, and also from the practical consideration that it is much easier to prove the uniqueness of a parameter estimation scheme in absence of noise.

In every real experiment perturbations of the measurements due to noise effects, originating for example from background radiation and reverberation but also from thermal effects in the receiver electronics, are inevitable. In this work we only consider additive noise contributions described by temporally and spatial white complex Gaussian noise according to the models in section 2.1. Further, in real applications the observation time is limited by the measurement setup and the coherence time, that is the time during which the parameters in the measurement setting can be regarded as stationary. The noise contained in the measurements and a limited sample support yields perturbed estimates of the true signal subspace. More specific, in the realistic case only perturbed estimates of the true signal and noise eigenvector matrices $E_S$ and $E_N$, denoted by $\hat{E}_S$ and $\hat{E}_N$, are available. These are obtained either from the sample covariance approach in (2.22) or the data domain approach in (2.44). The same statement holds for the estimates of the signal and noise eigenvalues contained on the main diagonals of $\hat{\Lambda}_S$ and $\hat{\Lambda}_N$, respectively.
The perturbations in the finite sample eigenvectors can be expressed as

\[ \hat{e}_k = e_k + \nu_k \]  
\[ \hat{\lambda}_{S,k} = \lambda_k + \mu_k \]

where \( \nu_k \) denotes the \( KL \times 1 \) vector representing the perturbations in the \( k \)th eigenvector in (2.22) (or \( k \)th singular vector in (2.44)) and \( \mu_k \) denotes the perturbations in the corresponding eigenvalues (or singular values). Obviously, the fewer the sample support from which the signal subspaces are estimated, the larger are the deviations of the estimated signal subspace from the true signal subspace. It is well known that the asymptotic distribution of the \( P \) largest eigenvectors \( \hat{E}_S \) of the sample covariance matrix 2.22 is Gaussian with mean \( \mathbb{E}\{\hat{E}_S\} = E_S + \mathcal{O}(N^{-1}) \) and with a variance of the eigenvectors that is commensurate with the closeness of the corresponding eigenvalues to the noise variance (see for example [VOK91, RH89, PGH00c, PGH00a, PGH00b] and references therein).

If only sample estimates of the signal subspace are available then the true signal eigenvectors in \( E_S \) are replaced by the estimate \( \hat{E}_S \) in the argument of the MPs of kind 1-7. In style of equations (3.6), (3.13), (3.24), (3.26), (3.31), (3.32), and (3.50) we define the following finite sample MPs:

\[ M_i(a, \hat{E}_S | \text{“p”}) = \mathcal{I}_a(a^{-1}) \left( I_P - \hat{E}_S \hat{E}_S^H \right) T_a(a) \]  
\[ M_2(a, \hat{E}_S | \text{“p”}) = I_P - \hat{E}_S^H T_a(a) \Omega^{-1} T_a(a^{-1}) \hat{E}_S \]  
\[ M_3(a, \hat{E}_S | \text{“d”}) = \left[ T_a(a) | \hat{E}_S \right] \]  
\[ M_4(a, \hat{E}_S | \text{“p”}) = \left[ T^T(a^{-1}) T(a) \quad T^T(a^{-1}) \hat{E}_S \right] \hat{E}_S^H T(a) \right] \]  
\[ M_5(a, \hat{E}_S | \text{“d”}) = \left[ \begin{array}{c} \hat{E}_{S,a,1} - \hat{E}_{S,a,1} a^1 \\ \hat{E}_{S,a,2} - \hat{E}_{S,a,2} a^2 \\ \vdots \\ \hat{E}_{S,a,K-1} - \hat{E}_{S,a,K-1} a^{(K-1)} \end{array} \right] \]  
\[ M_6(a, \hat{E}_S | \text{“d”}) = \sum_{k=1}^{K-1} \left( \hat{E}_{S,a,k}^H \hat{E}_{S,a,k} - \hat{E}_{S,a,k}^H \hat{E}_{S,a,k} a^k \right) \]  
\[ M_7(a, \hat{E}_S | \text{“d”}) = \sum_{k=1}^{K-1} \left( \hat{E}_{S,a,k}^H \hat{E}_{S,a,k} - \hat{E}_{S,a,k}^H \hat{E}_{S,a,k} a^{-k} \right) \]

where the “˘” sign above the identifiers \( M_i, i = 1, \ldots, 7 \) shall emphasize that in this MPs the polynomial coefficients are perturbed so that the developed rank properties only hold true approximately. Further in (5.21)-(5.23) we introduced the finite sample versions of (2.66) and
5.2 Noise and finite sample effects

\[ (2.67), \text{ hence} \]

\[ \hat{E}_{S,a,k} = (I_L \otimes J_{K,k}) \hat{E}_S \]  

\[ \hat{E}_{S,a,k} = (I_L \otimes J_{K,k}) \hat{E}_S \]  

respectively, for \( k = 1, \ldots, K - 1 \). The corresponding finite sample MPs in the generators along the \( b- \) and \( c- \) axis are listed in appendix F.

Noise effects and perturbations in (signal) eigenvectors prevent the direct extension of the rank properties of the MPs, originally developed assuming precise knowledge of the true signal subspace vectors, to the realistic case of perturbed eigenvectors. In the following we shall specify the difficulties emerging due to noise perturbations, illustrate their effects on the MP identities presented in the previous section and show how to exploit the MP rank properties to estimate the harmonics of interest in the noise and finite sample case.

1. **Subspace swap:** For low SNR and small sample sizes, perturbations in the eigenvalues eventually become so severe that in the covariance approach of section 2.1 the smallest eigenvalues \( \hat{\lambda} \), which in the ideal case are equal to the noise power \( \sigma^2 \) (2.25), become greater than the smallest signal eigenvalues (2.24). In this case the eigenvectors located in the noise subspace are erroneously assigned to the signal subspace, i.e. to the matrix of signal eigenvectors \( \hat{E}_S \). Similar effects can also be observed in the data domain approach of section 2.2. In case of a wrong eigenvector selection the estimated signal subspace is not only strongly perturbed but rather irrecoverably destroyed, as part of the signal components are missing while noise subspace components are added to the signal subspace. This makes it impossible to recover all signal parameters, i.e. the true signal manifold vectors, from the estimated subspace. This phenomenon is commonly referred to as *subspace-swap* and usually associated with a drastic performance breakdown in threshold domain. We shall come back to the subspace-swap in context of chapter 7, where the behavior of the estimators in threshold domain is studied by means of simulations.

2. **Deviation of signal and noise roots:** Even in the case that the true signal eigenvectors are selected in the subspace extraction step of (2.22) and (2.44), it is simple to see from the definitions of the MPs of kind 1-7 that the perturbations in the eigenvectors result in perturbations of the polynomial coefficients. It is well-known, i.e. from robustness analysis in control theory, that even small perturbations in the polynomial coefficients can have great impact on the root loci. In the realistic case, the signal roots are displaced from its ideal positions given by the true generator locations in the complex plane. Also the noise roots are subject to such displacements. Therefore in practical applications it is important to provide some means to efficiently separate the estimated signal roots from the noise solutions.
In the following we will consider some root selection procedures that are based on the rank properties of the MPs that were previously derived for the noise-free case. Therefore it appears reasonable to distinguish between the following three cases: a) pure HR in square MPs of kind 1, 2, and 4, b) damped HR in square MPs of kind 6 and 7, and c) damped HR in “tall” MPs of kind 3 and 5.

3. Root selection for pure HR in MPs of kind 1, 2, and 4: In the undamped harmonic case the signal roots of the exact MPs of kind 1, 2, and 4 are, according to corollary C2, located on the unit-circle. In contrast, the corresponding spurious or noise roots lie strictly inside and outside of the unit-circle. Recall that the noise (or signal) roots inside (or on) the unit-circle are conjugate-reciprocal to the corresponding roots outside (or on) the unit-circle. In the real case the roots are computed from the estimated MPs in (5.17), (5.18) and (5.20), respectively. Hence, the signal roots are displaced from their ideal positions on the unit-circle. A straightforward approach, that is successfully applied in virtually all root-MUSIC based approaches [Bar83, RH89, Tre02, KV96, PGH00c, PGW02a], is to simply select the signal roots as the $P$ largest (in terms of magnitude) roots inside or on the unit-circle.\(^2\)

4. Root selection for damped (and pure) HR in MPs of kind 6 and 7: In damped HR, the generators are assumed to be located inside or on the unit-circle. From theorem T3 we know that in the ideal case all the signal roots of the MP of kind 6 given by the true generators are located inside or on the unit circle, while the remaining spurious solutions are altogether located strictly outside. In the finite sample case solutions are obtained from rooting the sampled version of the MP defined in (5.22). As already mentioned before, signal and noise roots are subject to displacements from their true locations. According to the procedure proposed above for the pure HR, here the signal roots are computed as the $P$ smallest (in terms of magnitude) signal roots. The full benefit of the proposed estimation scheme using the MP of kind 6, compared to the estimation schemes using MPs of kind 1, 2, and 4, becomes apparent when considering the eigendecomposition based rooting technique introduced in section 5.1.2. Precisely because estimates of the true generators along the $a$-axis are obtained from the $P$ smallest roots of the finite sample MP $\hat{M}_6(a, \hat{E}_S \mid “d”)$, in this algorithm only the $P$ principal GEVs of the BCM pair $\mathcal{V}\{\hat{M}_6(a, \hat{E}_S \mid “d”)\}$, $\mathcal{T}\{\hat{M}_6(a, \hat{E}_S \mid “d”)\}$ need to be determined. It is important to note that the essential difference between the present approach and HR retrieval based on MPs of kind 1, 2, and 4 is that here the signal roots are separated from the noise roots prior to the rooting step, hence the undesired spurious roots (i.e. the remaining eigenvalues of the BCM pair) need not be computed. Efficient eigendecomposition techniques that yield only the $P$ smallest eigenvalues and eigenvectors without perform-

\(^2\)It is simple to show that conjugate-reciprocity of the roots still holds even in the realistic case. Therefore, only the roots inside the unit-circle need to be considered in the selection procedure.
5.2 Noise and finite sample effects

ing the full spectral decomposition are available in literature [Saa00, LSY98, PSBG05], so that the computational cost associated with the algorithm is significantly reduced. In solving the generalized eigenproblem of form (5.8), the Arnoldi-type algorithms, which exploit the sparsity of the generalized eigenpair \( (V\{M(a)\}, T\{M(a)\}) \) to further reduce the computational complexity, appear to be particularly useful. The same statements that have been made with respect to parameter estimation in the realistic case based on the MP of kind 6 can (with the help of corollary C3) can directly be transferred to estimation using the MP of kind 7 defined in (5.23). The only difference is that, according to the discussion in section 3.4, in the ideal case the true generators are obtained as the conjugate-reciprocal of the roots of \( M_7(a, E_S | "d") \) located outside (or on) the unit circle, while all signal roots are located strictly inside the unit circle. Hence in the realistic case where the roots are displaced from their true positions, we compute the signal parameter estimates as the conjugate-reciprocal of the \( P \) largest (in terms of magnitude) roots of \( \hat{M}_7(a, \hat{E}_S | "d") \). Thus we only need to compute the \( P \) largest GEVs of the BCM pair \( (V\{\hat{M}_7(a, \hat{E}_S | "d")\}, T\{\hat{M}_7(a, \hat{E}_S | "d")\}) \) and take its conjugate-reciprocals. The advantage in using the MP of kind 7 over using the MP of kind 6 is that the block diagonal matrix \( T\{M_7(a, E_S | "d")\} \) is generally non-singular while \( T\{M_6(a, E_S | "d")\} \) is generally rank deficient. It is clear that with invertible \( T\{\hat{M}_7(a, \hat{E}_S | "d")\} \) the GEV of the matrix pair \( (V\{\hat{M}_7(a, \hat{E}_S | "d")\}, T\{\hat{M}_7(a, \hat{E}_S | "d")\}) \) can be computed from the eigenvalues of the matrix

\[
T^{-1}\{M_7(a, E_S | "d")\}V\{M_7(a, E_S | "d")\}
\]

in a numerically stable manner.

5. Root selection for damped (and pure) HR in MPs of kind 3 and 5: The rectangular or “tall” MP of kind 3 and 5 that are defined in (3.24) and (3.31), respectively, appear to be only of limited use in practical applications and serve in this work merely to provide better understanding of the underlying subspace relations. We know from corollary C2 that in the ideal case the MPs in (3.24) and (3.31) become rank deficient for \( a \) equal to one of the true generators. However, in the case of random perturbations in the coefficients, this property holds only in an approximate sense. That is, in general the MP is full rank for all values of \( a \) and becomes close to low column-rank for some values of \( a \) close to a true generator. The difficulty that prohibits the practical use of the MPs (5.19) and (5.21) consists of a lack of robust procedures to reliably determine the values of \( a \) for which the “tall” MP with perturbed coefficients is close to low column-rank. An attempt to adopt existing robust GCD and GRD estimation algorithms to the specific rooting problem of the perturbed MP of kind 3 has been undertaken in [PGB03].

6. Parameter association problem: Noise and finite sample-effects also play an important role in the parameter association of the harmonic estimates that are separately obtained
from MPs along the various array axes (see the following chapter 6). Difficulties that arise in this context are described in detail in the following chapter, where robust and computationally efficient parameter association schemes are developed.
6 Parameter association and MD processing

In this chapter we study the parameter association problem arising in algorithms that decompose joint MD parameter estimation into multiple corresponding 1D estimation problems. The benefits of decoupled parameter estimation are at hand: First estimating the parameters separately along the different dimensions ensures the scalability of the algorithms. That is, with increasing the dimensionality of the estimation problem, the associated computational cost of the algorithm does not grow unreasonably. For example, if we move from a 2D HRP to a 4D HRP while keeping the overall sample support fixed, and if we only consider the parameter estimation and not the pairing task then it is simple to show that the overall computational cost of the root-MI-ESPRIT algorithm is generally equal or less than doubled, which definitely is a reasonable increase in this case. The second advantage of separable parameter estimation is that such scheme strongly supports parallel processing to speed up the implementation in real-time systems.

The benefits in computational complexity and efficiency of implementations are only valuable if simple and reliable parameter association procedures exist that assign the parameter estimates contained in the different sets, which were separately obtained along the various dimensions, to the correct MD-harmonics. Parameter association is a difficult problem that can easily become computationally more demanding than the parameter estimation itself, especially when the dimensionality of the estimation problem and the number of signals is high. In this case the number of possible signal constellations, in other words the number of possible parameter assignments or the number of permutations of parameters in the various sets, becomes prohibitively high. It is clear that for large problems, ad-hoc solutions like selection of the true parameter tuples according to a MD criteria like the MUSIC spectrum (2.72) is not feasible because the computational cost associated with the evaluation of the cost functions for all possible candidate constellations would be too high.

Apart from the computational complexity, major difficulties in such a simple assignment approach stem from the estimation errors in the finite sample parameter estimates, as mentioned at the end of the previous section. If the deviations of the signal estimates along the various dimensions from the corresponding true values are large, then the “correct” choice of signal M-tuples (i.e. the candidate M-tuples that are “closest” to the true M-tuples describing the MD harmonics in some mean square sense) does not necessarily yield the largest values of the MUSIC function. This is for example the case when two candidate M-tuples are located close to a true MD harmonic with corresponding large values in the MUSIC spectra and if these values exceed all remaining maxima of the MUSIC function. Then two candidate M-tuples are located
in the same main lobe of the MUSIC function. Therefore, in practice it is either necessary to check whether two of the selected M-tuples converge, in a gradient optimization procedure, to the same maxima of the MUSIC function. Alternatively, a joint criterion like for example the conditional or unconditional ML function [SN89, Böh91, Tre02] needs to be used as a cost function in the association procedure. However, a joint cost function would further increase the computational requirements because in this approach we need to jointly choose $P$ M-tuples from all permutations of $M$-tuples that can be formed from the parameter estimates obtained along the $M$ dimensions, rather than separately selecting the $P$ “best” $M$-tuples corresponding to the maxima of the MUSIC spectrum (2.72).

In summary, when using separate criteria for estimating the parameters along the various dimensions, there exists a strong need to develop fast, efficient and reliable pairing or parameter association strategies. This chapter proposes a variety of parameter association schemes that are based on the specific structure of the underlying MD HRP. We start our consideration on parameter association and joint MD HR in section 6.1, where a tree-structured estimation scheme is proposed in which the parameters along the different axes are sequentially estimated. The estimates along the dimensions that are already obtained are used to successively reduce the dimensionality of the underlying MD HRP. Proper selection procedures performed in each branch eventually yield the correctly associated M-tuples as estimates of the true MD harmonic parameters. In section 6.2 similarities between the nullspace vectors of the low-rank MPs associated with a true M-tuple are exploited to develop an algorithm that is free from error propagation. Finally in section 6.3 we extend the result of section 6.2 to present two particularly efficient implementations of the parameter association and joint MD HRP procedures.

6.1 MD tree-RARE

The tree-structured rank reduction procedure discussed in this subsection consists of sequential estimation of the parameters along the various sampling axes. Each parameter set that is obtained along a single dimension is kept fixed and the parameters are sequentially inserted back into the same MP they were originally obtained from. In each step the dimensionality of the original HRP is reduced by one. Substituting a subset of the unknown parameters for which estimates are available back into the original cost function is a popular trick to simplify complex MD optimization problems. Backsubstitution is applied in a large variety of MD estimation algorithms [PG01]. The successive dimensionality reduction method presented here is somehow related to alternating projection algorithms like [ZW88] in which known signal components and parameters are “projected out” of the data.

Under the framework of this section, we only describe the fundamental concepts and limitations
of the tree-structured estimation scheme. More sophisticated procedures that are free from such limitations are presented in the following sections. However, these algorithms shall rely on similar nullspace properties of the MPs as the algorithm presented in this sections. For a detailed description of the MD Tree-RARE estimator and for questions regarding its implementation we refer to [PMB03].

Let us start our considerations from the uniform pure 3D HRP case discussed in sections 1.2.1 and 4.1 in absence of noise. The procedures presented here easily generalize to the uniformly sampled, undamped MD HRP. From equation (4.1) we know that the signal matrix $H$ can be represented as the Khatri-Rao product of Vandermonde matrices $C$, $B$ and $A$. Further we assume, without loss of generality, that the set of true generators $\{a_1, \ldots, a_P\}$ along the first sampling axis is obtained as the roots of the MP of kind 1 and let $a_1 = a_2 = \ldots = a_m$ be a generator of multiplicity $m \leq LM$, then we know from corollary C2 that the MP $M_1(a_1, E_S | "p")$ is singular with $m$ denoting the dimension of the corresponding nullspace. Let us partition the signal matrix as

$$H = \begin{bmatrix} \tilde{H}_1 & \tilde{H}_2 \end{bmatrix}$$

with

$$\tilde{H}_1 = \tilde{C}_1 \circ \tilde{B}_1 \circ \tilde{A}_1 , \quad \tilde{H}_1 \in \mathbb{C}^{KLM \times m}$$

$$[\tilde{A}_1]_{k,p} = a_1^{(k-1)} , \quad \tilde{A}_1 \in \mathbb{C}^{K \times m}$$

$$[\tilde{B}_1]_{l,p} = b_p^{(l-1)} , \quad \tilde{B}_1 \in \mathbb{C}^{L \times m}$$

$$[\tilde{C}_1]_{m,p} = c_p^{(m-1)} , \quad \tilde{C}_1 \in \mathbb{C}^{M \times m}$$

containing only the signal vectors that correspond to the specific generator $a_1$. The matrix partition $\tilde{H}_2 \in \mathbb{C}^{KLM \times (P-m)}$ is composed of the remaining signal vectors of $H$.

From the discussion in section 3.2 on the relaxation approach it is clear that the following proposition holds true:

**Theorem T4:** The nullspace of $M_1(a_1, E_S | "p")$ (with $a_1$ specified as above) is spanned by the columns of the $LM \times m$ matrix $\tilde{C}_1 \circ \tilde{B}_1$ with $\tilde{B}_1$ and $\tilde{C}_1$ given in (6.4) and (6.5), respectively.

**Proof of T4:** According to assumption A1 (full rank of $H$) and with $m \leq LM$ the matrix $\tilde{H}_1$ has full column rank $m$. Hence its rank is equal to the rank of the nullspace of $M_1(a_1, E_S | "p")$. Next we emphasize that $(I_{KLM} - E_S E_S^H)$ denotes the orthogonal projector onto the noise subspace that is orthogonal to the signal subspace spanned by the columns of $H$. Then we have

$$0 = T_a^T(a_1^{-1}) \left( I_{KLM} - E_S E_S^H \right) \tilde{H}_1$$

$$= T_a^T(a_1^{-1}) \left( I_{KLM} - E_S E_S^H \right) T_a(a_1) \left( \tilde{C}_1 \circ \tilde{B}_1 \right)$$

$$= M_1(a_1, E_S | "p") \left( \tilde{C}_1 \circ \tilde{B}_1 \right)$$

$$= \begin{bmatrix} \tilde{C}_1 \circ \tilde{B}_1 \end{bmatrix}$$
In other words, for a given generator $a_1$ of multiplicity $m$ the original 3D HRP, which consists of finding the signal matrix $\tilde{H}_1$ located in the nullspace of $\mathbf{I}_{KLM} - \mathbf{E}_S \mathbf{E}_S^H$, reduces to the 2D HRP of finding the 2D signal matrix $\left(\tilde{C}_1 \circ \tilde{B}_1\right)$ (containing the contributions of the $m$ signals corresponding to $a_1$) in the nullspace of

$$\mathbf{M}_1(a_1, \mathbf{E}_S | \text{“}p\text{”}) = \mathbf{I}_{LM} - K^{-1} \mathbf{T}_a^T(a_1^{-1}) \mathbf{E}_S \mathbf{E}_S^H \mathbf{T}_a(a_1),$$

(6.7)

where we made use of $\mathbf{T}_a^T(a_1^{-1}) \mathbf{T}_a(a_1) = K \mathbf{I}_{LM}$. Hence the dimensionality of the HRP is reduced by one. For solving the resulting 2D HRP we shall again use the MP of first kind. For solving the original 3D HRP, i.e., for finding the signal matrix of $a_1$, $b_1$, $\tilde{C}_1$ and $\tilde{B}_1$, we obtain the MP $\mathbf{M}_1(a_1, \mathbf{E}_S | \text{“}p\text{”})$. Hence with (6.7) we define the MP

$$\mathbf{M}_1(b, \left(\tilde{C}_1 \circ \tilde{B}_1\right) | \text{“}p\text{”}) = \mathbf{T}_b^T(b^{-1}) \mathcal{P}^\perp \{ \tilde{C}_1 \circ \tilde{B}_1 \} \mathbf{T}_b(b)$$

(6.8)

where

$$\mathcal{P}^\perp \{ \mathbf{U} \} = \mathbf{I}_K - \mathbf{U} (\mathbf{U}^H \mathbf{U})^{-1} \mathbf{U}^H$$

(6.9)

denotes the orthogonal projector onto the nullspace of an arbitrary but nonsingular $K \times P$ matrix $\mathbf{U}$, $\mathbf{T}_b(b) = (\mathbf{I}_M \otimes \mathbf{b})$, and $\mathbf{b} = [1, b_1, b_2, \ldots, b_\text{L-1}]^T$. According to corollary $C2$ the generators $b_1, \ldots, b_m$ are uniquely obtained as the roots of the MP in (6.8). Since the signal subspace spanned by the columns of $\tilde{C}_1 \circ \tilde{B}_1$ is not directly accessible from the data (or from the signal eigenvectors in $\mathbf{E}_S$ respectively) it can, based on theorem $T4$, be estimated from the nullspace eigenvectors of $\mathbf{M}_1(a_1, \mathbf{E}_S | \text{“}p\text{”})$. However, to keep computations low and to avoid an additional eigendecomposition step the idea is to replace the projector in (6.9) directly by the low-rank matrix in $\mathbf{M}_1(a_1, \mathbf{E}_S | \text{“}p\text{”})$. Hence with (6.7) we define the MP

$$\tilde{\mathbf{M}}_1(b, a_1, \mathbf{E}_S | \text{“}p\text{”}) =$$

$$\mathbf{T}_b^T(b^{-1}) \left(\mathbf{I}_{LM} - K^{-1} \mathbf{T}_a^T(a_1^{-1}) \mathbf{E}_S \mathbf{E}_S^H \mathbf{T}_a(a_1)\right) \mathbf{T}_b(b).$$

(6.10)

Note that in the ideal case the replacement of the projector $\mathcal{P}^\perp \{ \tilde{C}_1 \circ \tilde{B}_1 \}$ in the original MP of kind 1 by the low-rank matrix $\mathbf{M}_1(a_1, \mathbf{E}_S | \text{“}p\text{”})$ does not change any of the statements concerning the rank of the MP for the different values of $b$. In fact this is because $\mathbf{M}_1(a_1, \mathbf{E}_S | \text{“}p\text{”})$ drops rank only if (for a specific value of $b$ on the unit circle) there exists a linear combination of the columns of $\mathbf{T}_b(b)$ that is located in the nullspace of $\mathcal{P}^\perp \{ \tilde{C}_1 \circ \tilde{B}_1 \}$. In other words, only the span of the nullspace and not the unit scaling of the nonzero eigenvectors of the projection matrix are of interest for the rank properties of the matrix polynomial (6.10).

Following the same procedure described above we now assume without loss of generality that the pair $(a_1, b_1)$ denotes the true 2D harmonic of multiplicity $n \leq M$ in the set of true generator pairs $(a_1, b_1), (a_2, b_2), \ldots, (a_P, b_P)$. Inserting the solutions obtained along the $b$-axis back into the MP (6.10) we obtain the $M \times M$ matrix of rank $M - n$

$$\tilde{\mathbf{M}}_1(b_1, a_1, \mathbf{E}_S | \text{“}p\text{”}) = \mathbf{T}_b^T(b_1^{-1}) \mathbf{T}_a^T(a_1^{-1}) \mathcal{P}^\perp \{ \mathbf{C} \circ \mathbf{B} \circ \mathbf{A} \} \mathbf{T}_a(a_1) \mathbf{T}_b(b_1).$$

(6.11)
Multiplying the matrix in (6.11) from the left and the right with the Vandermonde vectors $e^H$ and $c = [1, c^1, \ldots, c^{M-1}]^T$, respectively, we arrive at the original 3D root-MUSIC function (2.75) for fixed values of $a = a_1$ and $b = b_1$ and variable $c$, given by

$$f_{1-M}(a_1, b_1, c) = e^H \hat{M}_1(b_1, a_1, E_S | "p")c$$

$$= e^HT_k(b_1^{-1})T_a^{-1}(a_1^{-1})P_{\perp} \{C \circ B \circ A\}T_a(a_1)T_k(b_1)c$$

$$= (c \otimes b \otimes a)^H P_{\perp} \{C \circ B \circ A\} (c \otimes b \otimes a)$$

$$= h^H P_{\perp} \{C \circ B \circ A\} h \quad (6.12)$$

which is known to yield zero function values only if the triplet $(a_1, b_1, c)$ contains the true generators on the unit circle. Evaluating (6.12) on the unit circle, hence replacing $e^*$ by $e^{-1}$, we obtain a 1D root-MUSIC polynomial. The $n$ roots of this polynomial yield the true generators associated with the pair $(a_1, b_1)$.

So far, only the estimation of 3D harmonics corresponding to the partition $\hat{H}_1$ (6.2) has been considered. In order to determine the complete set of 3D harmonics the 3-step dimensionality reduction scheme described above needs to be performed in a tree-structured fashion. This is illustrated in figure 6.1. At the first stage, which marks the root of the tree-structured algorithm, the $P$ signal roots of the original MP $M_1(a, E_S | "p")$ are computed. The multiplicity of each distinct signal root is determined. Different roots (of given multiplicity) open new branches of the tree. At the second stage the generators along the $b$-axis corresponding to each branch (i.e., each generator along the $a$-axis) are computed. For $a_p$ denoting the generator of multiplicity $m_p$ associated with the $p$th branch, the corresponding parameters along the $b$-axis are obtained as the $m_p$ signal roots of the MP $\hat{M}_1(b, a_p, E_S | "p")$ (6.10) located on the unit-circle. Again the multiplicity of each distinct signal root is determined and different roots give rise to new sub-branches of the tree. At the final stage of the algorithm, the generators along the $c$-axis corresponding to each sub-branch (i.e., the generator pairs along the $a$- and $b$-axis) are computed. For $(a_q, b_q)$ denoting the generator pair of multiplicity $n_q$ associated with the $q$th subbranch of the $p$th branch, the corresponding parameters along the $c$-axis are obtained as the $n_q$ signal roots of the 1D root-MUSIC polynomial $f_{1-M}(a_1, b_1, c)$ (6.12). In conclusion, the proposed algorithm yields automatically associated 3D harmonic estimates from consequent backsubstitution and successive dimensionality reduction.

To illustrate the tree-structured estimation scheme, let us consider the following representative example. Given the 3D undamped HRP with 6 generators characterized by the triplets $(a_1, b_1, c_1), (a_1, b_1, c_2), (a_1, b_2, c_3), (a_1, b_3, c_4), (a_2, b_4, c_1), \text{ and } (a_2, b_4, c_5)$, the algorithm is performed as depicted in figure 6.1. At the first stage we obtain the signal roots $a_1$ and $a_2$ with multiplicities 4 and 2, respectively, as the roots of the MP of kind 1 (3.6) formulated along the $a$-axis. At stage two of the first branch, created by $a_1$, the associated parameters along the $b$-axis are obtained from rooting the MP $\hat{M}_1(b, a_1, E_S | "p")$ (6.10). In this example, we obtain the
generators $b_1$ of multiplicity 2 as well as the simple generators $b_2$ and $b_3$. Thus two branches depart from the knot associated with the pair $(a_1, b_1)$ and single branches depart from each of the pairs $(a_1, b_2)$ and $(a_1, b_3)$. At stage three, the different pairs are inserted into the root-MUSIC polynomial (6.12) to obtain the corresponding estimate along the $c$-axis. Returning to the second stage and now considering the second branch, created by $a_2$, we observe from figure 6.1 that a single root is obtained from rooting the MP $\tilde{M}_1(b, a_1, E_S \mid \text{“p”})$ (6.8), such that only a single branch is originating from the knot associated with the pair $(a_2, b_4)$. At stage three, this pair is again inserted into the root-MUSIC polynomial (6.12) to obtain the corresponding triplets $(a_2, b_3, c_1)$, and $(a_2, b_4, c_5)$ from the roots located on the unit circle.

We observe from the description of the procedure and also from the preceding example that in the realistic case when the measurements are corrupted by additive noise the following difficulties arise.

1. **Determination of multiplicities**: Harmonics of higher multiplicity corresponding to different 3D harmonics which nominally, hence in the noise-free case, correspond to identical generators along one (or multiple) axis (axes), are in the noisy case displaced from their ideal position on the unit circle. However, the random perturbations of the polynomial coefficients cause distinct displacements of the various signal roots. The effect is that signal roots obtained from polynomial rooting in the realistic case are usually distinct even if they stem from generators which in the ideal case are identical. The difficulty arising in this context is to reliably estimate the multiplicity of the signal roots. Sophisticated root clustering procedures\(^1\) are required to accomplish this task. Recall that the multiplicity of the estimated signal roots is of great importance for further estimation of

\(^1\)Alternatively, to determine the multiplicity of a signal root obtained from a MP it is also possible to estimate the approximate dimension of the nullspace of the MP evaluated at the root. The nullspace dimension corresponds to the multiplicity of the root.
the generators along the remaining axes since the multiplicity determines the number of signals that are obtained along a certain branch. Underestimation of the multiplicity has the effect that two branches eventually yield parameter tuples corresponding to the same harmonic while a different harmonic may not be contained in the solution set.

2. **Critical error propagation:** In the proposed tree-structured algorithm solutions obtained at a given stage are fixed and exploited in estimating the remaining parameters. This property makes the algorithm sensitive to error propagation. Defective estimates obtained along a single dimension at an early stage of the algorithm, where the signal components are not yet well separated, can significantly degrade the estimation performance along all remaining array axes.

The problems reported above become more severe in the case of closely separated 3D harmonics. In the following sections we shall provide simple and robust tools to obtain properly associated MD harmonic estimates of pure and undamped harmonics.

### 6.2 Eigenvector approach

In this section a different approach towards 3D uniform HR is taken. Instead of successive dimensionality reduction, all generator sets along the various array axes are separately estimated from any of the square MPs of kind 2, 5, 6 or 7 (formulated for the array axis under consideration, see chapter 3 and appendix E). In a second step we exploit specific nullspace properties of these MPs to efficiently associate corresponding estimates. The association procedure is based on the following theorem.

**Theorem T5:** Provided that \((a_p, b_p, c_p)\) characterizes a true 3D harmonic along the \(a\)-axis, \(b\)-axis, and \(c\)-axis, then the matrix polynomials \(M_5(a, E_S | \text{“d”})\), \(M_5(b, E_S | \text{“d”})\) and \(M_5(c, E_S | \text{“d”})\) evaluated at the true generators \(a_p, b_p\) and \(c_p\), respectively, share a common right nullspace vector. This vector is given by \(k_p\) and identical to the \(p\)th column of the full-rank mixing matrix \(K\) defined in (2.27).

**Proof of T5:** The proof follows immediately from (3.30), where the \(P \times 1\) vector \(k_p\) representing the \(p\)th column of the mixing matrix \(K\) lies in the right nullspace of \(M_5(a_p, E_S | \text{“d”})\) for \(a_p\) denoting the generator along the \(a\)-axis of the \(p\)th harmonic. Due to the symmetry of the MD uniform HRP problem with respect to the sampling axes, and making use of the row permutation methodology introduced in chapter 4, the same statement can be made about the MPs along the remaining axes. For \(b_p\) and \(c_p\) denoting the true generators of the \(p\)th signal observed along the
From the relations between the MP of kind 5 with the MP of kind 2 reported in (3.47), it is simple to show that $T5$ also extends to the square MPs of kind 2. For the true generators $a_p$, $b_p$ and $c_p$ associated with the $p$th harmonic and with $k_p$ as defined above we obtain

$$M_2(a_p, E_S | \text{“p”}) k_p = K^{-1} M_5^H(a_p^{-1}, E_S | \text{“p”}) M_5(a_p, E_S | \text{“d”}) k_p = 0 \quad (6.15)$$

$$M_2(b_p, E_S | \text{“p”}) k_p = L^{-1} M_5^H(b_p^{-1}, E_S | \text{“p”}) M_5(b_p, E_S | \text{“d”}) k_p = 0 \quad (6.16)$$

$$M_2(c_p, E_S | \text{“p”}) k_p = M^{-1} M_5^H(c_p^{-1}, E_S | \text{“p”}) M_5(c_p, E_S | \text{“d”}) k_p = 0 \quad . \quad (6.17)$$

From $T5$ and its extension to the square MP of kind 2 we deduce the following corollary.

**Corollary C4:** Let $\{a_1, \ldots, a_P\}$, $\{b_1, \ldots, b_P\}$, and $\{c_1, \ldots, c_P\}$ be the unsorted (or mutually un-associated) sets of signal roots obtained from the MP of kind 2 along the first, second and third sampling axis, respectively. Then the convex linear combination of MPs in $a$, $b$, and $c$ given by

$$\bar{M}_2(a, b, c) =$$

$$\kappa_1 M_2(a, E_S | \text{“p”}) + \kappa_2 M_2(b, E_S | \text{“p”}) + \kappa_3 M_2(c, E_S | \text{“p”}) \quad (6.18)$$

with $\kappa_i > 0 \in \mathbb{R}$, for $i = 1, \ldots, 3$ becomes singular if and only if the triplet $(a, b, c)$ with $a \in \{a_1, \ldots, a_P\}$, $b \in \{b_1, \ldots, b_P\}$, and $c \in \{c_1, \ldots, c_P\}$ represents the parameters of a true 3D harmonic.

**Proof of C4:** With $\bar{M}_2(a, b, c)$ representing a quadratic form for $a$, $b$, and $c$ located on the unit circle and from the rank properties of the MPs of kind 2 for the true generators it follows that the non-zero eigenvalues of $M_2(a, E_S | \text{“p”})$ are positive real, hence $M_2(a, E_S | \text{“p”})$ is positive semi-definite. The same statement holds true for $M_2(b, E_S | \text{“p”})$ and $M_2(c, E_S | \text{“p”})$. With positive semidefinite MPs on the right hand side of (6.18) and with strictly positive linear coefficients the nullspace of the matrix $\bar{M}_2(a, b, c)$ is spanned by the intersection of the three nullspaces, namely

$$\mathcal{N}\{\bar{M}_2(a, b, c)\} =$$

$$= \mathcal{N}\{M_2(a, E_S | \text{“p”})\} \cap \mathcal{N}\{M_2(b, E_S | \text{“p”})\} \cap \mathcal{N}\{M_2(c, E_S | \text{“p”})\} \quad . \quad (6.19)$$

From (6.15) and the rank properties of the MPs of kind 2 for the true generators we know that the nullspace of $M_2(a_p, E_S | \text{“p”})$ is spanned only by those columns of the mixing matrix $K$
(2.27) that are associated with the generator \( a_p \). The same statement holds true for the nullspace of \( \hat{M}_2(b_p, E_S \mid \text{“p”}) \) and \( \hat{M}_2(c_p, E_S \mid \text{“p”}) \). Assuming that the true 3D harmonics are separable along at least one dimension it follows that the intersection of the nullspaces corresponding to \( \hat{M}_2(a_p, E_S \mid \text{“p”}) \), \( \hat{M}_2(b_p, E_S \mid \text{“p”}) \), and \( \hat{M}_2(c_p, E_S \mid \text{“p”}) \) is given by the vector \( \hat{k}_p \) if and only if \((a_p, b_p, c_p)\) is a true generator and the nullspaces do not intersect otherwise. This completes the proof.

With corollary C4 we propose the following simple and powerful parameter association scheme for the parameter estimates separately obtained along the three sampling axes in the realistic case where the measurements are corrupted by additive noise [PMB04]. Let \( A = \{\hat{a}_1, \ldots, \hat{a}_P\} \), \( B = \{\hat{b}_1, \ldots, \hat{b}_P\} \), and \( C = \{\hat{c}_1, \ldots, \hat{c}_P\} \) be the sets of un-associated signal roots obtained from the MPs of kind 2 along the first, second and third sampling axis, in (5.18), (F.3) and (F.4) respectively. Then for a specific harmonic \( \hat{a}_p \) of the first set, the corresponding harmonics \( \hat{b}_q \) of the second set and \( \hat{c}_r \) of the third set are given by the elements of \( \{\hat{b}_1, \ldots, \hat{b}_P\} \), and \( \{\hat{c}_1, \ldots, \hat{c}_P\} \) that minimize the cost function

\[
F_{\text{assoc}}(p, q, r) = \\
= \lambda_{\min}\{\hat{M}_2(\hat{a}_p, \hat{b}_q, \hat{c}_r)\} \\
= \lambda_{\min}\{\kappa_1\hat{M}_2(\hat{a}_p, \hat{E}_S \mid \text{“p”}) + \kappa_2\hat{M}_2(\hat{b}_q, \hat{E}_S \mid \text{“p”}) + \kappa_3\hat{M}_2(\hat{c}_r, \hat{E}_S \mid \text{“p”})\}
\]

(6.20)

for appropriately chosen \( \kappa_1, \kappa_2, \kappa_3 > 0 \). Here \( \lambda_{\min}\{\hat{M}_2(\hat{a}_p, \hat{b}_q, \hat{c}_r)\} \) denotes the smallest eigenvalue of \( \hat{M}_2(\hat{a}_p, \hat{b}_q, \hat{c}_r) \).

In practice the parameter association scheme resulting from (6.20) consists of evaluating the cost function \( F_{\text{assoc}}(p, q, r) \) for all triplets \( (p, q, r) \) from integer set \( \{1, \ldots, P\} \times \{1, \ldots, P\} \times \{1, \ldots, P\} \). Given the estimated signal eigenvectors in \( \hat{E}_S \) from (2.22) or (2.47) the MD RARE algorithm consists of the following steps [PMB04].

**Step 1a:** If \( P \leq L'M \) then compute the \( P \) largest (signal) roots \( \hat{a}_1, \ldots, \hat{a}_P \) inside the unit circle (in terms of magnitude) of the MP \( \hat{M}_2(a, \hat{E}_S \mid \text{“p”}) \) (5.18) using one of the techniques given in chapter 5 and assign them to the set \( A \). Otherwise, compute the roots of the MP \( \hat{M}_1(a, \hat{E}_S \mid \text{“p”}) \) (5.17) and assign the largest (signal) roots \( \hat{a}_1, \ldots, \hat{a}_P \) inside the unit circle to the set \( A \).

**Step 1b:** If \( P \leq K'M \) then compute the \( P \) largest (signal) roots \( \hat{b}_1, \ldots, \hat{b}_P \) inside the unit circle (in terms of magnitude) of the MP \( \hat{M}_2(b, \hat{E}_S \mid \text{“p”}) \) (F.3) using one of the techniques given in chapter 5 and assign them to the set \( B \). Otherwise, compute the roots of the MP \( \hat{M}_1(b, \hat{E}_S \mid \text{“p”}) \) (F.1) and assign the largest (signal) roots \( \hat{b}_1, \ldots, \hat{b}_P \) inside the unit circle to the set \( B \).

\[2\]We say that the \( p \)th column of \( K \) is associated with a generator \( a_p \) if \( a_p \) is a generator of the \( p \)th 3D harmonic.
**Step 1c:** If \( P \leq KL' \) then compute the \( P \) largest (signal) roots \( \hat{c}_1, \ldots, \hat{c}_P \) inside the unit circle (in terms of magnitude) of the MP \( \hat{M}_2(c, \hat{E}_S | \text{“p”}) \) (F.4) using one of the techniques given in chapter 5 and assign them to the set \( C \). Otherwise, compute the roots of the MP \( \hat{M}_1(c\hat{E}_S | \text{“p”}) \) (F.2) and assign the largest (signal) roots \( \hat{c}_1, \ldots, \hat{c}_P \) inside the unit circle to the set \( C \).

**Step 2:** Select the generator of the sets \( A, B, \) or \( C \) that is best-separated in terms of its minimum angular distance to the remaining generators in the set. Without loss of generality we shall assume that this generator is given by \( a_p \) from set \( A \).

**Step 3:** Find the corresponding roots \( \hat{b}_q \) and \( \hat{c}_r \) for \( q, r = 1, \ldots, P \) that minimize the cost function (6.20).

**Step 4:** Store the generator triplet \((\hat{a}_p, \hat{b}_q, \hat{c}_r)\) as the \( n \)th estimate of the 3D harmonic.

**Step 5:** Remove the generators \( \hat{a}_p, \hat{b}_q, \) and \( \hat{c}_r \) from the sets \( A, B, \) and \( C, \) respectively.

**Step 6:** Repeat Steps 2 – 5 until all \( P \) harmonics are determined.

Note that in Step 2 we select the generator from the sets \( A, B, \) or \( C \) that is best-separated in terms of its minimum angular distance to the remaining generators in the set. This is to ensure best performance of the proposed association scheme in low SNR scenarios because in practice usually well-separated signal roots are estimated with higher precision than close roots. Thus well-separated signal roots shall be associated and removed from the set prior to the remaining roots in the set.

The association procedure described above has essential advantages over the MD Tree-RARE estimator described in the previous section. First of all in this algorithm it is not required to determine the multiplicity of the generators in the parameter sets, and second, the parameters along the three sample axes are obtained from separate MPs without backsubstitution of known estimates. Thus the algorithm does not suffer from error-propagation effects like in the tree-structured algorithm. However, the computational cost associated with the evaluation of (6.20), namely the computation of the smallest eigenvalue of a \( P \times P \) matrix for all combinations of parameter sets along the various dimensions, is considerably high. In the next section we shall derive a method that retains the benefits of the association algorithm proposed in this section at a significantly reduced computational load.

---

\(^3\)If a generator in set \( B, \) or \( C \) has a larger minimum angular distance to the remaining generators then this case amounts to consistent renaming of variables and sets.
6.3 Generalized eigendecomposition approach

The approach presented in this section is based on similar MP nullspace properties as implied by corollary C4. We consider the uniform 3D damped or undamped HR problem based on the MP of kind 6. From theorem T3 we have learned that the generators along the $a$-axis are obtained as the signal roots of $M_6(a, E_S \mid \text{“d”})$ located inside or on the unit-circle. These roots can efficiently be computed based on the BCM approach in section 5.1.2. In the following we shall deduce some useful properties of the GEV of the BCM associated with the MP of kind 6. Linearizing the $P \times P$ MP $M_6(a, E_S \mid \text{“d”})$ of degree $K-1$ according to (5.8) by inserting the polynomial coefficients into (5.9) and (5.10), we obtain

$$\mathcal{L}\{M_6(a, E_S \mid \text{“d”})\} = \mathcal{V}\{M_6(a, E_S \mid \text{“d”})\} - a^T \{M_6(a, E_S \mid \text{“d”})\} \quad (6.21)$$

where

$$\mathcal{V}\{M_6(a, E_S \mid \text{“d”})\} = \begin{bmatrix} 0 & I_p & 0 & \cdots & 0 \\ 0 & 0 & I_p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & I_p \\ - \sum_{k=1}^{K-1} E_{S,a,k}^H E_{S,a,k} E_{S,a,1}^H E_{S,a,1} \cdots E_{S,a,K-2}^H E_{S,a,K-2} & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \quad (6.22)$$

and

$$T \{M_6(a, E_S \mid \text{“d”})\} = \begin{bmatrix} I_p & 0 & \cdots & 0 & 0 \\ 0 & I_p & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_p & 0 \\ 0 & 0 & \cdots & 0 & E_{S,a,K-1}^H E_{S,a,K-1} \end{bmatrix} \quad (6.23)$$

Let $a_p$ denote a true signal generator of multiplicity $M_p$ in the generator set. Then we know from theorem T3 and the discussion in section 5.1.2 that $a_p$ is one of the $P$ principle generalized eigenvalues of the BCM pair $\mathcal{V}\{M_6(a, E_S \mid \text{“d”})\}$ and $T \{M_6(a, E_S \mid \text{“d”})\}$ defined above. Let $K_p$ denote the partition of the mixing matrix $K$ (2.27) that contains all vectors associated with the generator $a_p$, then according to C6 and T3 (3.32) this matrix spans the nullspace of $M_6(a, E_S \mid \text{“d”})$, i.e.

$$M_6(a_p, E_S \mid \text{“d”}) K_p = M_5^T(a, E_S \mid \text{“d”}) \big|_{a=0} M_5(a_p, E_S \mid \text{“d”}) K_p = 0. \quad (6.24)$$
Making use of (5.12), the \( M_p \) GEVs of the BCM pair

\[
(V\{ M_6(a, E_S \mid \text{“d”}) \}, T\{ M_6(a, E_S \mid \text{“d”}) \})
\]

form the \((K - 1)P \times M_p\) matrix

\[
V_{a,p} = \begin{bmatrix}
K_p \\
ap_pK_p \\
\vdots \\
ap_p^{K-2}K_p
\end{bmatrix}.
\]

Hence, taking into account all signal generators along the \( a \)-axis, i.e. the \( P \) principal generalized eigenvalues, then with the considerations above the corresponding characteristic equation reads

\[
V\{ M_6(a, E_S \mid \text{“d”}) \} V_a = T\{ M_6(a, E_S \mid \text{“d”}) \} V_a \Delta_a
\]

where the diagonal matrix \( \Delta_a \) defined in (2.64) contains the \( P \) principal generalized eigenvalues of the BCM pair on its main diagonal, and the \((K - 1)P \times P\) matrix

\[
V_a = \begin{bmatrix}
K \\
K \Delta_a \\
\vdots \\
K \Delta_a^{K-2}
\end{bmatrix}
\]

is formed from the corresponding GEVs [PSBG05]. If we partition the GEV matrix into \((K - 1)\) submatrices of identical dimensions and denote them as

\[
V_a = [v_{a,0}^T, v_{a,1}^T, \ldots, v_{a,K-2}^T]^T
= (K - 1)^{-1} [K^T, \Delta_a K^T, \ldots, \Delta_a^{K-2} K^T]^T
\]

and if we define

\[
K_a = \sum_{k=0}^{K-2} v_{a,k} \Delta_a^{-k},
\]

then, in the ideal case, it is simple to check that \( K_a = K \). Thus in absence of noise the sum of the partitions of the GEV in (6.30) is equal to the linear transformation matrix relating the signal matrix with the signal eigenvectors. Note, however, that depending on the definition of the GEV here it is clear that equality holds up to permutation and complex scaling of the columns.

Following the considerations above but now for the MPs of kind 6 in parameters \( b \) and \( c \), given by

\[
M_6(b, E_S \mid \text{“d”}) \quad (E.11) \quad \text{and} \quad M_6(c, E_S \mid \text{“d”}) \quad (E.12),
\]

respectively, then the characteristic equations corresponding to (6.27) read

\[
V\{ M_6(b, E_S \mid \text{“d”}) \} V_b = T\{ M_6(b, E_S \mid \text{“d”}) \} V_b \Delta_b
\]

\[
V\{ M_6(c, E_S \mid \text{“d”}) \} V_c = T\{ M_6(c, E_S \mid \text{“d”}) \} V_c \Delta_c
\]
6.3 Generalized eigendecomposition approach

with \( \Delta_b \) and \( \Delta_c \) defined in (E.27) and (E.28) containing the \( P \) principal generalized eigenvalues of the BCM pairs in (6.31) and (6.32) and the associate GEV contained in

\[
V_b = \left[ \bar{V}_{b,0}^T, \bar{V}_{b,1}^T, \ldots, \bar{V}_{b,L-2}^T \right]^T \in \mathbb{C}^{(L-1)P \times P}
\]

\[
V_c = \left[ \bar{V}_{c,0}^T, \bar{V}_{c,1}^T, \ldots, \bar{V}_{c,M-2}^T \right]^T \in \mathbb{C}^{(M-1)P \times P}
\]

respectively. According to (6.30) we define the \( P \times P \) matrices

\[
K_b = \sum_{l=0}^{L-2} \bar{V}_{b,l} \Delta_b^{-l}, \quad (6.35)
\]

\[
K_c = \sum_{m=0}^{M-2} \bar{V}_{c,m} \Delta_c^{-m}, \quad (6.36)
\]

which in the ideal case should be (up to permutation and complex scaling of the columns) equal to the true mixing matrix. That is

\[
K_a = K_b = K_c = K. \quad (6.37)
\]

6.3.1 Root-MI-ESPRIT

The property expressed in equation (6.37) in combination with definitions (6.30), (6.35), and (6.36) provides the mean by which we shall address in the following the parameter association problem of the generators separately obtained along the three dimensions.

If the diagonal elements of \( \Delta_a, \Delta_b \) and \( \Delta_c \) containing the true generators along the array axis have the correct association, then \( K_a = K_b = K_c = K \), otherwise \( K_a, K_b \) and \( K_c \) are column-wise permutations of each other.

In the later case, the permutation of the parameters in the three generator sets can be obtained from a correlation analysis of the individual columns in \( K_a, K_b, \) and \( K_c \). Specifically, if we set the elements with the maximum absolute value (corresponding to maximum correlation between the columns) in each particular row of the product \( \Gamma_{a,b} = K_a^H K_b \) equal to one and the remaining elements equal to zero then we obtain the permutation matrix relating the elements in \( A \) with the elements in \( B \). The row and column indices of the non-zero elements of the permutation matrix show, respectively, which columns of \( K_a \) and \( K_b \) should be paired. Similarly, we obtain a second permutation matrix relating the elements in \( A \) with the elements in \( C \) if we perform the same procedure on the product \( \Gamma_{a,c} = K_a^H K_c \).
In the realistic case, when noise effects are present, then the root-MI-ESPRIT algorithm consists of the following steps, given the finite sample estimates of the MPs of kind 6 along the $a$-, $b$- and $c$-axis defined in (3.32), (F.5), and (F.6).

**Step 1:** Find the principal $P$ GEV matrices $\hat{V}_a$, $\hat{V}_b$ and $\hat{V}_c$ and the corresponding eigenvalues on the main diagonal of $\hat{\Delta}_a$, $\hat{\Delta}_b$ and $\hat{\Delta}_c$, respectively, from the characteristic equations

\[
\begin{align*}
\mathcal{V}\{\hat{M}_6(a, \hat{E}_S | "d")\} \hat{V}_a &= T \{\hat{M}_6(a, \hat{E}_S | "d")\} \hat{V}_a \hat{\Delta}_a, \\
\mathcal{V}\{\hat{M}_6(b, \hat{E}_S | "d")\} \hat{V}_b &= T \{\hat{M}_6(b, \hat{E}_S | "d")\} \hat{V}_b \hat{\Delta}_b, \\
\mathcal{V}\{\hat{M}_6(c, \hat{E}_S | "d")\} \hat{V}_c &= T \{\hat{M}_6(c, \hat{E}_S | "d")\} \hat{V}_c \hat{\Delta}_c.
\end{align*}
\]

**Step 2:** Partition the GEV matrices according to (6.29), (6.33) and (6.34) in submatrices $\hat{V}_{a,k}$, $\hat{V}_{b,l}$, and $\hat{V}_{c,m}$ of size $P \times P$ for $k = 0, \ldots, K-2$, $l = 0, \ldots, L-2$, and $m = 0, \ldots, M-2$, respectively. Compute the matrices

\[
\begin{align*}
\hat{K}_b &= \sum_{k=0}^{K-2} \hat{V}_{a,k} \hat{\Delta}_b^{-l}, \\
\hat{K}_b &= \sum_{l=0}^{L-2} \hat{V}_{b,l} \hat{\Delta}_b^{-l}, \\
\hat{K}_c &= \sum_{m=0}^{M-2} \hat{V}_{c,m} \hat{\Delta}_c^{-m}.
\end{align*}
\]

**Step 3:** Compute the products $\hat{\Gamma}_{a,b} = \hat{K}_a^H \hat{K}_b$ and $\hat{\Gamma}_{a,c} = \hat{K}_a^H \hat{K}_c$.

**Step 4a:** Find the element in $\hat{\Gamma}_{a,b}$ with maximum magnitude.

**Step 5a:** The row and column indices of this element show which element on the main diagonal of $\hat{\Delta}_a$ and which element on the main diagonal of $\hat{\Delta}_b$ form a true pair. Store this pair of generators and set all elements in $\hat{\Gamma}_{a,b}$ corresponding to the same row indices and all elements corresponding to the same column indices equal to zero.

**Step 6a:** Repeat Steps 4a-5a until $\hat{\Gamma}_{a,b}$ contains only zero entries.

**Step 4b:** Find the element in $\hat{\Gamma}_{a,c}$ with maximum magnitude.

**Step 5b:** The row and column indices of this element show which element on the main diagonal of $\hat{\Delta}_a$ and which element on the main diagonal of $\hat{\Delta}_c$ form a true pair. Store this pair of generators and set the entries all elements in $\hat{\Gamma}_{a,c}$ corresponding to the same row indices and all elements corresponding to the same column indices equal to zero.

**Step 6b:** Repeat Steps 4b-5b until $\hat{\Gamma}_{a,c}$ contains only zero entries.
The procedure described above yields sets of estimates \(\{(\hat{a}_p, \hat{b}_p)\}_{p=1}^{P}\) and \(\{(\hat{a}_p, \hat{c}_p)\}_{p=1}^{P}\) of the true parameter pairs observed along first and second array axis, and first and third array axis, respectively. The elements in both sets, i.e. the pairs, are easily associated according to their first element, the \(a\)-axis parameters.

The parameter association scheme proposed in this section is computationally efficient and robust against noise and finite sample effects. Instead of expensive evaluation of the cost function (6.20) for all permutations of elements in the three generator sets, here the parameter association is obtained almost as a byproduct of the parameter estimation procedure (which at the same time contains the computationally most efficient procedure, the computation of the signal roots of interest, see section 5.1.2 for details). Unlike in the previous approach here it is always possible to use pairwise association of the parameters in the three generator sets, which further reduces the computation complexity of the association scheme. In conclusion with the proposed procedure the parameter association problem becomes, from a computation point a view, a negligible issue. Hence if we consider only the estimation of the harmonics along a single dimension to roughly estimate the computational complexity of root-MI-ESPRIT compared to the complexity of the conventional single invariance ESPRIT algorithm we obtain the following result. Provided that estimates of the signal subspace eigenvectors are given, then the single invariance ESPRIT algorithm requires to solve an \(P \times P\) eigenvalue problem which approximately requires \(O\{P^2\}\) operations (for each update of the iteration of the eigendecomposition algorithm). In the root-MI-ESPRIT algorithm the \(P\) smallest eigenvalues of a sparse generalized eigenproblem of size \(P(K-1) \times P(K-1)\) need to be computed. In an Arnoldi-type algorithm roughly \(O\{P^2(K-1)\}\) operations are required (for each update of the iteration of the Arnoldi-Modified Gram-Schmidt method) [LSY98, Saa00]. Therefore the computational complexity of root-MI-ESPRIT is increased approximately by the factor \((K-1)\) compared to the computational complexity of the 1D single invariance ESPRIT algorithm. In other words, both algorithms require a comparable number of operations.

### 6.3.2 Joint root-MI-ESPRIT

In this subsection we propose a slightly modified 3D harmonic estimation method. Recall that the MI polynomials of kind 6 introduced in section 3.4 stem back from the MI equations along the various axes. Apart from the numerous advantages that the rooting based solution of the MI equations has over the joint diagonalization approaches one drawback is that in some applications it might be desirable to obtain 3D estimates that are jointly obtained from the MPs along the various axes instead of estimates that are separately estimated along the various dimensions. This is for example the case when, due to differences in the sample support that is available along the various axes or due to close separation of the generators along a specific
array axis, the estimation performance obtained along a single axis is significantly lower than
the estimation performance obtained along the remaining axes. The idea that we propose in this
subsection is to exploit the property (6.37) to fulfill a twofold task: a) to determine associated
columns in the matrices $K_a$, $K_b$, and $K_c$ to solve the permutation problem, and b) to compute
a joint estimate of the mixing matrix $K$ (2.27). The joint estimate of $K$ then allows us to
trace back the true signal matrix $H$ through the subspace relation $H = E_S K$ from which the
true generators are readily obtained. The joint root-MI-ESPRIT algorithm is performed in the
following steps.

**Step 1:** Perform the Steps 1-6b of the root-MI-ESPRIT algorithm given in section 6.3.1.

**Step 2:** According to the 3D harmonic estimates obtained in the previous step permute
the columns of associated matrices $\hat{K}_a$, $\hat{K}_b$, and $\hat{K}_c$ such that afterwards corresponding
columns have the same column indices in all three matrices.

**Step 3:** Compute a joint estimate of the mixing matrix, for example\(^4\) as

$$\hat{K} = \frac{(K - 1)^3 \hat{K}_a + (L - 1)^3 \hat{K}_b + (M - 1)^3 \hat{K}_c}{(K - 1)^3 + (L - 1)^3 + (M - 1)^3}. \quad (6.44)$$

**Step 4:** Estimate the signal matrix as $\hat{H} = \hat{E}_S \hat{K}$.

**Step 5:** Form the row-reduced versions of the estimated signal matrix in Step 4 as

$$\overline{H}_{a,1} = (I_{LM} \otimes J_{K,1}) \hat{H} \quad (6.45)$$
$$\overline{H}_{b,1} = (I_{LM} \otimes J_{K,1}) \hat{H} \quad (6.46)$$
$$\overline{H}_{b,1} = (I_{LM} \otimes J_{K,1}) \hat{H} \quad (6.47)$$
$$\overline{H}_{c,1} = (I_{LM} \otimes J_{K,1}) \hat{H} \quad (6.48)$$
$$\overline{H}_{c,1} = (I_{LM} \otimes J_{K,1}) \hat{H} \quad (6.49)$$
$$\overline{H}_{c,1} = (I_{LM} \otimes J_{K,1}) \hat{H} \quad (6.50)$$

**Step 6:** Compute the following row vectors

$$\hat{\phi}_a = (K - 2)^{-1} 1_{K-2,1}^T (\overline{H}_{a,1} \odot \overline{H}_{a,1}) \quad (6.51)$$
$$\hat{\phi}_b = (L - 2)^{-1} 1_{L-2,1}^T (\overline{H}_{b,1} \odot \overline{H}_{b,1}) \quad (6.52)$$
$$\hat{\phi}_c = (M - 2)^{-1} 1_{M-2,1}^T (\overline{H}_{c,1} \odot \overline{H}_{c,1}) \quad (6.53)$$

where $1_{k,1}$ denotes the $k \times 1$ vector composed of ones in all entries.

**Step 7:** Corresponding estimates of the 3D generators along the $a$-, $b$-, and $c$-axis are
stored at corresponding positions in the vectors $\hat{\phi}_a$, $\hat{\phi}_b$, and $\hat{\phi}_c$, respectively.

\(^4\)Different scalings of the matrices in the sum of $\hat{K}$ can be used.
6.4 Non-uniform sampling case

Here, “⊙” stands for Hadamard-product (A.1), hence element-wise multiplication. Different estimation procedures to extract the signal parameters from the estimate of the signal matrix as proposed in Steps 5-7 are possible. The procedure presented here can also be found in [MSPM04] in context of the 3-MDF algorithm.

6.4 Non-uniform sampling case

To this point we have only considered the highly structured case of uniform sampling along all array axes. In this section an MD algorithm that applies to the “hybrid” case of non-uniform sampling along one or multiple axes and uniform samples along at least a single dimension. The uniform sampling axes are then used to estimate the mixing matrix \( K \). This approach is closely related to the procedure proposed in the preceding section. The major advantage accomplished by this procedure compared to the algorithm in section 4.2 is that an expensive spectral search can be omitted [SSJ01, SG04].

Consider the case of partly uniform HRP at the example 3D (damped or undamped) HRP obtained from uniform sampling with sample support \( K \) and \( M \) along \( a \)- and \( c \)-axis respectively, and non-uniform sampling with sample support \( L' \) along the \( b \)-axis. The estimation problem thus consists of determining the signal matrix of the form

\[
H = C \odot B \odot A
\]  

(6.54)

from the signal eigenvectors in \( E_S \) with Vandermonde matrices \( A \) and \( C \) defined in (2.7) and (2.51), respectively, and matrix \( B \) of known or unknown arbitrary structure. The estimation problem consist in determining the generators along the \( a \)- and \( c \)-axis in the signal matrix \( B \).

The general MD HRP in the “hybrid” case of uniform and non-uniform sampling can simply be deduced from this example. We propose the hybrid MI-ESPRIT algorithm consisting in the following steps.

**Step 1:** Estimate the generators along the \( a \)- and \( c \)-axis from Steps 1-3 and Steps 4b-6b of the root MI-ESPRIT algorithm given in section 6.3.1.

**Step 2:** According to the harmonic estimates obtained in the previous step permute the columns of the associated matrices \( \hat{K}_a \) and \( \hat{K}_c \) such that afterwards corresponding columns have the same column indices in both matrices.

**Step 3:** Compute a joint estimate of the mixing matrix, for example\(^5\) as

\[
\hat{K} = \frac{(K - 1)^3 \hat{K}_a + (M - 1)^3 \hat{K}_c}{(K - 1)^3 + (M - 1)^3}.
\]  

(6.55)

\(^5\)Different scalings of the matrices in the sum of \( \hat{K} \) can be used.
**Step 4:** Estimate the signal matrix as $\hat{H} = \hat{E}_S \hat{K}$.

**Step 5:** Estimate the generators along the $a$- and $c$-axis from the estimated signal matrix $\hat{H}$ as in Steps 5-7 of the joint root-MI-ESPRIT algorithm described in section 6.3.2.

**Step 6:** Compute the estimated signal matrix $\hat{B}$ according to (4.4) as the solution of

$$
\hat{H}_b = Q_b \hat{H} = \hat{A} \circ \hat{C} \circ \hat{B}.
$$

Interestingly, the algorithm presented above does not require any spectral search along the non-uniform sample directions because the signals are separated only according to the generators (and mixing matrices) obtained from the MPs along the uniform sample axis. It is obvious that the algorithm therefore ignores the MI relations expressed in (2.70) with respect to the generators $b_1, \ldots, b_P$. It is however important to note that in estimating the mixing matrix $K$ all samples taken along the three sample dimensions are incorporated. Thus only some structural prior information on the used non-uniform sampling scheme along the second axis is ignored. Note that improved estimates of the parameters along the $b$-axis (and also along the $a$- and $c$-axis) can for example be obtained if we use the estimates obtained from this method to initialize the spectral search methods described in sections 4.2 and 2.4.
7 Simulation results

In this chapter the estimation performance of the proposed methods is investigated using simulation results obtained from both synthetic data and real measurements. In simulations carried out with synthetic data, the described estimation procedures are tested under the ideal case that all model assumptions are perfectly satisfied. In such experiments the true parameters are known from design. This knowledge is used in measuring the accuracy of the estimators and to compute the corresponding asymptotical accuracy bounds which are conventionally used as a reference. Simulation results with real measurement data are inevitable not only to test the validity of the proposed signal model in real world applications but also to investigate the robustness of the proposed algorithms to existing model mismatches.

7.1 Synthetic data

We simulate several algorithms for the 2D and 3D case, which are a) the 2D and 3D root-MI-ESPRIT algorithm listed in section 6.3.1, b) the 2D and 3D RARE algorithm described in section 6.2, c) the tree-MD-RARE algorithm in section 6.1, d) the joint root-MI-ESPRIT algorithm in section 6.3.2, e) the 2D SPEC-MI-ESPRIT and 2D SPEC-RARE algorithms in section 4.2 and [SG04], f) the 3D hybrid MI-ESPRIT algorithm presented in section 6.4, g) the 2D ESPRIT algorithm in [ZHM96], h) the 2D and 3D unitary-ESPRIT algorithm [HN98], i) the MI-ESPRIT algorithm in [SORK92], j) the MI-MODE algorithm given in [SSJ01], k) the Trilinear Alternating Least Squares (TALS) algorithm [SBG00], l) the 3D MD Embedding (3D-MDE) algorithm in [SLS01], and m) the MD Folding (3D-MDF) algorithm in [MSPM04].

For later reference we briefly review the main features of the existing algorithms that are used for comparison in the simulations. It is important to note that the diverse algorithms exploit distinct prior-information and use slightly different model assumption. The 2D ESPRIT algorithm [ZHM96] solves a real-valued version of the single invariance equation jointly along first and second sampling axis through eigendecomposition of a complex matrix. Hence a common matrix of eigenvectors is sought, that approximately solves the single invariance equation along $a$ and $b$-axis. The popular MD unitary ESPRIT algorithm consists of jointly solving the set of single invariance equations taken along $a$ dimensions. This is accomplished by a joint Schur-decomposition algorithm for multiple real-valued non-symmetric matrices [ZHM96]. Hence in both the 2D ESPRIT and the MD unitary-ESPRIT algorithm only a single invariance is considered. The 2D and 3D MI-ESPRIT algorithm solve the MI equations along the $a$-axis simultaneously using Gauss-Newton iteration to minimize the corresponding joint cost function.
This method requires good initial estimates which in our simulations were obtained from the 1D ESPRIT algorithm performed only along the $a$-axis. The MI-MODE algorithm only estimates the parameters along the first array axis using a rooting-based subspace fitting approach. This algorithm is particularly interesting for comparison because it uses the same model assumptions as the new rank reduction algorithms proposed in this work and is further known to be asymptotically equivalent to the ML estimator [SSJ01] in this case. More specific the algorithm requires uniform sampling along the $a$-axis and similar to the rank reduction algorithms for estimating the harmonics along the $a$-axis, makes no assumptions on the sampling scheme along the remaining axes. Hence for estimating the $a$-axis parameters decoupled from the remaining parameters the same optimality bounds that apply for MI-MODE also apply to our new algorithms. The TALS algorithm contains an alternating LS estimation procedure to solve the MI equations jointly along all sampling axes. Similar to the unitary-ESPRIT algorithm this method also requires good initial estimates which in our simulations were obtained from the 1D ESPRIT algorithm performed only along the $a$-axis. The 3D-MDF and 3D-MDE algorithm both rely on a single invariance and compute the parameters from singular-value decomposition. It is important to note, as discussed previously in section 3.4, that all existing ESPRIT algorithm except the one presented in this thesis only exploit the fact that the MI equations share a common eigenvector matrix. However, the specific relation between the corresponding diagonal matrices of eigenvalues is ignored.

Further we stress that in all simulations for computing the parameter estimation errors the estimates are assigned to the corresponding true parameters according to the frequencies along the first sampling axis.

**Example 1**

Consider the pure uniform 2D HRP with 2 equi-powered pure harmonics characterized by the pairs $(a_1, b_1) = (e^{j0.13\pi}, e^{j0.09\pi})$ and $(a_2, b_2) = (e^{j0.16\pi}, e^{j0.12\pi})$. The sample support along the two sampling dimensions and the time axis is given by $K = 5$, $L = 5$, and $N = 1000$. The figures 7.1(a) and 7.1(b) show the root-mean-square-error (RMSE) of the frequency estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ along the $a$-axis obtained from the different algorithms versus the signal-to-noise ratio. Simulation results are averaged over 1000 simulation runs and compared to the deterministic Cramér-Rao bound (CRB) for 2D pure HR that is derived in appendix G [ZHM96].

From figure 7.1(a) we observe that the root-MI-ESPRIT algorithm has the best performance in threshold domain and clearly outperforms the RARE algorithm. This can be explained by the fact that the degree of the MI-ESPRIT polynomial is only half the degree of the RARE polynomial. It is clear that the numerical difficulties arising in rooting a MP with perturbed coefficients increase with the degree of the polynomial. Further we note that the rooting-based rank reduc-
7.1 Synthetic data

Synthetic data approaches have better resolution than its spectral search based implementations. This behavior has already been observed in literature [RH89], where it was shown that root-MUSIC outperforms spectral-MUSIC. An intuitive explanation is that in rooting approaches the radial errors, i.e., the error in estimating the magnitude of the roots, do not affect the estimation of the frequency parameters (provided that no subspace-swap occurs) and only the angular deviations of the signal roots from their original loci in the complex plane yields a frequency estimation error. In contrast, in spectral search based algorithms the solutions are forced to lie on the unit-circle. Finally we observe that SPEC-MI-ESPRIT yields notably better threshold performance than SPEC-RARE, which is accompanied by an significant increase in the computational cost of determining the smallest singular value of a “tall” \( (1/2KL(L-1)) \times P \) matrix in contrast to a square \( P \times P \) matrix in the cost functions.

In figure 7.1(b) the root-MI-ESPRIT algorithm is compared to other methods known from literature. Root-MI-ESPRIT shows similar performance as 2D ESPRIT, 2D MI-ESPRIT and TALS, however the computational complexity of the MI-ESPRIT algorithms and TALS is significantly larger than the cost associated with 2D root-MI-ESPRIT and 2D ESPRIT. This is due to the slow convergence of the gradient-based iterative optimization procedure of both algorithms that was observed in the simulations. Interestingly, root-MI-ESPRIT outperforms MI-MODE in threshold domain and loses only negligible performance compared this method asymptotically. Recall that MI-MODE is asymptotically equivalent to the ML solution for estimating the \( a \)-axis parameter decoupled from the remaining parameters.

![Figure 7.1: \( a \)-axis parameter for 2D harmonics](image-url)
Example 2

In this experiment we assume 3 equi-powered pure harmonics described by the triplets

\[
(a_1, b_1, c_1) = (e^{j0.16\pi}, e^{j0.86\pi}, e^{j0.16\pi}) \\
(a_2, b_2, c_2) = (e^{j0.34\pi}, e^{j0.16\pi}, e^{j0.28\pi}) \\
(a_3, b_3, c_3) = (e^{j0.58\pi}, e^{j0.34\pi}, e^{j0.24\pi}).
\]

The sample support along the three sampling dimensions and the time axis is given by \( K = 3 \), \( L' = 3 \), \( M = 3 \) and \( N = 100 \), respectively. The RMSE of the frequency estimates along the \( a \)-axis, \( b \)-axis and \( c \)-axis are displayed in figures 7.2(a), 7.2(b) and 7.2(c) versus the SNR. Simulation results are averaged over 100 Monte-Carlo runs and compared to the deterministic CRB for 3D pure HR provided in G.

From figures 7.2(a)-7.2(c) it becomes apparent that root-MI-ESPRIT shows the best average performance if we consider all three dimensions. We observe that the parameter association requirement in the rank reduction algorithms is not limiting the performance of the algorithms because along all three axes the threshold domain is located at around the same SNR value of about \(-7.5\,\text{dB}\). It is clear that in case that the parameter association fails the threshold domain along the first array axis, according to which the estimates are assigned to the true signal, should be located at significantly lower SNR values than the threshold domain along the remaining axis. Further along the \( c \)-axis, where the three harmonics are close together the tree-RARE algorithms shows best performance. This can be explained be the fact that at the third stage of the tree structured algorithm the signals are already well separated according to the well separated harmonics along the \( a \)- and \( b \)-axis. In this case the error propagation effects are moderate and the algorithm benefits from backsubstitution. The rooting-based algorithms outperform the joint ESPRIT algorithms as MI-ESPRIT and TALS, which only exploit that the MI equations share a common mixing matrix but do not account for the specific relation between the diagonal eigenvalue matrices (see section 3.4 for details).

Example 3

In the third experiment 3 equi-powered pure harmonics with generators contained in the triplets

\[
(a_1, b_1, c_1) = (e^{j0.24\pi}, e^{j0.26\pi}, e^{j0.16\pi}) \\
(a_2, b_2, c_2) = (e^{j0.34\pi}, e^{j0.16\pi}, e^{j0.28\pi}) \\
(a_3, b_3, c_3) = (e^{j0.42\pi}, e^{j0.34\pi}, e^{j0.24\pi})
\]

were considered. The sample support along the three sampling dimensions and the time axis is given by \( K = 6 \), \( L' = 6 \), \( M = 6 \) and \( N = 100 \), respectively. The RMSE of the frequency estimates along the \( a \)-axis, \( b \)-axis and \( c \)-axis are displayed in figures 7.3(a), 7.3(b) and 7.3(c) versus
the SNR. All results are averaged over 100 simulation runs and compared to the deterministic CRB for 3D pure HR in G.

In figures 7.3(a)-7.3(c) we observe similar results as in the previous example. We note that root-MI-ESPRIT attains the highest estimation precision, both asymptotically and in threshold domain. In all three sampling axes the algorithm asymptotically approaches the corresponding CRB asymptotically. Further we see that, according to the argumentation in the previous example, the parameter association task in root-MI-ESPRIT succeeds even in SNR regions close to threshold domain. However, some difficulties in this region are reported for the MD-RARE algorithm. The error propagation becomes critical in the tree-RARE algorithm when considering the parameter estimation of the $b$- and $c$-axis parameters. Root-MI-ESPRIT and RARE clearly outperform MI-ESPRIT and TALS at a significantly reduced computational complexity.
Figure 7.3: Parameter estimates for closely spaced 3D harmonics

Example 4

In this experiment we assume 3 equi-powered pure harmonics described by the triplets

\[(a_1, b_1, c_1) = (e^{j0.12\pi}, e^{j0.14\pi}, e^{j0.28\pi})\]  \hspace{2cm} (7.7)

\[(a_2, b_2, c_2) = (e^{j0.32\pi}, e^{j0.12\pi}, e^{j0.64\pi})\]  \hspace{2cm} (7.8)

\[(a_3, b_3, c_3) = (e^{j0.64\pi}, e^{j0.13\pi}, e^{j0.44\pi})\]  \hspace{2cm} (7.9)

The sample support along the three sampling dimensions and the time axis is given by \(K = 6\), \(L' = 2\), \(M = 6\) and \(N = 100\), respectively. The RMSE of the frequency estimates along the \(a\)-, \(b\)-, and \(c\)-axis are displayed in figures 7.4(a), 7.4(b) and 7.4(c) versus the SNR. Simulation results are averaged over 100 runs and compared to the deterministic CRB for 3D pure HR that is given in G.

Figures 7.4(a)-7.4(c) clearly demonstrate the high potential of the hybrid root-RARE algorithm
7.1 Synthetic data

which in this case does not exploit the prior-information that the second array axis is subject to uniform sampling (see section 6.3.2 for details). We also note the benefits of the joint root-MI-ESPRIT algorithm in the particular case where the sample support along the second array axis is significantly smaller than the sample support along the other dimensions. In particular both algorithms outperform root-MI-ESPRIT and RARE in estimating the parameters along the second array axis. However joint root-MI-ESPRIT loses some estimation performance in determining the frequency parameters along the first and third array axis. Furthermore 3D RARE shows remarkably reduced parameter estimation performance along the second array axis where the sample support is severely limited. This results from difficulties in associating the parameters according to the criteria in (6.18).

Figure 7.4: Parameter estimates for 3D harmonics
Example 5

In the fifth experiment we consider 3 damped harmonics generated by the triplets

\[
(a_1, b_1, c_1) = (e^{-0.10+0.16j}, e^{-0.20+0.34j}, e^{-0.00+0.16j})
\]  \hspace{1cm} (7.10)

\[
(a_2, b_2, c_2) = (e^{-0.20+0.28j}, e^{-0.10+0.16j}, e^{-0.04+0.28j})
\]  \hspace{1cm} (7.11)

\[
(a_3, b_3, c_3) = (e^{-0.03+0.30j}, e^{-0.00+0.46j}, e^{-0.06+0.24j})
\]  \hspace{1cm} (7.12)

The sample support along the three sampling dimensions and the time axis is given by \( K = 5 \), \( L' = 5 \), \( M = 5 \) and \( N = 100 \). The RMSE of the 3D estimates computed as

\[
\sqrt{\frac{1}{3} \sum_{p=1}^{P} \left( |\hat{a}_p - a_p|^2 + |\hat{b}_p - b_p|^2 + |\hat{c}_p - c_p|^2 \right)}
\]  \hspace{1cm} (7.13)

is displayed in figures 7.5 versus the SNR. The results are averaged over 100 simulation runs and compared to the deterministic CRB for 3D damped HR as given in G.

Example 5 represents a well-separated source scenario in which the number of harmonics is much smaller than the number of samples available. Figure 7.5 reveals that in the damped harmonic case 3D root-MI-ESPRIT clearly outperforms the TALS algorithm and asymptotically reaches performance close the the corresponding CRB. This can be explained by the fact that TALS does not exploit all information contained in the MI equations while root-MI-ESPRIT uses the specific relation between the diagonal eigenvalue matrices (see section 3.4 for details).

![Figure 7.5: Parameters for well separated 3D damped harmonics (all three axis)](image-url)
Example 6

Here we consider 3 damped harmonics with the generator-triplets:

\[(a_1, b_1, c_1) = (e^{-0.10+j0.24\pi}, e^{-0.20+j0.30\pi}, e^{-0.00+j0.22\pi})\] (7.14)
\[(a_2, b_2, c_2) = (e^{-0.20+j0.28\pi}, e^{-0.10+j0.22\pi}, e^{-0.04+j0.32\pi})\] (7.15)
\[(a_3, b_3, c_3) = (e^{-0.03+j0.30\pi}, e^{-0.00+j0.32\pi}, e^{-0.06+j0.24\pi})\] (7.16)

The sample support along the three sampling dimensions and the time axis is \(K = 3\), \(L' = 3\), \(M = 3\) and \(N = 100\), respectively. The RMSE of the 3D estimates computed as in (7.13) displayed in figures 7.6 versus the SNR. All results are averaged over 100 Monte-Carlo simulations and compared to the deterministic CRB for 3D damped HR G.

The sixth example contains a difficult estimation scenario. The sample support along the different dimensions is comparably small. Figure 7.5 reveals that under this setting the 3D root-MI-ESPRIT also outperforms the TALS algorithm. However, the CRB is not attained asymptotically. Further, the threshold domain lies at comparably large SNR values.

![Figure 7.6: Parameters for closely spaced 3D damped harmonics (all three axis)](image)

Example 7

In the seventh experiment we assume 3 damped harmonics with parameter-triplets

\[(a_1, b_1, c_1) = (e^{-0.10+j0.16\pi}, e^{-0.20+j0.28\pi}, e^{-0.04+j0.92\pi})\] (7.17)
\[(a_2, b_2, c_2) = (e^{-0.00+j0.46\pi}, e^{-0.03+j0.30\pi}, e^{-0.06+j0.46\pi})\] (7.18)
\[(a_3, b_3, c_3) = (e^{-0.20+j0.74\pi}, e^{-0.10+j0.20\pi}, e^{-0.00+j0.04\pi})\] (7.19)
The sample support along the three sampling dimensions and the time axis reads \( K = 4, L' = 4, M = 4 \) and \( N = 100 \), respectively. The RMSE of the 3D estimates computed as in (7.13) is displayed in figures 7.7 versus the SNR. Simulation results are averaged over 100 simulation runs and compared to the deterministic CRB for 3D damped HR provided in G.

Here we observe that the joint diagonalization approach (TALS) outperforms the rooting-based methods (root-MI-ESPRIT, hybrid MI-ESPRIT, and joint root-MI-ESPRIT) in threshold domain. This mainly results from the poor separation of the harmonics along the second array axis. In this case the decoupled estimation of the parameters as in root-MI-ESPRIT is not convenient to estimate the parameters along the \( b \)-axis. However asymptotically the joint root-MI-ESPRIT yields best estimation performance which, in this region, runs close to CRB. This is because the algorithm first computes three different estimates of the mixing matrices separately along the different dimensions and then, in a post processing step, performs an averaging procedure to obtain a single joint estimate. It is clear that in the case of poor separation along one dimension some benefit is taken from exploiting the joint nature of the estimation problem along the different dimensions.

![Figure 7.7: Parameters for closely spaced 3D damped harmonics (all three axis)](image_url)

**Example 8**

This experiment consists of 3 equi-powered pure harmonics characterized by the triplets

\[
(a_1, b_1, c_1) = (e^{j0.12\pi}, e^{j0.88\pi}, e^{j0.28\pi})
\]

\[
(a_2, b_2, c_2) = (e^{j0.46\pi}, e^{j0.40\pi}, e^{j0.66\pi})
\]

\[
(a_3, b_3, c_3) = (e^{j0.74\pi}, e^{j0.64\pi}, e^{j0.44\pi})
\]
The single snapshot case $N = 1$ is considered. The sample support along the three dimensions is $K = 8$, $L' = 8$, and $M = 8$. The RMSE of the frequency estimates along the $a$-axis is depicted in figure 7.8 versus the SNR. Note, however, that the parameter estimates along the $b$- and $c$-axis show a similar behavior. Simulation results are averaged over 100 independent runs and compared to the single-snapshot CRB for 3D pure HR in G.

From figure 7.8 we recognize that in the case where a small number of signals are contained in the MD mixture compared to the number of available samples along the three measurement axes, the root-MI-ESPRIT algorithm outperforms the MDF algorithm.

![Figure 7.8: $a$-axis parameter estimation in comparably large sample support.](image)

**Example 9**

In this experiment we assume 4 equi-powered pure harmonics with parameter-triplets

$$
(a_1, b_1, c_1) = (e^{j0.036\pi}, e^{j0.270\pi}, e^{j0.468\pi}) \quad (7.23)
$$

$$
(a_2, b_2, c_2) = (e^{j0.120\pi}, e^{j0.045\pi}, e^{j0.072\pi}) \quad (7.24)
$$

$$
(a_3, b_3, c_3) = (e^{j0.384\pi}, e^{j0.615\pi}, e^{j0.348\pi}) \quad (7.25)
$$

$$
(a_4, b_4, c_4) = (e^{j0.480\pi}, e^{j0.480\pi}, e^{j0.024\pi}) \quad (7.26)
$$

The single snapshot case is considered i.e. $N = 1$. The sample support along the three dimensions is $K = 3$, $L' = 3$, and $M = 3$. The RMSE of the frequency estimates along the $a$-axis is depicted in figure 7.9 versus the SNR. We remark that the parameter estimates along the $b$- and $c$-axis show a similar behavior. All Results are averaged over 100 simulation runs and compared to the single-snapshot CRB for 3D pure HR given in G.

In contrast to the preceding example the sample support is considerably small compared to the number of sources that are present. In this case the root-MI-ESPRIT algorithm yields signifi-
Simulation results

cantly better parameter estimates than the MDF algorithm, however the asymptotic performance that is attained is far from optimal as the comparison the corresponding CRB reveals.

![Figure 7.9: α-axis parameter estimation in comparably small sample support.](image)

**Example 10**

In this experiment we assume 3 damped harmonics described by the triplets

\[
(a_1, b_1, c_1) = (e^{-0.20+j0.24\pi}, e^{-0.05+j0.51\pi}, e^{-0.00+j0.24\pi})
\]

(7.27)

\[
(a_2, b_2, c_2) = (e^{-0.10+j0.42\pi}, e^{-0.10+j0.24\pi}, e^{-0.04+j0.42\pi})
\]

(7.28)

\[
(a_3, b_3, c_3) = (e^{-0.03+j0.45\pi}, e^{-0.00+j0.69\pi}, e^{-0.06+j0.36\pi})
\]

(7.29)

The single snapshot case is considered. The sample support along the three dimensions is \(K = 11\), \(L' = 11\), and \(M = 11\). The RMSE of the 3D estimates computed as in (7.13) are displayed in figure 7.10 versus the SNR. Results are averaged over 100 simulation runs and compared to the single-snapshot CRB for 3D damped HRG.

Here, similar results as in example 8 are obtained. In the case of a small number of signal compared to the number of available samples along the three array axes the root-MI-ESPRIT algorithm outperforms the MDE algorithm.
7.2 Measurement data

Example 11

In this experiment we assume 4 damped harmonics described by the triplets

\[
(a_1, b_1, c_1) = (e^{-0.20+j0.32\pi}, e^{-0.05+j0.68\pi}, e^{-0.00+j0.32\pi})
\]

\[
(a_2, b_2, c_2) = (e^{-0.10+j0.56\pi}, e^{-0.10+j0.32\pi}, e^{-0.04+j0.56\pi})
\]

\[
(a_3, b_3, c_3) = (e^{-0.03+j0.60\pi}, e^{-0.00+j0.92\pi}, e^{-0.06+j0.48\pi})
\]

\[
(a_4, b_4, c_4) = (e^{-0.00+j0.80\pi}, e^{-0.20+j0.00\pi}, e^{-0.02+j0.80\pi})
\]

(7.30) (7.31) (7.32) (7.33)

The single snapshot case is considered, hence \( N = 1 \). The sample support along the three dimensions is \( K = 3, L' = 3 \), and \( M = 3 \). The RMSE of the 3D estimates computed as in (7.13) are displayed in figure 7.11 versus the SNR. All results are averaged over 1000 simulation runs and compared to the corresponding single-snapshot CRB for 3D damped HR given in G.

Similarly as in example 9 here we observe that for small sample support along the various axes compared to the number of sources the root-MI-ESPRIT algorithm yields better parameter estimates than the MDE algorithm. However in contrast to pure HR case of example 9, here, in the damped case the asymptotic performance is slightly closer to the optimality bound (CRB).

7.2 Measurement data

Measurement data were recorded with the RUSK-ATM vector channel sounder, manufactured and marketed by MEDAV [THR+99, MED]. The measurement data used for the numerical experiments in this paper were recorded during a measurement run in Weikendorf, a suburban area in a small town approximately 50 km north of Vienna, Austria, in autumn 2001 [HVU02,
Simulation results

Figure 7.11: 3D parameter estimation in comparably small sample support.)

HMM\textsuperscript{+02}, Ftw]. The measurement area covers one-family houses with private gardens around them. The houses are typically one floor high. A rail-road track is present in the environment which breaks the structure of single placed houses. A small pedestrian tunnel passes below the railway. A map of the environment with the position of the receiver and transmitter is shown in figure 7.12.

The sounder was operated at a center frequency of 2000 MHz with an output power of 2 Watt and a transmitted signal bandwidth of 120 MHz. The transmitter emitted a periodically-repeated signal composed of 384 sub-carriers in the band 1940 ... 2060 MHz. The repetition period was 3.2 $\mu$s. The transmitter was the mobile station and the receiver was at a fixed location. The transmit array had a uniform circular geometry composed of 15 monopoles arranged on a ground plane at an inter-element spacing of 0.43$\lambda \approx 6.45$ cm. The mobile transmitter was mounted on top of a small trolley together with the uniform circular array at a height of approx. 1.5 m above ground level. At the receiver site a ULA\textsuperscript{1} composed of 8 elements with half wavelength distance (7.5 cm) between adjacent patch-elements was mounted on a lift in approx. 20m height.

With this experimental arrangement, consecutive sets of the $(15 \times 8)$ individual transfer functions, cross-multiplexed in time, were acquired. The receiver calculates the discrete Fourier-transform over data blocks of duration 3.2 $\mu$s and deconvoloves the data in the frequency domain with the known transmit signal. The effects from mutual coupling between Rx antenna elements are reduced by multiplying the measurement snapshots $y(i)$ with a complex-symmetric correction matrix [SHK\textsuperscript{+01}]. The acquisition period of 3.2 $\mu$s corresponds to a maximum path length of approx. 1 km. During the measurements the receiver moved at speeds of approx. 5 km/h on the sidewalk. Rx-position and Tx-position, as well as the motion of the transmitter are marked in the site map in figure 7.12. The transmitter passed through a pedestrian tunnel approximately

\textsuperscript{1}provided by T-Systems NOVA, Darmstadt, Germany.
between times $t = 25\,\text{s}$ and $t = 30\,\text{s}$ of the measurement run. We estimated the data covariance matrix from $J = 10$ consecutive MIMO snapshots in time. The measurement system in this experiment differs from the data acquisition model described in the introduction (1.1-d) in that a uniform circular array instead of a ULA was used at the transmitter side. Therefore we can not simply apply the estimation procedure for the 3D parameter estimation problem described in section 1.1 to estimate the directions-of-departure. In this experiment we only consider a 2D model instead of the general 3D model (1.1-d). In specific we are interested in estimating only the directions of arrival and the time delays. In order to still exploit the complete 3D measurement block that was recorded as described above we use averaging over Tx samples and smoothing over frequency bins in order to increase the number of snapshots and to obtain a full rank covariance matrix estimate of reduced variance. Due to the smoothing over frequency bins, the original sample support of $K = 384$ frequency bins along the $a$-axis is reduced to a sample support of $K' = 12$. For further variance reduction we apply FB averaging introduced in section 2.2.1. Making use of the notation of the general 3D model in (1.3) the smoothed FB sample covariance matrix corresponding to the 2D model reads

$$
\hat{R} = \frac{1}{D} \sum_{i=1}^{J} \sum_{k=1}^{K-K'} \sum_{m=1}^{M} \left( |\tilde{Y}_{k,m}(i)| |\tilde{Y}_{k,m}^H(i) + \Pi_{96}|\tilde{Y}_{k,m}^*(i)|\tilde{Y}_{k,m}(i)\Pi_{96} \right)
$$

(7.34)
where $D = J(K - K')M$,

$$\hat{Y}_{k,m}(i) = \text{vec}\left\{ \begin{bmatrix} Y_{k,1,m}(i) \\ Y_{k+1,1,m}(i) \\ \vdots \\ Y_{k+K',1,m}(i) \\ Y_{k,2,m}(i) \\ Y_{k+1,2,m}(i) \\ \vdots \\ Y_{k+K',2,m}(i) \\ \vdots \\ Y_{k,L,m}(i) \\ Y_{k+1,L,m}(i) \\ \vdots \\ Y_{k+K',L,m}(i) \end{bmatrix} \right\}, \quad (7.35)$$

$M = 15$, $L = 8$ and $\Pi_{96}$ denotes the $96 \times 96$ exchange matrix. In the first experiment the propagation delay and DOA estimates obtained with 2D RARE are displayed in figure 7.13 and figure 7.14 relative to the orientation of the array. We have assumed $P = 10$ paths and applied 2D RARE for the joint estimation of propagation delay and DOA. In these two figures, the estimates are plotted as colored marks (dots ‘.’ and ‘*’) versus measurement time in seconds. The pairing of the estimates is indicated by the chosen mark and its color. In these figures, the circles (‘◦’) mark the line of sight path, dots (‘.’) mark the consecutive early arrivals whereas the asterisks (‘∗’) mark the late ones.

Figure 7.13: Estimates of the propagation delay versus snapshot in time obtained from 2D RARE [PMB04].

We see that the propagation scenario is dominated by a strong line-of-sight (LOS) component surrounded by local scattering paths from trees and buildings during the first 25 seconds of the experiment (shown with the ‘◦’ mark in the figures). The trace of the DOA estimates in
Figure 7.14: DOA estimates versus snapshot in time obtained from 2D RARE [PMB04].

Figure 7.14 and the corresponding propagation delay estimates in figure 7.13 match the motion of the transmitter depicted in figure 7.12 for the direct path. At time 25s the trolley reaches the pedestrian tunnel and a second path resulting from scattering at the building (see figure 7.14) appears at a DOA of approximately $-3^\circ$. This path corresponds to a significantly larger access delay of approx. 0.55...0.58 $\mu$s. By the time the Tx moves out of the tunnel the dominant LOS component with local scattering is newly tracked by the 3D-RARE algorithm. In figure 7.14 we observe a path emerging at a constant DOA of approx. $22^\circ$ between snapshot time 0s and 25s. Similarly, a path emerging at a constant DOA of approx. $17^\circ$ between time 28s and 52s. These paths are interpreted as contributions from the two ends of the pedestrian tunnel. Furthermore, it is interesting to observe that those propagation paths with large delay estimates generally yield corresponding DOA estimates with large angular deviations from the line of sight.

In the second experiment, displayed in figures 7.15 and figure 7.16 the propagation delay and DOA estimates were obtained from the root-MI-ESPRIT algorithm. A variable model order was used. From time $t = 0$ s to $t = 1.5$ s we assumed $P = 17$ paths, from $t = 1.5$ s to $t = 2.8$ s we considered $P = 20$ paths, during the time the transmitter passed through a pedestrian tunnel, between times $t = 25$ s and $t = 30$ s, we assumed $P = 11$ paths and for the remaining time intervals as much as $P = 24$ paths were considered.

In these two figures, the estimates are plotted as colored marks (small dots and fat dots) versus
measurement time in seconds. The pairing of the estimates is indicated by the chosen mark and its color. Due to the high model order the color mapping in the figure is not unique. In these figures, the fat dots mark the first 7 early arrivals whereas the small dots mark the late ones. Similar to the preceding example we see that the propagation scenario is dominated by a strong LOS component surrounded by local scattering paths from trees and buildings during the first 25 seconds of the experiment. The direct path is shown with fat blue dots in the figures. Also in this experiment the trace of the DOA estimates in figure 7.16 and the corresponding propagation delay estimates in figure 7.15 match the motion of the transmitter. We observe that the direct path is blocked during the time the trolley passes the pedestrian tunnel and is newly tracked by the root-MI-ESPRIT algorithm when the Tx moves out of the tunnel.

In figure 7.16 we observe paths emerging at constant DOAs of approx. $-6^\circ$, $-2^\circ$ and $22^\circ$ between snapshot time 0s and 25s. Similarly, paths emerging at a constant DOAs of approx. $-6^\circ$, $17^\circ$, and $60^\circ$ appear between time 28s and 52s. These paths are interpreted as constant scatterers that are illuminated by the Tx over a comparably long period of time. Similar to the preceding example, it is notable that those propagation paths that show large propagation delay generally yield corresponding DOA estimates with large angular deviations from the line of sight.
The second experiment was repeated for the 2D unitary-ESPRIT algorithm. The model order was assumed as above. The propagation delay and DOA estimates are displayed in figures 7.17 and figure 7.18. Apparently the 2D unitary-ESPRIT algorithm has difficulties to resolve the large number of discrete propagation paths. The estimation results are contradictory and only allow limited physical interpretation. A line of sight matching the motion of the trolley is also visible in 7.18, however the corresponding estimates do not correspond to the first arrivals as can be observed from figure 7.17.
7 Simulation results

Figure 7.17: TDOA estimates versus snapshot in time obtained from 2D ESPRIT [ZHM96].

Figure 7.18: DOA estimates versus snapshot in time obtained from 2D ESPRIT [ZHM96].
8 Conclusions and Outlook

In this work a variety of subspace methods for MD HR has been proposed. The novel procedures stem back from a suitable parameterization of the manifold vector that allows to separate the parameters along one dimension from the parameters along the remaining dimensions. The original MD estimation problem is thus solved based on multiple one-dimensional rank criteria. This procedure makes the estimation problem computationally tractable while retaining much of the benefits inherent in the multidimensional nature of the measurement data such as, for example, relatively mild uniqueness conditions and high resolution capability compared to one dimensional data. In the case of uniform sampling along one or multiple axes the proposed rank reduction algorithms exploits the regular structure of the estimation problem to estimate the harmonics along the various dimensions separately from the roots of univariate MPs. The rank criteria are interpreted in diverse contexts, which are a) a relaxation approach in minimizing the classic root-MUSIC criterion, b) in a Gaussian-elimination framework, and c) as a rooting-based solution of the multiple invariance equations. From the different viewpoints new stochastic uniqueness conditions for the rank reduction methods are derived. Further, a nullspace-relation is discovered between the rank criteria formulated along the diverse dimensions, from which efficient parameter association strategies to correctly group the parameters of a specific multidimensional harmonic signal are obtained. A link between the popular ESPRIT-type methods and the rooting based methods is revealed that allows to reformulate the rank reduction idea in terms of a set of related generalized eigenproblems. The parameters of interests along the distinct dimensions are uniquely obtained from the \( P \) smallest generalized eigenvalues of a BCM pair. The associated generalized eigenvectors allow simple and reliable parameter association. The idea to cast the MD HRP as a set of related eigenproblems not only reduces the computational cost of the rank reduction methods but also makes the algorithm equally applicable to the cases of pure and damped HR.

Simulation results obtained from synthetic data for the single and multiple snapshot case are presented and illustrate that the proposed algorithms are competitive with other existing methods from both a numerical viewpoint and also in terms of estimation performance. Further, in the example of parametric MIMO channel identification, it is demonstrated that the novel algorithms perform well if applied to real measurement data obtained from a channel-sounding campaign.

It is understood that the separation of the parameter estimation along the various dimensions is attractive from a computational point of view, because it allows parallel processing and makes the MD estimation algorithm scalable. However, in some cases improvements in terms of estimation accuracy can be expected when considering the parameter estimation jointly. In the
algorithms proposed in this work the joint character of the estimation problem is reflected in the nullspace-relation that exists between the different rank reduction criteria along the various dimensions. These relations were primarily used to mutually associate the parameter estimates. A challenging task consists in the attempt to solve the generalized eigenproblems of the root-MI-ESPRIT algorithm jointly while taking into account the specific structure of the generalized eigenvectors.

Another open question that requires further research is the detection problem. In this work we assumed that the true model order, i.e. the correct number of signals associated with the sum-of-harmonic mixture, is known. In practice this is usually not the case. In subspace-based algorithms a popular approach is to use some function of the eigenvalues of the covariance matrix as the data component in the detection algorithm. Typical examples of detection criteria are the Akaike Information Criterion (AIC) and the Minimum Description Length (MDL) [WK85]. The question arising in this context is how to extend the eigenvalue-based detection criteria to the single snapshot case considered in section 2.2 and also to the case of smoothed covariance matrices that were used in the real measurement experiment of section 7.2.

In this work we discounted the incomplete data HRP that was formulated in section 1.2.5. The rank reduction algorithms developed for the complete data case also apply to the incomplete data case, where some observations along the uniform sampling axis are missing. However, the uniqueness results obtained for these algorithms explicitly rely on the fact that all data samples are available and thus can not directly be transferred to the incomplete harmonic case. To determine the number of harmonics that can uniquely be identified from the rank reduction algorithms in this case is still an open problem that requires further research.
A Useful properties of vector algebra

Hadamard Product: The Hadamard-product of two $N \times M$ matrices $A$ and $B$ with $a_{i,j} = [A]_{ij}$ and $b_{i,j} = [B]_{ij}$ is defined as element-wise multiplication

$$A \odot B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1M}b_{1M} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2M}b_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1}b_{N1} & a_{N2}b_{N2} & \cdots & a_{NM}b_{NM} \end{bmatrix} \quad (A.1)$$

Kronecker-product: If $A$ is a $N \times M$ with $a_{i,j} = [A]_{ij}$ and $B$ is a $K \times L$ matrix, the Kronecker-product is defined to be the $NM \times ML$ matrix,

$$A \otimes B = \begin{bmatrix} a_{11}B & B & \cdots & a_{1M}B \\ a_{21}B & a_{22}B & \cdots & a_{2M}B \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1}B & a_{N2}B & \cdots & a_{NM}B \end{bmatrix} \quad (A.2)$$

Khatri-Rao product: The Khatri-Rao product of matrix $A = [a_1, \ldots, a_M] \in \mathbb{C}^{N \times M}$ and matrix $B = [b_1, \ldots, b_M] \in \mathbb{C}^{P \times M}$ is defined as

$$A \circ B = [a_1 \otimes b_1 | a_2 \otimes b_2 | \cdots | a_M \otimes b_M] \in \mathbb{C}^{(PN) \times M} \quad (A.3)$$

Let $U_i$ denote a arbitrary complex matrix of dimension $U_i \times P$ for $i = 1, \ldots, R$. Then

$$U_2 \circ U_1 = [I_{M_1} \otimes i_{1,M_2}, I_{M_1} \otimes i_{2,M_2}, \ldots, I_{M_1} \otimes i_{M_2,M_2}] (U_1 \circ U_2) \in \mathbb{C}^{(U_1U_2) \times M} \quad (A.4)$$

From equations (A.4) we conclude that cyclic commutation of the matrices in a series of Khatri-Rao products $U_R \circ U_{R-1} \circ \cdots \circ U_2 \circ U_1$ (i.e. moving the first matrix factor $U_R$ in the product to the end and leaving the ordering of the remaining matrix factors $U_{R-1} \circ \cdots \circ U_2 \circ U_1$ unchanged) amounts to matrix multiplication of the original series of Khatri-Rao products with a permutation matrix of the form $[I_{\tilde{M}} \otimes i_{1,M_R}, I_{\tilde{M}} \otimes i_{2,M_R}, \ldots, I_{\tilde{M}} \otimes i_{M_R,M_R}]$, where $\tilde{M} = \prod_{r=1}^{R-1} M_r$ denotes the number of rows in unchanged Khatri-Rao product, hence

$$U_{R-1} \circ U_{R-2} \circ \cdots \circ U_2 \circ U_1 \circ U_R = [I_{\tilde{M}} \otimes i_{1,M_R}, I_{\tilde{M}} \otimes i_{2,M_R}, \ldots, I_{\tilde{M}} \otimes i_{M_R,M_R}] (U_R \circ U_{R-1} \circ \cdots \circ U_2 \circ U_1) \quad (A.5)$$
Vectorization operator:

\[
\text{vec}\{ABC\} = (C^T \otimes A) \text{vec}B
\]  
(A.6)

Block determinant lemma:

\[
\det \begin{bmatrix}
A & B \\
C & D
\end{bmatrix}
= \det \{ A - BD^{-1}C \} \det \{ D \}
\]  
(A.7)

Sylvester inequality: [Zha99, GvL96] Given two matrices \( A \in \mathbb{C}^{p \times n} \) and \( B \in \mathbb{C}^{n \times q} \), the following inequality holds true:

\[
\text{rank}\{A\} + \text{rank}\{B\} - n \leq \text{rank}\{AB\} \leq \min (\text{rank}\{A\}, \text{rank}\{B\})
\]  
(A.8)

Rank equality: Given two matrices \( A \in \mathbb{C}^{p \times n} \) and \( B \in \mathbb{C}^{p \times n} \) and a full-rank matrix \( K \in \mathbb{C}^{n \times n} \) such that \( B = AK \), the following equality holds true:

\[
\text{rank}\{A^H B\} = \text{rank}\{A\} = \text{rank}\{B\}.
\]  
(A.9)

Proof: According to the assumptions we can write

\[
A^H B = A^H AK
\]  
(A.10)

Thus applying Sylvester’s inequality and with full-rank matrix \( K \) it is immediate that \( \text{rank}\{A^H B\} = \text{rank}\{A^H AK\} = \text{rank}\{A^H A\} = \text{rank}\{A\} \). Similarly we have from full-rank matrix \( K^{-H} \) that \( \text{rank}\{A^H B\} = \text{rank}\{K^{-H} B^H B\} = \text{rank}\{B^H B\} = \text{rank}\{B\} \).

Equivalence of eigenvalues: Given \( A \in \mathbb{C}^{m \times n} \) and \( B \in \mathbb{C}^{n \times m} \). Then \( AB \) and \( BA \) have the same nonzero eigenvalues, counting multiplicity [Zha99].

Conjugate-reciprocity of MPs: Given a \( K \times L \) MP \( M(a) \) of degree \( M \) with

\[
M(a) = \sum_{m=0}^{M} M_m a^m
\]  
(A.11)

with polynomial coefficients denoted by the \( K \times L \) matrices \( M_1, \ldots, M_M \). Define the quadratic from

\[
G(a, a^*) = M^H(a)M(a) = \sum_{m=0}^{M} \sum_{l=0}^{M} M_m^H M_l (a^m)^* a^l.
\]  
(A.12)
If evaluating $G(a, a^*)$ on the unit circle, hence for $|a| = 1$, then we can replace $a^*$ by $a^{-1}$. Hence we obtain

$$G(a) = G(a, a^*) \big|_{|a|=1} = \sum_{l=-M}^{M} G_l a^l = \sum_{m=0}^{M} \sum_{l=0}^{M} M_m^H M_l a^{l-m}. \quad (A.13)$$

The MF in (A.13) represents a $L \times L$ MP of degree $2M - 1$ with polynomial coefficients $G_{-M}, \ldots, G_0, \ldots, G_M$. It is simple to check that the polynomial coefficients are Hermitian-symmetric with respect to the center coefficient $G_0$. In other words $G_m^H = G_m$ for $m = 1, \ldots, M$.

$$G_m^H(a^*) = \sum_{m=-M}^{M} G_m^H a^m = \sum_{m=-M}^{M} G_{-m} a^m = \sum_{m=-M}^{M} G_m a^{-m} = G(a^{-1}) \quad (A.14)$$

Since the MP $G(a)$ and it Hermitian version $G^H(a)$ drop rank for the same values of $a$, we thus obtain from (A.14) that if $a^{-1}$ is a root of $G(a)$ then it is immediate that $a^*$ is also a root of $G(a)$. This identity is commonly referred to as the conjugate – reciprocity property of the MP $G(a)$.
B Proof of T2

Without loss of generality, we consider the limiting case that \( P = (K - 1)L \). The augmented matrix in (3.22) then becomes \( KL \times KL \) square. The case of \( P < (K - 1)L \) follows immediately from deletion of columns of the augmented matrix in (3.22). In order to determine the singularities of \( M_3(a, H \mid \text{“d”}) \) we apply appropriate elementary matrix operations on its rows. More precisely, we exploit the property that adding a multiple of a row of a matrix to any other row does not change the determinant of the matrix. Similar to the procedure used in Gaussian elimination, we wish to bring the first \( L \) columns of the augmented matrix \( M_3(a, H \mid \text{“d”}) \) to triangular form. Towards this aim, we subtract \( a \) times the \( (k - 1)\)-st row from the \( k\)-th row of \( M_3(a, H \mid \text{“d”}) \) (3.22), for \( k = 2, \ldots, K, K + 2, \ldots, 2K, 2K + 2, \ldots, 3K, \ldots, (L - 1)K, (L - 1)K + 2, \ldots, LK \), i.e. \( \forall k \in \{1, \ldots, KL\} \) such that \((k)_K \neq 1\), where \((k)_K\) denotes \( k \) modulo \( K \). The \( k\)-th row of the resulting matrix, denoted by \( W_{\text{tri}}(a) \), is then given by

\[
[W_{\text{tri}}(a)]_k = [0_K, \ldots, 0_K]_{L} = \begin{bmatrix} b_1 \frac{1}{a_1^{(k)_K - 2}} (a_1 - a), \ldots, b_P \frac{1}{a_P^{(k)_K - 2}} (a_P - a) \end{bmatrix}
\] (B.1)

for \((k)_K \neq 1\). For \((k)_K = 1\) the rows of \( W_{\text{tri}}(a) \) remain unchanged and identical to the corresponding rows of \( M_3(a, H \mid \text{“d”}) \). Note that \( \det\{W_{\text{tri}}(a)\} = \det\{M_3(a, H \mid \text{“d”})\} \).

It can readily be verified that each of the \( L \) first columns of \( W_{\text{tri}}(a) \) contains only a single non-zero element. These columns form a matrix

\[
T_0 = T_a(a)|_{a=0} = [e_1, e_{K+1}, e_{2K+1}, \ldots, e_{(L-1)K+1}]
\] (B.2)

where \( e_k \) denotes the \( k \)th column of a \( KL \times KL \) identity matrix \( I_{KL} \). Making use of a well-known expansion rule for determinants it is immediate to show that

\[
\det\{M_3(a, H \mid \text{“d”})\} = \\
= \det\{W_{\text{tri}}(a)\} \\
= \det\{|T_0 \mid H (\Delta_a - I_P a)|\} \\
= \pm \det\{H_{a,1} (\Delta_a - I_P a)\} \\
= \pm \det\{H_{a,1}\} \det\{(\Delta_a - I_P a)\} \\
= \pm \det\{H_{a,1}\} \prod_{p=1}^{P}(a_p - a)
\] (B.3)

where “±” indicates that equality holds up to “+” or “-” sign, and the row-reduced upper signal matrix \( H_{a,1} \) is defined in (2.58).

Provided that \( H_{a,1} \) has full column-rank we observe from (B.3) that for \( a \neq a_p \), \( p = 1, \ldots, P \), \( P \leq L(K - 1) \) the determinant \( \det\{M_3(a, H \mid \text{“d”})\} \neq 0 \) and \( \det\{M_3(a, H \mid \text{“d”})\} = 0 \).
otherwise. Furthermore we observe that

\[
\begin{align*}
\text{rank}\{M_3(a, \mathbf{H} \mid \text{“d”})\} &= L + \text{rank}\{H_{a,1} (\Delta_a - I_P \a)\} \\
&= L + \text{rank}\{(\Delta_a - I_P \a)\} \\
&= \begin{cases} 
P + L & \text{for } a \notin \{a_1, \ldots, a_P\} \\
P + L - \text{mult}\{a|\mathcal{H}_{a,1}\} & \text{otherwise}.
\end{cases}
\end{align*}
\]  

(B.4)
C Proof of equivalence between \( M_3(a, H \mid "d") \) and \( M_5(a, H \mid "d") \)

In order to prove that \( M_3(a, H \mid "d") \) and \( M_5(a, H \mid "d") \) are equivalent in terms of their signal and noise roots it is sufficient to show that \( M_3(a, H \mid "d") \) and \( M_5(a, H \mid "d") \) have identical roots. To prove the last statement we shall for example show that the augmented matrix

\[
\begin{bmatrix}
0 & M_5(a, H \mid "d") \\
I_L & B
\end{bmatrix}
\]  
(C.1)

can be formed from

\[
\begin{bmatrix}
M_3(a, H \mid "d") \\
0
\end{bmatrix}
\]  
(C.2)

through elementary row operations. We recall that the matrix \( M_5(a, H \mid "d") \) in (3.30) consists of individual blocks of the form

\[
H_{a,k}(I - \Delta^{-k}a^k)
\]  
(C.3)

for \( k = 1, \ldots, K - 1 \). Hence the matrix in (C.1) can equivalently be written as

\[
\begin{bmatrix}
0 & H_{a,1}(I - \Delta^{-1}a^1) \\
0 & H_{a,2}(I - \Delta^{-2}a^2) \\
\vdots & \vdots \\
0 & H_{a,K-1}(I - \Delta^{-K+1}a^{K-1}) \\
I_L & B
\end{bmatrix}
\]  
(C.4)

Note that the last \( L \) rows of the matrix in (C.1) are identical to the first, \( (K + 1) \)st, \( (2K + 1) \)st, \( \ldots \), \((L - 1)K + 1)\)st row of \( M_3(a, H \mid "d") \). Next consider the remaining row-blocks in (C.4) which are of the form

\[
\begin{bmatrix}
0 & H_{a,k}(I - \Delta^{-k}a^k)
\end{bmatrix}
\]  
(C.5)

for \( k = 1, \ldots, K - 1 \). It is simple to check that the \( m \)th row of the matrix in (C.5) evaluated for a specific \( k = 1, \ldots, K - 1 \) is formed by subtracting \( a^k \) times the \((m - k)\)th row of \( M_3(a, H \mid "d") \) from the \( m \)th row of \( M_3(a, H \mid "d") \) (3.22). Note that the definition of the selection matrices in (2.57) and (2.60) implies that the integer \( m \) takes only values \( k + 1, \ldots, K, K + k + 1, \ldots, 2K, 2K + k + 1, \ldots, 3K, \ldots, (L - 1)K + k + 1, \ldots, 2K \). Finally we remark that in forming the set of matrices in (C.5) each row of \( M_3(a, H \mid "d") \) is used at least once. Therefore the “tall” matrix in (C.1) is entirely formed from \( M_3(a, H \mid "d") \) and hence both matrices have identical singularities.

\[\blacksquare\]
D Proof of (3.39)

Recall that
\[
H_{a,k}^H H_{a,k} = (B_1 \circ A_{1,k})^H (F_1 \circ A_{1,k}) = (A_{1,k} \circ F)^H (A_{1,k} \circ F)
\]
\[
= \left[ \Delta^k_a F^H, \Delta^{k+1}_a F^H, \ldots, \Delta^{K-1}_a F^H \right]
\]
\[
= \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a ,
\]  
(D.1)

so that (3.34) can also be written as
\[
W_{\text{res}}(a) = M_6(a, H \mid \text{“d”})(I - \Delta^{-1}_a a)^{-1}
\]
\[
= \sum_{k=1}^{K-1} H_{a,k}^H H_{a,k} (I - \Delta^{-k}_a a^k) (I - \Delta^{-1}_a a)^{-1}
\]
\[
= \sum_{k=1}^{K-1} H_{a,k}^H H_{a,k} \left( \sum_{l=0}^{k-1} \Delta^{-l}_a a^l \right)
\]
\[
= \sum_{k=1}^{K-1} \sum_{l=0}^{k-1} \left( \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a \right) \Delta^{-l}_a a^l .
\]  
(D.2)

With (D.2) and for \( k \geq 2 \) we have
\[
\left( \sum_{l=0}^{k-1} (\Delta^{-1}_a a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a \right) \left( \sum_{n=0}^{k-1} (\Delta^{-1}_a a)^n \right) =
\]
\[
= \left( \sum_{l=0}^{k-1} (\Delta^{-1}_a a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a \right)
\]
\[
+ \left( \sum_{l=0}^{k-1} (\Delta^{-1}_a a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a \right) \left( \sum_{n=1}^{k-1} (\Delta^{-1}_a a)^n \right)
\]
\[
+ \left( \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a \right) \left( \sum_{n=0}^{k-1} (\Delta^{-1}_a a)^n \right) - \left( \sum_{m=k}^{K-1} \Delta^m_a F^H F \Delta^m_a \right) \left( \sum_{n=0}^{k-1} (\Delta^{-1}_a a)^n \right)
\]

113
Inserting (D.3) into (3.37) reveals that

\[
\begin{align*}
&\quad = \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) + \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) \left(\sum_{n=0}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad - \sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a} - \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) \left(\sum_{n=1}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad + \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right) \left(\sum_{n=1}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad + \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) \left(\sum_{n=1}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad = \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right) + \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right) \left(\sum_{n=0}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad + \sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a} - \left(\sum_{m=k}^{K-1} \Delta^m_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) \left(\sum_{n=1}^{k-1} (\Delta^a_{-1} a)^n\right) (D.3)
\end{align*}
\]

Hence, for \( k \geq 2 \) the following identity holds

\[
\begin{align*}
&\quad = \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) + \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) \left(\sum_{n=0}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad + \sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a} - \left(\sum_{m=k}^{K-1} \Delta^m_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{-1} F^H F \Delta^m_{a}\right) \left(\sum_{n=1}^{k-1} (\Delta^a_{-1} a)^n\right) (D.3)
\end{align*}
\]

Inserting (D.3) into (3.37) reveals that

\[
2W_{\text{res},h}(a) = W_{\text{res}}(a) + W_{\text{res}}^H(a) =
\]

\[
\begin{align*}
&\quad = \sum_{k=1}^{K-1} \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right) \\
&\quad + \sum_{k=1}^{K-1} \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right) \left(\sum_{n=0}^{k-1} (\Delta^a_{-1} a)^n\right) \\
&\quad = \sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a} + \sum_{k=2}^{K-1} \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right) \\
&\quad + \sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a} - \sum_{k=2}^{K-1} \left(\sum_{l=0}^{k-1} (\Delta^a_{-1} a^*)^l\right) \left(\sum_{m=k}^{K-1} \Delta^m_{a} F^H F \Delta^m_{a}\right)
\end{align*}
\]
\[
\begin{align*}
&= \sum_{m=1}^{K-1} \Delta_a^m F^H F \Delta_a^m \\
&+ \sum_{k=2}^{K-1} \left( \sum_{l=0}^{k-1} (\Delta_a^{-1} a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \right) \left( \sum_{n=0}^{k-1} (\Delta_a^{-1} a)^n \right) \\
&- \sum_{k=2}^{K-1} \left( \sum_{l=0}^{k-1} (\Delta_a^{-1} a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \right) \left( \sum_{n=0}^{k-1} (\Delta_a^{-1} a)^n \right) \\
&+ \sum_{k=2}^{K-1} \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m + \sum_{m=1}^{K-1} \Delta_a^m F^H F \Delta_a^m \\
&= \sum_{k=1}^{K-1} \left( \sum_{l=0}^{k-1} (\Delta_a^{-1} a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \right) \left( \sum_{n=0}^{k-1} (\Delta_a^{-1} a)^n \right) \\
&- \sum_{k=1}^{K-1} \sum_{l=0}^{k-1} (\Delta_a^{-1} a^*)^l \left( \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \right) \left( \sum_{n=0}^{k-1} (\Delta_a^{-1} a)^n \right) \\
&+ \sum_{k=1}^{K-1} \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \\
&= \sum_{k=1}^{K-1} \left( \sum_{l=0}^{k-1} (\Delta_a^{-1} a^*)^l \right) \Delta_a^{K-1} F^H F \Delta_a^{K-1} \left( \sum_{n=0}^{k-1} (\Delta_a^{-1} a)^n \right) \\
&+ \left( \sum_{l=0}^{K-2} (\Delta_a^{-1} a^*)^l \right) \Delta_a^{K-1} F^H F \Delta_a^{K-1} \left( \sum_{n=0}^{K-2} (\Delta_a^{-1} a)^n \right) \\
&+ \sum_{k=1}^{K-2} \left( 1 - |a|^2 \right) \left( \sum_{l=0}^{k-1} (\Delta_a^{-1} a^*)^l \right) \left( \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \right) \left( \sum_{n=0}^{k-1} (\Delta_a^{-1} a)^n \right) \\
&+ \sum_{k=1}^{K-1} \sum_{m=k}^{K-1} \Delta_a^m F^H F \Delta_a^m \\&= (D.4)
\end{align*}
\]
E  MPs along remaining dimensions and properties

Making use of the definitions introduced in chapter 3, the MP criteria for estimating the generators along the $a$-axis naturally extend to the estimation of the generators $b$ and $c$. Then the following definitions of MPs are in order:

\begin{align*}
M_1(b, E_{S,b} | \text{“p”}) &= T_b^T (b^{-1}) (I_P - E_{S,b} E_{S,b}^H) T_b(b) \quad (E.1) \\
M_1(c, E_{S,c} | \text{“p”}) &= T_c^T (c^{-1}) (I_P - E_{S,c} E_{S,c}^H) T_c(c) \quad (E.2) \\
M_2(b, E_{S,b} | \text{“p”}) &= I_P - E_{S,b}^T T_b(b) \Omega^{-1}_b T_b^T (b^{-1}) E_{S,b} \quad (E.3) \\
M_2(c, E_{S,c} | \text{“p”}) &= I_P - E_{S,c}^T T_c(c) \Omega^{-1}_c T_c^T (c^{-1}) E_{S,c} \quad (E.4) \\
M_3(b, E_{S,b} | \text{“p”}) &= [T_b(b) | E_{S,b}] \quad (E.5) \\
M_3(c, E_{S,c} | \text{“p”}) &= [T_c(c) | E_{S,c}] \quad (E.6) \\
M_4(b, E_{S,b} | \text{“p”}) &= \begin{bmatrix} T_b^T (b^{-1}) T_b(b) & T_b^T (b^{-1}) E_{S,b} \\ E_{S,b}^H T_b(b) & I_P \end{bmatrix} \quad (E.7) \\
M_4(c, E_{S,c} | \text{“p”}) &= \begin{bmatrix} T_c^T (c^{-1}) T_c(c) & T_c^T (c^{-1}) E_{S,c} \\ E_{S,c}^H T_c(c) & I_P \end{bmatrix} \quad (E.8) \\
M_5(b, E_{S,b} | \text{“d”}) &= \begin{bmatrix} \overline{E}_{S,b,1} - \overline{E}_{S,b,1} b^1 \\
\overline{E}_{S,b,2} - \overline{E}_{S,b,1} b^2 \\
\vdots \\
\overline{E}_{S,b,L-1} - \overline{E}_{S,b,L-1} b^{(L-1)} \end{bmatrix} \quad (E.9) \\
M_5(c, E_{S,c} | \text{“d”}) &= \begin{bmatrix} \overline{E}_{S,c,1} - \overline{E}_{S,c,1} c^{1} \\
\overline{E}_{S,c,2} - \overline{E}_{S,c,1} c^{2} \\
\vdots \\
\overline{E}_{S,c,M-1} - \overline{E}_{S,c,M-1} c^{(M-1)} \end{bmatrix} \quad (E.10) \\
M_6(b, E_{S,b} | \text{“d”}) &= \sum_{l=1}^{L-1} (E_{S,b,l}^H \overline{E}_{S,b,l} - E_{S,b,l}^H \overline{E}_{S,b,l} b^{l}) \quad (E.11) \\
M_6(c, E_{S,c} | \text{“d”}) &= \sum_{m=1}^{M-1} (E_{S,c,m}^H \overline{E}_{S,c,m} - E_{S,c,m}^H \overline{E}_{S,c,m} c^{m}) \quad (E.12) \\
M_7(b, E_{S,b} | \text{“d”}) &= \sum_{l=1}^{L-1} (\overline{E}_{S,b,l}^H \overline{E}_{S,b,l} - \overline{E}_{S,b,l}^H \overline{E}_{S,b,l} a^{-l}) \quad (E.13) \\
M_7(c, E_{S,c} | \text{“d”}) &= \sum_{m=1}^{M-1} (\overline{E}_{S,c,m}^H \overline{E}_{S,c,m} - \overline{E}_{S,c,m}^H \overline{E}_{S,c,m} a^{-m}) \quad . \quad (E.14)
\end{align*}
Consistently with the permutation of the rows in the signal matrix $H$ we define the “tall” sparse MPs as

$$T_c(c) = Q_c T_b(c) = (c \otimes I_{KL'})$$

(E.15)

$$T_b(b) = Q_b T_c(b) = (I_M \otimes b \otimes I_K)$$

(E.16)

for $T_n(a)$ defined in (3.3) (recall that $L = L'M$) and

$$c = [1, c, c^2, \ldots, c^{M-1}]^T$$

(E.17)

$$b = [1, b, b^2, \ldots, b^{L'-1}]^T.$$  

(E.18)

Further, according to (2.58), (2.61) and (2.68)-(2.69):

$$H_{b,l} = (I_{KL'} \otimes J_{L',l}) H_b$$

(E.19)

$$H_{c,m} = (I_{KL'} \otimes J_{M,m}) H_c$$

(E.20)

$$H_{b,l} = (I_{KL'} \otimes J_{L',l}) H_b$$

(E.21)

$$H_{c,m} = (I_{KL'} \otimes J_{M,m}) H_c$$

(E.22)

$$E_{S,b,l} = (I_{KL'} \otimes J_{L',l}) E_{S,b}$$

(E.23)

$$E_{S,b,l} = (I_{KL'} \otimes J_{L',l}) E_{S,b}$$

(E.24)

$$E_{S,c,m} = (I_{KL'} \otimes J_{M,m}) E_{S,c}$$

(E.25)

$$E_{S,c,m} = (I_{KL'} \otimes J_{M,m}) E_{S,c}$$

(E.26)

denote the (upper and lower) row-reduced versions of the signal matrices and signal eigenvector matrices, respectively. In equations (E.1)-(E.14)

$$\Delta_b = \text{diag}\{b_1, b_2, \ldots, b_P\}$$

(E.27)

$$\Delta_c = \text{diag}\{c_1, c_2, \ldots, c_P\}$$

(E.28)

denote the diagonal MPs containing the true generators on its main diagonal. Finally, the constant diagonal matrices $\Omega_b$ and $\Omega_c$ read

$$\Omega_b = T_b^T (b^{-1}) T_b(b)$$

(E.29)

$$\Omega_c = T_c^T (c^{-1}) T_c(c)$$

(E.30)

For completeness we list the MI equations along the $b$- and $c$-axis that are obtained in accordance to (2.65) and (2.70) as

$$\overline{H}_{b,l} \Delta_b^l = H_{b,l}$$

(E.31)

$$\overline{H}_{c,m} \Delta_c^m = H_{c,m}$$

(E.32)

$$E_{S,b,l} K \Delta_b^l = E_{S,b,l} K$$

(E.33)

$$E_{S,c,m} K \Delta_c^m = E_{S,c,m} K$$

(E.34)

where $l = 1, \ldots, L' - 1$ and $m = 1, \ldots, M - 1$. 
| MP | description | relations | on the UC (|a| = 1) | inside / outside the UC (|a| ≠ 1) |
|----|-------------|-----------|------------------|-------------------------------|
| M_{1}(a | “p”) in (3.6) | square $L \times L$ MP of degree $2K - 1$, pure HRP | $\sim M_2(a)$, $\sim M_4(a)$, $\sim M_6(a)$, $\sim M_8(a) + M_7(a)$ | no NRs, $\text{rank} \{M_1(a | “p”)\} = \begin{cases} L - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ L, & \text{otherwise}. \end{cases}$ | complex-reciprocal NRs |
| M_{2}(a | “p”) in (3.13) | square $P \times P$ MP of degree $2K - 1$, pure HRP | $\sim M_2(a)$, $\sim M_4(a)$, $\sim M_6(a)$, $\sim M_8(a) + M_7(a)$ | no NRs, $\text{rank} \{M_2(a | “p”)\} = \begin{cases} P - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P, & \text{otherwise}. \end{cases}$ | conjugate-reciprocal NRs |
| M_{3}(a | “d”) in (3.24) | tall $KL \times (P + L)$ MP of degree $K - 1$, damped HRP | $\sim M_3(a)$ | no NRs, $\text{rank} \{M_3(a | “d”)\} = \begin{cases} P + L - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P + L, & \text{otherwise}. \end{cases}$ | no NRs |
| M_{4}(a | “p”) in (3.26) | square $(P + L) \times (P + L)$ MP of degree $2K - 1$, pure HRP | $\sim M_2(a)$, $\sim M_4(a)$, $\sim M_6(a)$, $\sim M_8(a) + M_7(a)$ | no NRs, $\text{rank} \{M_4(a | “p”)\} = \begin{cases} P + L - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P + L, & \text{otherwise}. \end{cases}$ | conjugate-reciprocal NRs |
| M_{5}(a | “d”) in (3.31) | tall $(K - 1)K \times P$ tall MP of degree $K - 1$, damped HRP | $\sim M_3(a)$ | no NRs, $\text{rank} \{M_5(a | “d”)\} = \begin{cases} P - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P, & \text{otherwise}. \end{cases}$ | no NRs |
| M_{6}(a | “d”) in (3.32) | square $P \times P$ MP of degree $K - 1$, damped HRP | $\sim M_2(a)$, $\sim M_4(a)$, $\sim M_6(a)$, $\sim M_8(a) + M_7(a)$ | no NRs, $\text{rank} \{M_6(a | “d”)\} = \begin{cases} P - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P, & \text{otherwise}. \end{cases}$ | $M_{6,h}(a) = (M_6^H(a) + M_6(a)) / 2 \geq 0$ for $|a| < 1$ |
| M_{7}(a | “d”) in (3.50) | square $P \times P$ MP of degree $K - 1$, damped HRP | $\sim M_2(a)$, $\sim M_4(a)$, $\sim M_6(a)$, $\sim M_8(a) + M_7(a)$ | no NRs, $\text{rank} \{M_7(a | “d”)\} = \begin{cases} P - \text{mult} \{a|\mathcal{H}_a\}, & \text{for } a \in \mathcal{H}_a; \\ P, & \text{otherwise}. \end{cases}$ | $M_{7,h}(a) = (M_7^H(a) + M_7(a)) / 2 \geq 0$ for $|a| > 1$ |

Table E.1: Rank properties of MPs along $a$-axis.
| MP       | description                | relations                                                                 | on the UC ($|b| = 1$)                                                                 | inside / outside the UC ($|b| \neq 1$) |
|----------|----------------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------|
| $M_1(b \mid \text{“p”})$ in (E.1) | square $KM \times KM$ MP of degree $2L' - 1$, pure HRP | $\sim M_2(b), \sim M_4(b), \sim M_2(b^{-1})M_3(b), \sim M_6(b) + M_7(b)$ | no NRs, $\text{rank}\{M_1(b \mid \text{“p”})\} = \begin{cases} KM - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ KM, & \text{otherwise.} \end{cases}$ | complex-reciprocal NRs |
| $M_2(b \mid \text{“p”})$ in (E.3) | square $P \times P$ MP of degree $2L' - 1$, pure HRP                 | $\sim M_4(b), \sim M_4(b), \sim M_2(b^{-1})M_3(b), \sim M_6(b) + M_7(b)$ | no NRs, $\text{rank}\{M_2(b \mid \text{“p”})\} = \begin{cases} P - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P, & \text{otherwise.} \end{cases}$ | conjugate-reciprocal NRs |
| $M_3(b \mid \text{“d”})$ in (E.5) | $(L'-1)KL'M \times (P + KM)$ tall MP of degree $2L' - 1$, damped HRP | $\sim M_2(b), \sim M_2(b^{-1})$ | no NRs, $\text{rank}\{M_3(b \mid \text{“d”})\} = \begin{cases} P + KM - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P + KM, & \text{otherwise.} \end{cases}$ | no NRs |
| $M_4(b \mid \text{“p”})$ in (E.7) | square $(P + KM) \times (P + KM)$ MP of degree $2L' - 1$, pure HRP | $\sim M_4(b), \sim M_4(b), \sim M_2(b^{-1})M_3(b), \sim M_6(b) + M_7(b)$ | no NRs, $\text{rank}\{M_4(b \mid \text{“p”})\} = \begin{cases} P + KM - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P + KM, & \text{otherwise.} \end{cases}$ | conjugate-reciprocal NRs |
| $M_5(b \mid \text{“d”})$ in (E.9) | $(L'-1)KL'M \times P$ tall MP of degree $L' - 1$, damped HRP | $\sim M_5(b), \sim M_5(b), \sim M_2(b^{-1})$ | no NRs, $\text{rank}\{M_5(b \mid \text{“d”})\} = \begin{cases} P - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P, & \text{otherwise.} \end{cases}$ | no NRs |
| $M_6(b \mid \text{“d”})$ in (E.11) | square $P \times P$ MP of degree $L' - 1$, damped HRP | $\sim M_5(b) |_{b=0} M_5(b), \sim M_5(b^{-1})$ | no NRs, $\text{rank}\{M_6(b \mid \text{“d”})\} = \begin{cases} P - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P, & \text{otherwise.} \end{cases}$ | $\text{rank}\{M_6(b \mid \text{“d”})\} = \begin{cases} P - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P, & \text{for } |b| < 1. \end{cases}$ for $|b| < 1$ |
| $M_7(b \mid \text{“d”})$ in (E.13) | square $P \times P$ MP of degree $L' - 1$, damped HRP | $\sim M_5(b) |_{b=0} M_5(b^{-1}), \sim M_5(b^{-1})$ | no NRs, $\text{rank}\{M_7(b \mid \text{“d”})\} = \begin{cases} P - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P, & \text{otherwise.} \end{cases}$ | $\text{rank}\{M_7(b \mid \text{“d”})\} = \begin{cases} P - \text{mult}\{b|H_b\}, & \text{for } b \in \mathcal{H}_b; \\ P, & \text{for } |b| > 1. \end{cases}$ for $|b| > 1$ |
| MP | description | relations | on the UC (|c| = 1) | inside / outside the UC (|c| ≠ 1) |
|---|---|---|---|---|
| $M_1(c \mid \text{“p”})$ in (E.2) | square $KL' \times KL'$ MP of degree $2M-1$, pure HRP | $\sim M_2(c), \sim M_4(c), \sim M_2^H(c-1)M_2(c), \sim M_6(c) + M_7(c)$ | no NRs, $\text{rank} \{ M_1(c \mid \text{“p”}) \} = \begin{cases} KM - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ KM, & \text{otherwise.} \end{cases}$ | complex-reciprocal NRs |
| $M_2(c \mid \text{“p”})$ in (E.4) | square $P \times P$ MP of degree $2M-1$, pure HRP | $\sim M_1(c), \sim M_4(c), \sim M_2^H(c-1)M_2(c), \sim M_6(c) + M_7(c)$ | no NRs, $\text{rank} \{ M_2(c \mid \text{“p”}) \} = \begin{cases} P - \text{mult} \{ b|H_c \}, & \text{for } c \in H_c; \\ P, & \text{otherwise.} \end{cases}$ | conjugate-reciprocal NRs |
| $M_3(c \mid \text{“d”})$ in (E.6) | $KL'M \times (P + KL')$ tall MP of degree $M - 1$, damped HRP | $\begin{bmatrix} M_3(c) \\ 0 \end{bmatrix} \sim \begin{bmatrix} I_{KM} & 0 \\ 0 & 0 \\ 0 & M_5(c) \end{bmatrix}$ | no NRs, $\text{rank} \{ M_3(c \mid \text{“d”}) \} = \begin{cases} P + KM - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P + KM, & \text{otherwise.} \end{cases}$ | no NRs |
| $M_4(c \mid \text{“p”})$ in (E.8) | square $(P + KL') \times (P + KL')$ MP of degree $2M-1$, pure HRP | $\sim M_1(c), \sim M_2(c), \sim M_2^H(c-1)M_2(c), \sim M_6(c) + M_7(c)$ | no NRs, $\text{rank} \{ M_4(c \mid \text{“p”}) \} = \begin{cases} P + KM - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P + KM, & \text{otherwise.} \end{cases}$ | conjugate-reciprocal NRs |
| $M_5(c \mid \text{“d”})$ in (E.10) | $(M-1)KL'M \times P$ tall MP of degree $M - 1$, damped HRP | $\begin{bmatrix} M_3(c) \\ 0 \end{bmatrix} \sim \begin{bmatrix} I_{KM} & 0 \\ 0 & 0 \\ 0 & M_5(c) \end{bmatrix}$ | no NRs, $\text{rank} \{ M_5(c \mid \text{“d”}) \} = \begin{cases} P - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P, & \text{otherwise.} \end{cases}$ | no NRs |
| $M_6(c \mid \text{“d”})$ in (E.12) | square $P \times P$ MP of degree $M - 1$, damped HRP | $\sim M_5^H(c) \mid_{c=0} M_5(c), \sim M_7^H(c-1)$ | no NRs, $\text{rank} \{ M_6(c \mid \text{“d”}) \} = \begin{cases} P - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P, & \text{for } |c| < 1. \end{cases}$ | $\text{rank} \{ M_6(c \mid \text{“d”}) \} = \begin{cases} P - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P, & \text{for } |c| < 1. \end{cases}$ |
| $M_7(c \mid \text{“d”})$ in (E.14) | square $P \times P$ MP of degree $M - 1$, damped HRP | $\sim M_5^T(c) \mid_{b=0} M_5^T(c-1), \sim M_6^T(c-1)$ | no NRs, $\text{rank} \{ M_7(c \mid \text{“d”}) \} = \begin{cases} P - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P, & \text{otherwise.} \end{cases}$ | $\text{rank} \{ M_7(c \mid \text{“d”}) \} = \begin{cases} P - \text{mult} \{ c|H_c \}, & \text{for } c \in H_c; \\ P, & \text{for } |c| > 1. \end{cases}$ |

Table E.3: Rank properties of MPs along remaining dimensions and properties.
Finite sample MPs along remaining dimensions

Making use of the definitions introduced in appendix E, we define the following finite sample estimates of the MPs of kind 1, 2, 6, and 7 in the parameters $b$ and $c$:

\[
\hat{M}_1(b, \hat{E}_{S,b} \mid \text{“p”}) = T_b^T (b^{-1}) \left( I_P - \hat{E}_{S,b} \hat{E}_{S,b}^H \right) T_b(b) \quad (F.1)
\]

\[
\hat{M}_1(c, \hat{E}_{S,c} \mid \text{“p”}) = T_c^T (c^{-1}) \left( I_P - \hat{E}_{S,c} \hat{E}_{S,c}^H \right) T_c(c) \quad (F.2)
\]

\[
\hat{M}_2(b, \hat{E}_{S,b} \mid \text{“p”}) = I_P - \hat{E}_{S,b}^H T_b(b) \Omega_b^{-1} T_b^T (b^{-1}) \hat{E}_{S,b} \quad (F.3)
\]

\[
\hat{M}_2(c, \hat{E}_{S,c} \mid \text{“p”}) = I_P - \hat{E}_{S,c}^H T_c(c) \Omega_c^{-1} T_c^T (c^{-1}) \hat{E}_{S,c} \quad (F.4)
\]

\[
\hat{M}_6(b, \hat{E}_{S,b} \mid \text{“d”}) = L' \left( \hat{E}_{S,b,l}^H \hat{E}_{S,b,l} - \hat{E}_{S,b,l} \hat{E}_{S,b,l}^H b_l \right) \quad (F.5)
\]

\[
\hat{M}_6(c, \hat{E}_{S,c} \mid \text{“d”}) = M \left( \hat{E}_{S,c,m}^H \hat{E}_{S,c,m} - \hat{E}_{S,c,m} \hat{E}_{S,c,m}^H c_m \right) \quad (F.6)
\]

\[
\hat{M}_7(b, \hat{E}_{S,b} \mid \text{“d”}) = \sum_{l=1}^{L'-1} \left( \hat{E}_{S,b,l}^H \hat{E}_{S,b,l} - \hat{E}_{S,b,l} \hat{E}_{S,b,l}^H a_l^{-1} \right) \quad (F.7)
\]

\[
\hat{M}_7(c, \hat{E}_{S,c} \mid \text{“d”}) = \sum_{m=1}^{M-1} \left( \hat{E}_{S,c,m}^H \hat{E}_{S,c,m} - \hat{E}_{S,c,m} \hat{E}_{S,c,m}^H a_m^{-1} \right) \quad (F.8)
\]

where

\[
\hat{E}_{S,b} = Q_b \hat{E}_S \quad (F.9)
\]

\[
\hat{E}_{S,c} = Q_c \hat{E}_S \quad (F.10)
\]

and

\[
\hat{E}_{S,b,l} = \left( I_{KM} \otimes J_{L,l} \right) \hat{E}_{S,b} \quad (F.12)
\]

\[
\hat{E}_{S,b,l} = \left( I_{KM} \otimes J_{L,l} \right) \hat{E}_{S,b} \quad (F.13)
\]

\[
\hat{E}_{S,c,m} = \left( I_{KL} \otimes J_{M,m} \right) \hat{E}_{S,c} \quad (F.14)
\]

\[
\hat{E}_{S,c,m} = \left( I_{KL} \otimes J_{M,m} \right) \hat{E}_{S,c} \quad (F.15)
\]

with $l = 1, \ldots, L' - 1$ and $m = 1, \ldots, M - 1$.
### G Deterministic CRB for pure and damped HR

The derivation follows the steps in [SL01]. Consider the signal model in (2.11) for a single snapshot, hence

\[ x = H(\mu, \alpha, \theta, \vartheta) w + n \]  
\[ (G.1) \]

Here, the vectors

\[ \theta = [\theta_1^T, \ldots, \theta_P^T]^T \]  
\[ (G.2) \]

contain the nuisance parameters of the \( P \) harmonics with \( \theta_p \in \mathbb{R}^r, \vartheta \in \mathbb{R}^q, \alpha = [\alpha_1, \ldots, \alpha_P] \in \mathbb{R}^P \) and \( \mu = [\mu_1, \ldots, \mu_P] \in \mathbb{R}^P \) is defined according to section 1.2. The KL \( \times P \) signal matrix is given by

\[ H(\mu, \alpha, \theta, \vartheta) = [h(\theta_1, \alpha_1, \vartheta_1), \ldots, h(\theta_P, \alpha_P, \vartheta_P)] \]  
\[ (G.3) \]

Let \((\cdot)_r\) and \((\cdot)_i\) denote real and imaginary part, respectively. The CRB matrix for the signal parameter vector \([\{w\}_r^T, \{w\}_i^T, \mu^T, \alpha^T, \theta^T, \vartheta^T]^T\) in (G.1) is given by

\[ \text{CRB} = \frac{\sigma^2}{2} [(G^*G)_r]^{-1} \]  
\[ (G.4) \]

where

\[ G = \begin{bmatrix} \frac{\partial H w}{\partial \mu_1}, \ldots, \frac{\partial H w}{\partial \mu_P} \\ \frac{\partial (\{w\}_r^T)}{\partial (\{w\}_i^T)}, \frac{\partial H w}{\partial \alpha_1}, \ldots, \frac{\partial H w}{\partial \alpha_P} \\ \frac{\partial H w}{\partial \theta_{1r}}, \ldots, \frac{\partial H w}{\partial \theta_{Pr}} \\ \frac{\partial H w}{\partial \vartheta_1}, \ldots, \frac{\partial H w}{\partial \vartheta_q} \end{bmatrix} \]  
\[ (G.5) \]

with

\[ D_\mu = \begin{bmatrix} \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)}{\partial \mu_1} \cdots \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)}{\partial \mu_P} \\ \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)}{\partial \alpha_1} \cdots \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)}{\partial \alpha_P} \end{bmatrix} \]  
\[ (G.6) \]

\[ D_\alpha = \begin{bmatrix} \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)}{\partial \alpha_1} \cdots \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)}{\partial \alpha_P} \\ \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)}{\partial \theta_{1r}} \cdots \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)}{\partial \theta_{Pr}} \end{bmatrix} \]  
\[ (G.7) \]

\[ D_{\theta_{ir}} = \begin{bmatrix} \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)}{\partial \theta_{ir}} \cdots \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)}{\partial \theta_{ir}} \\ \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)}{\partial \vartheta} \cdots \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)}{\partial \vartheta} \end{bmatrix} \]  
\[ (G.8) \]

\[ D_\vartheta = \begin{bmatrix} \frac{\partial H w}{\partial \theta_{1r}} \cdots \frac{\partial H w}{\partial \theta_{Pr}} \\ \frac{\partial H w}{\partial \vartheta_1} \cdots \frac{\partial H w}{\partial \vartheta_q} \end{bmatrix} \]  
\[ (G.9) \]

Define

\[ D = [D_\mu, D_\alpha, D_{\theta_{ir}}, D_{\theta_P}, D_\vartheta] \]  
\[ (G.10) \]
and the parameter vector
\[ \tau = [\mu, \alpha, \theta, \phi]. \]  
(G.11)

Introducing the new parameter vector
\[ \left[ ((w)_r + (\Upsilon)_r \tau)^T, ((w)_i + (\Upsilon)_i \tau)^T, \tau^T \right]^T \] (G.12)

where
\[ \Upsilon = (H^H H)^{-1} H^H D \] (G.13)

The CRB for the new parameter vector in (G.12) is related to the original CRB as follows
\[ \text{CRB}_{\text{new}} = \frac{\sigma^2}{2} F \left[ (G^H G)_r \right]^{-1} F^T \] (G.14)

where
\[ F = \begin{bmatrix} I_P & 0 & (\Upsilon)_r \\
0 & I_P & (\Upsilon)_i \\
0 & 0 & I_{(P+2+r)+q} \end{bmatrix} \] (G.15)
\[ F^{-1} = \begin{bmatrix} I_n & 0 & -(\Upsilon)_r \\
0 & I_n & -(\Upsilon)_i \\
0 & 0 & I_{(P+2+r)+q} \end{bmatrix}. \] (G.16)

It is easy to see that
\[ F[((w)_r^T, (w)_i^T, \tau^T)^T = \left[ ((w)_r + (\Upsilon)_r \tau)^T, ((w)_i + (\Upsilon)_i \tau)^T, \tau^T \right]^T \] (G.17)

and that
\[ GF^{-1} = [H, jH, D] F^{-1} = [H, jH, D - H\Upsilon] \]
\[ = [H, jH, \Pi_H D] \] (G.18)

where \( \Pi_H = I - H (H^H H)^{-1} H^H \) Inserting (G.18) into (G.14) yields directly
\[ \text{CRB}_{\text{new}} = \frac{\sigma^2}{2} \begin{bmatrix} (H^H H)_r & -(H^H H)_i & 0 \\
(H^H H)_i & (H^H H)_r & 0 \\
0 & 0 & (D^H \Pi_H^+ D)_r \end{bmatrix}^{-1} \] (G.19)

whose bottom-right corner corresponds to the parameter vector \( \tau \). Note that the CRB only exists if \( (D^H \Pi_H^+ D)_r \) is invertible.

It is clear that in the case of \( N \) independent snapshot the deterministic CRB corresponding to the parameter vector \( \tau \) becomes
\[ \text{CRB}_\tau = \frac{\sigma^2}{2} \begin{bmatrix} \sum_{n=1}^N (D^H (n) \Pi_H^+ D(n))_r \end{bmatrix}^{-1} \] (G.20)
with

\[ D(n) = [D_\mu(n), D_\alpha(n), D_{\theta_1}(n), \ldots, D_{\theta_P}(n), D_\vartheta(n)] \] (G.21)

\[ D_\mu(n) = \left[ \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)w_1(n)}{\partial \mu_1}, \ldots, \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)w_P(n)}{\partial \mu_P} \right] \] (G.22)

\[ D_\alpha(n) = \left[ \frac{\partial h(\mu_1, \alpha_1, \theta_1, \vartheta)w_1(n)}{\partial \alpha_1}, \ldots, \frac{\partial h(\mu_P, \alpha_P, \theta_P, \vartheta)w_P(n)}{\partial \alpha_P} \right] \] (G.23)

\[ D_{\theta_i}(n) = \left[ \frac{\partial h(\mu_i, \alpha_i, \theta_i, \vartheta)w_i(n)}{\partial \theta_{i,1}}, \ldots, \frac{\partial h(\mu_i, \alpha_i, \theta_i, \vartheta)w_i(n)}{\partial \theta_{i,r}} \right] \] (G.24)

\[ D_\vartheta(n) = \left[ \frac{\partial Hw(n)}{\partial \vartheta_1}, \ldots, \frac{\partial Hw(n)}{\partial \vartheta_q} \right] \] (G.25)

Obviously the CRB only exists if \((D^H \Pi_H D)_r\) is invertible.
### H Notation and symbols

#### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>BCM</td>
<td>block companion matrix</td>
</tr>
<tr>
<td>CRB</td>
<td>Cramér-Rao bound</td>
</tr>
<tr>
<td>DFT</td>
<td>discrete Fourier transformation</td>
</tr>
<tr>
<td>DOA</td>
<td>direction-of-arrival</td>
</tr>
<tr>
<td>DOD</td>
<td>direction-of-departure</td>
</tr>
<tr>
<td>FB</td>
<td>forward-backward</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transformation</td>
</tr>
<tr>
<td>GCD</td>
<td>greatest common divisor</td>
</tr>
<tr>
<td>GEV</td>
<td>generalized eigenvector</td>
</tr>
<tr>
<td>GRD</td>
<td>greatest right divisor</td>
</tr>
<tr>
<td>HR</td>
<td>harmonic retrieval</td>
</tr>
<tr>
<td>HRP</td>
<td>harmonic retrieval problem</td>
</tr>
<tr>
<td>LS</td>
<td>least-squares</td>
</tr>
<tr>
<td>LOS</td>
<td>line-of-sight</td>
</tr>
<tr>
<td>TLS</td>
<td>total-least-squares</td>
</tr>
<tr>
<td>MD</td>
<td>multidimensional</td>
</tr>
<tr>
<td>MDL</td>
<td>Minimum Description Length</td>
</tr>
<tr>
<td>MF</td>
<td>matrix function</td>
</tr>
<tr>
<td>MI</td>
<td>multiple invariance</td>
</tr>
<tr>
<td>MIMO</td>
<td>multiple-input multiple-output</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum-Likelihood</td>
</tr>
<tr>
<td>MP</td>
<td>matrix polynomial</td>
</tr>
<tr>
<td>NRs</td>
<td>noise roots</td>
</tr>
<tr>
<td>NSD</td>
<td>nullspace dimension</td>
</tr>
<tr>
<td>RF</td>
<td>radio frequency</td>
</tr>
<tr>
<td>RMSE</td>
<td>root-mean-square-error</td>
</tr>
<tr>
<td>Rx</td>
<td>receive antenna elements</td>
</tr>
<tr>
<td>SRs</td>
<td>signal roots</td>
</tr>
<tr>
<td>Tx</td>
<td>transmit antenna elements</td>
</tr>
<tr>
<td>UC</td>
<td>unit circle</td>
</tr>
<tr>
<td>ULA</td>
<td>uniform linear array</td>
</tr>
</tbody>
</table>
Algorithms

ESPRIT  Estimation of Signal Parameters via Rotation Invariance Techniques
MODE    Method Of Direction-of-arrival Estimation
MUSIC   Multiple-Signal-Classification
MDE     Multi-Dimensional Embedding algorithm
MDF     Multi-Dimensional Folding algorithm
RARE    Rank-Reduction Estimator
SPEC-RARE Spectral Rank-Reduction Estimator
TALS    Trilinear Alternating Least Squares algorithm
tree-MD-RARE tree structured MD Rank-Reduction Estimator
WSF     Weighted Subspace Fitting

Operators and Transformations

\((\cdot)^*\)  complex conjugate
\((\cdot)^T\)  Transpose
\((\cdot)^H\)  Hermitian
\((\cdot)^{-1}\) Inverse
\((\cdot)^\dagger\) Generalized inverse, (Moore-Penrose-Pseudo Inverse)
\(|\cdot|\)  magnitude
\(\|\cdot\|\)  norm of a vector
\(\|\cdot\|_2\)  euclidian norm of a vector
\(\partial(\cdot)/\partial(\cdot)\)  Derivative
\([a]_n\)  \(n\)th element of a vector \(a\)
\([A]_{jn}\)  \(n\)th row of a matrix \(A\)
\(\hat{\cdot}\)  estimate
\(E\)  statistical expectation
\(\text{Re}\)  real part
\(\text{Im}\)  imaginary part
\(\text{vec}\{\cdot\}\)  vectorization
\(\text{diag}\)  diagonal matrix
\((M)_K\)  \(M\) modulo \(K\)
\([k]\)  smallest integer greater or equal \(k\)
\([k]\)  greatest integer smaller or equal \(k\)
\(\mathcal{O}\{\cdot\}\)  Landau-Symbol
rank  Rank of a matrix
\(*\)  convolution
\(\otimes\)  Kronecker-product (A.2)
⊙ Hadamard-product (A.1)
⊙ Khatri-Rao product (A.3)

\[ \mathcal{Z}\{p(n)\}(a) \] z-transform of a sequence \( p(n) \)

\[ \text{FFT}\{p_1(n)\}(k) \] FFT of a sequence \( p_1(n) \)

\[ \text{IFFT}\{P_1(k)\}(n) \] IFFT of a discrete sequence \( P_1(k) \)

\[ \mathcal{V}\{M(a)\}, \mathcal{T}\{M(a)\} \] BCM pair of a MP \( M(a) \) (5.9)-(5.9)

\[ \mathcal{L}\{M(a)\} \] linearized form of a MP \( M(a) \) (5.8)

\[ \mathcal{P}^{\perp}\{U\} \] orthogonal projector onto the nullspace of a matrix \( U \) (6.9)
Constants

- $e$: Euclidian number
- $j$: imaginary unit
- $\pi$: Pi
- $1_{k,1}$: $k \times 1$ vector composed of ones in all entries
- $i_{M,m}$: $m$th column of $I_M$ (4.5)
- $I_M$: $M \times M$ identity matrix
- $0_{M,N}$: $M \times N$ zero matrix
- $0_M$: $M \times 1$ zero vector
- $\Pi_K$: $K \times K$ exchange matrix

Vector spaces, sets and manifolds

- $\mathbb{C}^{m \times n}$: $m \times n$ dimensional vector-space of complex numbers
- $\mathcal{H}_a$: set of true generators along $a$-axis (3.23)
- $\mathcal{H}_b$: set of true generators along $b$-axis
- $\mathcal{H}_c$: set of true generators along $c$-axis
- $\mathcal{M}$: signal subspace spanned by the columns of $H$
- $\mathcal{M}_{\text{org}}$: original signal manifold (3.14)
- $\mathcal{M}_{\text{RARE}}$: RARE manifold (3.15)
- $\mathcal{N}\{U\}$: null space of a matrix $U$
- $\mathcal{P}$: space spanned by the nuisance parameter vector $\theta$
- $\mathcal{Q}$: space spanned by the nuisance parameter vector $\vartheta$
- $\mathbb{R}^{m \times n}$: $m \times n$ dimensional vector-space of real numbers
- $\mathcal{R}\{U\}$: range-space spanned by the columns of a matrix $U$
- $\mathcal{S}$: signal subspace (2.26)
- $\mathbb{Z}^{m \times n}$: $m \times n$ dimensional vector-space of integer numbers

Functions

- $\delta_{n,m}$: Kronecker delta
- $d_M(n-1)$: $n$th polynomial coefficients corresponding to the determinant of $MP M(a)$
- $f_M(\theta, \vartheta, \mu, \alpha)$: inverse MUSIC spectrum (2.72)
- $f_{r-M}(a, b, c)$: root-MUSIC function (2.75)
- $F_{\text{assoc.}}(p, q, r)$: cost function for parameter association (6.20)
- $\text{mult}\{a|\mathcal{H}_a\}$: multiplicity of $a$ in set $\mathcal{H}_a$
- $p(n-1)$: $n$th polynomial coefficient of $P(a)$, i.e. coefficient corresponding to $a^{n-1}$
- $P(a)$: RARE scalar polynomial of degree $2L(K - 1) - 1$ (3.9)
\text{sinc}(\cdot) \quad \text{sinc function}

\lambda_{\text{min}}\{M\} \quad \text{minimum eigenvalue of a matrix } M (6.20)

\sigma_{\text{min}}\{M\} \quad \text{minimum singular value of a matrix } M (4.23)

\textbf{Symbols}

\begin{align*}
    a_p &= e^{\mu_p+j\alpha_p} \quad \text{pth harmonic along the first array axis (1.8)} \\
    b_p &= e^{\nu_p+j\beta_p} \quad \text{pth harmonic along b-axis (1.9)} \\
    c_p &= e^{\xi_p+j\gamma_p} \quad \text{pth harmonic along c-axis} \\
    d_R &= \text{elemental spacings of transmit antenna} \\
    d_T &= \text{elemental spacings of receive antenna} \\
    d_a &= \text{elemental spacings along first array axis} \\
    d_b &= \text{elemental spacings along first array axis} \\
    f_l(\theta_p, \vartheta, \theta_p, \alpha_p) &= \text{the } l\text{th sample of } p\text{th harmonic along the second array axis (1.8).} \\
    K &= \text{sample support along the first array axis} \\
    K_1 &= \text{sample support along first array axis on left side of } \tilde{Y} (2.32) \\
    K_2 &= \text{sample support along first array axis in right side of } \tilde{Y} (2.32) \\
    l_k &= \text{complex factor} \\
    L &= \text{sample support along the second array axis (2D HRP)} \\
    L' &= \text{sample support along the second array axis (3D HRP)} \\
    L_1 &= \text{sample support along second array axis on left side of } \tilde{Y} (2.33) \\
    L_2 &= \text{sample support along second array axis in right side of } \tilde{Y} (2.33) \\
    M &= \text{sample support along the third array axis} \\
    M_1 &= \text{sample support along third array axis on left side of } \tilde{Y} (2.34) \\
    M_2 &= \text{sample support along third array axis in right side of } \tilde{Y} (2.34) \\
    M_a &= \text{multiplicity of } a \text{ in generator set } \{a_1, \ldots, a_P\} \\
    N &= \text{number of time snapshots available} \\
    P &= \text{number of harmonics} \\
    w_p &= \text{complex signal weight of } p\text{th signal} \\
    w_{B,p} &= \text{complex signal weight of } p\text{th signal in backwards approach (2.53)} \\
    \alpha_p &= \text{frequency of } p\text{th harmonic along } a\text{-axis} \\
    \alpha_p &= \text{frequency of } p\text{th signal along evolution axis} \\
    \alpha_p &= \text{propagation delay corresponding to } p\text{th harmonic} \\
    \alpha_p &= \text{azimuth angle corresponding to } p\text{th harmonic} \\
    \beta_p &= \text{frequency of } p\text{th harmonic along } b\text{-axis} \\
    \beta_p &= \text{DOA corresponding to } p\text{th harmonic} \\
    \beta_p &= \text{frequency of } p\text{th signal along detection axis} \\
    \gamma_p &= \text{elevation angle corresponding to } p\text{th harmonic} \\
    \gamma_p &= \text{frequency of } p\text{th harmonic along } c\text{-axis}
\end{align*}
\( \tilde{\gamma}_p \)  
DOD corresponding to \( p \)th harmonic

\( \varepsilon_{a,l} \)  
a-axis displacement of sensor \( (1, l) \) w.r.t. origin \((1.6)\)

\( \varepsilon_{b,l} \)  
b-axis displacement of sensor \( (1, l) \) w.r.t. origin \((1.6)\)

\( \kappa_i \)  
real linear coefficients \((6.18)\)

\( \mu_p \)  
damping factor of \( p \)th harmonic along \( a \)-axis

\( \dot{\mu}_p \)  
damping factor of \( p \)th signal along evolution axis

\( \nu_p \)  
damping factor of \( p \)th harmonic along \( b \)-axis

\( \dot{\nu}_p \)  
damping of \( p \)th signal along detection axis

\( \xi_p \)  
damping factor of \( p \)th harmonic along \( c \)-axis

\( \dot{\xi}_p \)  
damping of \( p \)th signal along detection axis

\( \sigma^2 \)  
noise variance

\( \Omega \)  
constant scalar \((3.12)\)

\( T_s \)  
sampling period

\( T_e \)  
sampling period of evolution phase

\( T_d \)  
sampling period of detection phase

\( a_p \)  
\( K \times 1 \) Vandermonde vector in generator \( a_p \) \((2.3)\)

\( f(\theta_p, \vartheta) \)  
\( p \)th signal component vector along the second array axis \((2.4)\)

\( h(a, \theta, \vartheta) \)  
\( KL \times 1 \) signal (manifold) vector \((3.2)\)

\( k_l \)  
lth column of \( K \)

\( n \)  
\( KL \times 1 \) noise vector \((2.11)\)

\( w \)  
complex signal weight vector \((2.6)\)

\( \hat{w}_{LS} \)  
LS estimate of the signal weight vector \((2.77)\)

\( \hat{w}_{LS,H}(t) \)  
fine sample estimate of the signal weight vector \((2.79)\)

\( x \)  
\( KL \times 1 \) data vector \((2.1)\)

\( \alpha \)  
\( P \times 1 \) vector containing the frequencies of the \( P \) harmonics along the \( a \)-axis.

\( \alpha \)  
\( P \times 1 \) vector containing the frequencies of the \( P \) harmonics along the \( a \)-axis.

\( \mu \)  
\( P \times 1 \) vector containing the damping factors of the \( P \) harmonics along the \( a \)-axis.

\( \hat{\phi}_a \)  
vector containing the estimated generators along \( a \)-axis \((6.51)\)

\( \vartheta \)  
nuisance vector parameterizing the measurement setup \((1.8)\)

\( \theta_p \)  
nuisance parameter vector corresponding to the \( p \)th harmonic \((1.8)\)

\( \hat{\phi}_b \)  
vector containing the estimated generators along \( b \)-axis \((6.52)\)

\( \hat{\phi}_c \)  
vector containing the estimated generators along \( c \)-axis \((6.53)\)

\( A \)  
\( K \times P \) Vandermonde matrix in generator \( a_1, \ldots, a_P \) \((2.7)\)

\( A_i \)  
\( K_i \times P \) Vandermonde matrix in generator \( a_1, \ldots, a_P, i=1,2 \) \((2.37)\)

\( \tilde{A}_k \)  
\( (K - k) \times P \) \( k \)-rows-reduced upper Vandermonde matrix \((2.59)\)

\( \tilde{A}_k \)  
\( (K - k) \times P \) \( k \)-rows-reduced lower Vandermonde matrix \((2.62)\)

\( \tilde{A}_1 \)  
partition of the signal matrix \((6.3)\)
\[ B \]
\( L \times P \) unstructured signal matrix (1.13)
\( L' \times P \) Vandermonde matrix in generator \( b_1, \ldots, b_P \) (2.50)
\[ \tilde{B}_1 \]
partition of the signal matrix (6.4)
\[ B_i \]
\( L_i \times P \) Vandermonde matrix in generator \( b_1, \ldots, b_P \) (2.38)
\[ C \]
\( M \times P \) Vandermonde matrix in generator \( c_1, \ldots, c_P \) (2.51)
\[ \tilde{C}_1 \]
partition of the signal matrix (6.5)
\[ C_i \]
\( M_i \times P \) Vandermonde matrix in generator \( c_1, \ldots, c_P \) (2.40)
\[ D \]
diagonal \( P \times P \) matrix containing singular values of \( \tilde{Y} \) (2.44)
\[ \tilde{D} \]
diagonal \( P \times P \) matrix containing singular values of \( \tilde{Y} \) (2.47)
\[ E_N \]
\( KL \times (KL - P) \) matrix containing the noise eigenvectors (2.19)
\[ \tilde{E}_N \]
\( KL \times (KL - P) \) estimated noise eigenvector matrix (2.22)
\[ E_S \]
\( KL \times P \) matrix containing the signal eigenvectors (2.19)
\[ E_{S,a,k} \]
\( L \times P \) matrix containing specific rows of \( E_S \) (3.43)
\[ E_{S,a,k} \]
\( (K - k)L \times P \) \( k \)-rows reduced signal eigenvector matrix (2.66)
\[ \tilde{E}_{S,a,k} \]
\( (K - k)L \times P \) \( k \)-rows reduced signal eigenvector matrix (2.67)
\[ F \]
signal component matrix along the second array axis (2.8)
\[ H \]
\( KL \times P \) signal matrix (2.9) in 2D HRP
\[ H \]
\( KL'M \times P \) signal matrix (2.48) in 3D HRP
\[ \tilde{H}_1 \]
partition of the signal matrix (6.1)
\[ \tilde{H}_2 \]
partition of the signal matrix (6.1)
\[ H_1 \]
\( K_1L_1M_1 \times P \) left signal matrix (2.36)
\[ H_2 \]
\( K_2L_2M_2 \times P \) right signal matrix (2.36)
\[ H_{n,k} \]
\( (K - k)L \times P \) \( k \)-rows reduced signal matrix (2.61)
\[ \tilde{H}_{a,k} \]
\( (K - k)L \times P \) \( k \)-rows reduced signal matrix (2.9)
\[ J_{K,k} \]
\( K \times K \) selection matrix (2.60)
\[ \tilde{J}_{K,k} \]
\( K \times K \) selection matrix (2.57)
\[ K \]
full rank mixing matrix (2.27)
\[ L_{K,k} \]
\( K \times K \) selection matrix (3.43)
\[ N_{\text{residual}} \]
residual noise term in (2.47)
\[ N_B \]
\( K \times L' \times M \) three-way backwards noise matrix (2.52)
\[ \tilde{N} \]
\( (K_1L_1M_1) \times (K_2L_2M_2) \) reassembled noise matrix (2.41)
\[ N_{l,m} \]
\( K_1 \times K_2 \) noise reassembling matrix (2.43)
\[ N_m \]
\( (K_1L_1) \times (K_2L_2) \) noise reassembling matrix (2.42)
\[ \tilde{N}_B \]
\( (K_1L_1M_1) \times (K_2L_2M_2) \) reassembled backwards noise matrix (2.54)
\[ M_{a,i}(a) \]
MP of kind \( i \) in generator \( a \) (see table E.1).
\[ M_{b,i}(b) \]
MP of kind \( i \) in generator \( b \) (see table E.2).
\[ M_{c,i}(c) \]
MP of kind \( i \) in generator \( c \) (see table E.3).
\[ \tilde{M}_1(b, a_1, E_S) \]
MP of first kind after backsubstitution of \( a_1 \) (6.10)
\( \tilde{M}_1(b_1, a_1, E_S) \) MP of first kind after backsubstitution of \( a_1 \) and \( b_1 \) (6.11)
\( \tilde{M}_2(a, b, c) \) linear combination of MPs along \( a-, b-, \) and \( c-\)axis (6.18)

\( N \) three-way array of size \( K \times L' \times M \) containing noise entries (2.28)

\( P \) weight vector covariance matrix (2.20)

\( \hat{P} \) weight vector sample covariance matrix (2.18)

\( R_t \) single snapshot data covariance matrix at time instance \( t \) (2.16)

\( R \) multiple snapshot data covariance matrix (2.19)

\( \hat{R} \) multiple snapshot sample covariance matrix (2.22)

\( R_B \) backwards covariance matrix (2.56)

\( R_{FB} \) FB covariance matrix (2.56)

\( T(a) \) sparse matrix polynomial of size \( KL \times L \) and degree \( K-1 \) (3.4)

\( T_0 \) matrix polynomial \( T(a) \) evaluated at \( a = 0 \) (B.2)

\( U_1 \) \( K_1 L_1 M_1 \times P \) matrix containing left singular vectors of \( \tilde{Y} \) (2.44)

\( U_2 \) \( K_2 L_2 M_2 \times P \) matrix containing right singular values of \( \tilde{Y} \) (2.44)

\( \hat{U}_1 \) \( K_1 L_1 M_1 \times P \) matrix containing left singular vectors of \( \tilde{Y} \) (2.47)

\( \hat{U}_2 \) \( K_2 L_2 M_2 \times P \) matrix containing right singular values of \( \tilde{Y} \) (2.47)

\( W \) \( P \times P \) matrix containing signal weights on main diagonal (2.40)

\( W_B \) \( P \times P \) diagonal matrix containing backwards signal weights (2.55)

\( W_{res}(a) \) residual MP (3.34)

\( W_{res,h}(a) \) Hermitian part of residual MP (3.37)

\( X \) \( K \times L \) data matrix (1.8)

\( Y \) \( K \times L' \times M \) three-way array containing data samples (1.9)

\( \hat{Y} \) MIMO snapshot in time domain

\( \tilde{Y} \) \( (K_1 L_1 M_1) \times (K_2 L_2 M_2) \) reassembled data matrix (2.29)

\( Y_{l,m} \) \( K_1 \times K_2 \) data reassembling matrix (2.31)

\( Y_m \) \( (K_1 L_1) \times (K_2 L_2) \) data reassembling matrix (2.30)

\( Y_B \) \( K \times L' \times M \) three-way array containing backwards data samples (2.52)

\( \tilde{Y}_B \) \( (K_1 L_1 M_1) \times (K_2 L_2 M_2) \) reassembled backwards data matrix (2.54)

\( \Delta_a \) \( P \times P \) diagonal matrix with true generators on main diagonal (2.64)

\( \Delta_{a,b} \) diagonal matrix to center the origin of the sampling scheme (3.51)

\( \hat{\Lambda}_N \) \( (KL-P) \times (KL-P) \) estimated noise eigenvalue matrix (2.22)

\( \Lambda_N \) \( (KL-P) \times (KL-P) \) matrix containing noise eigenvalues (2.19)

\( \Lambda_S \) \( P \times P \) diagonal matrix containing the signal eigenvalues (2.19)

\( \hat{\Lambda}_S \) \( P \times P \) estimated signal eigenvalue matrix (2.22)

\( \Omega \) constant \( L \times L \) diagonal matrix (3.11)

\( \Omega_b \) constant \( L \times L \) diagonal matrix (E.29)

\( \Omega_c \) constant \( L \times L \) diagonal matrix (E.30)
Bibliography


[ftw] ftw’s MIMO measurements, on-line documentation and selected data sets for download, http://www.ftw.at/measurements/.


Bibliography


