UNIVERSITY OF MINNESOTA

This is to certify that I have examined this copy of a doctoral thesis by

Tao Jiang

and have found that it is complete and satisfactory in all respects, and that any and all revisions required by the final examining committee have been made.

Name of Faculty Advisor

Signature of Faculty Advisor

Date

GRADUATE SCHOOL

Low-Rank Decomposition of Multi-Way Arrays with Applications in Signal Processing and Wireless Communications

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Tao Jiang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Professor Nikolaos D. Sidiropoulos, Advisor

July 2003

To my parents and grandparents

Acknowledgments

Abstract

The proliferation of broadband wireless communications calls for novel signal processing and communications techniques to tackle many new challenging problems, for which traditional approaches either do not provide satisfactory solutions, or cannot solve certain problems at all. Multi-way array analysis is a promising set of tools that has recently been introduced in the context of signal processing and communications. Many successful applications of multi-way analysis in diverse types of problems have demonstrated the potential advantages of multi-way array analysis.

This thesis documents some of the recent developments of multi-way array analysis in the field of signal processing and wireless communications. From the theory perspective, we derive two equivalent necessary and sufficient conditions for unique low rank decomposition of certain multi-way arrays, and point out a strong similarity between the conditions for unique decomposition of bilinear models subject to Constant Modulus (CM) constraints and unique low rank decomposition of multi-way arrays. We also derive the most general identifiability conditions to date for multi-dimensional harmonic retrieval in arbitrary dimensions, with important applications in wireless channel sounding. From the application perspective, based on the theory of multi-way array analysis, we develop a novel receiver to deal with the blind identification of out-of-cell users in Direct-Sequence Code-Division Multiple Access (DS-CDMA) systems. This receiver not only detects the in-cell users' data symbols reliably, but also helps identify out-of-cell transmissions, the steering vectors of all active users and spreading codes of out-of-cell users. We also design an effective blind reception scheme for single input multiple output (SIMO) and multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) subject to unknown frequency offset and multipath.

Contents

List of Figures

List of Acronyms

AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase-Shift-Keying
BS	Base Station
CDMA	Code Division Multiple Access
CM	Constant Modulus
CP	Cyclic Prefix
CSI	Channel State Information
FDMA	Frequency Division Multiple Access
FER	Frame Error Rate
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
HIPERLAN	HIgh PERformance LAN
IBI	Inter Block Interference
IFFT	Inverse Fast Fourier Transform
IS-95	EIA Interim Standard for U.S. code division multiple access
ISI	Inter-Symbol Interference
MC	MultiCarrier
MIMO	Multi-Input Multi-Output
ML	Maximum Likelihood

MLSE	Maximum Likelihood Sequence Estimation
MMSE	Minimum Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
MS	Mobile Station
PAM	Pulse Amplitude Modulation
PAR	Peak-to-Average Power Ratio
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
SIMO	Single-Input Multi-Output
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
ST	Space-Time
TDMA	Time Division Multiple Access
WLAN	Wireless Local Area Network
ZF	Zero Forcing

Notation

Bold upper letters (e.g. $\mathbf{H}$)	matrices;
Bold lower letters (e.g. $\mathbf{h}$)	column vectors;
$(\cdot)^*$	conjugate;
$(\cdot)^T$	transpose;
$(\cdot)^H$	Hermitian transpose;
$(\cdot)^\dagger$	pseudo-inverse;
$\mid \cdot \mid$	absolute value of a scalar;
$\det(\cdot)$	determinant of a matrix;
$r_{\mathbf{A}}, rank(\mathbf{A})$	rank of matrix $\mathbf{A}$;
$k_{\mathbf{A}}, k - rank(\mathbf{A})$	the Kruskal-rank of matrix $\mathbf{A}$, the maximum number of linearly independent columns that can be drawn from $\mathbf{A}$ in an arbitrary fashion;
$\Vert \cdot \Vert$	Frobenius norm;
$\delta[\cdot]$	Kronecker's delta;
$\otimes$	Kronecker product;
$\star$	convolution;
$\lceil \cdot \rceil$	integer ceiling;
$\lfloor \cdot \rfloor$	integer floor;
$\mathbf{E}\{\cdot\}$	expectation;
$\mathrm{tr}\{\cdot\}$	trace of a matrix;
$\mathrm{diag}(\mathbf{x})$	diagonal matrix with $\mathbf{x}$ on its diagonal;

$\mathbf{I}_N$	the identity matrix of size N;
$\mathbf{F}_N$	$N \times N$ FFT matrix with $(p+1, q+1)$st entry $(1/\sqrt{N})e^{-j2\pi pq/N}$; $\forall p, q \in [0, N-1]$;
$[\cdot]_p$	the p-th entry of a vector;
$[\cdot]_{p,q}$	the (p, q)-th entry of a matrix;
$\mathbb{C}$	the field of complex numbers;
$\mathbb{R}$	the field of real numbers;
$\mathbf{a}_i$	the ith column of $\mathbf{A}$;
$\mathbf{A}(i, f)$	the (i, f)th element of $\mathbf{A}$, which is also denoted by $a_{i,f}$;
$\mathbf{A}^{(m)}$	a sub-matrix of $\mathbf{A}$, formed by its first m rows;
$\mathbf{D}_i(\mathbf{A})$	a diagonal matrix constructed from the ith row of $\mathbf{A}$;
$\mathbf{A} \otimes \mathbf{B}$	the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$;
$\mathbf{A} \odot \mathbf{B}$	the Khatri-Rao (column-wise Kronecker) product of $\mathbf{A}$ and $\mathbf{B}$: $\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \cdots \mathbf{a}_F \otimes \mathbf{b}_F]$, where $\mathbf{A} \in \mathbb{C}^{I \times F}$, $\mathbf{B} \in \mathbb{C}^{J \times F}$;
$\mathbf{A} \diamond \mathbf{B}$	the Hadamard (element-wise) product of $\mathbf{A}$ and $\mathbf{B}$;
$\omega(\mathbf{x})$	the number of non-zero elements of vector $\mathbf{x}$;
$< f, g >$	the L^2 inner product between functions f and g.

Chapter 1

Introduction

1.1 Motivation and Background

In the past decade, the unprecedented growth in demand for reliable high-rate wireless multimedia services has triggered considerable research efforts to develop new effective and efficient signal processing schemes for improving the quality and spectral efficiency of broadband wireless links in the presence of channel fading. Broadband wireless services must cope with many critical performance-limiting challenges that include multiuser interference (MUI) caused by simultaneous transmission in the presence of inter-symbol interference (ISI), multi-carrier interference (MCI) due to unknown carrier frequency offset (CFO) and time-selective and frequency-selective fading induced by Doppler and multipath propagation. Traditional approaches either do not always give satisfactory solutions to those challenges, or cannot solve certain problems at all. Thus novel techniques in signal processing and wireless communications are called for tackling those challenges effectively and efficiently. In recent years, so-called *diversity* techniques have been introduced to help mitigate those impairments. Receive diversity, for example, can provide multiple independently faded copies of the same signal. The basic idea is that receive diversity increases the ratio of the number of independent equations over the number of unknowns. Spreading or frequency diversity yields a similar effect, albeit at the expense of bandwidth. Transmit diversity, on the other hand, helps to average-out fading effects, thereby enhancing average signal-to-noise ratio (SNR) at the receiver. Transmit diversity is different from other kinds of diversity, in the sense that it does not yield more equations, but rather "better-conditioned" equations. All other commonly used forms of diversity improve the ratio of the number of independent equations over the number of unknowns. When multiple such forms of diversity are simultaneously present, the baseband-equivalent data constitutes a multi-way data array. More often than not, the signal part of this data array can be modeled as a low-rank multi-way array whose rank is equal to the number of co-channel signals.

Multi-way array analysis in general, and low-rank multi-way array decomposition in particular, are promising sets of tools that have recently been introduced in modern signal processing and communications. Many successful applications of multi-way analysis in diverse types of problems have shown strong evidence of the advantages of multi-way array analysis. The primary purpose of this thesis was to tackle certain challenging problems in the field of signal processing and wireless communications via multi-way array analysis. However, the field of multi-way array analysis is still far from mature, despite its theoretical and practical significance. Hence, a fair part of this thesis is devoted to basic research and methodological developments related to multi-way array analysis.

1.2 Multi-way Array Analysis

Matrices (two-dimensional, or *two-way* arrays) are the most commonly encountered arrays. Low rank matrix decomposition is not unique in general, since inserting a product of an arbitrary invertible matrix and its inverse in between two matrix factors preserves their product, yet generates a different decomposition. This is known as the rotational indeterminacy problem in factor analysis. Low rank matrix decomposition can be unique only if one imposes additional problem-specific structural properties on the model or constraints on data to obtain an unique parameterization. Examples include orthogonality (as in singular value decomposition), Vandermonde, Toeplitz, constant modulus (CM) or finite-alphabet (FA) constraints.

Compared to the case of matrices, low-rank decomposition for three-way arrays has a distinctive and attractive feature: it is often unique [30], and the situation actually improves in higher dimensions [51]. Low rank decomposition of three-dimensional arrays was developed by Harshman [17–19] under the name PARAllel FACtor (PARAFAC) analysis, and independently by Carroll and Chang [7] under the name CANonical DECOMPosition (CANDECOMP). More recently, the term CP (CANDECOMP/PARAFAC) decomposition is often preferred [7, 17]. The deepest piece of work on uniqueness of CP decomposition of three-way arrays is due to J. B. Kruskal [30, 31]. Since then, the conditions for unique CP decomposition of three-way arrays have been gradually improved by various authors [28, 48, 51, 54, 62].

The uniqueness of low-rank decomposition for three- and higher-way arrays is of great practical significance since it enables unambiguous interpretation of the estimated model parameters, which is crucial in many applications. Therefore, after its introduction as a data analysis tool in Psychometrics, CP decomposition has gradually gained popularity in many fields and disciplines such as statistics, arithmetic complexity, and chemometrics [1, 4, 7, 19, 30, 32] over the past thirty years. In recent years many intriguing applications have been found in problems of interest in signal processing and wireless communications. In the latter context, three-way methods are often naturally applicable for the analysis of multi-dimensional data sets encountered in multiple-invariance sensor array processing [53], blind beamforming in specular multipath [35, 55], multiuser signal separation [23, 37, 54], and diversity systems [24, 52].

1.3 Applications

The primary purpose of this thesis was to tackle certain challenging problems emerging in the field of signal processing and wireless communications via multi-way array analysis. The following sections provide some background on the applications addressed in this thesis.

1.3.1 Multi-Dimensional Harmonic Retrieval

The problem of harmonic retrieval is commonly encountered under different disguises in diverse applications in the sciences and engineering [58]. Although one-dimensional harmonic retrieval is most common, many applications of multi-dimensional harmonic retrieval can be found in radar (e.g., [21, 34] and references therein), passive range-angle localization [57], joint 2-D angle and carrier frequency estimation [68, 69], and wireless channel sounding [13–16]. In wireless channel sounding, for example, one is interested in jointly estimating several multipath signal parameters like azimuth, elevation, delay, and Doppler, all of which can often be viewed as or transformed into frequency parameters.

A plethora of one-dimensional as well as multi-dimensional harmonic retrieval techniques have been developed, ranging from non-parametric Fourier-based methods, to modern parametric methods which are not bound by the Fourier resolution limit. In the high signal-to-noise ratio (SNR) regime, parametric methods work well with only a limited number of samples.

One important issue with parametric methods is to determine the maximum number of harmonics that can be resolved for a given total sample size; another is to determine the sample size needed to meet performance specifications.

Identifiability-imposed bounds on sample size are often not the issue in time series analysis, because samples are collected along the temporal dimension (hence "inexpensive"), and performance considerations dictate many more samples than what is needed for identifiability. The maximum number of resolvable harmonics comes back into play in situations where data samples along the harmonic mode come at a premium, e.g., in spatial sampling for direction-of-arrival estimation using a uniform linear array (ULA), in which case one can meet performance requirements with few spatial samples but many temporal samples [59].

Determining the maximum number of resolvable harmonics is a parameter identifiability problem, whose solution for the case of one-dimensional harmonics goes back to Carathéodory [6]; see also [45, 60]. In two or higher dimensions, the identifiability problem is considerably harder, but also more interesting. The reason is that, in many applications of higher-dimensional harmonic retrieval, one is constrained in the number of samples that can be taken along certain dimensions, usually due to hardware and/or cost limitations. Examples include ultrasound imaging [10] and direction of arrival (spatial frequency) estimation. The question that arises is whether the number of samples taken along any particular dimension bounds the overall number of resolvable harmonics or not.

1.3.2 Out-of-cell Interference in DS-CDMA

In the context of uplink reception for cellular DS-CDMA systems, interference can be classified as either (i) interchip (ICI) and intersymbol (ISI) self-interference, (ii) in-cell multiuser access interference (commonly referred to as MUI or MAI), or (iii) out-of-cell multiuser access interference. The latter is typically ignored, or treated as noise; however, it has been reported [49] that in IS-95 other cells account for a large percentage of the interference relative to the interference coming from within the cell. MUI is usually a side-effect of propagation through dispersive multipath channels. The conceptual difference between in-cell and out-of-cell interference boils down to what the base station (BS) can assume about the nature of interfering signals. Typically, the codes of interfering in-cell users are known to the BS, whereas those of out-of-cell users are not. Specifically, in the presence of ICI, the receive-codes of the in-cell users can be estimated via training or subspace techniques (e.g., cf. [39]), using the fact that the transmit-codes are known. This is not the case for out-of-cell users.

Appealing to the central limit theorem, the total interference from out-of-cell users is usually treated as Gaussian noise. In IS-95, a long random cell-specific code is overlaid on top of symbol spreading, and cell despreading is used at the BS to randomize out-of-cell interference. This helps mitigate out-of-cell interference in a statistical fashion. To see how random cell codes work, consider the simplified synchronous flat-fading baseband-equivalent received data model

$$\mathbf{x} = \mathbf{D}_{in}\mathbf{C}_{in}\mathbf{s}_{in} + \mathbf{D}_{out}\mathbf{C}_{out}\mathbf{s}_{out} + \mathbf{n},$$

where $\mathbf{x}$ holds the received data corresponding to one symbol period, $\mathbf{C}_{in}$ (resp. $\mathbf{C}_{out}$) is the spreading code matrix, $\mathbf{s}_{in}$ (resp. $\mathbf{s}_{out}$) is the symbol vector, $\mathbf{D}_{in}$ (resp. $\mathbf{D}_{out}$) is a diagonal matrix that holds a portion of the random cell code for the in-cell (resp. out-of-cell) users, and $\mathbf{n}$ models receiver noise. For simplicity, assume that the in-cell symbol-periodic codes are orthogonal of length P, and all codes and symbols are BPSK. Let $\mathbf{c}_1$ stand for the code of an in-cell user of interest. Then

$$z_1 := \frac{1}{P}\mathbf{c}_1^T\mathbf{D}_{in}\mathbf{x} = \mathbf{s}_{in}(1) + \frac{1}{P}\mathbf{c}_1^T\mathbf{D}_{in}\mathbf{D}_{out}\mathbf{C}_{out}\mathbf{s}_{out} + \widetilde{\mathbf{n}}.$$

The interference term is zero-mean; under certain conditions, its variance is $O(\frac{1}{P})$. This is easy to see for a single out-of-cell user. It follows that random cell codes work reasonably well in relatively underloaded systems with large spreading gain (e.g., 128 chips/symbol), but performance can suffer from near-far effects, and cell codes cannot help identify out-of-cell transmissions. Although the latter may seem of little concern in commercial applications, it can be important for tracking, hand-off, and monitoring.

In a way, a structured approach towards the explicit identification of out-of-cell users is the next logical step beyond in-cell multiuser detection, and is motivated by considerations similar to those that stimulated research that took us from matched filtering to multiuser detection. Note that, unlike the case of in-cell interference, out-of-cell interference cannot be mitigated by power control, simply because the BS does not have the authority to exercise power control over out-of-cell users. For a power-controlled in-cell population, near-far effects may be chiefly due to out-of-cell interference. Unfortunately, out-of-cell detection is compounded by the fact that it has to be blind, since the BS has no control and usually no prior information on out-of-cell users. This places limitations on the number and nature of out-of-cell transmissions that can be identified.

1.3.3 Orthogonal Frequency Division Multiplexing

The proliferation of broadband multimedia wireless communication systems such as Digital Audio Broadcasting (DAB), Digital Video Broadcasting (DVB), North-American IEEE 802.11a, and European HIPERLAN-2, has spawned considerable research on the design of wireless Orthogonal Frequency Division Multiplexing (OFDM) transceivers. OFDM can be used for frequency-selective point-to-point links, possibly in conjunction with transmit and receive diversity, but also for multiplexing. Regardless of the particular transmission modality, OFDM affords relatively easy equalization, and the ability to exploit frequency selectivity via symbol and power loading.

One of the weak points of OFDM is sensitivity to Carrier Frequency Offset (CFO), which is induced by local oscillator frequency mismatch between the transmitter and the receiver, drift, or Doppler effects due to mobility in the case of wireless systems. The net result is that orthogonality among OFDM subcarriers is destroyed, and the resulting inter-carrier interference (ICI) significantly degrades OFDM system performance. With the FFT-based OFDM receiver, severe degradation occurs even for relatively small CFO - about $2-5\%$ of subcarrier spacing. For this reason, accurate CFO estimation is necessary.

1.4 How to read this Thesis

This thesis consists of two parts. Part I (Chapters 2 - 3) is devoted to theory, while Part II (Chapters 4 - 5) contains the applications.

1.4.1 Part I: Theory

Chapter 2 - Multi-way Array Analysis. Here we introduce the CP model and study its uniqueness properties. Our main result consists of two equivalent necessary and sufficient conditions for unique identification of certain CP models. A strong similarity between the conditions for unique decomposition of bilinear models subject to CM constraints and certain CP models is pointed out. Our results yield the first necessary and sufficient conditions for unique CP decomposition and unique bilinear decomposition under CM constraints. Algorithms for fitting the CP model are also presented and discussed.

Chapter 3 - Multi-dimensional Harmonic Retrieval. We derive the most general identifiability conditions to date for multidimensional harmonic retrieval in arbitrary dimensions. Our proof has subsequently provided the backbone for a new and effective algebraic 2-D harmonic retrieval algorithm [37].

1.4.2 Part II: Applications

Chapter 4 - Blind Identification of Out-of-cell Users in DS-CDMA. Based on the theory of multi-way array analysis, we develop an algebraic solution under the premise that the codes of the in-cell users are known. The codes of out-of-cell users and all array steering vectors are unknown. In this pragmatic scenario, we show that in addition to algebraic solution, better identifiability is possible. Our approach yields the best known identifiability result for three-dimensional low-rank decomposition when one of the three component matrices is partially known, albeit non-invertible. A pertinent asymptotic (symbol-independent) Cramér-Rao bound is also presented.

Chapter 5 - Direct Blind Receiver for SIMO and MIMO OFDM Subject to Frequency Offset. We show that by employing two or more antennas at the receiver affords not only a direct receive-diversity benefit, but also important side-benefits as well: in fact CFO can be blindly estimated and the transmitted symbols can be directly recovered, under very relaxed blind identifiability conditions. The results are general enough to cover both single input multiple output (SIMO) and multiple input multiple output (MIMO) OFDM systems with multiple users or multiple transmit antennas.

1.4.3 Conclusions

Chapter 6 - Conclusions and Future Work. In this chapter, we give a brief summary of the thesis, discuss some open questions, and point out directions for future work.

The results of this thesis have been reported in Journal papers [25–28]; and conference publications [22–24].

Part I

Theory

Chapter 2

Multi-way Array Analysis

2.1 CANDECOMP/PARAFAC (CP) Model and Problem Statement

The CP model was independently introduced in 1970 by two different groups, as CANonical DECOMPosition [7] and PARAllel FACtor analysis [17], respectively. Let $\underline{\mathbf{X}}$ be a three way array of order $I \times J \times K$. Analogous to matrix rank, which can be defined as the smallest number of rank-one matrices (outer products of two vectors) that generate the said matrix as their sum, the rank of a three-way array $\underline{\mathbf{X}}$ is defined as the smallest number of rank-one three-way arrays that generate $\underline{\mathbf{X}}$ as their sum. A rank-one three-way array is the outer product of three vectors. The CP model decomposes $\underline{\mathbf{X}}$ into a sum of F rank-one three-way factors as follows

$$x_{i,j,k} = \sum_{f}^{F} a_{i,f} b_{j,f} c_{k,f} \tag{2.1}$$

where the residual terms have been ignored because they play no role in identification. Define an $I \times F$ matrix $\mathbf{A}$ with typical element $\mathbf{A}(i,f) = a_{i,f}$, $J \times F$ matrix $\mathbf{B}$ with $\mathbf{B}(j,f) = b_{i,f}$, $K \times F$ matrix $\mathbf{C}$ with $\mathbf{C}(k,f) = c_{k,f}$, and $J \times K$ matrices $\mathbf{X}_i$ with

$\mathbf{X}_i(j,k) = x_{i,j,k}$. The model in (2.1) can be written as

$$\mathbf{X}_i = \mathbf{B}\,diag(\mathbf{A}_i)\mathbf{C}^T, \tag{2.2}$$

where $\mathbf{A}_i$ stands for the i-th row of $\mathbf{A}$. If we stack the $\mathbf{X}_i$ one over another, a compact matrix representation of the model in (2.1) is possible by employing the Khatri-Rao (column-wise Kronecker) product

$$\mathbf{X}^{JI\times K} := \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_I \end{bmatrix} = \begin{bmatrix} \mathbf{B}\,diag(\mathbf{A}_i) \\ \mathbf{B}\,diag(\mathbf{A}_2) \\ \vdots \\ \mathbf{B}\,diag(\mathbf{A}_I)) \end{bmatrix} \mathbf{C}^T = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T. \tag{2.3}$$

By symmetry, $\mathbf{A}, \mathbf{B}, \mathbf{C}$ may switch their places in (2.3) if the modes of the array are switched accordingly.

Suppose we have two different decompositions of the same array $\underline{\mathbf{X}}$, namely, there is a triple $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$ other than $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ such that $(\mathbf{A} \odot \mathbf{B})\mathbf{C}^T = (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T$. Note that if $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{\Lambda_A}, \mathbf{\Lambda_B}, \mathbf{\Lambda_C}$ are diagonal matrices such that $\mathbf{\Lambda_A}\mathbf{\Lambda_B}\mathbf{\Lambda_C} = \mathbf{I}$, then $(\mathbf{A}\mathbf{\Pi}\mathbf{\Lambda_A}, \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda_B}, \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda_C})$ will yield the same array given by $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. A CP decomposition of (2.3) is therefore said to be unique if for every other decomposition of (2.3), $\mathbf{X}^{JI\times K} = (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T$, we have $\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}\mathbf{\Lambda_A}, \bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda_B}, \bar{\mathbf{C}} = \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda_C}$ for some permutation matrix $\mathbf{\Pi}$ and diagonal matrices $\mathbf{\Lambda_A}, \mathbf{\Lambda_B}, \mathbf{\Lambda_C}$ with $\mathbf{\Lambda_A}\mathbf{\Lambda_B}\mathbf{\Lambda_C} = \mathbf{I}$.

The key problem addressed in this chapter is to pursue necessary and sufficient conditions under which CP decomposition is unique. Kruskal has shown [30] that uniqueness holds under relatively mild conditions. A conjecture that these conditions are generally necessary and sufficient has been upheld until the recent work [62]. An approach of finding alternative CP decompositions for any given CP model has been developed in [62] to study the uniqueness of CP decomposition and Kruskal's conjecture was failed by simple counterexample. However, [62] did not further qualify the uniqueness conditions. In what follows, we will derive two equivalent necessary and sufficient uniqueness conditions when one of the component matrices involved in the decomposition is full column rank, and explain the examples in [62]. As a bonus, we will establish a link to uniqueness of bilinear factorization under CM constraints. Last but not least, we will provide a more palatable proof of a cornerstone result in the area, Kruskal's Permutation Lemma, that at least we would have appreciated being readily available several years ago.

2.2 Roadmap of Uniqueness Results

Recall that CP decomposition, when unique, it is unique up to a common permutation and non-singular scaling/counter-scaling of columns of the component matrices. On hindsight, it is therefore natural to ask under what conditions two matrices are the same up to permutation and scaling of columns. This is precisely the subject of Kruskal's Permutation Lemma [30]:

Lemma 2.1 *[30] Suppose we are given two matrices* $\mathbf{A}$ *and* $\bar{\mathbf{A}}$*, which are* $I \times F$ *and* $I \times \bar{F}$*. Suppose* $\mathbf{A}$ *has no zero columns. If for any vector* $\mathbf{x} \in \mathbb{C}^N$ *such that*

$$\omega(\mathbf{x}^H \bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}} + 1, \tag{2.4}$$

where $\omega(\mathbf{x})$ *stands for the number of non-zero elements of* $\mathbf{x}$*, we have*

$$\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}}),$$

then $F \leq \bar{F}$*; if also* $F \geq \bar{F}$*, then* $F = \bar{F}$*, and there exist a permutation matrix* $\mathbf{P}_{\bar{\mathbf{A}}}$ *and a non-singular diagonal matrix* $\mathbf{\Lambda}$ *such that* $\mathbf{A} = \bar{\mathbf{A}}\mathbf{P}_{\bar{\mathbf{A}}}\mathbf{\Lambda}$*.*

This lemma is *the* key tool in the area of CP analysis and the cornerstone for Kruskal's proof of uniqueness of CP decomposition. Since the statement reads as a sufficient condition, it is tempting to attempt to improve upon Kruskal's condition for uniqueness of CP decomposition by sharpening the condition in the permutation lemma. However, Kruskal's proof of the permutation lemma is ingenious but also largely inaccessible. We managed to re-prove Kruskal's Permutation Lemma using a systematic basic linear algebra and induction approach (see the Appendix). The new proof suggests that the condition in Kruskal's Permutation Lemma is sharp, hence the aforementioned attempt is unlikely to succeed.

Necessary conditions for CP uniqueness are worth recounting at this point. One is that neither $\mathbf{A}$, nor $\mathbf{B}$, nor $\mathbf{C}$ has a pair of proportional columns [29]. Another is that the Khatri-Rao product of any two component matrices must be full column rank [36].

On hindsight, the proof of uniqueness of CP decomposition can be decoupled into three separate steps. Given $(\mathbf{A} \odot \mathbf{B})\mathbf{C}^T = (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T$, the first step is to show that $\bar{\mathbf{A}} = \mathbf{A}\boldsymbol{\Pi}_{\mathbf{A}}\boldsymbol{\Lambda}_{\mathbf{A}}, \bar{\mathbf{B}} = \mathbf{B}\boldsymbol{\Pi}_{\mathbf{B}}\boldsymbol{\Lambda}_{\mathbf{B}}$; the second step is to show that $\boldsymbol{\Pi}_{\mathbf{A}} = \boldsymbol{\Pi}_{\mathbf{B}} = \boldsymbol{\Pi}$; the last step is to show $\bar{\mathbf{C}} = \mathbf{C}\boldsymbol{\Pi}(\boldsymbol{\Lambda}_{\mathbf{A}}\boldsymbol{\Lambda}_{\mathbf{B}})^{-1} = \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{C}}$. This last step is straightforward once the previous steps are finished:

$$(\mathbf{A} \odot \mathbf{B})\mathbf{C}^T = (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T = ((\mathbf{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{A}}) \odot (\mathbf{B}\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{B}}))\bar{\mathbf{C}}^T = (\mathbf{A} \odot \mathbf{B})\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{A}}\boldsymbol{\Lambda}_{\mathbf{B}}\bar{\mathbf{C}}^T,$$

and since $\mathbf{A} \odot \mathbf{B}$ is full column rank (recall this is one of the necessary conditions for uniqueness), we have

$$\mathbf{C}^T = \boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{A}}\boldsymbol{\Lambda}_{\mathbf{B}}\bar{\mathbf{C}}^T,$$

or

$$\bar{\mathbf{C}} = \mathbf{C}\boldsymbol{\Pi}(\boldsymbol{\Lambda}_{\mathbf{A}}\boldsymbol{\Lambda}_{\mathbf{B}})^{-1}.$$

When one of the component matrices, say $\mathbf{C}$, is full column rank, the aforementioned procedure can be further simplified. One can first show that $\bar{\mathbf{C}} = \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{C}}$, then obtain

$$(\mathbf{A} \odot \mathbf{B})\mathbf{C}^T = (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T = (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\boldsymbol{\Lambda}_{\mathbf{C}}\boldsymbol{\Pi}^T\mathbf{C}^T,$$

and since $\mathbf{C}$ is full column rank,

$$\bar{\mathbf{A}} \odot \bar{\mathbf{B}} = (\mathbf{A} \odot \mathbf{B})\mathbf{\Pi}\mathbf{\Lambda}_{\mathbf{C}}^{-1},$$

it then follows that $\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}\mathbf{\Lambda}_{\mathbf{A}}, \bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda}_{\mathbf{B}}$ for some $\mathbf{\Lambda}_{\mathbf{A}}$ and $\mathbf{\Lambda}_{\mathbf{B}}$, such that $\mathbf{\Lambda}_{\mathbf{A}}\mathbf{\Lambda}_{\mathbf{B}}\mathbf{\Lambda}_{\mathbf{C}} = \mathbf{I}$. Therefore, when $\mathbf{C}$ is full column rank, showing $\bar{\mathbf{C}} = \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}_{\mathbf{C}}$ is the key step. In Section 2.3, we will derive conditions under which this step can be accomplished. When those conditions do not hold, alternative CP decompositions will be constructed.

2.3 Main Results

We now focus on proving uniqueness for restricted CP models, meaning those with full column rank $\mathbf{C}$. As we have seen in the previous section, one way to show that the decomposition of restricted CP models is unique is to prove that the full column rank component matrix $\mathbf{C}$ is unique up to permutation and column scaling. This entails conditions on both $\mathbf{A}$ and $\mathbf{B}$. $k_{\mathbf{A}} + k_{\mathbf{B}} \geq F + 2$, as a special case of Kruskal's conditions [30], can achieve the goal, but as shown in [62] this condition is not necessary.

The following condition will be proven to be necessary and sufficient to show that $\mathbf{C}$ is unique up to permutation and column scaling:

Condition A: None of the non-trivial linear combinations of columns of $\mathbf{A} \odot \mathbf{B}$ *can be written as a tensor product of two vectors.*

By *non-trivial* linear combinations we mean those involving *at least two* columns of $\mathbf{A} \odot \mathbf{B}$.

Clearly, Condition A implies that $\mathbf{A} \odot \mathbf{B}$ is full column rank, since if $\mathbf{A} \odot \mathbf{B}$ is rank deficient, a non-trivial linear combination of columns of $\mathbf{A} \odot \mathbf{B}$ would constitute a zero vector and this zero vector can be given in the form of tensor product of a zero vector and another vector.

Condition A also implies that neither $\mathbf{A}$ nor $\mathbf{B}$ has a pair of proportional columns. Otherwise, one can arrange a non-trivial linear combination of columns of $\mathbf{A}\odot\mathbf{B}$ such that the resulting vector is in the form of tensor product of two vectors.

Now, let us see why this condition is sufficient for the identification of restricted CP models. Suppose we have another decomposition of the same array $\underline{\mathbf{X}}$, $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$, such that $(\mathbf{A}\odot\mathbf{B})\mathbf{C}^T = (\bar{\mathbf{A}}\odot\bar{\mathbf{B}})\bar{\mathbf{C}}^T$. Thanks to the non-singularity of $\mathbf{A}\odot\mathbf{B}$ implied by Condition A, it can be seen that $\mathbf{C}^T\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$ such that $\bar{\mathbf{C}}^T\mathbf{x} = \mathbf{0}$. This implies that $r_C \leq r_{\bar{\mathbf{C}}}$. Since $\mathbf{C}$ is assumed full column rank, $\bar{\mathbf{C}}$ has to be full column rank as well.

To further proceed to show that $\bar{\mathbf{C}}$ is the same as $\mathbf{C}$ up to permutation and column scaling, we resort to Kruskal's Lemma. It suffices to show that $\omega(\mathbf{x}^H\mathbf{C}) \leq \omega(\mathbf{x}^H\bar{\mathbf{C}})$ for all $\omega(\mathbf{x}^H\bar{\mathbf{C}}) \leq 1$. Clearly, we only need to verify that $\omega(\mathbf{x}^H\mathbf{C}) \leq \omega(\mathbf{x}^H\bar{\mathbf{C}})$ holds for all $\omega(\mathbf{x}^H\bar{\mathbf{C}}) = 1$.

Given all $\mathbf{x}$ such that $\omega(\mathbf{x}^H\bar{\mathbf{C}}) = 1$, we have

$$(\bar{\mathbf{A}}\odot\bar{\mathbf{B}})\bar{\mathbf{C}}^T(\mathbf{x}^H)^T = (\mathbf{A}\odot\mathbf{B})\mathbf{C}^T(\mathbf{x}^H)^T.$$

Note that $(\bar{\mathbf{A}}\odot\bar{\mathbf{B}})\bar{\mathbf{C}}^T(\mathbf{x}^H)^T$ is nothing but a scaled tensor product of a column of $\bar{\mathbf{A}}$ and the corresponding column of $\bar{\mathbf{B}}$. Therefore, $\omega(\mathbf{x}^H\mathbf{C})$ must be less or equal to 1, otherwise, Condition A will be violated. Invoking Lemma 2.1, $\mathbf{C}$ and $\bar{\mathbf{C}}$ are the same up to permutation and column scaling. The result therefore follows.

To show necessity, we proceed by contradiction. Without loss of generality, we assume that $\mathbf{C}$ is an identity matrix[1], $\mathbf{I}_F$, $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_F]$, $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_F]$, and a linear combination of the first two columns of $\mathbf{A}\odot\mathbf{B}$ constitutes a vector in the form

[1]Under our working assumption of full column rank $\mathbf{C}$, this is without loss of generality in so far as uniqueness is concerned. This has been shown by ten Berge via suitable pre-transformation of the data; see, e.g., [62].

of a tensor product of $\bar{\mathbf{a}}_1$ and $\bar{\mathbf{b}}_1$, i.e.,

$$\mathbf{a}_1 \otimes \mathbf{b}_1 + \mathbf{a}_2 \otimes \mathbf{b}_2 = \bar{\mathbf{a}}_1 \otimes \bar{\mathbf{b}}_1.$$

It is easy to see that

$$\begin{aligned}(\mathbf{A} \odot \mathbf{B})\mathbf{C}^T &= [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \cdots, \mathbf{a}_F \otimes \mathbf{b}_F]\, \mathbf{I}_F \\ &= [\bar{\mathbf{a}}_1 \otimes \bar{\mathbf{b}}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \cdots, \mathbf{a}_F \otimes \mathbf{b}_F] \begin{bmatrix} 1 & 0 & \mathbf{0} \\ -1 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{F-2} \end{bmatrix} \\ &= (\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T, \end{aligned} \tag{2.5}$$

where

$$\begin{aligned} \bar{\mathbf{A}} &= [\bar{\mathbf{a}}_1, \mathbf{a}_2, \cdots, \mathbf{a}_F] \\ \bar{\mathbf{B}} &= [\bar{\mathbf{b}}_1, \mathbf{b}_2, \cdots, \mathbf{b}_F] \\ \bar{\mathbf{C}} &= \begin{bmatrix} 1 & 0 & \mathbf{0} \\ -1 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{F-2} \end{bmatrix}^T, \end{aligned} \tag{2.6}$$

Hence $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$ constitutes an alternative decomposition. This completes the necessity part for Condition A.

Although Condition A has helped us intuitively understand the nature of the identification of restricted CP models, it has two limitations. The first is that Condition A is not easily verifiable. Second and more important is that the techniques used in the proof do not readily generalize to general CP models (rank-deficient $\mathbf{C}$). In the following, we shall derive an alternative equivalent condition that is often better suited for verification, and can be extended to cover general CP models.

We first define a set of $F \times F$ symmetric matrices $\mathbf{W}_{i_1,i_2,j_1,j_2}$ determined by the second-order minors of $\mathbf{A}$ and $\mathbf{B}$ as follows:

$$\mathbf{W}_{i_1,i_2,j_1,j_2}(f_1, f_2) = \frac{1}{2} \begin{vmatrix} a_{i_1,f_1} & a_{i_2,f_1} \\ a_{i_1,f_2} & a_{i_2,f_2} \end{vmatrix} \begin{vmatrix} b_{j_1,f_1} & b_{j_2,f_1} \\ b_{j_1,f_2} & b_{j_2,f_2} \end{vmatrix}, \tag{2.7}$$

for $i_1 = 1, \ldots, I, i_2 = 1, \ldots, I, j_1 = 1, \ldots, J, j_2 = 1, \ldots, J$.

We are now ready to derive the equivalent condition for identification of restricted CP models. As discussed before, it suffices that $\omega(\mathbf{x}^H\mathbf{C}) \leq \omega(\mathbf{x}^H\bar{\mathbf{C}})$ for all $\omega(\mathbf{x}^H\bar{\mathbf{C}}) \leq 1$. In particular, $\omega(\mathbf{x}^H\mathbf{C}) \leq \omega(\mathbf{x}^H\bar{\mathbf{C}})$ for $\omega(\mathbf{x}^H\bar{\mathbf{C}}) = 1$.

Since

$$(\bar{\mathbf{A}} \odot \bar{\mathbf{B}})\bar{\mathbf{C}}^T(\mathbf{x}^H)^T = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T(\mathbf{x}^H)^T,$$

invoking the identity [2] $vec(\mathbf{A}\,diag(\mathbf{x}^T)\mathbf{B}^T) = (\mathbf{A} \odot \mathbf{B})\mathbf{x}$, we have

$$\mathbf{A}\,diag(\mathbf{x}^H\mathbf{C})\mathbf{B}^T = \bar{\mathbf{A}}\,diag(\mathbf{x}^H\bar{\mathbf{C}})\bar{\mathbf{B}}^T.$$

Since $\omega(\mathbf{x}^H\bar{\mathbf{C}}) = 1$, we know

$$\mathbf{A}\,diag(\mathbf{x}^H\mathbf{C})\mathbf{B}^T = \bar{\mathbf{a}}_{f_0}\bar{\mathbf{b}}_{f_0}^T. \tag{2.8}$$

for some $f_0 \in \{1, \cdots, F\}$, and therefore

$$r_{\mathbf{A}\,diag(\mathbf{x}^H\mathbf{C})\mathbf{B}^T} \leq 1,$$

which is equivalent to all the 2nd order minors of $\mathbf{A}\,diag(\mathbf{x}^H\mathbf{C})\mathbf{B}^T$ being zero.

Let $[y_1, \cdots, y_F] := \mathbf{x}^H\mathbf{C}$, we have

$$\mathbf{M} := \mathbf{A}\,diag(\mathbf{x}^H\mathbf{C})\mathbf{B}^T = \sum_{f=1}^{F} y_f\mathbf{a}_f\mathbf{b}_f^T.$$

Since all the 2nd order minors of $\mathbf{A}\,diag(\mathbf{x}^H\mathbf{C})\mathbf{B}^T$ are equal to zero, this is *equivalent* to

$$\begin{vmatrix} m_{i_1,j_1} & m_{i_1,j_2} \\ m_{i_2,j_1} & m_{i_2,j_2} \end{vmatrix} = \begin{vmatrix} \sum_{f=1}^F y_f a_{i_1,f}b_{j_1,f} & \sum_{f=1}^F y_f a_{i_1,f}b_{j_2,f} \\ \sum_{f=1}^F y_f a_{i_2,f}b_{j_1,f} & \sum_{f=1}^F y_f a_{i_2,f}b_{j_2,f} \end{vmatrix} = 0, \tag{2.9}$$

for $i_1 = 1, \cdots, I, i_2 = 1, \cdots, I, j_1 = 1, \cdots, J, j_2 = 1, \cdots, J$.

(2.9) can be written as

$$(\sum_{f=1}^F y_f a_{i_1,f}b_{j_1,f})(\sum_{f=1}^F y_f a_{i_2,f}b_{j_2,f}) - (\sum_{f=1}^F y_f a_{i_1,f}b_{j_2,f})(\sum_{f=1}^F y_f a_{i_2,f}b_{j_1,f}) = 0$$

which is nothing but

$$\sum_{g \neq h} y_g y_h (a_{i_1,g} a_{i_2,h} b_{j_1,g} b_{j_2,h} - a_{i_1,g} a_{i_2,h} b_{j_1,h} b_{j_2,g}) = 0. \tag{2.10}$$

Further simplifying (2.10), we obtain

$$\sum_{1=g<h=F}^{F} y_g y_h (a_{i_1,g} a_{i_2,h} b_{j_1,g} b_{j_2,h} + a_{i_1,h} a_{i_2,g} b_{j_1,h} b_{j_2,g} - a_{i_1,g} a_{i_2,h} b_{j_1,h} b_{j_2,g} - a_{i_1,h} a_{i_2,g} b_{j_1,g} b_{j_2,h}) = 0. \tag{2.11}$$

(2.11) can be written as

$$\sum_{1=g<h=F}^{F} y_g y_h \begin{vmatrix} a_{i_1,g} & a_{i_2,g} \\ a_{i_1,h} & a_{i_2,h} \end{vmatrix} \begin{vmatrix} b_{j_1,g} & b_{j_2,g} \\ b_{j_1,h} & b_{j_2,h} \end{vmatrix} = 0, \tag{2.12}$$

each of which is equivalent to a bilinear form as follows

$$[y_1, \cdots, y_F] \mathbf{W}_{i_1,i_2,j_1,j_2} \begin{bmatrix} y_1 \\ \vdots \\ y_F \end{bmatrix} = 0, \tag{2.13}$$

for $i_1 = 1, \ldots, I, i_2 = 1, \ldots, I, j_1 = 1, \ldots, J, j_2 = 1, \ldots, J$.

We are now ready to state the equivalent condition on identification of restricted CP models.

Condition B: The set of equations in (2.13) only admits solutions satisfying $\omega([y_1, \cdots, y_F]) \leq 1$.

Note that any $\mathbf{y}$ with $\omega([y_1, \cdots, y_F]) \leq 1$ will automatically satisfy the equations in (2.13).

When Condition B holds, it is easily seen that $\omega(\mathbf{x}^H \mathbf{C}) \leq \omega(\mathbf{x}^H \bar{\mathbf{C}})$ for all $\omega(\mathbf{x}^H \bar{\mathbf{C}}) \leq 1$, therefore CP decomposition is unique. On the other hand, when Condition B is not satisfied, it is easy to show that either $\mathbf{A} \odot \mathbf{B}$ is rank deficient or a non-trivial linear combination of columns of $\mathbf{A} \odot \mathbf{B}$ constitutes a vector in the form of tensor product of two vectors. In either case, an alternative decomposition of the given array can be constructed.

Sometimes, solving a system of bilinear equations such as (2.13) is not as complicated as it appears. If all $\mathbf{W}_{i_1,i_2,j_1,j_2}$ are real positive semi-definite matrices, then the solutions to the system of bilinear equations can be obtained by solving a suitable linear equation. Unfortunately, this is not the case for our problem. More often than not, $\mathbf{W}_{i_1,i_2,j_1,j_2}$ are in-definite complex matrices. This poses difficulties in checking whether solutions to (2.13) adhere to the constraint in Condition B. We don't have a general tool for handling this verification yet, but, as will be shown shortly, some instructive simple cases can be worked out by hand, and the issue is currently under investigation.

We note that since $\mathbf{W}_{i_1,i_2,j_1,j_2} = \mathbf{0}$ if $i_1 = i_2$ or $j_1 = j_2$, the number of active bilinear equations can be reduced. It is also worth mentioning that instead of casting (2.12) into (2.13), we can "linearize" (2.12) as follows:

$$\mathbf{U}\begin{bmatrix} y_1 y_2 \\ \vdots \\ y_g y_h \\ \vdots \\ y_{F-1} y_F \end{bmatrix} = 0, \tag{2.14}$$

where the entries of $\mathbf{U}$ are determined by (2.12).

In this way, we deal with a linear equation that involves a structured vector. Note that $\omega([y_1, \cdots, y_F]) \leq 1$ is *equivalent* to $[y_1 y_2, \cdots, y_g y_h, \cdots, y_{F-1} y_F] = \mathbf{0}$. $\mathbf{U}$ being full column rank guarantees that $\omega([y_1, \cdots, y_F]) \leq 1$. However, a rank-deficient $\mathbf{U}$ does *not* necessarily imply that $\omega([y_1, \cdots, y_F]) \geq 2$ since $[y_1 y_2, \cdots, y_{F-1} y_F]$ is a *structured* vector. In particular, simulations show that $\mathbf{U}$ can be rank-deficient even when $\mathbf{A}$ and $\mathbf{B}$ are drawn from a continuous distribution.

2.4 Discussion

2.4.1 ten Berge's example and counter-example

One of the motivations of this chapter is to explain the puzzle brought by the counter-example to necessity of Kruskal's condition given in [62]. In [62], two simple examples illustrate that the uniqueness of CP decomposition depends on the particular joint pattern of zeros in the component matrices. The first example is

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & b_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & b_3 \end{bmatrix},$$

with a_1, a_2, b_1 and b_3 nonzero, and $\mathbf{C} = \mathbf{I}_4$. The second example in [62] is given by changing the first example slightly to have the zero entry in the last columns of $\mathbf{A}$ and $\mathbf{B}$ in the same place as follows

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & b_1 \\ 0 & 1 & 0 & b_2 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

with a_1, a_2, b_1 and b_2 nonzero, and a common $\mathbf{C}$. It has been proven in [62] that the decomposition of the array given by $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ in the first example is unique, whereas alternative decompositions arise in the second example. However, no explanation on this interesting phenomenon was provided. Equipped with the main results of the previous Section, we are now in position to offer such explanation.

In the first example, we know that

$$\sum_{f=1}^{F} y_f \mathbf{a}_f \mathbf{b}_f^T = \begin{bmatrix} y_1 + a_1 b_1 y_4 & 0 & a_1 b_3 y_4 \\ a_2 b_1 y_4 & y_2 & a_2 b_3 y_4 \\ 0 & 0 & y_3 \end{bmatrix} \tag{2.15}$$

Therefore, following (2.12), we have

$$\begin{cases} y_1 y_2 + a_1 b_1 y_2 y_4 = 0 \\ y_1 y_3 + a_1 b_1 y_3 y_4 = 0 \\ y_2 y_4 = 0 \\ y_2 y_3 = 0 \\ y_1 y_4 = 0 \end{cases} \tag{2.16}$$

(2.16) can be written as $\mathbf{U}\mathbf{w} = \mathbf{0}$, with $\mathbf{U}$ being 5×6 and $\mathbf{w} := [y_1y_2, y_1y_3, y_1y_4, y_2y_3, y_2y_4, y_3y_4]^T$. (2.16) admits a solution of weight larger than 1 if and only if there is a nonzero $\mathbf{w}$ orthogonal to the five rows of $\mathbf{U}$. For the particular $\mathbf{U}$ in (16), the only possibility for this is to have $\mathbf{w}$ proportional to $[0,\ a_1b_1,\ 0,\ 0,\ 0,\ -1]$ with y_1, y_3, and y_4 nonzero. Because $\mathbf{w}_3 = y_1 y_4 = 0$, this is not possible, and $\mathbf{w}$ is the zero vector after all. Therefore, (2.16) does not admit a solution with $\omega([y_1, y_2, y_3, y_4]) \geq 2$.

On the other hand, in the second example, we have

$$\begin{cases} y_1 y_2 + a_2 b_2 y_1 y_4 + a_1 b_1 y_2 y_4 = 0 \\ y_1 y_3 = 0 \\ y_2 y_3 = 0 \\ y_3 y_4 = 0 \end{cases} \tag{2.17}$$

It is easy to see that (2.17) admits the solution $y_1 = -\frac{a_1 b_1}{1 + a_2 b_2}, y_2 = 1, y_3 = 0, y_4 = 1$ where we assume $1 + a_2 b_2 \neq 0$. If $1 + a_2 b_2 = 0$, we can modify the values of y_2 and y_4

accordingly and still have a solution with $\omega([y_1, y_2, y_3, y_4]) \geq 2$. As it has been shown in [62], the number of such solutions is infinite.

Furthermore, in the second example,

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & a_1b_1 \\ 0 & 0 & 0 & a_1b_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_2b_1 \\ 0 & 1 & 0 & a_2b_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

we can see that a linear combination of the first, the second, and the forth columns of $\mathbf{A} \odot \mathbf{B}$ constitutes a vector in the form of tensor product of two vectors as follows,

$$-\frac{a_1b_1}{1+a_2b_2}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_1b_1 \\ a_1b_2 \\ 0 \\ a_2b_1 \\ a_2b_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{a_1a_2b_1b_2}{1+a_2b_2} \\ a_1b_2 \\ 0 \\ a_2b_1 \\ 1+a_2b_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a_2b_1 \\ 1+a_2b_2 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} \frac{a_1b_2}{1+a_2b_2} \\ 1 \\ 0 \end{bmatrix}. \tag{2.18}$$

2.4.2 Bilinear decomposition under CM constraints

Although bilinear decomposition is not unique in general, bilinear decomposition with constant modulus (CM) constraints can be unique [33, 61, 65]. Interestingly, as pointed out next, the identification condition on bilinear decomposition with CM constraints is very similar to Condition B derived herein for the identification of restricted CP models . Any progresses on identification of bilinear decomposition with CM constraints might be beneficial to better understand Condition B and vice versa.

Let $\mathbf{C} = \mathbf{A}\mathbf{B}^T$, with full column rank $\mathbf{B}$ and a CM constraint on $\mathbf{A}$: that is, without loss of generality, $|a_{i,f}| = a_{i,f}a_{i,f}^H = 1$.

Like CP decomposition, bilinear decomposition with CM constraints, when unique, it is unique up to column permutation and scaling. Therefore, Kruskal's Permutation Lemma can again be taken as the cornerstone for uniqueness. Earlier work on the identification of bilinear mixtures under CM constraints [33, 61] has yielded sufficient conditions, but necessity has been left open to the best of our knowledge. Equipped with Kruskal's Permutation Lemma, we are ready to give a necessary and sufficient condition for unique bilinear decomposition under CM constraints. In this context, uniqueness means that if there is another pair $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$, with $\bar{\mathbf{A}}$ having CM elements, such that $\mathbf{A}\mathbf{B}^T = \bar{\mathbf{A}}\bar{\mathbf{B}}^T$, then there exist a permutation matrix $\mathbf{\Pi}$ and two non-singular diagonal matrices $\mathbf{\Lambda}_A$, $\mathbf{\Lambda}_B$ with $\mathbf{\Lambda_A}\mathbf{\Lambda_B} = \mathbf{I}$ such that $\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}\mathbf{\Lambda_A}, \bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda_B}$. Note that the scaling indeterminacy remains despite the CM constraint, due to the possibility of rotation in the complex plane. Also, since we have assumed a full column rank $\mathbf{B}$, it suffices to show that $\bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda_B}$ for a permutation matrix $\mathbf{\Pi}$ and a diagonal matrix $\mathbf{\Lambda}_B$; the result for $\bar{\mathbf{A}}$ then follows by simple inversion. This is the usual route taken to show uniqueness in this context.

Note that

$$|\sum y_f a_{i_1,f}| = |\sum y_f a_{i_2,f}| \Rightarrow \sum_{g\neq h}^{F} \begin{vmatrix} a_{i_1,g} & a_{i_2,h}^H \\ a_{i_2,g} & a_{i_1,h}^H \end{vmatrix} y_g y_h^H = 0, \qquad (2.19)$$

or

$$[y_1, \cdots, y_F]\mathbf{W}_{i_1,i_2} \begin{bmatrix} y_1^H \\ \vdots \\ y_F^H \end{bmatrix} = 0,$$

$$\mathbf{W}_{i_1,i_2}(g,h) := \begin{vmatrix} a_{i_1,g} & a_{i_2,h}^H \\ a_{i_2,g} & a_{i_1,h}^H \end{vmatrix},$$

for $i_1 = 1, \cdots, I, i_2 = 1, \cdots, I$. The necessary and sufficient condition for unique bilinear decomposition under CM constraints can now be stated:

Condition C: The set of equations in (2.19) only admits solutions satisfying $\omega([y_1, \cdots, y_F]) \leq 1$.

Let us show that Condition C guarantees $\bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Lambda}_\mathbf{B}$ for a permutation matrix $\mathbf{\Pi}$ and a diagonal matrix $\mathbf{\Lambda}_B$. Invoking the Permutation Lemma 2.1, it suffices to show that $\omega(\mathbf{x}^H\mathbf{B}) \leq \omega(\mathbf{x}^H\bar{\mathbf{B}})$ for all $\omega(\mathbf{x}^H\bar{\mathbf{B}}) \leq 1$. To see this, note that Condition C guarantees $\mathbf{A}$ being full rank. This is because Condition C is equivalent to none of the non-trivial linear combinations of columns of $\mathbf{A}$ can be written as a vector comprising of constant modulus entries, and a zero vector is a constant modulus vector. Then, from the hypothesis $\mathbf{A}\mathbf{B}^T = \bar{\mathbf{A}}\bar{\mathbf{B}}^T$ and the assumption that $\mathbf{B}$ is full column rank, it follows that $\bar{\mathbf{B}}$ is full column rank as well, in which case (cf. statement of Lemma 2.1) we only need to verify that $\omega(\mathbf{x}^H\mathbf{B}) \leq \omega(\mathbf{x}^H\bar{\mathbf{B}})$ holds for all $\omega(\mathbf{x}^H\bar{\mathbf{B}}) = 1$.

From the hypothesis $\mathbf{A}\mathbf{B}^T = \bar{\mathbf{A}}\bar{\mathbf{B}}^T$, we have

$$\bar{\mathbf{A}}\bar{\mathbf{B}}^T(\mathbf{x}^H)^T = \mathbf{A}\mathbf{B}^T(\mathbf{x}^H)^T,$$

for all $\mathbf{x}$. In particular, for all those $\mathbf{x}$ such that $\omega(\mathbf{x}^H\bar{\mathbf{B}}) = 1$, the left hand side is a column drawn from $\bar{\mathbf{A}}$, and thus a vector comprising of constant modulus entries.

The first element of the right hand side is a linear combination of the elements in the first row of $\mathbf{A}$, the second is a linear combination of the elements in the second row of $\mathbf{A}$, and so on. All these row-combinations should have equal modulus. If the only way for this to happen is that a single column is selected from $\mathbf{A}$, as per Condition C, then it must be that $\omega(\mathbf{x}^H\mathbf{B}) = 1$. This shows that Condition C is sufficient for uniqueness. For the converse, suppose that Condition C is violated. Without loss of generality, we may assume that $\mathbf{B}$ is an identity matrix $\mathbf{I}_F$ and a linear combination of the first two columns of $\mathbf{A}$ constitutes a constant modulus vector $\bar{\mathbf{a}}_1$, i.e.,

$$\mathbf{a}_1 + \mathbf{a}_2 = \bar{\mathbf{a}}_1.$$

and the modulus of each entry of $\bar{\mathbf{a}}_1$ is equal to a, a constant, not necessarily equal to one.

If a is zero, we know $\mathbf{A}$ is rank deficient. Then adding any non-zero null vectors of $\mathbf{A}$ to the first column of $\mathbf{B}^T$ preserves $\mathbf{A}\mathbf{B}^T$, but generates a different solution for $\mathbf{B}$.

If a is not zero, it is easy to see that

$$\begin{aligned} \mathbf{A}\mathbf{B}^T &= [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_F]\ \mathbf{I}_F \\ &= [\frac{\bar{\mathbf{a}}_1}{\sqrt{a}}, \mathbf{a}_2, \cdots, \mathbf{a}_F] \begin{bmatrix} \sqrt{a} & 0 & \mathbf{0} \\ -1 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{F-2} \end{bmatrix} \\ &= \bar{\mathbf{A}}\bar{\mathbf{B}}^T, \end{aligned} \tag{2.20}$$

where

$$\begin{aligned} \bar{\mathbf{A}} &= [\frac{\bar{\mathbf{a}}_1}{\sqrt{a}}, \mathbf{a}_2, \cdots, \mathbf{a}_F] \\ \bar{\mathbf{B}} &= \begin{bmatrix} \sqrt{a} & 0 & \mathbf{0} \\ -1 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{F-2} \end{bmatrix}^T . \end{aligned} \tag{2.21}$$

Clearly, the modulus of each entry of $\bar{\mathbf{A}}$ is one. Hence $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$ constitutes an alternative decomposition. This completes the necessity part for Condition C.

We now can see that Condition B for the identification of restricted CP models and Condition C for the identification of bilinear models subject to CM constraints are very similar. While both have been derived using Kruskal's Permutation Lemma, they stem from conceptually very different structural constraints on the equivalent bilinear models. More specifically, CP can be viewed as a bilinear model with Khatri-Rao product structure along the one dimension; whereas CM is a bilinear model with a modulus constraint on the elements of one matrix factor.

When the CM constraint is imposed along one or more modes of CP, identifiability naturally improves in terms of the number of available equations. For instance, given a CP model $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ with full column rank $\mathbf{C}$ and CM constraints on both $\mathbf{A}$ and $\mathbf{B}$, the following is a necessary and sufficient set of uniqueness conditions:

Both

$$\sum_{1=g<h=F}^{F} \begin{vmatrix} a_{i_1,g} & a_{i_2,g} \\ a_{i_1,h} & a_{i_2,h} \end{vmatrix} \begin{vmatrix} b_{j_1,g} & b_{j_2,g} \\ b_{j_1,h} & b_{j_2,h} \end{vmatrix} y_g y_h = 0$$

and

$$\sum_{g \neq h}^{F} \begin{vmatrix} a_{i_1,g} b_{j_1,g} & a_{i_2,h}^H b_{j_2,h}^H \\ a_{i_2,g} b_{j_2,g} & a_{i_1,h}^H b_{j_1,h}^H \end{vmatrix} y_g y_h^H = 0,$$

only admit joint solutions with $\omega([y_1, \cdots, y_F]) \leq 1$.

A concise unifying treatment of necessary and sufficient uniqueness conditions for the identification of general[2] CP models subject to CM constraints along one or more modes is not available at this point. Nevertheless, individual cases can be dealt with, given the tools developed herein. Even discarding CM constraints, stating and checking necessary and sufficient uniqueness conditions for unrestricted CP models is possible but cumbersome. When none of the component matrices is full column rank,

[2]Meaning: without the full column rank restriction along one mode.

following the road map provided in Section 2.2, one has to show that $\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}_{\mathbf{A}}\mathbf{\Lambda}_{\mathbf{A}}$ and $\bar{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}_{\mathbf{B}}\mathbf{\Lambda}_{\mathbf{B}}$ separately. High-order minors of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ must be exploited, and the condition for the identification of general CP models boils down to a number of multi-linear equations with particular constraints on common solutions. We defer this pursuit at this point, pending further understanding of Condition B, which we hope to develop in on-going work.

2.5 Conclusions

Two equivalent necessary and sufficient conditions for unique decomposition of restricted CP models where at least one of the component matrices is full column rank have been derived. These conditions explain the puzzle in [62]. A strong similarity between the conditions for unique decomposition of bilinear models subject to CM constraints and certain restricted CP models has been pointed out. It is hoped that this link will facilitate cross-fertilization and unification of associated uniqueness results. Last but not least, Kruskal's Permutation Lemma has been demystified. The new proof should be accessible to a much wider readership than Kruskal's original proof.

Appendix 2.A Kruskal's Permutation Lemma: Redux

Kruskal's Permutation Lemma 2.1:

We are given two matrices $\mathbf{A}$ and $\bar{\mathbf{A}}$, which are $I \times F$ and $I \times \bar{F}$. Suppose $\mathbf{A}$ has no zero columns. If for any vector $\mathbf{x} \in \mathbb{C}^N$ such that

$$\omega(\mathbf{x}^H \bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}} + 1,$$

we have

$$\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}}),$$

then $F \leq \bar{F}$; if also $F \geq \bar{F}$, then $F = \bar{F}$, and there exist a permutation matrix $\mathbf{P}_{\bar{\mathbf{A}}}$ and a non-singular matrix $\mathbf{\Lambda}$ such that $\mathbf{A} = \bar{\mathbf{A}}\mathbf{P}_{\bar{\mathbf{A}}}\mathbf{\Lambda}$.

Remark 2.1 *Kruskal's condition is equivalent to*

> *If a certain vector is orthogonal to $c \geq r_{\bar{\mathbf{A}}} - 1$ columns of $\bar{\mathbf{A}}$, then it must be orthogonal to at least c columns of $\mathbf{A}$*

which implies

> *For every collection of $c \geq r_{\bar{\mathbf{A}}} - 1$ columns of $\bar{\mathbf{A}}$, there exists a collection of at least c columns of $\mathbf{A}$ such that*
>
> $$\mathit{Span}(\text{these } c \geq r_{\bar{\mathbf{A}}} - 1 \text{ columns of } \bar{\mathbf{A}}) \supseteq \mathit{Span}(c \text{ or more columns of } \mathbf{A})$$

To show why the first statement implies the second statement, we proceed by contradiction. Suppose that there is a collection of $c_0 \geq r_{\bar{\mathbf{A}}} - 1$ columns of $\bar{\mathbf{A}}$, say, $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\}$, and there are *only* $(c_0 - k)$ columns of $\mathbf{A}$, say, $\{\mathbf{a}_1, \cdots, \mathbf{a}_{c_0-k}\}$, such that

$$\mathit{Span}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\}) \supseteq \mathit{Span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_{c_0-k}\}), \tag{2.22}$$

where $1 \le k \le c_0$.

Note the each of the remaining columns of $\mathbf{A}$, i.e., $\{\mathbf{a}_{c_0-k+1}, \cdots, \mathbf{a}_F\}$, is linearly independent with the column set $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\}$, otherwise k can be reduced by 1; this implies that for every $i = c_0 - k + 1, \ldots, F$, there exists a certain non-zero vector $\mathbf{x}_i \in \mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\})$ such that

$$\mathbf{x}_i^H \mathbf{a}_i \neq 0;$$

otherwise, if for every $\mathbf{x} \in \mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\})$, $\mathbf{x} \in \mathit{Null}(\{\mathbf{a}_{i_0}\})$ for a certain i_0, this implies

$$\mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\}) \subseteq \mathit{Null}(\{\mathbf{a}_{i_0}\}),$$

i.e.,

$$\mathit{Span}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\}) \supseteq \mathit{Span}(\{\mathbf{a}_{i_0}\}),$$

which means that k can be reduced by 1 as well.

Let us assume that $\mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\})$ is an m-dimensional linear subspace, $m \geq 1$. $m = 0$ means that $\mathit{Span}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\}) = \mathbf{C}^I$; this further implies that all columns of $\mathbf{A}$ belong to $\mathit{Span}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\})$.

Now, consider

$$\mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}, \mathbf{a}_i\})$$

for each i. Due to the existence of aforementioned $\mathbf{x}_i$, $\mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}, \mathbf{a}_i\})$ is a proper linear subspace of $\mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\})$ with dimension $m-1$. Since the union of a countable number of $(m-1)$-dimensional linear subspaces of $\mathbb{C}^I$ cannot cover an m-dimensional linear subspace of $\mathbb{C}^I$, we are able to find a non-zero vector $\mathbf{x}_0 \in \mathit{Null}(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{c_0}\})$, such that

$$\mathbf{x}_0^H \mathbf{a}_i \neq 0, \ \forall i = c_0 - k + 1, \ldots, F.$$

The existence of such $\mathbf{x}_0$ contradicts the first statement.

Unlike the statement of Lemma 2.1, where the columns sizes of $\bar{\mathbf{A}}$ and $\mathbf{A}$ might be different, we assume that both $\bar{\mathbf{A}}$ and $\mathbf{A}$ are $I \times F$ matrices; furthermore, without loss of generality, we assume both do not contain zero columns.

Remark 2.2 *Given two non-trivial vectors* $\bar{\mathbf{y}}$ *and* $\mathbf{y}$*, they are linearly dependent if and only if* $\omega(\mathbf{x}^H\mathbf{y}) = 0$ *for all* $\mathbf{x}$ *satisfying* $\omega(\mathbf{x}^H\bar{\mathbf{y}}) = 0$*. This can be easily checked using the testing vector* $\mathbf{x} = [a, 0, \cdots, 0, b, 0, \cdots, 0]$ *with* a*,* b *chosen such that* $\omega(\mathbf{x}^H\bar{\mathbf{y}}) = 0$.

Lemma 2.2 *Given* $\bar{\mathbf{A}} \in \mathbb{C}^{I\times F}$ *and* $\mathbf{A} \in \mathbb{C}^{I\times F}$*,* $\mathbf{A} = \bar{\mathbf{A}}\mathbf{P}_{\bar{\mathbf{A}}}\mathbf{\Lambda}$ *if and only if* $\omega(\mathbf{x}\mathbf{A}) \leq \omega(\mathbf{x}\bar{\mathbf{A}})$ *for all* $\mathbf{x}$.

Proof of Lemma 2.2: It suffices to prove the "if" part, and we will prove this by induction on the number of columns of $\bar{\mathbf{A}}$, namely F.

When $F = 1$, the condition in Lemma 2.2 implies that $\omega(\mathbf{x}^H\mathbf{A}) = 0$ for all $\mathbf{x}$ satisfying $\omega(\mathbf{x}^H\bar{\mathbf{A}}) = 0$. From Remark 2.2, this implies that $\bar{\mathbf{A}}$ and $\mathbf{A}$ are linearly dependent.

Assume that Lemma 2.2 holds true for all $F \leq K$. Now, consider $F = K+1$. Let $\bar{\mathbf{a}}_i$ denote the i-th column of $\bar{\mathbf{A}}$, and $\mathbf{a}_j$ denote the j-th column of $\mathbf{A}$.

We claim that under the condition in Lemma 2, there must exist at least one column of $\mathbf{A}$, $\mathbf{a}_{j_0}$, which is linearly dependent with $\bar{\mathbf{a}}_1$. We will prove this by contradiction. Suppose that this claim is not true; then, based on Remark 2.2 and the assumption that $\mathbf{A}$ does not contain zero columns, we know that for every j, there exists a $\mathbf{x}_j$ such that

$$\omega(\mathbf{x}_j^H\bar{\mathbf{a}}_1) = 0, \;\; \omega(\mathbf{x}_j^H\mathbf{a}_j) = 1, \; \forall j = 1, \ldots, F. \tag{2.23}$$

Then we will show that in fact there exists a *common* $\mathbf{x}_0$ such that

$$\omega(\mathbf{x}_0^H\bar{\mathbf{a}}_1) = 0, \;\; \omega(\mathbf{x}_0^H\mathbf{a}_j) = 1, \; \forall j = 1, \ldots, F. \tag{2.24}$$

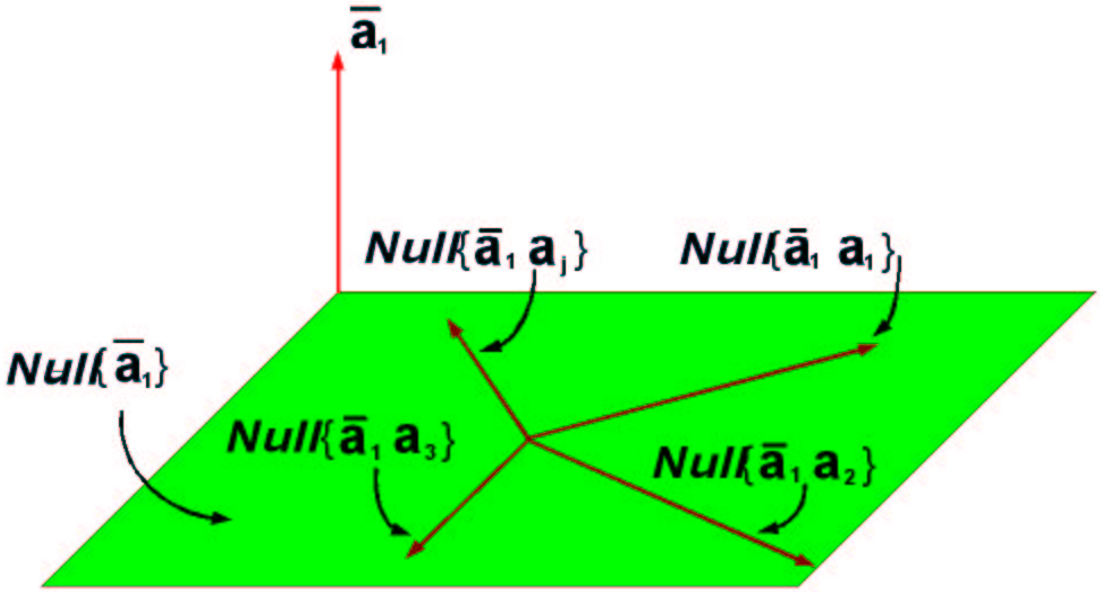

Figure 2.1: Geometric Illustration

$\bar{\mathbf{a}}_1 \neq \mathbf{0}$, hence the null space of $\bar{\mathbf{a}}_1$, $Null(\bar{\mathbf{a}}_1)$, is an $(I-1)$-dimensional linear space.

Now consider $Null(\{\bar{\mathbf{a}}_1, \mathbf{a}_j\})$ for all j. Clearly, all $Null(\{\bar{\mathbf{a}}_1, \mathbf{a}_j\})$ are covered by $Null(\bar{\mathbf{a}}_1)$. It is clearly seen that the existence of $\mathbf{x}_0$ in (2.24) is equivalent to $\bigcup_{j=1}^F Null(\{\bar{\mathbf{a}}_1, \mathbf{a}_j\}) \neq Null(\bar{\mathbf{a}}_1)$.

Recall that for every j, there exists a $\mathbf{x}_j$ such that

$$\omega(\mathbf{x}_j^H \bar{\mathbf{a}}_1) = 0, \ \ \omega(\mathbf{x}_j^H \mathbf{a}_j) = 1, \ \forall j,$$

which implies that $Null(\{\bar{\mathbf{a}}_1, \mathbf{a}_j\})$ cannot be the same as $Null(\bar{\mathbf{a}}_1)$, but rather a proper linear subspace of $Null(\bar{\mathbf{a}}_1)$ with dimension $I-2$. Furthermore, the union of a countable number of $(I-2)$-dimensional linear subspaces of $\mathbf{C}^I$ cannot cover an $(I-1)$-dimensional subspace of $\mathbf{C}^I$, also see Fig 2.1. Therefore, $\bigcup_{j=1}^F Null(\{\bar{\mathbf{a}}_1, \mathbf{a}_j\})$ does not cover $Null(\bar{\mathbf{a}}_1)$, hence, we do have a $\mathbf{x}_0$ such that (2.24) holds.

This implies that

$$\omega(\mathbf{x}_0^H \bar{\mathbf{A}}) \leq F - 1 < F = \omega(\mathbf{x}_0^H \mathbf{A}),$$

which contradicts the condition in Lemma 2.2. Therefore, we can claim there exists at least one column of $\mathbf{A}$, which is linearly dependent with $\bar{\mathbf{a}}_1$. Without loss generality, we say this column is $\mathbf{a}_{j_0}$. Clearly,

$$\omega(\mathbf{x}^H \bar{\mathbf{a}}_1) = \omega(\mathbf{x}^H \mathbf{a}_{j_0}), \forall \mathbf{x}. \tag{2.25}$$

Now, construct a submatrix of $\bar{\mathbf{A}}$ by removing column $\bar{\mathbf{a}}_1$ from $\bar{\mathbf{A}}$, and denote this matrix $\bar{\mathbf{A}}_0$; similarly construct a submatrix of $\mathbf{A}$ by removing column $\mathbf{a}_{j_0}$ from $\mathbf{A}$, and denote this matrix $\mathbf{A}_0$.

From $\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}})$ for all $\mathbf{x}$ (condition in statement of Lemma 2), and (2.25), it follows that

$$\omega(\mathbf{x}^H \mathbf{A}_0) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}}_0), \forall \mathbf{x}.$$

But $\bar{\mathbf{A}}_0$ and $\mathbf{A}_0$ are K-column matrices; the result then follows from the induction hypothesis. That is, the $(K+1)$-column matrices $\bar{\mathbf{A}}$, $\mathbf{A}$ are the same up to permutation and scaling of columns. This completes the proof. □

Remark 2.3 *The proof of Lemma 2.2 can be also applied to the following corollary.*

Corollary 2.1 *Given $\bar{\mathbf{A}} \in \mathbb{C}^{I \times F}$ and $\mathbf{A} \in \mathbb{C}^{I \times F}$, $\mathbf{A} = \bar{\mathbf{A}} \mathbf{P}_{\bar{\mathbf{A}}} \mathbf{\Lambda}$ if and only if $\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}})$ for any $\mathbf{x}$ such that*

$$\omega(\mathbf{x}^H \bar{\mathbf{A}}) \leq F - 1.$$

Compared with Kruskal's result, the conditions in both Lemma 2.2 and Corollary 2.1 appear more restrictive. There is a gap between the results presented in this Appendix so far and Kruskal's result. If $r_{\bar{\mathbf{A}}} = 2$, this gap has been filled by Corollary 2.1. For the general case, we have the following Lemma:

Lemma 2.3 *Given $\bar{\mathbf{A}} \in \mathbb{C}^{I \times F}$ and $\mathbf{A} \in \mathbb{C}^{I \times F}$, $\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}})$ for all $\mathbf{x}$ if and only if $\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}})$ for any vector $\mathbf{x}$ such that $\omega(\mathbf{x}^H \bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}} + 1$.*

The interesting case occurs when $r_{\bar{\mathbf{A}}}$ is strictly less than F. Without loss of generality, we assume $r_{\bar{\mathbf{A}}} < F$.

With an additional condition, namely, $k_{\bar{\mathbf{A}}} = r_{\bar{\mathbf{A}}}$, where $k_{\bar{\mathbf{A}}}$ stands for Kruskal rank of $\bar{\mathbf{A}}$, a relatively simpler proof can be obtained as follows.

Proof of Lemma 2.3 - Case of $k_{\bar{\mathbf{A}}} = r_{\bar{\mathbf{A}}}$:

It suffices to prove the "if" part, and we prove it by contradiction. Suppose there exists a non-zero vector $\mathbf{x}_0$, such that $\omega(\mathbf{x}_0^H\mathbf{A}) > \omega(\mathbf{x}_0^H\bar{\mathbf{A}})$, and $\omega(\mathbf{x}_0^H\bar{\mathbf{A}}) > F - r_{\bar{\mathbf{A}}} + 1$, and suppose $\omega(\mathbf{x}_0^H\bar{\mathbf{A}})$ is the *smallest such number* bigger than $F - r_{\bar{\mathbf{A}}} + 1$ in the sense that $\omega(\mathbf{x}^H\mathbf{A}) \leq \omega(\mathbf{x}^H\bar{\mathbf{A}})$ for any vector $\mathbf{x}$ such that

$$\omega(\mathbf{x}^H\bar{\mathbf{A}}) < \omega(\mathbf{x}_0^H\bar{\mathbf{A}}).$$

Without loss of generality, we assume

$$\omega(\mathbf{x}_0^H\bar{\mathbf{A}}) = F - r_{\bar{\mathbf{A}}} + k,$$

where $2 \leq k < r_{\bar{\mathbf{A}}}$, and

$$\omega(\mathbf{x}_0^H\mathbf{A}) = F - r_{\bar{\mathbf{A}}} + k + l,$$

where $1 \leq l \leq r_{\bar{\mathbf{A}}} - k$.

With such $\mathbf{x}_0$, we know that there exist $(r_{\bar{\mathbf{A}}} - k)$ columns of $\bar{\mathbf{A}}$, say, $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$, and $(r_{\bar{\mathbf{A}}} - k - l)$ columns of $\mathbf{A}$, say, $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, such that

$$\{\mathbf{x}_0, \mathbf{0}\} \subseteq Null(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}) \bigcap Null(\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}).$$

Since $\omega(\mathbf{x}_0^H\mathbf{A}) > \omega(\mathbf{x}_0^H\bar{\mathbf{A}})$, $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ are the only columns of $\mathbf{A}$ that can possibly belong to $Span(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\})$; otherwise, if there is one more column, say $\mathbf{a}_F$, belonging to $Span(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\})$, then, $\mathbf{x}_0^H\mathbf{a}_F = 0$, which implies that $\omega(\mathbf{x}_0^H\mathbf{A}) = F - r_{\bar{\mathbf{A}}} + k + l - 1$, and contradicts $\omega(\mathbf{x}_0^H\mathbf{A}) = F - r_{\bar{\mathbf{A}}} + k + l$.

The remaining $P - r_{\bar{\mathbf{A}}} + k$ columns of $\bar{\mathbf{A}}$ are $\{\bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k+1}, \cdots, \bar{\mathbf{a}}_P\}$.

Recall that, by definition of $\mathbf{x}_0$,

$$\omega(\mathbf{x}^H\mathbf{A}) \leq \omega(\mathbf{x}^H\bar{\mathbf{A}}) \;\; \forall \mathbf{x} \;\; s.t. \;\; \omega(\mathbf{x}^H\bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}} + k - 1 < \omega(\mathbf{x}_0^H\bar{\mathbf{A}}). \qquad (2.26)$$

Similar to Remark 2.1, we can show that (2.26) implies that for every $r_{\bar{\mathbf{A}}} - k + 1$ or more columns chosen from $\bar{\mathbf{A}}$, there must exist at least as many columns from $\mathbf{A}$, such that each of those from $\mathbf{A}$ is a linear combination of the said columns of $\bar{\mathbf{A}}$.

Now consider the following $F - r_{\bar{\mathbf{A}}} + k$ column sets drawn from $\bar{\mathbf{A}}$,

$$\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}, \bar{\mathbf{a}}_i\},$$

where $i = r_{\bar{\mathbf{A}}} - k + 1, \ldots, P$. Each of them has $r_{\bar{\mathbf{A}}} - k + 1$ distinct columns of $\bar{\mathbf{A}}$.

According to (2.26), for each column set $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}, \bar{\mathbf{a}}_i\}$, there exist at least $r_{\bar{\mathbf{A}}} - k + 1$ columns $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_{r_{\bar{\mathbf{A}}}-k+1}}\}$ such that each column from the latter set is a linear combination of those in the former set.

Recall that except for $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, there is no other column of $\mathbf{A}$, which belongs to $Span(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\})$. This implies that *at least* $(l+1) = (r_{\bar{\mathbf{A}}} - k + 1) - (r_{\bar{\mathbf{A}}} - k - l)$ columns from $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_{r_{\bar{\mathbf{A}}}-k+1}}\}$, other than those in $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, must be such that each is a linear combination of $\bar{\mathbf{a}}_i$ and some or all of $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$. Let ϕ_i denote the column set consisting of those $l+1$ columns from $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_{r_{\bar{\mathbf{A}}}-k+1}}\}$.

We claim that every two ϕ_i and ϕ_j are disjoint for $i \neq j$; for if there exists a common element between ϕ_i and ϕ_j, say $\mathbf{a}_j^i \in \phi_i \bigcap \phi_j$, then,

$$\mathbf{a}_j^i = \sum_{n=1}^{r_{\bar{\mathbf{A}}}-k} c_n \bar{\mathbf{a}}_n + c_i \bar{\mathbf{a}}_i = \sum_{m=1}^{r_{\bar{\mathbf{A}}}-k} d_m \bar{\mathbf{a}}_m + d_j \bar{\mathbf{a}}_j, \;\; c_i \neq 0, \;\; d_j \neq 0,$$

which in turn implies that the $r_{\bar{\mathbf{A}}} - k + 2$ column set

$$\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}, \bar{\mathbf{a}}_i, \bar{\mathbf{a}}_j\}$$

is a linearly dependent set of columns with distinct indices. Since $k \geq 2$ and $k_{\bar{\mathbf{A}}} = r_{\bar{\mathbf{A}}}$, we have a contradiction. Therefore, every two ϕ_i and ϕ_j are disjoint for $i \neq j$. In addition, it is easily seen that ϕ_i is disjoint with $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ as well.

The remainder is a counting problem. The number of all columns of $\mathbf{A}$ should not be less than the number of columns in all the above disjoint column subsets of $\mathbf{A}$. However, from $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, we have $r_{\bar{\mathbf{A}}} - k - l$ columns, from each ϕ_i, we have at least $l+1$ columns, and we have $P - r_{\bar{\mathbf{A}}} + k$ such ϕ_i, therefore, the total number of columns from all disjoint column subsets of $\mathbf{A}$ is not less then

$$r_{\bar{\mathbf{A}}} - k - l + (l+1)(F - r_{\bar{\mathbf{A}}} + k) = l(F - r_{\bar{\mathbf{A}}}) + F + (k-1)l,$$

which is strictly greater than F for $l \geq 1$, and $k \geq 2$ whereas $\mathbf{A}$ has F columns only. We have a contradiction. $\square$

The above proof of the special case of Lemma 2.3 provides helpful intuition. Armed with this insight, the following proof of Lemma 2.3 becomes natural.

Proof of Lemma 2.3 - General Case:

The spirit of the proof follows the earlier argument for the special case. In particular, we argue by contradiction.

Suppose that there exists a $\mathbf{x}_0$, such that $\omega(\mathbf{x}_0^H \mathbf{A}) > \omega(\mathbf{x}_0^H \bar{\mathbf{A}})$, and $\omega(\mathbf{x}_0^H \bar{\mathbf{A}}) > F - r_{\bar{\mathbf{A}}} + 1$, i.e.,

$$\omega(\mathbf{x}_0^H \bar{\mathbf{A}}) = F - r_{\bar{\mathbf{A}}} + k > F - r_{\bar{\mathbf{A}}} + 1, \quad 2 \leq k < r_{\bar{\mathbf{A}}}$$

$$\omega(\mathbf{x}_0^H \mathbf{A}) = F - r_{\bar{\mathbf{A}}} + k + l, \quad 1 \leq l \leq r_{\bar{\mathbf{A}}} - k,$$

and suppose $\omega(\mathbf{x}_0^H \bar{\mathbf{A}}) = F - r_{\bar{\mathbf{A}}} + k$ is the *smallest number* bigger than $F - r_{\bar{\mathbf{A}}} + 1$ in the sense that $\omega(\mathbf{x}^H \mathbf{A}) \leq \omega(\mathbf{x}^H \bar{\mathbf{A}})$ for any vector $\mathbf{x}$ such that

$$\omega(\mathbf{x}^H \bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}} + k - 1 < \omega(\mathbf{x}_0^H \bar{\mathbf{A}}),$$

which implies that for every $r_{\bar{\mathbf{A}}} - k + 1$ or more columns chosen from $\bar{\mathbf{A}}$, there must exist at least as many columns from $\mathbf{A}$, such that each of those from $\mathbf{A}$ is a linear combination of the said columns of $\bar{\mathbf{A}}$.

As before, with such $\mathbf{x}_0$, we know that there exist $(r_{\bar{\mathbf{A}}} - k)$ columns of $\bar{\mathbf{A}}$, say, $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$, and $(r_{\bar{\mathbf{A}}} - k - l)$ columns of $\mathbf{A}$, say, $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, such that

$$\{\mathbf{x}_0, \mathbf{0}\} \subseteq Null(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}) \bigcap Null(\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}),$$

and since $\omega(\mathbf{x}_0^H \mathbf{A}) > \omega(\mathbf{x}_0^H \bar{\mathbf{A}})$, $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ are the only columns of $\mathbf{A}$ that can possibly belong to $Span(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\})$.

What we are going to do next is different from the previous proof. We are going to partition the remaining $F - r_{\bar{\mathbf{A}}} + k$ columns of $\bar{\mathbf{A}}$, namely $\{\bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k+1}, \cdots, \bar{\mathbf{a}}_P\}$. Notice that none of those remaining columns of $\bar{\mathbf{A}}$ is going to be linearly dependent with $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$; otherwise, k can be reduced by 1. We will partition those remaining $P - r_{\bar{\mathbf{A}}} + k$ columns into[3] $M \geq 2$ non-empty disjoint subsets in the sense that each subset contains one particular remaining column and all the other columns that are the linear combinations of this particular remaining column and $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$. Let $S_i \geq 1$ denote the number of columns in the i-th partition set. Clearly,

$$\sum_{i=1}^{M} S_i = F - r_{\bar{\mathbf{A}}} + k.$$

Now, add each partition set to $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$ to form a concatenated set. Each concatenated column set of $\bar{\mathbf{A}}$ has $S_i + r_{\bar{\mathbf{A}}} - k \geq r_{\bar{\mathbf{A}}} - k + 1$ columns. Recall that for every $r_{\bar{\mathbf{A}}} - k + 1$ or more columns chosen from $\bar{\mathbf{A}}$, there must exist at least as many columns from $\mathbf{A}$, such that each of those from $\mathbf{A}$ is a linear combination of the said columns of $\bar{\mathbf{A}}$. Then, there must exist at least $(S_i + r_{\bar{\mathbf{A}}} - k)$ columns of $\mathbf{A}$ such that each of those from $\mathbf{A}$ is a linear combination of those columns of the concatenated

[3] M can be equal to 1 only if $r_{\bar{\mathbf{A}}} = 2$, however, the case $r_{\bar{\mathbf{A}}} = 2$ has been solved by Corollary 2.1. For $r_{\bar{\mathbf{A}}} \geq 3$, M cannot be 1. Suppose $M = 1$, then, according to the definition of our partition, each remaining column of $\bar{\mathbf{A}}$ is the linear combination of $\bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k+1}$ and $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$. Therefore, adding all remaining columns of $\bar{\mathbf{A}}$ to $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$ is only expected to increase the rank of $[\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}]$ by at most one. This implies the rank of $\bar{\mathbf{A}}$ is bounded by $r_{\bar{\mathbf{A}}} - k + 1$ which is less than $r_{\bar{\mathbf{A}}}$ since $k \geq 2$. Hence, $M \geq 2$.

set. We already know that $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ are the only columns of $\mathbf{A}$ that can possibly belong to $Span(\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\})$, therefore, every such $(S_i + r_{\bar{\mathbf{A}}} - k)$-column subset of $\mathbf{A}$ must have at least $S_i + l = (S_i + r_{\bar{\mathbf{A}}} - k) - (r_{\bar{\mathbf{A}}} - k - l)$ columns, other than those in $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, such that each column is a linear combination of at least one column from i-th partition set and $\{\bar{\mathbf{a}}_1, \cdots, \bar{\mathbf{a}}_{r_{\bar{\mathbf{A}}}-k}\}$.

Let ϕ_i denote the column set consisting of those $S_i + l$ columns of $\mathbf{A}$.

We claim that all ϕ_i and $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ are mutually disjoint.

Suppose this is not true. Recall that no element of ϕ_i belongs to $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$, hence there is no common element between $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ and any particular ϕ_{i0}; meanwhile, if there is a column belonging to two different ϕ_i, this will contradict the way we partition the remaining columns of $\bar{\mathbf{A}}$.

The remainder is again a counting problem. Each ϕ_i contributes at least $S_i + l$ columns. $\{\mathbf{a}_1, \cdots, \mathbf{a}_{r_{\bar{\mathbf{A}}}-k-l}\}$ also contributes $r_{\bar{\mathbf{A}}} - k - l$ columns. Summing up, we know $\mathbf{A}$ should have *at least*

$$r_{\bar{\mathbf{A}}} - k - l + \sum_{i=1}^{M}(S_i + l) = F + (M-1)l$$

columns. Since $M \geq 2$ and $l \geq 1$,

$$F + (M-1)l > F,$$

whereas $\mathbf{A}$ only has F columns. Hence, we have a contradiction. □

One natural question that arises at this point is whether one can further improve Lemma 2.1, in the sense that Lemma 2.1 can be viewed as an improved version of Lemma 2.2. Does the conclusion of Lemma 2.1 hold if we pose a smaller bound on the right hand side of (2.4)? The answer is *no* in general. It is known that $k_{\mathbf{A}} = r_{\mathbf{A}}$ almost surely when $\mathbf{A}$ is drawn from a continuous distribution. With the aid of Remark 2.1, it can be seen that given a matrix $\bar{\mathbf{A}}$ with $k_{\bar{\mathbf{A}}} = r_{\bar{\mathbf{A}}}$, even if $\omega(\mathbf{x}^H\mathbf{A}) \leq \omega(\mathbf{x}^H\bar{\mathbf{A}})$ for any vectors with $\omega(\mathbf{x}^H\bar{\mathbf{A}}) \leq F - r_{\bar{\mathbf{A}}}$, $\mathbf{A}$ and $\bar{\mathbf{A}}$ are *not* necessarily equivalent up to permutation and scaling. The Lemma can be relaxed when $k_{\bar{\mathbf{A}}} = 1$, but this is not the case of interest.

Chapter 3

Multi-dimensional Harmonic Retrieval

Determining the maximum number of resolvable harmonics is a parameter identifiability problem, whose solution for the case of one-dimensional harmonics goes back to Carathéodory [6]; see also [45, 60]. In two or higher dimensions, the identifiability problem is considerably harder, but also more interesting. The reason is that, in many applications of higher-dimensional harmonic retrieval, one is constrained in the number of samples that can be taken along certain dimensions, usually due to hardware and/or cost limitations. Examples include ultrasound imaging [10] and direction of arrival (spatial frequency) estimation. The question that arises is whether the number of samples taken along any particular dimension bounds the overall number of resolvable harmonics or not.

Essentially all of the work to date on identifiability of multidimensional harmonic retrieval deals with the 2-D case (e.g., [34, 67]), and provides sufficient identifiability conditions that are constrained by $\min(I, J)$, where I denotes the number of samples taken along one dimension, and J likewise for the other dimension. To the best of our knowledge, the most relaxed condition to date has been derived in [50], which shows that identifiability is determined by the *sum* $I+J$. The result of [50] is deterministic, in the sense that no statistical assumptions are needed aside from the requirement that the frequencies along *each* dimension are distinct. Furthermore, it generalizes naturally to N dimensions for arbitrary N, and shows that identifiability improves with increasing N, which is intuitively pleasing. However, the sufficient condition in [50] improves with the *sum* of $I, J, K, \cdots$, whereas total sample size grows with the *product* of $I, J, K, \cdots$. This indicates that significantly stronger results are *possible*.

The contribution of this chapter is the derivation of stochastic identifiability results for multidimensional harmonic retrieval which fulfill this potential. Our tools allow us to treat the general case of multidimensional complex exponentials that incorporate real exponential components (e.g., decay rates). We thus make no distinction between the terms *harmonic* and *exponential*. We show that if the number of 2-D harmonics is less than or equal to roughly $IJ/4$, then, assuming sampling at the Nyquist rate or above, the parameterization (including the pairing of parameters) is $P_{\mathcal{L}}(\mathbb{C}^{2F})$-almost surely identifiable, where F is the number of harmonics and $P_{\mathcal{L}}(\mathbb{C}^{2F})$ is the distribution used to draw the $2F$ complex decay/frequency parameters, assumed continuous with respect to the Lebesgue measure in $\mathbb{C}^{2F}$. In plain words, this means that if F is under roughly $IJ/4$, then the model parameters (amplitudes, phases, decay rates, and frequencies, including pairing thereof) that give rise to the observed noiseless data are unique *for almost every selection of complex decay/frequency parameters*, or, if one draws the complex decay/frequency parameters from a continuous distribution over $\mathbb{C}^{2F}$, then the probability that one encounters a non-identifiable model is zero. This result is subsequently generalized to N dimensions for arbitrary N.

The rest of this chapter is structured as follows. We begin with a discussion of notation and preliminaries. Section 3.2 summarizes earlier deterministic identifiability results, while section 3.3 illuminates the rank properties of the Khatri-Rao matrix product. Both are needed to prove the stochastic identifiability results presented herein. In particular, section 3.3 proves that the Khatri-Rao product is full rank almost surely[1]. Our main contributions are presented in sections 3.4 and 3.5. Section 3.4 contains the 2-dimensional result, whereas section 3.5 contains its generalization to arbitrary number of dimensions. The proof of the latter is highly technical, and therefore deferred to the Appendix, along with other proofs of auxiliary results. Some comments and extensions of the main results are collected in section 3.6. Conclusions are drawn in section 3.7.

3.1 Some preliminaries

$\mathbb{C}$ denotes the complex numbers, $\mathbf{U} = \{x \in \mathbb{C} \mid |x| = 1\}$ denotes the unit circle, $\mathbb{C}^F = \overbrace{\mathbb{C} \times \mathbb{C} \times \cdots \times \mathbb{C}}^{F}$, and $\mathbf{U}^F = \overbrace{\mathbf{U} \times \mathbf{U} \times \cdots \times \mathbf{U}}^{F}$. Matrices (vectors) are denoted by boldface capital (lowercase) letters. N denotes the number of dimensions, whereas I_n denotes the number of (equispaced) samples along the n-th dimension. An N-dimensional (also known as N-way) array is a dataset that is indexed by N indices: $x_{i_1,\cdots,i_N}$, where $i_n \in \{1, \cdots, I_n\}$, $n = 1, \cdots, N$. We do not follow the usual convention of using i or j to denote $\sqrt{-1}$; instead we explicitly write $\sqrt{-1}$ when needed, and use i (j) as row (respectively, column) index, in accordance with common practice in matrix algebra. We also make extensive use of superscripts to denote variables stemming from a given variable.

The rank of a matrix (2-way array) $\mathbf{A}$ is the smallest number of rank-one matrices needed to decompose $\mathbf{A}$ into a sum of rank-one factors. Each rank-one factor is the outer product of two vectors. Matrix rank can be equivalently defined as the maximum number of linearly independent columns (or rows) that can be drawn from $\mathbf{A}$. We will use $r_{\mathbf{A}}$ to denote the rank of $\mathbf{A}$. The rank of an N-way array is defined as the smallest number of rank-one N-way factors needed to decompose it [30].

[1]This statement has to be interpreted properly; see section 3.3.

Each rank-one N-way factor is the "outer product" of N vectors, meaning that its $(i_1, \cdots, i_N)$-th element is given by $a_{f,1,i_1} \cdots a_{f,N,i_N}$, where f is a factor index. Thus, an N-way array of rank F can be written as:

$$x_{i_1,\cdots,i_N} = \sum_{f=1}^{F} c_f \prod_{n=1}^{N} a_{f,n,i_n}.$$

The Kruskal-rank or k-rank [30] of a matrix $\mathbf{A}$ (denoted by $k_{\mathbf{A}}$) is r if *every* r columns of $\mathbf{A}$ are linearly independent, and either $\mathbf{A}$ has r columns or $\mathbf{A}$ contains a set of $r+1$ linearly dependent columns. The k-rank of $\mathbf{A}$ is therefore the maximum number of linearly independent columns that can be drawn from $\mathbf{A}$ in an *arbitrary fashion*. Note that k-rank is generically asymmetric: the k-rank of a matrix need not be equal to the k-rank of its transpose. k-rank is always less than or equal to rank.

A constant-envelope 1-dimensional discrete-time exponential is written as $x_i = ce^{(\sqrt{-1})\omega(i-1)}$, $i = 1, \cdots, I$, where $c \in \mathbb{C}$ accounts for both amplitude and phase. A non-constant-envelope 1-dimensional exponential is written as $x_i = ca^{i-1}$, $i = 1, \cdots, I$, where $a \in \mathbb{C}$ accounts for both decay (or growth) rate and frequency. A 2-dimensional exponential is simply the product of two 1-dimensional exponentials indexed by different independent variables, i.e., $x_{i_1,i_2} = ca_1^{i_1-1}a_2^{i_2-1}$, $i_1 = 1, \cdots, I_1$, $i_2 = 1, \cdots, I_2$; and likewise for higher dimensions.

An $m \times \rho$ Vandermonde matrix with generators $\alpha_1, \alpha_2, \cdots, \alpha_\rho \in \mathbb{C}$ is given by

$$\mathbf{V} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_\rho \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_\rho^2 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{m-1} & \alpha_2^{m-1} & \cdots & \alpha_\rho^{m-1} \end{bmatrix}.$$

If the generators are distinct, then $\mathbf{V}$ is full rank [58] and also full k-rank [55]: $k_{\mathbf{V}} = r_{\mathbf{V}} = \min(m, \rho)$.

3.2 Deterministic identifiability results

We will make use of the following results.

Theorem 3.1 *(Identifiability of low-rank decomposition of N-way arrays [51, 52]) Consider the F-component N-linear model*

$$x_{i_1,\cdots,i_N} = \sum_{f=1}^{F} c_f \prod_{n=1}^{N} a_{f,n,i_n},$$

for $i_n = 1, \cdots, I_n \geq 2$, $n = 1, \cdots, N$, *with* $c_f \in \mathbb{C}$, $a_{f,n,i_n} \in \mathbb{C}$. *Let* $\mathbf{A}^{(n)}$ *denote the* $I_n \times F$ *matrix with* (i_n, f) *element* a_{f,n,i_n}. *If*

$$\sum_{n=1}^{N} k_{\mathbf{A}^{(n)}} \geq 2F + (N-1),$$

then given the N-way array $x_{i_1,\cdots,i_N}$, $i_n = 1, \cdots, I_n$, $n = 1, \cdots, N$, *its F rank-one N-way factors*

$$c_f \prod_{n=1}^{N} a_{f,n,i_n}, \; f = 1, \cdots, F$$

are unique.

Kruskal was the one who developed the backbone result for $N = 3$ and array elements drawn from $\mathbb{R}$ [30]. See also [53–55] for other related results.

Theorem 3.2 *(Deterministic identifiability of N-dimensional harmonic retrieval [50]) Given a sum of F exponentials in N-dimensions*

$$x_{i_1,\cdots,i_N} = \sum_{f=1}^{F} c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1},$$

for $i_n = 1, \cdots, I_n \geq 2$, $n = 1, \cdots, N$, *with* $c_f \in \mathbb{C}$ *and* $a_{f,n} \in \mathbb{C}$ *such that* $a_{f_1,n} \neq a_{f_2,n}$, $\forall f_1 \neq f_2$ *and all* n, *if*

$$\sum_{n=1}^{N} I_n \geq 2F + (N-1),$$

then there exist unique $(a_{f,n},\ n = 1, \cdots, N;\ c_f)$, $f = 1, \cdots, F$ *that give rise to* $x_{i_1,\cdots,i_N}$. *If an additional* M *non-exponential dimensions are available,*

$$x_{i_1,\cdots,i_N,j_1,\cdots,j_M} = \sum_{f=1}^{F} c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1} \prod_{m=1}^{M} b_{f,m,j_m}, \tag{3.1}$$

for $j_m = 1, \cdots, J_m \geq 2$, $m = 1, \cdots, M$, *with* $b_{f,m,1} = 1$, $\forall f, m$ *by convention, then uniqueness (including the associated component vectors along non-exponential dimensions) holds provided that*

$$\sum_{n=1}^{N} I_n + \sum_{m=1}^{M} k_{\mathbf{B}^{(m)}} \geq 2F + (N + M - 1),$$

where $\mathbf{B}^{(m)}$ *denotes the* $J_m \times F$ *matrix with* (j_m, f) *element* b_{f,m,j_m}.

3.3 On rank and k-rank of the Khatri-Rao product

Consider two Vandermonde matrices

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_F \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_F^2 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{I-1} & \alpha_2^{I-1} & \cdots & \alpha_F^{I-1} \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_F \\ \beta_1^2 & \beta_2^2 & \cdots & \beta_F^2 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1^{J-1} & \beta_2^{J-1} & \cdots & \beta_F^{J-1} \end{bmatrix}, \end{aligned} \tag{3.2}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_F$ and $\beta_1, \beta_2, \cdots, \beta_F$ are complex generators. The Khatri-Rao product of $\mathbf{A}$ and $\mathbf{B}$ is

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_F \\ \beta_1^2 & \beta_2^2 & \cdots & \beta_F^2 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1^{J-1} & \beta_2^{J-1} & \cdots & \beta_F^{J-1} \\ \alpha_1 & \alpha_2 & \cdots & \alpha_F \\ \alpha_1\beta_1 & \alpha_2\beta_2 & \cdots & \alpha_F\beta_F \\ \alpha_1\beta_1^2 & \alpha_2\beta_2^2 & \cdots & \alpha_F\beta_F^2 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^i\beta_1^j & \alpha_2^i\beta_2^j & \cdots & \alpha_F^i\beta_F^j \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{I-1}\beta_1^{J-1} & \alpha_2^{I-1}\beta_2^{J-1} & \cdots & \alpha_F^{I-1}\beta_F^{J-1} \end{bmatrix}.$$

One can show that full rank (even full k-rank) of both $\mathbf{A}$ and $\mathbf{B}$ does not necessarily guarantee that the Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ is full rank (let alone full k-rank). For example, let $F = 6$. The generators can be chosen as follows:

$$\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 3, \alpha_4 = 4, \alpha_5 = 5, \alpha_6 = 6.$$

$$\beta_1 = 1, \beta_2 = \sqrt{2}, \beta_3 = \sqrt{3}, \beta_4 = \sqrt{4}, \beta_5 = \sqrt{5}, \beta_6 = \sqrt{6}.$$

With this choice of generators, $\mathbf{A}$ and $\mathbf{B}$ are full k-rank. When $I = 3$ and $J = 2$, the 6×6 Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ is full rank, hence also full k-rank: $k_{\mathbf{A}\odot\mathbf{B}} = r_{\mathbf{A}\odot\mathbf{B}} = 6$. Now set $I = 2$ and $J = 3$; the Khatri-Rao product is still 6×6, but[2] its rank is 5.

Irrespective of Vandermonde structure, it is simple to show that

$$r_{\mathbf{A}\odot\mathbf{B}} \leq r_{\mathbf{A}} r_{\mathbf{B}},$$

[2] It will be shown that, with proper random sampling, this phenomenon is a measure-zero event; see Theorem 3.3, and Corollary 3.1.

e.g., by noting that the Khatri-Rao product of **A** and **B** is a selection of columns drawn from the Kronecker product of **A** and **B**. The rank of the Kronecker product is the product of ranks of the constituent matrices [2].

The following result provides a deterministic lower bound on the k-rank of the Khatri-Rao product, irrespective of Vandermonde structure. Note that since rank ≥ k-rank, it also provides a lower bound on rank.

Lemma 3.1 *[55] Given two matrices* $\mathbf{A} \in \mathbb{C}^{I\times F}$ *and* $\mathbf{B} \in \mathbb{C}^{J\times F}$*, if* $k_{\mathbf{A}} \geq 1$ *and* $k_{\mathbf{B}} \geq 1$*, then it holds that*

$$k_{\mathbf{A}\odot\mathbf{B}} \geq \min(k_{\mathbf{A}} + k_{\mathbf{B}} - 1,\ F) \tag{3.3}$$

Other researchers have noted that the Khatri-Rao product appears to exhibit full rank in essentially all cases of practical interest [66], but no rigorous argument has been given to justify this observation to date. The following two results settle this issue[3].

Theorem 3.3 *For a pair of Vandermonde matrices* $\mathbf{A} \in \mathbb{C}^{I\times F}$ *and* $\mathbf{B} \in \mathbb{C}^{J\times F}$

$$r_{\mathbf{A}\odot\mathbf{B}} = k_{\mathbf{A}\odot\mathbf{B}} = \min(IJ, F), \quad P_{\mathcal{L}}(\mathbb{C}^{2F}) - a.s., \tag{3.4}$$

where $P_{\mathcal{L}}(\mathbb{C}^{2F})$ *is the distribution used to draw the* $2F$ *complex generators for* **A** *and* **B***, assumed continuous with respect to the Lebesgue measure in* $\mathbb{C}^{2F}$*.*

As an almost direct by-product, we obtain

Corollary 3.1 *For a pair of matrices* $\mathbf{A} \in \mathbb{C}^{I\times F}$ *and* $\mathbf{B} \in \mathbb{C}^{J\times F}$*,*

$$r_{\mathbf{A}\odot\mathbf{B}} = k_{\mathbf{A}\odot\mathbf{B}} = \min(IJ, F), \quad P_{\mathcal{L}}(\mathbb{C}^{(I+J)F}) - a.s., \tag{3.5}$$

where $P_{\mathcal{L}}(\mathbb{C}^{(I+J)F})$ *is the distribution used to draw the* $(I+J)F$ *complex elements of* A *and* B*, assumed continuous with respect to the Lebesgue measure in* $\mathbb{C}^{(I+J)F}$*.*

Equipped with these results, we proceed to address the main problem of interest herein.

[3]Proofs can be found in the Appendix.

3.4 Almost sure identifiability of 2-D harmonic retrieval

Proposition 1 [4] *Given a sum of F 2-D exponentials*

$$x_{i,j} = \sum_{f=1}^{F} c_f a_f^{i-1} b_f^{j-1}, \tag{3.6}$$

for $i = 1, \ldots, I \geq 4$, and $j = 1, \ldots, J \geq 4$, the parameter triples (a_f, b_f, c_f), $f = 1, \cdots, F$ are $P_{\mathcal{L}}(\mathbb{C}^{2F})$-a.s. unique[5]*, where $P_{\mathcal{L}}(\mathbb{C}^{2F})$ is the distribution used to draw the $2F$ complex exponential parameters (a_f, b_f), $f = 1, \cdots, F$, assumed continuous with respect to the Lebesgue measure in $\mathbb{C}^{2F}$, provided that there exist four integers, I_1, I_2, J_1, J_2 such that*

$$I - I_1 - I_2 + \min(I_1 J_1, F) + \min(I_2 J_2, F) \geq 2F, \tag{3.7}$$

subject to

$$I_1 + I_2 \leq I, \quad J_1 + J_2 = J + 1, \quad \min(I_1, I_2, J_1, J_2) \geq 2. \tag{3.8}$$

Proof:

We first define a 5-way array with typical element

$$\begin{aligned} \widehat{x}_{i_1,i_2,i_3,j_1,j_2} &:= x_{i_1+i_2+i_3-2, j_1+j_2-1} \\ &= \sum_{f=1}^{F} c_f a_f^{i_1+i_2+i_3-1-1-1} b_f^{j_1+j_2-1-1} \\ &= \sum_{f=1}^{F} c_f a_f^{i_1-1} a_f^{i_2-1} a_f^{i_3-1} b_f^{j_1-1} b_f^{j_2-1} \end{aligned} \tag{3.9}$$

where $i_\alpha = 1, \cdots, I_\alpha \geq 2$, and $j_\beta = 1, \cdots, J_\beta \geq 2$, for $\alpha = 1, 2, 3$, $\beta = 1, 2$. Since $\min(I, J) \geq 4$ has been assumed in the statement of the proposition, such extension

[4] The result holds true if we switch I and J.

[5] We assume throughout that sampling is at the Nyquist rate or higher, to avoid spectral folding. This allows us to restrict attention to discrete-time frequencies in $(-\pi, \pi]$.

to 5 ways is always feasible. Define matrices

$$\mathbf{A}_\alpha = (a_f^{i_\alpha - 1}) \in \mathbb{C}^{I_\alpha \times F}, \qquad \mathbf{B}_\beta = (b_f^{j_\beta - 1}) \in \mathbb{C}^{J_\beta \times F}. \tag{3.10}$$

The next step is to nest the 5-way array $\widehat{x}$ into a three-way array $\bar{x}$ by collapsing two pairs of dimensions as follows

$$\begin{aligned}
\bar{x}_{i_3,k,l} &:= \widehat{x}_{\lceil \frac{k}{J_1} \rceil, \lceil \frac{l}{J_2} \rceil, i_3, k-(\lceil \frac{k}{J_1} \rceil - 1)J_1, l-(\lceil \frac{l}{J_2} \rceil - 1)J_2} \\
&= \sum_{f=1}^{F} c_f \; (\; a_f^{\lceil \frac{k}{J_1} \rceil - 1} a_f^{\lceil \frac{l}{J_2} \rceil - 1} a_f^{i_3 - 1} b_f^{k-(\lceil \frac{k}{J_1} \rceil - 1)J_1 - 1} \\
&\quad \times b_f^{l-(\lceil \frac{l}{J_2} \rceil - 1)J_2 - 1} \;) \\
&= \sum_{f=1}^{F} c_f a_f^{i_3 - 1} \; (\; a_f^{\lceil \frac{k}{J_1} \rceil - 1} b_f^{k-(\lceil \frac{k}{J_1} \rceil - 1)J_1 - 1} \\
&\quad \times a_f^{\lceil \frac{l}{J_2} \rceil - 1} b_f^{l-(\lceil \frac{l}{J_2} \rceil - 1)J_2 - 1} \;) \\
&= \sum_{f=1}^{F} c_f a_f^{i_3 - 1} d_{k,f} e_{l,f},
\end{aligned} \tag{3.11}$$

for $k = 1, \cdots, I_1 J_1$, $l = 1, \cdots, I_2 J_2$, with $d_{k,f}$ and $e_{l,f}$ given by

$$\begin{aligned}
d_{k,f} &:= a_f^{\lceil \frac{k}{J_1} \rceil - 1} b_f^{k-(\lceil \frac{k}{J_1} \rceil - 1)J_1 - 1}, \\
e_{l,f} &:= a_f^{\lceil \frac{l}{J_2} \rceil - 1} b_f^{l-(\lceil \frac{l}{J_2} \rceil - 1)J_2 - 1}.
\end{aligned} \tag{3.12}$$

Define matrices

$$\mathbf{D} = (d_{k,f}) \in \mathbb{C}^{I_1 J_1 \times F}, \qquad \mathbf{E} = (e_{l,f}) \in \mathbb{C}^{I_2 J_2 \times F}. \tag{3.13}$$

D and **E** are nothing but

$$\mathbf{D} = \mathbf{A}_1 \odot \mathbf{B}_1, \qquad \mathbf{E} = \mathbf{A}_2 \odot \mathbf{B}_2. \tag{3.14}$$

Since $\mathbf{A}_3$ is Vandermonde, Theorem 3.2 can be invoked to claim uniqueness, provided

$$I_3 + k_{\mathbf{D}} + k_{\mathbf{E}} \geq 2F + 3 - 1. \tag{3.15}$$

Note that for any particular i_3, k and l, the product $c_f a_f^{i_3-1} d_{k,f} e_{l,f}$ is equal to $c_f a_f^{i-1} b_f^{j-1}$ with the following choice of i and j:

$$
\begin{aligned}
i &= i_3 + \lceil \frac{k}{J_1} \rceil + \lceil \frac{l}{J_2} \rceil - 2, \\
j &= k - (\lceil \frac{k}{J_1} \rceil - 1) J_1 + l - (\lceil \frac{l}{J_2} \rceil - 1) J_2 - 1.
\end{aligned}
$$

As i_3, k and l span their range, the corresponding i and j span their respective range. It follows that uniqueness of the F rank-one 3-D factors $c_f a_f^{i_3-1} d_{k,f} e_{l,f}$ is equivalent to uniqueness of the F rank-one 2-D factors $c_f a_f^{i-1} b_f^{j-1}$, $f = 1, \cdots, F$. Therefore, the rank-one factors $c_f a_f^{i-1} b_f^{j-1}$ and hence the triples (a_f, b_f, c_f), $f = 1, \cdots, F$, are unique provided that (3.15) holds true. Invoking Theorem 3.3, almost sure uniqueness holds provided there exist integers, $I_1, I_2, I_3, J_1, J_2 \geq 2$ such that

$$I_3 + \min(I_1 J_1, F) + \min(I_2 J_2, F) \geq 2F + 2, \tag{3.16}$$

subject to[6]

$$
\begin{cases}
I_1 + I_2 + I_3 = I + 2, \\
J_1 + J_2 = J + 1, \\
\min(I_1, I_2, I_3, J_1, J_2) \geq 2.
\end{cases} \tag{3.17}
$$

Setting $I_3 = I + 2 - I_1 - I_2$, we obtain

$$I - I_1 - I_2 + \min(I_1 J_1, F) + \min(I_2 J_2, F) \geq 2F,$$

subject to

$$
\begin{cases}
I_1 + I_2 \leq I, \\
J_1 + J_2 = J + 1, \\
\min(I_1, I_2, J_1, J_2) \geq 2,
\end{cases}
$$

and the proof is complete. □

[6]The first two conditions assure that we do not index beyond the available data sample.

Theorem 3.4 [7] *Given a sum of F 2-D exponentials*

$$x_{i,j} = \sum_{f=1}^{F} c_f a_f^{i-1} b_f^{j-1}, \tag{3.18}$$

for $i = 1, \ldots, I \geq 4$, and $j = 1, \ldots, J \geq 4$, the parameter triples (a_f, b_f, c_f), $f = 1, \cdots, F$ are $P_{\mathcal{L}}(\mathbb{C}^{2F})$-a.s. unique, where $P_{\mathcal{L}}(\mathbb{C}^{2F})$ is the distribution used to draw the $2F$ complex exponential parameters (a_f, b_f), $f = 1, \cdots, F$, assumed continuous with respect to the Lebesgue measure in $\mathbb{C}^{2F}$, provided that

$$F \leq \lfloor \frac{I}{2} \rfloor \lceil \frac{J}{2} \rceil. \tag{3.19}$$

Proof: If both I and J are even numbers , pick $I_1 = I_2 = \frac{I}{2}$, $J_1 = \frac{J}{2}$ and $J_2 = \frac{(J+2)}{2}$ (thereby satisfying condition (3.8)), and (3.7) becomes

$$\min(\frac{IJ}{4}, F) + \min(\frac{I(J+2)}{4}, F) \geq 2F, \tag{3.20}$$

which is satisfied for any $F \leq IJ/4$. If I is even, J is odd, pick $I_1 = I_2 = \frac{I}{2}$, and $J_1 = J_2 = \frac{J+1}{2}$ (thereby satisfying condition (3.8)), and (3.7) becomes

$$\min(\frac{I(J+1)}{4}, F) + \min(\frac{I(J+1)}{4}, F) \geq 2F, \tag{3.21}$$

which is satisfied for any $F \leq I(J+1)/4$. If both I and J are odd, pick $I_1 = \frac{(I-1)}{2}$, $I_2 = \frac{(I+1)}{2}$, $J_1 = J_2 = \frac{(J+1)}{2}$ (satisfying (3.8)), and (3.7) becomes

$$\min(\frac{(I-1)(J+1)}{4}, F) + \min(\frac{(I+1)(J+1)}{4}, F) \geq 2F, \tag{3.22}$$

satisfied for any $F \leq (I-1)(J+1)/4$. Finally, if I is odd and J is even, pick $I_1 = \frac{I-1}{2}$, $I_2 = \frac{I+1}{2}$, $J_1 = \frac{J}{2}$ and $J_2 = \frac{J+2}{2}$ (satisfying (3.8)), and (3.7) becomes

$$\min(\frac{(I-1)(J)}{4}, F) + \min(\frac{(I+1)(J+2)}{4}, F) \geq 2F, \tag{3.23}$$

satisfied for any $F \leq (I-1)J/4$. Invoking Proposition 1 completes the proof. □

[7] The Theorem holds true if I and J are switched.

Remark 3.1 *Some reflection reveals that the argument behind the proof of Theorem 3.4 (and also its N-D generalization, Theorem 3.5) is in fact constructive, leading to an eigenvalue solution that recovers everything exactly under only the model identifiability condition in the Theorem, in the noiseless case. Our results have subsequently led to the development of an effective algebraic identification algorithm for N-D harmonic retrieval [37].*

Remark 3.2 *It is interesting to note that equations-versus-unknowns considerations indicate a bound of $IJ/3$, without taking the pairing issue into consideration. To see this, note that each of the F 2-D exponential components is parameterized by 3 complex parameters, and a total of IJ complex data points are given. If the equations-versus-unknowns bound is violated, then, under certain conditions, the implicit function theorem indicates that infinitely many ambiguous solutions exist in the neighborhood of the true solution.*

3.5 The N-Dimensional Case

The result can be generalized to the N-dimensional case. Although the spirit of the associated proof is clear, the mathematical argument is highly technical. This is so primarily because one is forced to use a recursive dimensionality-embedding argument to preserve generality. We therefore state the result and defer the proof to the end of the Appendix, noting that Figures 3.1, 3.2 help convey the essence of the proof to the interested reader.

Theorem 3.5 [8] *Given a sum of F N-D exponentials*

$$x_{i_1,\cdots,i_N} = \sum_{f=1}^{F} c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1}, \tag{3.24}$$

[8] The Theorem holds true for any permutation of $\{I_n\}_{n=1}^N$

for $i_n = 1, \cdots, I_n \geq 4$, $n = 1, \cdots, N$, *the parameter* $(N+1)$*-tuples* $(a_{f,1}, \cdots, a_{f,N}, c_f)$, $f = 1, \cdots, F$, *are* $P_{\mathcal{L}}(\mathbb{C}^{NF})$*-a.s. unique, where* $P_{\mathcal{L}}(\mathbb{C}^{NF})$ *is the distribution used to draw the* NF *complex exponential parameters* $(a_{f,1}, \cdots, a_{f,N})$*, for* $f = 1, \cdots, F$*, assumed continuous with respect to Lebesgue measure in* $\mathbb{C}^{NF}$*, provided that*

$$F \leq \lfloor \frac{I_1}{2} \rfloor \prod_{n=2}^{N} \lceil \frac{I_n}{2} \rceil. \tag{3.25}$$

3.6 Comments and Extensions

The restriction of at least 4 samples per dimension is an artifact of the proof. In fact, we can also treat cases with less than 4 samples in any dimension(s). However, in the 2-D case with less than 4 samples per dimension, our approach does not yield anything significant. In the N-D case, having less than 4 samples along certain dimensions breaks the symmetry of the problem, forcing us to separately consider cases, depending on the number and sample size distribution of dimensions having less than 4 samples. This prohibits a concise unifying treatment. Nevertheless, individual cases can be easily dealt with, given the tools developed herein.

3.6.1 Constant-Envelope Exponentials

So far, we have considered multidimensional complex exponentials that incorporate real exponential components. In many applications, one deals with constant-envelope complex exponentials. The proof of Theorem 3.4 carries through verbatim in this case, except that one needs to ensure that Theorem 3.3 holds for generators drawn from the unit circle, $\mathbf{U}$. This is easy, because the generic example which shows that the determinant is nontrivial in the proof of Theorem 3.3 was actually constructed using generators drawn from the unit circle. We therefore have the following Corollary:

Corollary 3.2 *Given a sum of F 2-D constant-envelope complex exponentials*

$$x_{i,j} = \sum_{f=1}^{F} c_f e^{\sqrt{-1}\omega_f(i-1)} e^{\sqrt{-1}\nu_f(j-1)},$$

for $i = 1, \ldots, I \geq 4$, and $j = 1, \ldots, J \geq 4$, the parameter triples $(e^{\sqrt{-1}\omega_f}, e^{\sqrt{-1}\nu_f}, c_f)$, $f = 1, \cdots, F$ are $P_{\mathcal{L}}(\mathbf{U}^{2F})$-a.s. unique, provided that

$$F \leq \lfloor \frac{I}{2} \rfloor \lceil \frac{J}{2} \rceil.$$

The same argument holds for Proposition 4 and Theorem 3.5 in the case of constant-envelope complex exponentials; we skip the corresponding statements for brevity.

3.6.2 Common Frequency Mode

In most applications, having two or more *identical* frequencies along a certain dimension is a measure zero event. Having two frequencies close to each other is very common, but this affects performance, rather than identifiability. In certain applications, identical frequencies along one or two dimensions are in fact a modeling assumption, motivated by proximity of actual frequencies and compactness of model parameterization [34]. For this reason, it is of interest to investigate identifiability subject to common frequency constraints. This can be handled using the tools developed herein, but one needs to check on a case-by-case basis, depending on the "common mode configuration": how many distinct frequencies ("batches") per dimension, how many components per batch, and what is the pairing across dimensions. In general, the problem is combinatorial and a unified treatment does not seem to be possible. The reason is that one needs to construct a "generic" example (cf. the proof of Theorem 3.3) to demonstrate that the determinant of the associated Khatri-Rao product is nontrivial, for each common mode configuration. We illustrate how this situation can be handled in the 2-D case with a pair of 2-D exponentials having one frequency in common. Interestingly, we obtain exactly the same identifiability condition as before. The proof of the following result can be found in the Appendix.

Proposition 2 *Given a sum of F 2-D exponentials*

$$x_{i,j} = \sum_{f=1}^{F} c_f a_f^{i-1} b_f^{j-1},$$

for $i = 1, \ldots, I \geq 4$, *and* $j = 1, \ldots, J \geq 4$, *with* $b_2 = b_1$, *the parameter triples* (a_f, b_f, c_f), $f = 1, \cdots, F$ *are* $P_{\mathcal{L}}(\mathbb{C}^{2F-1})$*-a.s. unique, where* $P_{\mathcal{L}}(\mathbb{C}^{2F-1})$ *is the distribution used to draw the* $(2F-1)$ *complex exponential parameters* $(a_1, a_2, \cdots, a_F, b_1, b_3, \cdots, b_F)$, *assumed continuous with respect to the Lebesgue measure in* $\mathbb{C}^{2F-1}$, *provided that*

$$F \leq \lfloor \frac{I}{2} \rfloor \lceil \frac{J}{2} \rceil.$$

3.6.3 Non-Exponential Dimension(s)

In certain situations, the signals along one dimension are not exponentials, e.g., in uniform rectangular sensor array processing with two exponential (spatial) dimensions and a non-exponential temporal dimension. Our results can be extended to handle this case as well. As an example, we have the following result[9].

Proposition 3 *Consider*

$$x_{i,j,k} = \sum_{f=1}^{F} c_f a_f^{i-1} b_f^{j-1} s_{k,f},$$

for $i = 1, \ldots, I$, *and* $j = 1, \ldots, J$, *where* $k = 1, \ldots, K$, *is a temporal index, and assume that the temporal signal matrix* $\mathbf{S} = (s_{k,f}) \in \mathbb{C}^{K \times F}$ *is full column rank* F. *If* $\max(I, J) \geq 3$, *and*

$$F \leq IJ - \min(I, J),$$

[9]Note that, assuming sufficiently many temporal samples and persistence of excitation, and taking $M_3 = L_3 = L = 1$ in equation (22) of [13], yields $F \leq \min(I(J-1), J(I-1)) = IJ - \max(I, J)$; this is worse but close to our result in Proposition 3, albeit [13] contains no proof.

then the parameterization in terms of $\left(a_f, b_f, c_f, \{s_{k,f}\}_{k=1}^K\right)$, $f = 1, \cdots, F$, *is* $P_{\mathcal{L}}(\mathbb{C}^{2F})$*-a.s. unique, where* $P_{\mathcal{L}}(\mathbb{C}^{2F})$ *is the distribution used to draw the* $2F$ *complex exponential parameters* (a_f, b_f), $f = 1, \ldots, F$, *assumed continuous with respect to the Lebesgue measure in* $\mathbb{C}^{2F}$.

3.7 Conclusions

We have derived stochastic identifiability results for multidimensional harmonic retrieval. The sufficient conditions provided are the most relaxed to date. The sufficient condition for the 2-D case is not far from equations-versus-unknowns considerations - hence additional improvements, if any, will be marginal. In the N-D case, the resolvability bound is proportional to total sample size, but the proportionality factor is dependent on N. Although this is not a serious limitation, it does indicate that one moves further from the equations-versus-unknowns bound in higher dimensions. It remains to be seen whether a significantly tighter bound can be found in higher dimensions.

Appendix 3.A Proofs

We will need to invoke the following Lemma.

Lemma 3.2 *Consider an analytic function $h(\mathbf{x})$ of several complex variables $\mathbf{x} = [x_1, \cdots, x_n]^T \in \mathbb{C}^n$. If h is non-trivial in the sense that there exists $\mathbf{x}_0 \in \mathbb{C}^n$ such that $h(\mathbf{x}_0) \neq 0$, then the zero set of $h(\mathbf{x})$*

$$\mathcal{Z} := \{\mathbf{x} \in \mathbb{C}^n | h(\mathbf{x}) = 0\}$$

is of measure (Lebesgue measure in $\mathbb{C}^n$) zero.

This Lemma is known (e.g., [60]), but we have not been able to find a satisfactory proof in the literature. We therefore include a simple proof for completeness.

Proof of Lemma 3.2: If $n = 1$, it is well known that $\mathcal{Z}$ is countable (e.g., see [9] Theorem 3.7[10]). For $n > 1$, define $g(\mathbf{x}) = 1$ if $h(\mathbf{x}) = 0$, and $g(\mathbf{x}) = 0$, otherwise. The measure of $\mathcal{Z}$ is the integral of $g(\mathbf{x})$ over $\mathbb{C}^n$. Fix $x_2, x_3, \cdots, x_n$, and consider the single-variable function $h(x_1, x_2, ..., x_n)$. This is analytic in x_1, hence its zero set is of measure zero. This means that, for any fixed $x_2, \cdots, x_n$,

$$\int g(x_1, x_2, \cdots, x_n) dx_1 = 0.$$

Hence

$$\begin{aligned}\int \cdots \int g dx_1 dx_2 \cdots dx_n &= \int \cdots \left(\int g dx_1 \right) dx_2 \cdots dx_n \\ &= \int \cdots \int 0 \, dx_2 \cdots dx_n \\ &= 0.\end{aligned}$$

Note that the argument works irrespective of order of integration - hence the multi-dimensional integral is indeed zero, by Fubini's Theorem. This completes the proof. □

[10] Any uncountable set in the complex plane must have at least one limit point because any complex Cauchy sequence must have one and only one complex limit.

Proof of Theorem 3.3: We will show that

$$r_{\mathbf{A}\odot\mathbf{B}} = k_{\mathbf{A}\odot\mathbf{B}} = \min(IJ, F), \quad P_{\mathcal{L}}(\mathbb{C}^{2F}) - a.s.$$

The general case can be reduced to the $IJ = F$ case. If $IJ \leq F$, it suffices to prove that the result holds for an *arbitrary selection* of IJ columns; if $IJ \geq F$, then it suffices to prove that the result holds for any row-reduced square submatrix.

When $IJ = F$, full rank and full k-rank can be established by showing that the determinant of $\mathbf{A} \odot \mathbf{B}$ is nonzero. Define

$$\begin{aligned} &H(\alpha_1, \cdots, \alpha_F, \beta_1, \cdots, \beta_F) \\ &= det(\mathbf{A}(\alpha_1, \cdots, \alpha_F) \odot \mathbf{B}(\beta_1, \cdots, \beta_F)). \end{aligned}$$

H is a polynomial in several variables, hence analytic. In order to establish the desired result, it suffices to show that H is non-trivial. This requires a "generic" example, that works for any I, J, F. This can be constructed as follows. For any given I, J, F with $2 \leq I \leq F$ and $2 \leq J \leq F$, $IJ = F$, define the generators $\alpha_f = e^{\sqrt{-1}\frac{2\pi}{F}J(f-1)}$, and $\beta_f = e^{\sqrt{-1}\frac{2\pi}{F}(f-1)}$ for $f = 1, \cdots, F$. It can be verified that, with this choice of generators for $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \odot \mathbf{B}$ is itself a Vandermonde matrix with generators $(1, e^{\sqrt{-1}\frac{2\pi}{F}}, \cdots, e^{\sqrt{-1}\frac{2\pi}{F}(F-1)})$, and therefore full rank. This shows that $H(\alpha_1, \cdots, \alpha_F, \beta_1, \cdots, \beta_F)$ is a non-trivial polynomial in $\mathbb{C}^{2F}$. Invoking the analytic function Lemma 3.2, $H(\alpha_1, \cdots, \alpha_F, \beta_1, \cdots, \beta_F)$ is non-zero almost everywhere, except for a measure zero subset of $\mathbb{C}^{2F}$. □

Remark 3.3 *An alternative proof of Theorem 3.3 can be constructed by using the theory of Lagrange interpolation in several variables [12], [40], [46]. The advantage of such an approach is that it affords geometric insight which facilitates the construction of full-rank examples and counter-examples. The disadvantage is that the proof requires a long and delicate argument.*

Proof of Corollary 1: It is again sufficient to consider the case $IJ = F$. The generic example provided for a pair of Vandermonde matrices can be used here also to show that the determinant of the square Khatri-Rao product of two matrices of appropriate dimensions (but otherwise arbitrary) is a non-trivial polynomial in $(I+J)F$ complex variables, therefore the analytic function Lemma 3.2 applies. □

We will need the following preparatory results to prove Theorem 3.5.

Proposition 4 *Given N Vandermonde matrices $\mathbf{A}_n \in \mathbb{C}^{I_n \times F}$ for $n = 1, \cdots, N \geq 2$.*

$$r_{\mathbf{A}_1 \odot \cdots \odot \mathbf{A}_N} = k_{\mathbf{A}_1 \odot \cdots \odot \mathbf{A}_N} = \min(\prod_{n=1}^{N} I_n,\ F), \quad P_{\mathcal{L}}(\mathbb{C}^{NF}) - a.s., \tag{3.26}$$

where $P_{\mathcal{L}}(\mathbb{C}^{NF})$ is the distribution used to draw the NF complex generators for $\mathbf{A}_n$, $n = 1, \cdots, N$, assumed continuous with respect to the Lebesgue measure in $\mathbb{C}^{NF}$.

Proof:

The general case can be reduced to the $\prod_{n=1}^{N} I_n = F$ case. When $\prod_{n=1}^{N} I_n = F$, the full rank and full k-rank of $(\mathbf{A}_1 \odot \cdots \odot \mathbf{A}_N)$ is equivalent to its determinant being nonzero. Define

$$\begin{aligned} &H(\alpha_{1,1}, \cdots, \alpha_{1,F}, \cdots, \alpha_{N,1}, \cdots, \alpha_{N,F}) \\ &= det(\mathbf{A}_1(\alpha_{1,1}, \cdots, \alpha_{1,F}) \odot \cdots \odot \mathbf{A}_N(\alpha_{N,1}, \cdots, \alpha_{N,F})). \end{aligned}$$

where $\alpha_{n,f}$ is the f-th generator of $\mathbf{A}_n$, $n = 1, \cdots, N$, $f = 1, \cdots, F$. H is a polynomial in NF variables, hence analytic in $\mathbb{C}^{NF}$. It therefore suffices to show that H is non-trivial. The following generic example works for any F and $I_n \geq 2$, $n = 1, \cdots, N$, showing that H is non-trivial:

$$\alpha_{n,f} = e^{(\sqrt{-1})\frac{2\pi}{F}(\prod_{k=1}^{n-1} I_k)(f-1)}$$

for $n = 1, \cdots, N$, $f = 1, \cdots, F$. It can be verified that, with this choice of generators for $\mathbf{A}_n$, $n = 1, \cdots, N$, $\mathbf{A}_1 \odot \cdots \odot \mathbf{A}_N$ is a Vandermonde matrix with generators $(1, e^{(\sqrt{-1})\frac{2\pi}{F}}, \cdots, e^{(\sqrt{-1})\frac{2\pi}{F}(F-1)})$, therefore full rank. □

Proposition 5 [11] *Given a sum of F N-D exponentials*

$$x_{i_1,\cdots,i_N} = \sum_{f=1}^{F} c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1}, \tag{3.27}$$

for $i_n = 1,\cdots,I_n \geq 4$, $n = 1,\cdots,N$, the parameter $(N+1)$-tuples $(a_{f,1},\cdots,a_{f,N},c_f)$, $f = 1,\cdots,F$, are $P_{\mathcal{L}}(\mathbb{C}^{NF})$-a.s. unique, where $P_{\mathcal{L}}(\mathbb{C}^{NF})$ is the distribution used to draw the NF complex exponential parameters $(a_{f,1},\cdots,a_{f,N})$, for $f = 1,\cdots,F$, assumed continuous with respect to the Lebesgue measure in $\mathbb{C}^{NF}$, provided that there exist $2N$ integers, $I_{n,j}$ for $n = 1,\cdots,N$, $j = 1,2$ such that

$$I_1 - I_{1,1} - I_{1,2} + \min\left(\prod_{n=1}^{N} I_{n,1}, F\right) + \min\left(\prod_{n=1}^{N} I_{n,2}, F\right) \geq 2F \tag{3.28}$$

subject to

$$\begin{cases} I_{1,1} + I_{1,2} \leq I_1, \\ I_{n,1} + I_{n,2} = I_n + 1, \quad n = 2,\cdots,N, \\ I_{n,j} \geq 2, \quad \forall n,\ j. \end{cases} \tag{3.29}$$

Proof:

We first extend the given N-way array to a $(2N+1)$-way array with typical element

$$\begin{aligned} &\widehat{x}_{i_{1,1},i_{1,2},i_{1,3},i_{2,1},i_{2,2},\cdots,i_{N,1},i_{N,2}} \\ &:= x_{i_{1,1}+i_{1,2}+i_{1,3}-2,\ i_{2,1}+i_{2,2}-1,\ \cdots,\ i_{N,1}+i_{N,2}-1} \\ &= \sum_{f=1}^{F} c_f \left(a_{f,1}^{i_{1,1}+i_{1,2}+i_{1,3}-3} \prod_{n=2}^{N} a_{f,n}^{i_{n,1}+i_{n,2}-2} \right) \\ &= \sum_{f=1}^{F} c_f \left(a_{f,1}^{i_{1,3}-1} \prod_{n=1}^{N} a_{f,n}^{i_{n,1}+i_{n,2}-2} \right) \\ &= \sum_{f=1}^{F} c_f \left(a_{f,1}^{i_{1,3}-1} \prod_{n=1}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=1}^{N} a_{f,n}^{i_{n,2}-1} \right), \end{aligned} \tag{3.30}$$

where $i_{n,j} = 1,\cdots,I_{n,j} \geq 2$, $i_{1,3} = 1,\cdots,I_{1,3} \geq 2$, $n = 1,\cdots,N$, $j = 1,2$. Such extension is always possible under our working assumption that $I_n \geq 4,\ \forall n$. We also

[11] The Proposition holds true for any permutation of $\{I_n\}_{n=1}^{N}$.

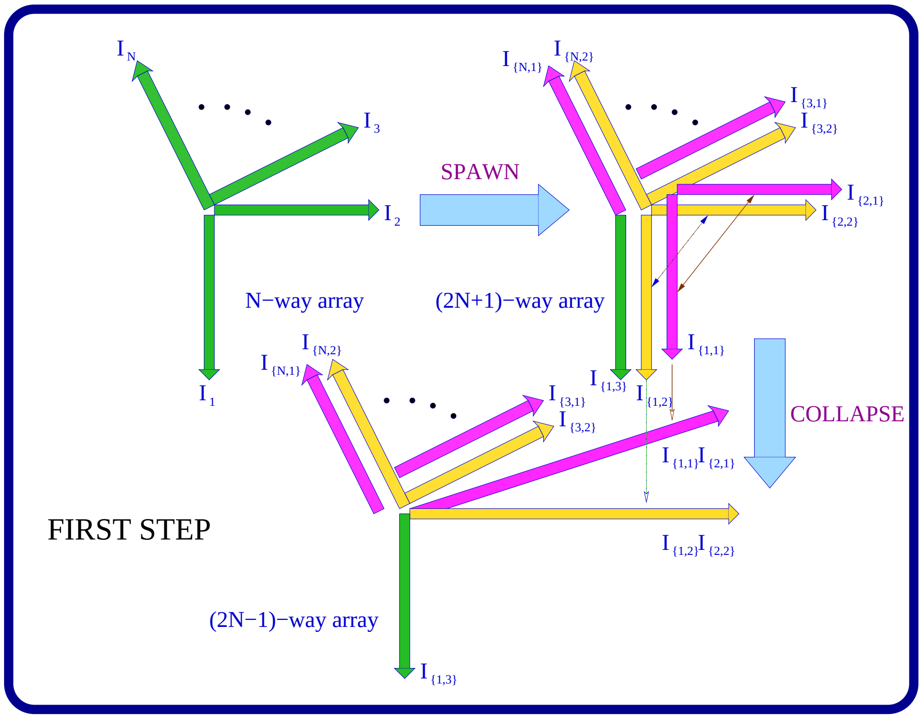

Figure 3.1: The first step in the proof of the N-dimensional case

need the following constraints to avoid indexing beyond the available data sample.

$$\begin{cases} I_{1,1} + I_{1,2} + I_{1,3} = I_1 + 2, \\ I_{n,1} + I_{n,2} = I_n + 1, \quad n = 2, \cdots, N. \end{cases} \tag{3.31}$$

Define matrices

$$\mathbf{A}_{n,j} = (a_{f,n}^{i_{n,j}-1}) \in \mathbb{C}^{I_{n,j} \times F}, \qquad \mathbf{A}_{1,3} = (a_{f,1}^{i_{1,3}-1}) \in \mathbb{C}^{I_{1,3} \times F}, \tag{3.32}$$

for $n = 1, \cdots, N$, $j = 1, 2$. Next, we compress the $(2N+1)$-way $\widehat{x}$ array into a three-way array $\bar{x}$. We do this in two steps for clarity. The first step is to nest $\widehat{x}$ into a $(2N-1)$-way array $\widehat{x}^{(1)}$. This process is illustrated in Figure 3.1.

$$\begin{aligned}
&\widehat{x}^{(1)}_{i_{1,3},k^{(1)},l^{(1)},i_{3,1},i_{3,2},\cdots,i_{N,1},i_{N,2}} \\
&= \widehat{x}_{\lceil \frac{k^{(1)}}{I_{2,1}} \rceil,\ \lceil \frac{l^{(1)}}{I_{2,2}} \rceil,\ i_{1,3},\ k^{(1)}-(\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1)I_{2,1},\ l^{(1)}-(\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1)I_{2,2},} \\
&\qquad {}_{i_{3,1},i_{3,2},\cdots,i_{N,1},i_{N,2}} \\
&= \sum_{f=1}^{F} c_f \ (\ a_{f,1}^{\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1} a_{f,1}^{\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1} a_{f,1}^{i_{1,3}-1} a_{f,2}^{k^{(1)}-(\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1)I_{2,1}-1} \\
&\quad \times a_{f,2}^{l^{(1)}-(\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1)I_{2,2}-1} \prod_{n=3}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=3}^{N} a_{f,n}^{i_{n,2}-1}\ ) \qquad (3.33) \\
&= \sum_{f=1}^{F} c_f \ (\ a_{f,1}^{i_{1,3}-1} a_{f,1}^{\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1} a_{f,2}^{k^{(1)}-(\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1)I_{2,1}-1} \\
&\quad \times a_{f,1}^{\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1} a_{f,2}^{l^{(1)}-(\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1)I_{2,2}-1} \prod_{n=3}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=3}^{N} a_{f,n}^{i_{n,2}-1}\ ) \\
&= \sum_{f=1}^{F} c_f a_{f,1}^{i_{1,3}-1} d^{(1)}_{k^{(1)},f} e^{(1)}_{l^{(1)},f} \prod_{n=3}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=3}^{N} a_{f,n}^{i_{n,2}-1}
\end{aligned}$$

for $k^{(1)} = 1, \cdots, I_{1,1}I_{2,1}$, $l^{(1)} = 1, \cdots, I_{1,2}I_{2,2}$, with $d^{(1)}_{k^{(1)},f}$ and $e^{(1)}_{l^{(1)},f}$ given by

$$\begin{aligned}
d^{(1)}_{k^{(1)},f} &:= a_{f,1}^{\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1} a_{f,2}^{k^{(1)}-(\lceil \frac{k^{(1)}}{I_{2,1}} \rceil -1)I_{2,1}-1}, \\
e^{(1)}_{l^{(1)},f} &:= a_{f,1}^{\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1} a_{f,2}^{l^{(1)}-(\lceil \frac{l^{(1)}}{I_{2,2}} \rceil -1)I_{2,2}-1}.
\end{aligned} \qquad (3.34)$$

Define matrices

$$\begin{aligned}
\mathbf{D}^{(1)} &= (d^{(1)}_{k^{(1)},f}) \in \mathbb{C}^{I_{1,1}I_{2,1}\times F}, \\
\mathbf{E}^{(1)} &= (e^{(1)}_{l^{(1)},f}) \in \mathbb{C}^{I_{1,2}I_{2,2}\times F},
\end{aligned} \qquad (3.35)$$

and note that

$$\mathbf{D}^{(1)} = \mathbf{A}_{1,1} \odot \mathbf{A}_{2,1}, \qquad \mathbf{E}^{(1)} = \mathbf{A}_{1,2} \odot \mathbf{A}_{2,2}. \qquad (3.36)$$

The next step is to show that, starting from $m = 1$, we can recursively nest the $(2(N-m)+1)$-way array $\widehat{x}^{(m)}$ into a $(2(N-(m+1))+1)$-way array $\widehat{x}^{(m+1)}$,

$m = 1, \cdots, (N-2)$. This step is illustrated in Figure 3.2.

$$\begin{aligned}
&\widehat{x}^{(m+1)}_{i_{1,3},\ k^{(m+1)},\ l^{(m+1)},\ i_{(m+3),1},\ i_{(m+3),2},\cdots,\ i_{N,1},\ i_{N,2}} \\
&= \widehat{x}^{(m)}_{i_{1,3},\ \lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil,\ \lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil,\ k^{(m+1)} - (\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil - 1) I_{(m+2),1},} \\
&\qquad {}_{l^{(m+1)} - (\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil - 1) I_{(m+2),2},\ i_{(m+3),1},\ i_{(m+3),2},\cdots, i_{N,1}, i_{N,2}} \\
&= \sum_{f=1}^{F} c_f \ (\ a_{f,1}^{i_{1,3}-1} \ d^{(m)}_{\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil, f} \ e^{(m)}_{\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil, f} \\
&\quad \times a_{f,(m+2)}^{k^{(m+1)} - (\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil - 1) I_{(m+2),1} - 1} \\
&\quad \times a_{f,(m+2)}^{l^{(m+1)} - (\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil - 1) I_{(m+2),2} - 1} \\
&\quad \times \prod_{n=(m+3)}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=(m+3)}^{N} a_{f,n}^{i_{n,2}-1} \ ) \qquad (3.37) \\
&= \sum_{f=1}^{F} c_f (a_{f,1}^{i_{1,3}-1} d^{(m)}_{\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil, f} a_{f,(m+2)}^{k^{(m+1)} - (\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil - 1) I_{(m+2),1} - 1} \\
&\quad \times e^{(m)}_{\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil, f} \ a_{f,(m+2)}^{l^{(m+1)} - (\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil - 1) I_{(m+2),2} - 1} \\
&\quad \times \prod_{n=(m+3)}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=(m+3)}^{N} a_{f,n}^{i_{n,2}-1} \ ) \\
&= \sum_{f=1}^{F} c_f \ (\ a_{f,1}^{i_{1,3}-1} \ d^{(m+1)}_{k^{(m+1)}, f} \ e^{(m+1)}_{l^{(m+1)}, f} \\
&\quad \times \prod_{n=(m+3)}^{N} a_{f,n}^{i_{n,1}-1} \prod_{n=(m+3)}^{N} a_{f,n}^{i_{n,2}-1} \ )
\end{aligned}$$

for

$$k^{(m+1)} = 1, \cdots, \prod_{n=1}^{(m+2)} I_{n,1},$$

$$l^{(m+1)} = 1, \cdots, \prod_{n=1}^{(m+2)} I_{n,2},$$

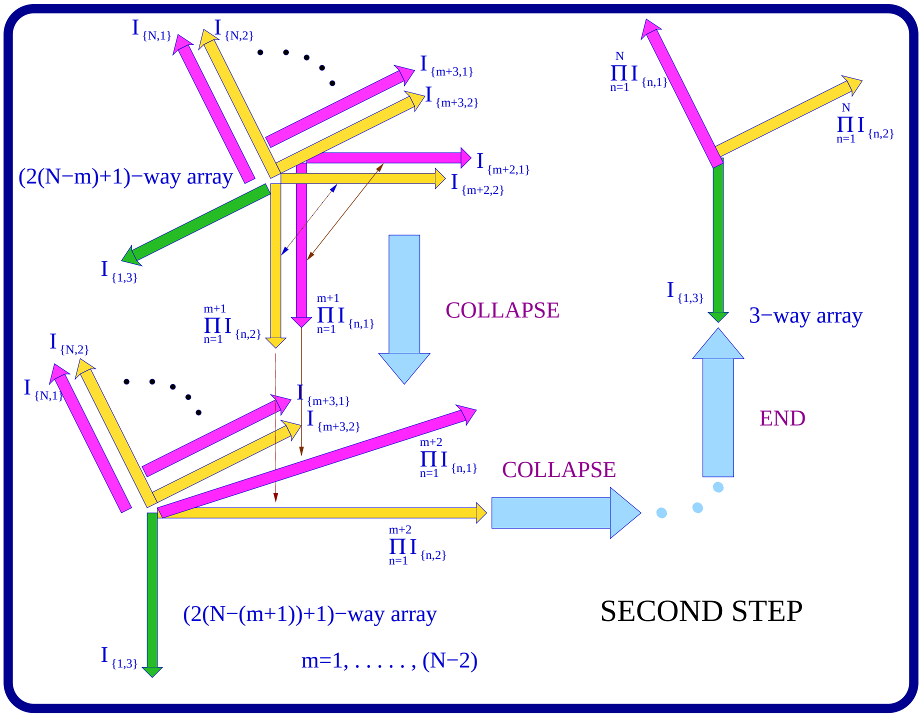

Figure 3.2: The second step in the proof of the N-dimensional case

with $d^{(m+1)}_{k^{(m+1)},f}$ and $e^{(m+1)}_{l^{(m+1)},f}$ given by

$$
\begin{aligned}
d^{(m+1)}_{k^{(m+1)},f} &:= d^{(m)}_{\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil, f} \; a_{f,(m+2)}^{k^{(m+1)} - (\lceil \frac{k^{(m+1)}}{I_{(m+2),1}} \rceil - 1) I_{(m+2),1} - 1}, \\
e^{(m+1)}_{l^{(m+1)},f} &:= e^{(m)}_{\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil, f} \; a_{f,(m+2)}^{l^{(m+1)} - (\lceil \frac{l^{(m+1)}}{I_{(m+2),2}} \rceil - 1) I_{(m+2),2} - 1}.
\end{aligned} \tag{3.38}
$$

Define matrices

$$
\begin{aligned}
\mathbf{D}^{(m+1)} &= (d^{(m+1)}_{k^{(m+1)},f}) \in \mathbb{C}^{\prod_{n=1}^{(m+2)} I_{n,1} \times F}, \\
\mathbf{E}^{(m+1)} &= (e^{(m+1)}_{l^{(m+1)},\ f}) \in \mathbb{C}^{\prod_{n=1}^{(m+2)} I_{n,2} \times F}.
\end{aligned} \tag{3.39}
$$

$\mathbf{D}^{(m+1)}$ and $\mathbf{E}^{(m+1)}$ can be written as

$$
\begin{aligned}
\mathbf{D}^{(m+1)} &= \mathbf{D}^{(m)} \odot \mathbf{A}_{(m+2),1} \\
&= \mathbf{A}_{1,1} \odot \cdots \odot \mathbf{A}_{(m+1),1} \odot \mathbf{A}_{(m+2),1}, \\
\mathbf{E}^{(m+1)} &= \mathbf{E}^{(m)} \odot \mathbf{A}_{(m+2),2} \\
&= \mathbf{A}_{1,2} \odot \cdots \odot \mathbf{A}_{(m+1),2} \odot \mathbf{A}_{(m+2),2}.
\end{aligned} \tag{3.40}
$$

The recursion finally terminates at $\widehat{x}^{(N-1)}$, which we are going to denote by $\bar{x}$,

$$\begin{aligned}\bar{x}_{i_{1,3},k^{(N-1)},l^{(N-1)}} &:= \widehat{x}^{(N-1)}_{i_{1,3},k^{(N-1)},l^{(N-1)}} \\ &= \sum_{f=1}^{F} c_f \ (\ a_{f,1}^{i_{1,3}-1} \ d^{(N-1)}_{k^{(N-1)},f} \ e^{(N-1)}_{l^{(N-1)},f} \ ),\end{aligned} \tag{3.41}$$

for $k^{(N-1)} = 1, \cdots, \prod_{n=1}^{N} I_{n,1}$, $l^{(N-1)} = 1, \cdots, \prod_{n=1}^{N} I_{n,2}$. We have :

$$\begin{aligned}\mathbf{D}^{(N-1)} &= (d^{(N-1)}_{k^{(N-1)},f}) = \mathbf{A}_{1,1} \odot \cdots \odot \mathbf{A}_{N,1}, \\ \mathbf{E}^{(N-1)} &= (e^{(N-1)}_{l^{(N-1)},f}) = \mathbf{A}_{1,2} \odot \cdots \odot \mathbf{A}_{N,2}.\end{aligned} \tag{3.42}$$

Since $\mathbf{A}_{1,3}$ is Vandermonde, Theorem 3.2 can be invoked to claim uniqueness, provided that

$$I_{1,3} + k_{\mathbf{D}^{(N-1)}} + k_{\mathbf{E}^{(N-1)}} \geq 2F + 3 - 1. \tag{3.43}$$

Similar to the 2-D case, each product form

$$c_f \ (\ a_{f,1}^{i_{1,3}-1} \ d^{(N-2)}_{k^{(N-2)},f} \ e^{(N-2)}_{l^{(N-2)},f} \ )$$

can be put in one-to-one correspondence with $c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1}$, $f = 1, \cdots, F$. Therefore, uniqueness of the F rank-one 3-D factors $c_f \ (\ a_{f,1}^{i_{1,3}-1} \ d^{(N-2)}_{k^{(N-2)},f} \ e^{(N-2)}_{l^{(N-2)},f} \ )$ is equivalent to uniqueness of the F rank-one N-D factors $c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1}$. It follows that the rank-one factors $c_f \prod_{n=1}^{N} a_{f,n}^{i_n-1}$ and hence the parameter $(N+1)$-tuples $(a_{f,1}, \cdots, a_{f,N}, c_f)$, $f = 1, \cdots, F$, are unique provided that (3.43) holds true. Invoking Proposition 4, almost sure uniqueness holds provided there exist $2N+1$ integers, $I_{1,3} \geq 2$ and $I_{n,j} \geq 2$, for $n = 1, \cdots, N$ and $j = 1, 2$ such that

$$I_{1,3} + \min(\prod_{n=1}^{N} I_{n,1}, F) + \min(\prod_{n=1}^{N} I_{n,2}, F) \geq 2F + 2, \tag{3.44}$$

subject to

$$I_{1,1} + I_{1,2} + I_{1,3} = I_1 + 2, \quad I_{n,1} + I_{n,2} = I_n + 1, \quad n = 2, \cdots, N,$$

or, equivalently

$$I_1 - I_{1,1} - I_{1,2} + \min(\prod_{n=1}^{N} I_{n,1}, F) + \min(\prod_{n=1}^{N} I_{n,2}, F) \geq 2F,$$

subject to

$$\begin{cases} I_{1,1} + I_{1,2} \leq I_1, \\ I_{n,1} + I_{n,2} = I_n + 1, \quad n = 2, \cdots, N, \\ I_{n,j} \geq 2, \quad \forall n, j. \end{cases}$$

and the proof is complete. □

Proof of Theorem 3.5: If I_1 is even, pick $I_{1,1} = I_{1,2} = \frac{I_1}{2}$, otherwise pick $I_{1,1} = \frac{I_1-1}{2}$ and $I_{1,2} = \frac{I_1+1}{2}$ (thereby satisfying condition (3.29)).

If I_n is even, pick $I_{n,1} = \frac{I_n}{2}$, $I_{n,2} = \frac{I_n+2}{2}$, otherwise, let $I_{n,1} = I_{n,2} = \frac{I_n+1}{2}$ (hence satisfying (3.29)), for all $n = 2, \cdots, N$.

Once we pick all $2N$ integers following the above rules, condition (3.25) assures that inequality (3.28) holds. Invoking Proposition 5 completes the proof. □

Proof of Proposition 2: It suffices to show that, when $IJ = F$,

$$\begin{aligned} & H(\alpha_1, \alpha_2, \cdots, \alpha_F, \beta_1, \beta_2, \cdots, \beta_{F-1}) \\ & = det(\mathbf{A}(\alpha_1, \alpha_2, \cdots, \alpha_F) \odot \mathbf{B}(\beta_1, \beta_1, \cdots, \beta_{F-1})), \end{aligned}$$

is non-trivial analytic function in $\mathbb{C}^{2F-1}$, where both $\mathbf{A}$ and $\mathbf{B}$ are Vandermonde matrices defined by (3.2). For any given I, J, F with $2 \leq I \leq F$, $2 \leq J \leq F$, and $IJ = F$, define the generators $\alpha_1 = 0$, $\alpha_f = e^{\sqrt{-1}\frac{2\pi}{F-1}J(f-2)}$, for $f = 2, \cdots, F$, and $\beta_f = e^{\sqrt{-1}\frac{2\pi}{F-1}(f-1)}$, for $f = 1, \cdots, F-1$. It can be verified that, with this choice of generators for $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A} \odot \mathbf{B} =$$

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & e^{\sqrt{-1}\frac{2\pi}{F-1}} & \cdots & e^{\sqrt{-1}\frac{2\pi}{F-1}(F-2)} \\ 1 & 1 & e^{\sqrt{-1}\frac{2\pi}{F-1}2} & \cdots & e^{\sqrt{-1}\frac{2\pi}{F-1}(F-3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & e^{\sqrt{-1}\frac{2\pi}{F-1}(J-1)} & \cdots & e^{\sqrt{-1}\frac{2\pi}{F-1}(F-J)} \\ 0 & 1 & e^{\sqrt{-1}\frac{2\pi}{F-1}J} & \cdots & e^{\sqrt{-1}\frac{2\pi}{F-1}(F-J-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & e^{\sqrt{-1}\frac{2\pi}{F-1}(F-2)} & \cdots & e^{\sqrt{-1}\frac{2\pi}{F-1}} \\ 0 & 1 & 1 & \cdots & 1 \end{bmatrix},$$

which is full rank, hence H is nontrivial in $\mathbb{C}^{2F-1}$. □

Proof of Proposition 3: Assume $I \le J$, without loss of generality. Spawn two dimensions out of J: $J_1 = J - 1$, $J_2 = 2$. Collapse I and $J - 1$. We are now in $2 \times I(J-1) \times K$ 3-D space, with the dimension corresponding to $I(J-1)$ being full k-rank almost surely. Theorem 3.2 then yields $2F + 2 \le \min(IJ - I, F) + k_{\mathbf{S}} + 2$. Since $k_{\mathbf{S}} = F$ has been assumed, the desired result follows. □

Part II

Applications

Chapter 4

Blind Identification of Out-of-cell Users in DS-CDMA

As mentioned earlier in the introduction of this thesis, out-of-cell interference accounts for a large percentage of the interference relative to the interference coming from within the cell, yet typically ignored, or treated as noise. Unlike the case of in-cell interference, out-of-cell interference cannot be mitigated by power control, simply because the BS does not have the authority to exercise power control over out-of-cell users. For a power-controlled in-cell population, near-far effects may be chiefly due to out-of-cell interference. Unfortunately, out-of-cell detection is compounded by the fact that it has to be blind, since the BS has no control and usually no prior information on out-of-cell users. This places limitations on the number and nature of out-of-cell transmissions that can be identified.

The literature on out-of-cell blind identification is scarce. Assuming that (i) the codes of the in-cell users are known, (ii) the total number of (in-cell plus out-of-cell) users is less than the spreading gain and the combined spreading code matrix is full column rank, and (iii) given the correlation matrix of the vector of chip samples taken over a symbol interval, it is possible to cancel out the effect of out-of-cell users [56], then adopt linear or nonlinear solutions for in-cell detection. This approach is appealing, but it has two drawbacks. First, it can be unrealistic to assume that the *total* number of users is less than the spreading gain. This is especially so in loaded systems and urban areas. Second, in practice one uses sample estimates of the correlation matrix. This yields cancellation errors for finite samples, even in the noiseless case.

Recently, a novel code-blind identification approach has been proposed, exploiting uniqueness of low-rank decomposition of three-way arrays [54]. This requires the use of a BS antenna array, but in return allows the identification of both in-cell and out-of-cell users without requiring knowledge of the code or steering vector of *any* user. More users than spreading and antenna elements can be supported. There are two drawbacks to this approach. First, a direct algebraic solution is generally not possible, thus iterative estimation techniques must be employed. Although these iterative methods generally work very well, they are computationally intensive. Second, in-cell code information, which may well be available, is not directly exploited (except numerically, by constraining certain parameters during the iterations). In this chapter, we develop an algebraic solution that exploits the fact that the codes of the in-cell users are known. In this scenario, we show that in addition to algebraic solution, better identifiability is possible. Our approach yields the best known identifiability result for three-dimensional low-rank decomposition when one of the three component matrices is partially known, albeit non-invertible.

Note that the group-blind multiuser detection approach of [56] can be easily extended to handle multiple BS antennae, but this requires that the array steering vectors, in addition to the spreading codes[1] of all the in-cell users are known. Esti-

[1] In the literature, it is common to use the term "(spreading) codes" for the transmit codes, and "signatures" for the effective receive codes. For brevity and to avoid confusion with spatial signatures, we adopt the term spreading codes throughout, with the understanding that, in the

mating steering vectors is more difficult than estimating codes, partly because they are generally unstructured, but also due to mobility-induced fast-fading. Note that the approach developed herein (see also [54]) does not assume any parameterization of the manifold vectors.

For clarity of exposition, we will begin our analysis by assuming that both in-cell and out-of-cell user transmissions are synchronized at the BS. In practice, this can be approximately true in synchronous CDMA systems, like CDMA2000[2]. Quasi-synchronism (i.e., timing offsets in the order of a few chips) can be handled by dropping a short chip prefix at the receiver. We shall refer to both cases as *synchronous CDMA* for brevity. Synchronization is usually achieved via pilot tones emitted from the BS, or a GPS-derived timing reference for synchronous networks involving multiple cells. Out-of-cell transmissions will typically not be synchronized with in-cell transmissions. Notable exceptions include synchronous micro-cellular networks for "hotspot" coverage, and calls undergoing hand-off at cell boundaries (hence approximately equidistant from the two base stations). As we will see, when delay spread is small relative to the symbol duration, this can be handled by treating each out-of-cell user as two virtual users. Hence our analysis generalizes to the interesting case of a quasi-synchronous in-cell population plus asynchronous out-of-cell interference, as in Wideband CDMA (WCDMA). We shall refer to this situation as *asynchronous CDMA*.

The rest of the chapter is organized as follows. The main ideas and concepts are exposed in Section 4.1.1, which treats the idealized case of a synchronous DS-CDMA uplink subject to flat fading. This is then extended to frequency-selective multipath and quasi-synchronous transmissions in Section 4.2, which also discusses a

presence of ICI/ISI, the term codes means the receive codes.

[2]CDMA2000 uses UTC (universal coordinated time) system time reference, derived from GPS. Mobile stations use the same system time, offset by the propagation delay from the base station to the mobile station

suitable *admission protocol* that avoids explicit code estimation for the in-cell users. Note that in the presence of strong out-of-cell interference and frequency selectivity, estimating the codes of the in-cell users is a difficult task in itself. Section 4.3 discusses issues related to our choice of a pertinent symbol-independent asymptotic Cramér-Rao Bound (CRB) to benchmark performance of steering vector and spreading code estimation. Associated derivations are deferred to the Appendix. Section 4.4 provides analytical and simulated performance comparisons, and Section 4.5 summarizes our conclusions.

4.1 Blind Identification Of Out-Of-Cell Users

4.1.1 Data Model

Consider a DS-CDMA uplink with M users (in-cell plus out-of-cell), normalized chip waveform ψ of duration T_c and spreading gain P (chips per symbol). The m-th user is assigned a binary chip sequence $(c_m(1), \cdots, c_m(P))$. The resulting signature waveform for the m-th user is

$$\phi_m(t) = \sum_{i=1}^{P} c_m(i)\psi(t - iT_c), 0 \leq t \leq T_s,$$

where $T_s = PT_c$ is the symbol duration. All spreading codes are assumed short (symbol-periodic).

The baseband-equivalent signal received at the BS for a burst of L transmitted symbols can be written as

$$x(t) = \sum_{m=1}^{M} \sum_{l=1}^{L} \alpha_m \sqrt{E_m} s_m(l) \phi_m(t - lT_s - \tau_m) + w(t)$$

where M is the total number of active users, α_m is the complex path gain, E_m is the incident power for the m-th user loaded at the transmitter, $s_m(l)$ is the l-th transmitted symbol associated with the m-th user, τ_m is the delay of the m-th user's signal, and $w(\cdot)$ is aditive white Gaussian noise (AWGN). Since in-cell users are synchronized with the BS, the delays τ_m for all in-cell users are taken to be zero. For out-of-cell users, the associated delays can be assumed to lie in $[0, T_s]$, without loss of generality.

If K receive antennas are employed at the BS, the baseband signal at the output of the chip-matched filter of the k-th antenna for the p-th chip in the n-th symbol interval can be written as

$$\begin{aligned} x_{k,n,p} &=< x(t), \beta_k \psi(t - nT_s - pT_c) > \\ &= \sum_{m=1}^{M_{in}} \alpha_{k,m}\beta_k\sqrt{E_m}s_m(n)c_m(p) + \sum_{m=M_{in}+1}^{M}\sum_{l=1}^{L} \alpha_{k,m}\beta_k\sqrt{E_m}s_m(l)\nu_{pm}(n,l) + w(k,n,p) \\ &= \sum_{m=1}^{M_{in}} \alpha_{k,m}\beta_k\sqrt{E_m}s_m(n)c_m(p) \\ &+ \sum_{m=M_{in}+1}^{M} \alpha_{k,m}\beta_k\sqrt{E_m}[s_m(n)\nu_{pm}(n,n) + s_m(n-1)\nu_{pm}(n,n-1)] + w(k,n,p), \end{aligned} \tag{4.1}$$

where M_{in} ($\leq P$) denotes the number of in-cell users and M_{out} the number of out-of-cell users ($M = M_{in} + M_{out}$); β_k is the antenna gain associated with the k-th antenna; $\nu_{pm}(n,l) = \sum_{i=1}^{P} c_m(i) \int_0^{T_c} \psi(t + (n-l)T_s + (p-i)T_c - \tau_m)\psi^H(t)dt$; $w(k,n,p) = \int_0^{T_c} w(t + nT_s + pT_c)\psi^H(t)dt$.

Note that, due to asynchronism, each out-of-cell user is viewed by the BS as two synchronous users, whose symbol sequences are time-shifted versions of one-another. The associated spreading codes are given by $\nu_{pm}(\cdot,\cdot)$.

From (4.1), in a frequency-flat block-fading scenario, the baseband-equivalent chip-rate sampled data model for a synchronous DS-CDMA system with short symbol-periodic spreading codes and K receive antennas at the BS can be written as

$$x_{k,n,p} = \sum_{m=1}^{M} a_m(k)c_m(p)s_m(n) + w_{k,n,p}, \tag{4.2}$$

for $k = 1, \cdots, K,\ n = 1, \cdots, N,\ p = 1, \cdots, P$, where N is the number of symbol snapshots, $x_{k,n,p}$ denotes the baseband output of the k-th antenna element for symbol ("time") n and chip p. $a_m(k)$ is the compound flat fading/antenna gain associated with the response of the k-th antenna to the m-th user.

It is useful to recast this model in matrix form. Let us define P received data matrices $\mathbf{X}_p \in \mathbb{C}^{K \times N}$ with (k, n)-element given by $x_{k,n,p}$, and AWGN matrices $\mathbf{W}_p \in \mathbb{C}^{K \times N}$ with (k, n)-element given by $w_{k,n,p}$. Let us also define the steering matrix $\mathbf{A} \in \mathbb{C}^{K \times M}$ with m-th column $[\ a_m(1)\ \ldots\ a_m(K)\ ]^T$, the spreading code matrix $\mathbf{C} \in \mathbb{C}^{P \times M}$ with m-th column $[\ c_m(1)\ \ldots\ c_m(P)\ ]^T$, and the signal matrix $\mathbf{S} \in \mathbb{C}^{N \times M}$ with m-th column $[\ s_m(1)\ \ldots\ s_m(N)\ ]^T$. Without loss of generality, we assume that the sub-matrices $\mathbf{A}_{in} \in \mathbb{C}^{K \times M_{in}}, \mathbf{C}_{in} \in \mathbb{C}^{P \times M_{in}}, \mathbf{S}_{in} \in \mathbb{C}^{N \times M_{in}}$, consisting of the first M_{in} columns of $\mathbf{A}, \mathbf{C}, \mathbf{S}$, respectively, correspond to the in-cell users; and similarly for $\mathbf{A}_{out}, \mathbf{C}_{out}$, and $\mathbf{S}_{out}$. Thus, we have $\mathbf{A} = [\ \mathbf{A}_{in}\ \mathbf{A}_{out}\ ]$, $\mathbf{C} = [\ \mathbf{C}_{in}\ \mathbf{C}_{out}]$, $\mathbf{S} = [\ \mathbf{S}_{in}\ \mathbf{S}_{out}\ ]$.

$\mathbf{X}_p$ admits the factorization:

$$\begin{aligned} \mathbf{X}_p &= \mathbf{A}\mathbf{D}_p(\mathbf{C})\mathbf{S}^T + \mathbf{W}_p \\ &= \mathbf{A}_{in}\mathbf{D}_p(\mathbf{C}_{in})\mathbf{S}_{in}^T + \mathbf{A}_{out}\mathbf{D}_p(\mathbf{C}_{out})\mathbf{S}_{out}^T + \mathbf{W}_p \\ &= \mathbf{X}_p^{in} + \mathbf{X}_p^{out} + \mathbf{W}_p, \end{aligned} \tag{4.3}$$

for $p = 1, 2, \ldots, P$.

It is also worth mentioning that we can write the above set of matrix equations into more compact form if we introduce the Khatri-Rao product. Stacking the matrices in (4.3), we obtain:

$$\mathbf{X}^{KP\times N} := \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_P \end{bmatrix} = \begin{bmatrix} \mathbf{AD}_1(\mathbf{C}) \\ \mathbf{AD}_2(\mathbf{C}) \\ \vdots \\ \mathbf{AD}_P(\mathbf{C}) \end{bmatrix} \mathbf{S}^T + \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_P \end{bmatrix} = (\mathbf{C} \odot \mathbf{A})\mathbf{S}^T + \mathbf{W}^{KP\times N}$$
$$= (\mathbf{C}_{in} \odot \mathbf{A}_{in})\mathbf{S}_{in}^T + (\mathbf{C}_{out} \odot \mathbf{A}_{out})\mathbf{S}_{out}^T + \mathbf{W}^{KP\times N}. \tag{4.4}$$

Due to the symmetry of the model (4.2), we may also recast (4.4) in the following form

$$\widetilde{\mathbf{X}}^{PN\times K} = (\mathbf{S} \odot \mathbf{C})\mathbf{A}^T + \widetilde{\mathbf{W}}^{PN\times K}, \tag{4.5}$$

where $\widetilde{\mathbf{W}}^{PN\times K}$ is a re-shuffled AWGN matrix (see [54]).

In what follows, we consider detecting the signal matrix $\mathbf{S}$ transmitted from *all* active users given only knowledge of $\mathbf{C}_{in}$ and M. As a byproduct, we will be able to recover the steering matrix $\mathbf{A}$ and the unknown spreading code matrix $\mathbf{C}_{out}$ from the received data $\mathbf{X}$ as well.

4.1.2 Preliminaries

We will make use of the following results in the next subsection to derive our main identifiability result.

Eigenanalysis

Consider two matrices $\mathbf{X}_1 = \mathbf{AD}_1(\mathbf{C})\mathbf{S}^\mathbf{T}$, $\mathbf{X}_2 = \mathbf{AD}_2(\mathbf{C})\mathbf{S}^\mathbf{T}$ where both $\mathbf{A} \in \mathbb{C}^{K\times M}$ and $\mathbf{S} \in \mathbb{C}^{N\times M}$ are full column rank (M), $\mathbf{C} \in \mathbb{C}^{2\times M}$ contains no zero entry, and all elements on the diagonal of $\mathbf{D} := \mathbf{D}_2(\mathbf{C})\mathbf{D}_\mathbf{1}^{-\mathbf{1}}(\mathbf{C})$ are distinct. Consider the singular value decomposition (SVD) of the stacked data matrix

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{AD} \end{bmatrix} \mathbf{D}_1(\mathbf{C})\mathbf{S}^{\mathbf{T}} = \mathbf{U\Sigma V^H}.$$

The linear space spanned by the columns of $\mathbf{U}$ is same as the space spanned by the columns of $\begin{bmatrix} \mathbf{A} \\ \mathbf{AD} \end{bmatrix}$ since $\mathbf{SD}_1(\mathbf{C})$ has full column rank; hence there exists a nonsingular matrix $\mathbf{P}$ such that

$$\mathbf{UP} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} \mathbf{P} = \begin{bmatrix} \mathbf{A} \\ \mathbf{AD} \end{bmatrix}.$$

Next, construct the auto- and cross-product matrices

$$\begin{aligned} \mathbf{R}_0 &= \mathbf{U}_1^H \mathbf{U}_1 = \mathbf{P}^{-H}\mathbf{A}^H\mathbf{A}\mathbf{P}^{-1} := \mathbf{Q}\mathbf{P}^{-1}, \\ \mathbf{R}_1 &= \mathbf{U}_1^H \mathbf{U}_2 = \mathbf{P}^{-H}\mathbf{A}^H\mathbf{AD}\mathbf{P}^{-1} := \mathbf{QD}\mathbf{P}^{-1}. \end{aligned} \tag{4.6}$$

Note that all matrices in (4.6) are square and full rank. Solving the first equation in (4.6) for $\mathbf{Q}$, then substituting the result into the second, it follows that

$$(\mathbf{R}_0^{-1}\mathbf{R}_1)\mathbf{P} = \mathbf{PD}, \tag{4.7}$$

which is a standard eigenvalue problem with distinct eigenvalues. $\mathbf{P}$ can therefore be determined up to permutation and scaling of columns based on the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. After that, $\mathbf{A}$ can be obtained as $\mathbf{A} = \mathbf{U}_1\mathbf{P}$, $\mathbf{CD}_1^{-1}(\mathbf{C})$ can be retrieved with all ones in the first row, and the entire second row taken from the diagonal of $\mathbf{D}$, and finally $\mathbf{SD}_1(\mathbf{C})$ can be recovered as $\mathbf{SD}_1(\mathbf{C}) = (\mathbf{A}^\dagger\mathbf{X}_1)^{\mathbf{T}}$, all under the same permutation and scaling of columns, which carries over from the solution of the eigenvalue problem in (4.7).

Repeated values along the diagonal of $\mathbf{D}_2(\mathbf{C})\mathbf{D}_1^{-1}(\mathbf{C})$ give rise to eigenvalues of multiplicity higher than one. In this case, the span of eigenvectors corresponding to each distinct eigenvalue can still be uniquely determined. This will be important when we discuss the case of asynchronous out-of-cell users later in Section 4.2.

In general, given matrices $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{C})\mathbf{S}^{\mathbf{T}}$ for $p = 1, \cdots, P \geq 2$, $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{S}$ can be found up to permutation and scaling of columns provided that both $\mathbf{A}$ and $\mathbf{S}$ are full column rank, and $k_{\mathbf{C}} \geq 2$.

Since $k_{\mathbf{C}} \geq 2$, we know that the spreading code matrix $\mathbf{C}$ does not contain any zero columns. Note that $k_{\mathbf{C}} \geq 2$ does not necessarily imply that there always exists a sub-matrix of $\mathbf{C}$ which comprises of two rows of $\mathbf{C}$ such that the k-rank of this sub-matrix is two. For instance, consider

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

It can be seen that $r_{\mathbf{C}} = k_{\mathbf{C}} = 3$ whereas none of 2×3 sub-matrices of $\mathbf{C}$ has k-rank great than 1. On the other hand, we have the following:

Claim: Given $\mathbf{C} \in \mathbb{C}^{P \times M}$ with $k_{\mathbf{C}} \geq 2$, there always exists a $2 \times P$ matrix $\mathbf{G}$ such that the k-rank of $\mathbf{GC}$ is two.

For a proof, note that the objective can been easily shown equivalent to proving that there exists a $2 \times P$ matrix $\mathbf{G}$ such that the determinants of all 2×2 sub-matrices of $\mathbf{GC}$ are not zero. $\mathbf{G}$ is determined by its $2P$ complex entries. The determinant of each 2×2 sub-matrix of $\mathbf{GC}$ is a polynomial in those $2P$ variables, and hence analytic. Since $k_{\mathbf{C}} \geq 2$, for each specific 2×2 sub-matrix of $\mathbf{GC}$, for instance, the sub-matrix comprising of the first two columns of $\mathbf{GC}$, it is not hard to show that there always exists a $\mathbf{G}_0$ such that the determinant of the corresponding sub-matrix of $\mathbf{G}_0\mathbf{C}$ is not zero. Invoking Lemma 2 in [25], we conclude that the set of $\mathbf{G}$'s which yield zero determinant for any specific sub-matrix of $\mathbf{GC}$ constitutes a measure zero set in $\mathbb{C}^{2P}$. The number of all 2×2 sub-matrices of $\mathbf{GC}$ is finite, and any finite union of measure zero sets is of measure zero. The existence of the desired $\mathbf{G}$ thus follows. Not only does such a $\mathbf{G}$ exist, but in fact a randomly drawn $\mathbf{G}$ will do with probability one.

The existence of such $\mathbf{G}$ implies that the elements on the diagonal of $\mathbf{D}_2(\mathbf{GC})\mathbf{D}_1^{-1}(\mathbf{GC})$ will be distinct. Therefore, the eigen-analysis steps can be carried through to solve for $\mathbf{A}$ and $\mathbf{S}$ from the two mixed slabs $\mathbf{AD}_1(\mathbf{GC})\mathbf{S}^T$ and $\mathbf{AD}_2(\mathbf{GC})\mathbf{S}^T$. With the recovered $\mathbf{A}$ and $\mathbf{S}$, $\mathbf{C}$ can be computed from $\mathbf{X}_p$.

In the proof of our main Theorem, we will need the following:

Lemma 4.1 *Given*

$$\begin{bmatrix} 1 & 0 & * & \cdots & * \\ 0 & 1 & * & \cdots & * \end{bmatrix} \in \mathbb{C}^{2\times M},$$

*where * stands for a non-zero entry, it holds that for almost every* $(\mu_1, \mu_2) \in \mathbf{R}^2$ *(i.e., except for a set of Lebesgue measure zero), the matrix*

$$\mathbf{E} := \begin{bmatrix} 1 & 1 \\ \mu_1 & \mu_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & * & \cdots & * \\ 0 & 1 & * & \cdots & * \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 1 & \bullet & \cdots & \bullet \\ \mu_1 & \mu_2 & * & \cdots & * \end{bmatrix}$$

contains no zero entry in the second row; and the first two elements on the diagonal of $\mathbf{D}_1(\mathbf{E})\mathbf{D}_2^{-1}(\mathbf{E})$ *are distinct and distinct from the remaining elements.*

Proof: Having a zero entry in the second row occurs when (μ_1, μ_2) lies on the union of M lines. Since a finite union of lines cannot cover the plane, zeros in the second row are excluded almost surely. The second claim can be proven in the same manner. □

4.1.3 Main Theorem On Identifiability

Without loss of generality, we assume that $\mathbf{C}_{in}$ is in *canonical form*. The general case can be reduced to canonical form as explained in the following section.

Theorem 4.1 *Given* $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{C})\mathbf{S}^{\mathbf{T}}$, $p = 1, \ldots, P$, $2 \leq M_{in} \leq P$, *where* $\mathbf{A} \in \mathbb{C}^{K\times M}$, $\mathbf{C} \in \mathbb{C}^{P\times M}$, $\mathbf{S} \in \mathbb{C}^{N\times M}$, *and* $\mathbf{C}$ *in canonical form* :

$$\mathbf{C} = \left[\begin{array}{cc} \mathbf{I}_P(1:M_{in}) & \mathbf{C}_{out}, \end{array} \right] \tag{4.8}$$

where $\mathbf{I}_P(1:M_{in})$ *denotes the first* M_{in} *columns of* $\mathbf{I}_P$, *if the first* M_{in} *rows of* $\mathbf{C}_{out}$ *contain no zero entries, and* $k_{\mathbf{C}} \geq 2$, $\min\{k_{\mathbf{A}},\ k_{\mathbf{S}}\} \geq M_{out} + 2$, *then the matrices* $\mathbf{A}$, $\mathbf{C}$, *and* $\mathbf{S}$ *are unique up to permutation and scaling of columns.*

Proof: We will show that we can first recover $\mathbf{A}_{in}$ and $\mathbf{S}_{in}$ up to permutation and scaling of columns from the given $\mathbf{X}_p$, and then obtain $\mathbf{A}_{out}$, $\mathbf{C}_{out}$, and $\mathbf{S}_{out}$ afterwards.

We begin by recovering the first two columns of $\mathbf{A}_{in}$ and $\mathbf{S}_{in}$. Start from

$$\begin{aligned}
\mathbf{X}_1 &= \mathbf{A}\mathbf{D}_1(\mathbf{C})\mathbf{S}^{\mathbf{T}} \\
&= \mathbf{A}\,diag\left[1\ 0\ \overbrace{0\ \cdots\ 0}^{M_{in}-2}\ \overbrace{*\ \cdots\ *}^{M_{out}}\right]\mathbf{S}^T \\
&= \bar{\mathbf{A}}\,diag\left[1\ 0\ *\ \cdots\ *\right]\bar{\mathbf{S}}^T, \\
\mathbf{X}_2 &= \mathbf{A}\mathbf{D}_2(\mathbf{C})\mathbf{S}^{\mathbf{T}} \\
&= \mathbf{A}\,diag\left[0\ 1\ 0\ \cdots\ 0\ *\ \cdots\ *\right]\mathbf{S}^T \\
&= \bar{\mathbf{A}}\,diag\left[0\ 1\ *\ \cdots\ *\right]\bar{\mathbf{S}}^T.
\end{aligned}$$

Recall that $*$ stands for a non-zero entry; $\bar{\mathbf{A}}$ ($\bar{\mathbf{S}}$) is a column-reduced sub-matrix of $\mathbf{A}$ ($\mathbf{S}$). Invoking Lemma 4.1, we always can pick a pair $(\mu_1, \mu_2\ ) \in \mathbf{R}^2$ such that

$$\begin{aligned}
\mathbf{E} :&= \begin{bmatrix} 1 & 1 \\ \mu_1 & \mu_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & * & \cdots & * \\ 0 & 1 & * & \cdots & * \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & \bullet & \cdots & \bullet \\ \mu_1 & \mu_2 & * & \cdots & * \end{bmatrix}
\end{aligned}$$

contains no zero entry in the second row; and the first two elements on the diagonal of $\mathbf{D}_1(\mathbf{E})\mathbf{D}_2^{-1}(\mathbf{E})$ are distinct and distinct from the remaining elements. We also note that both $\bar{\mathbf{A}}$ and $\bar{\mathbf{S}}$ have $M_{out}+2$ columns from the original $\mathbf{A}$ and $\mathbf{S}$; by definition of k-rank, it follows that

$$k_{\bar{\mathbf{A}}} \geq \min(k_{\mathbf{A}}, M_{out}+2),$$

$$k_{\bar{\mathbf{S}}} \geq \min(k_{\mathbf{S}}, M_{out}+2).$$

Due to fact that $\min\{k_{\mathbf{A}}, k_{\mathbf{S}}\} \geq M_{out}+2$, both $\bar{\mathbf{A}}$ and $\bar{\mathbf{S}}$ are full column rank. Therefore, eigenanalysis in subsection 4.1.2 can be applied to the following mixed slabs

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{X}_2 = \bar{\mathbf{A}} diag\,[1 \;\; 1 \;\; \bullet \;\; \cdots \;\; \bullet]\, \bar{\mathbf{S}}^T,$$

$$\mathbf{Y}_2 = \mu_1 \mathbf{X}_1 + \mu_2 \mathbf{X}_2 = \bar{\mathbf{A}} diag\,[\mu_1 \;\; \mu_2 \;\; \bullet \;\; \cdots \;\; \bullet]\, \bar{\mathbf{S}}^T,$$

to recover the first two columns of $\mathbf{A}$ and $\mathbf{S}^T$ up to permutation and scaling. We can repeat this procedure with $\mathbf{X}_i$ and $\mathbf{X}_{i+1}$ to recover the i-th and the $(i+1)$-th columns of $\mathbf{A}_{in}$ and $\mathbf{S}_{in}$ for $i = 2, \cdots, M_{in}-1$ until both $\mathbf{A}_{in}$ and $\mathbf{S}_{in}$ are recovered. The matrices $\mathbf{X}_p^{in} := \mathbf{A}_{in}\mathbf{D}_p(\mathbf{I}_P(1 : M_{in}))\mathbf{S}_{in}^T$ corresponding to the in-cell users can be constructed, and we thus obtain the matrices $\mathbf{X}_p^{out}$ by subtracting $\mathbf{X}_p^{in}$ from $\mathbf{X}_p$ for $p = 1, \ldots, P$.

$\mathbf{X}_p^{out}$ is nothing but $\mathbf{A}_{out}\mathbf{D}_p(\mathbf{C}_{out})\mathbf{S}_{out}$. Since $\mathbf{A}_{out}$, $\mathbf{C}_{out}$, and $\mathbf{S}_{out}$ are all M_{out}-column sub-matrices of $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{S}$, respectively, we have

$$k_{\mathbf{A}_{out}} \geq \min(k_{\mathbf{A}}, M_{out}) = M_{out},$$

$$k_{\mathbf{S}_{out}} \geq \min(k_{\mathbf{S}}, M_{out}) = M_{out},$$

$$k_{\mathbf{C}_{out}} \geq \min(k_{\mathbf{C}}, M_{out}) = \min(2, M_{out}).$$

The first two inequalities hold due to the condition that $\min\{k_{\mathbf{A}}, k_{\mathbf{S}}\} \geq M_{out}+2$, and imply that both $\mathbf{A}_{out}$ and $\mathbf{S}_{out}$ are full column rank matrices.

If $M_{out} \geq 2$, we know that $k_{\mathbf{C}_{out}} \geq 2$, therefore eigenanalysis (the general result) can be applied to recover $\mathbf{A}_{out}$, $\mathbf{C}_{out}$, and $\mathbf{S}_{out}$, up to permutation and scaling of columns.

When $M_{out} = 1$, it is known that rank-one matrix decomposition is unique up to scaling. □

Remark 4.1 *A similar result can be derived for* $M_{in} = 1$*, with slightly restricted condition on* $\mathbf{C}_{out}$*. We require both* $\mathbf{C}_{in}$ *and* $\mathbf{C}_{out}$ *to be full column k-rank in the proof of Theorem 4.1, but note that* $\mathbf{C}$ *can be a fat matrix.*

Remark 4.2 *The assumption that the first* M_{in} *rows of* $\mathbf{C}_{out}$ *contain no zero entries is posed mainly for simplicity of proof of Theorem 4.1. Theorem 4.1 holds provided that none of the columns of the sub-matrix comprising of the first* M_{in} *rows of* $\mathbf{C}_{out}$ *is proportional to a column of* $\mathbf{I}_{M_{in}}$*. We chose to prove the slightly restricted Theorem 4.1 due to space considerations.*

Remark 4.3 *The model identifiability conditions of Theorem 4.1 are usually met in practice deterministically or statistically with proper system parameters. For instance, if we assume that* $\mathbf{A}$ *and* $\mathbf{C}$ *are drawn from a continuous distribution, and* $\mathbf{S}$ *drawn from i.i.d. BPSK source, it can be shown that* $k_{\mathbf{A}} \geq M_{out} + 2, k_{\mathbf{C}} \geq 2$ *holds almost surely, while* $k_{\mathbf{S}} \geq M_{out} + 2$ *occurs with very high probability provided that* $K \geq M_{out} + 2, N \geq M, P \geq 2$*.*

4.1.4 Algorithms

The proof of Theorem 4.1 is constructive; it directly yields a sequential eigenvalue-based solution that recovers everything exactly in the noiseless case, under only the model identifiability condition in the Theorem. In the noisy scenario, this eigenvalue approach can be coupled with an iterative LS-based refinement algorithm that yields good estimation performance for moderate SNR and beyond.

Assuming that $\mathbf{C}_{in}$ is known, the two major steps of our algorithm are summarized next :

1) Algebraic Initialization : Arrange the received noisy data $x_{k,n,p}$ into a set of matrices, $\widetilde{\mathbf{X}}_k \in \mathbb{C}^{P\times N}$, for $k = 1, \cdots, K$. The (p, n) entry of $\widetilde{\mathbf{X}}_k$ is $x_{k,n,p}$. It can be shown

that

$$\widetilde{\mathbf{X}}_k = \mathbf{C}\mathbf{D}_k(\mathbf{A})\mathbf{S}^{\mathbf{T}} + \widetilde{\mathbf{W}}_{\mathbf{k}},$$

where $\widetilde{\mathbf{W}}_k$ is the AWGN noise matrix. Left-multiply by the pseudo-inverse of $\mathbf{C}_{in}$ to get $\widetilde{\mathbf{Z}}_k \in \mathbb{C}^{M_{in} \times N}$:

$$\widetilde{\mathbf{Z}}_k = \mathbf{C}_{in}^{\dagger}\widetilde{\mathbf{X}}_k. \tag{4.9}$$

Form another set of matrices $\mathbf{X}_m \in \mathbb{C}^{K \times N}$, for $m = 1, \cdots, M_{in}$ such that the (k, n) entry of $\mathbf{X}_m$ is equal to the (m, n) entry of $\widetilde{\mathbf{Z}}_k$. It can be shown that

$$\mathbf{X}_m = \mathbf{A}\mathbf{D}_m(\mathbf{C}_{in}^{\dagger}\mathbf{C})\mathbf{S}^T + \mathbf{W}_m,$$

where $\mathbf{W}_m$ is the rearranged Gaussian noise matrix. Note that $\mathbf{C}_{in}^{\dagger}\mathbf{C}$ is in canonical form, and thus we may apply the approach described in the proof of Theorem 4.1 to estimate $\mathbf{A}$, $\mathbf{C}_{in}^{\dagger}\mathbf{C}_{out}$, and $\mathbf{S}$. $\mathbf{C}$ can also be estimated as

$$\mathbf{C} = \left[\begin{bmatrix} \mathbf{A}\mathbf{D}_1(\mathbf{S}) \\ \vdots \\ \mathbf{A}\mathbf{D}_N(\mathbf{S}) \end{bmatrix}^{\dagger} \begin{bmatrix} \overline{\mathbf{X}}_1 \\ \vdots \\ \overline{\mathbf{X}}_N \end{bmatrix} \right]^T,$$

where the (k, p) element of $\overline{\mathbf{X}}_n \in \mathbb{C}^{K \times P}$ is given by $x_{k,n,p}$ (cf. [54] for details).

2) Joint Constrained Least Squares (LS) Refinement : Use the $\mathbf{A}$, $\mathbf{C}_{out}$, and $\mathbf{S}$ obtained in the first step and the known $\mathbf{C}_{in}$ as initialization for Constrained Trilinear Alternating Least Squares (CTALS) regression applied to the original data $x_{k,n,p}$. The basic idea behind TALS is to compute a conditional LS update of $\mathbf{A}$ given $\mathbf{C}$, $\mathbf{S}$, then repeat for $\mathbf{S}$, etc in a circular fashion until convergence [54]. For CTALS, the $\mathbf{C}_{in}$ part of $\mathbf{C}$ is fixed, and only $\mathbf{C}_{out}$ is updated in the iterations.

4.2 Extension to Quasi-Synchronous Systems and Multipath Channels

There are two issues that must be addressed in order to establish the usefulness of our algorithm in a realistic cellular CDMA environment. One is synchronization; the other is frequency selectivity.

In so-called Quasi-Synchronous CDMA (QS-CDMA) the symbol timing of the in-cell users may be off by as much as a few chips. This causes ISI, but, as already mentioned, it can be circumvented by dropping a short chip-prefix for each symbol at the receiver - the associated performance degradation is negligible when the prefix is short relative to the spreading gain.

Quasi-synchronism is a reasonable assumption for the in-cell user population in the context of 3G systems (e.g., CDMA2000), but much less so for out-of-cell users, who actually attempt to synchronize with a *different* BS. The key here is (4.1): asynchronous out-of-cell users appear as two virtual synchronous users, with "split" code pieces, and symbol sequences that are offset by one symbol. Note that splitting and offset generally preserve linear independence; however, the steering vectors (spatial responses) will be co-linear for each such pair of virtual users. Fortunately, by exchanging the roles of **A** and **C** and invoking the remark on repeated eigenvalues in Section 4.1.2, it can be shown that the parameters of all in-cell users can still be uniquely determined, along with the span of each pair of virtual out-of-cell users.

Frequency selectivity is realistically modeled by convolution with a relatively short chip-rate FIR filter that models the discrete-time baseband-equivalent channel impulse response, including transmit chip pulse-shaping and receive chip-matched filtering. The effective spreading codes seen at the receiver are the convolution of the transmit-codes with the corresponding multipath channels. This means that the in-cell receive-codes must be estimated before our basic approach developed in the above section can be applied. This estimation is compounded by the co-channel out-of-cell interference, which is not under the control of the base station. In order to deal with the problem of receive-code estimation for the in-cell users, we propose the following *admission protocol*:

> As new in-cell users come into the system, they are initially treated as out-of-cell: their receive-codes are thereby estimated blindly, and they are subsequently added to the list of in-cell users. Initially, the process is started by solving a blind problem, as in [54].

In this way, the problem of receive-code estimation for the in-cell users is never explicitly solved. Once the in-cell receive-codes have been estimated at the base station, the proposed algorithm can be carried over to the quasi-synchronous frequency-selective DS-CDMA systems.

4.3 Asymptotic Cramér-Rao Bound

In order to benchmark the performance of our estimation algorithm, it is useful to derive pertinent bounds. While low bit error rate is of primary concern, accurate estimates of the out-of-cell user's receive-codes and both in-cell and out-of-cell steering vectors are also of interest. Cramér-Rao bounds can be developed for the latter, owing to the fact that, unlike symbols, steering vectors and receive-codes are continuous parameters.

The conditional Cramér-Rao Bound (CRB) for low-rank decomposition of multi-dimensional arrays has been derived in [36], assuming all matrices are fixed unknowns. In our present context, however, we are more interested in bounds that are independent of the symbol matrix $\mathbf{S}$. Towards this end, we can aim for one of two options: Computing an averaged (or modified) CRB, or an asymptotic CRB. The former turns out being far more complicated to derive in closed form; we therefore opt for the latter.

In the Appendix, wherein the detailed CRB derivations can be found, we begin by developing a compact form of the conditional CRB in [36]. The new compact form is much simpler to compute than the expression given in [36]. Then, following the approach developed in [59], we work out the asymptotic CRB as the number of symbols, N, goes to infinity. The key to this computation is that the limit and the CRB operator can be exchanged, since the latter is continuous; and when N tends to infinity, the sample estimate of the correlation matrix of $\mathbf{S}$ approaches the exact correlation matrix of $\mathbf{S}$. For the sake of brevity, in what follows, we assume the entries of $\mathbf{S}$ are drawn from and i.i.d. BPSK constellation. This implies that

$$E(s_{m_1}(n_1)s_{m_2}(n_2)) = \delta_{m_1,m_2}\delta_{n_1,n_2}. \tag{4.10}$$

Note that the asymptotic CRB derived in the Appendix is valid for arbitrary $\mathbf{C}$ - it is not necessary to have $\mathbf{C}_{in}$ in canonical form. The main limitation of the asymptotic CRB is that it is valid for large enough N, but for small N there will be some mismatch. In the Section 4.4, we will compare the performance of the proposed approach against the asymptotic CRB. Throughout, the asymptotic CRB is first normalized in an element-wise fashion, i.e., each unknown parameter's CRB is weighed with weight proportional to the inverse modulus square of the respective parameter. The average weighted CRB of all the unknown parameters is then used as a single performance metric. The average mean squared error (MSE) for all free model parameters is treated in a similar fashion.

4.4 Simulation Results

In this section, we provide computer simulation results to demonstrate the performance of the proposed algorithm.

As per Theorem 4.1, scaling ambiguity for all active users and the permutation ambiguity among out-of-cell users is inherent to this blind separation problem. We remove the column scaling ambiguity among the estimated symbol matrix $\mathbf{S}$ via differential encoding, and assume differentially encoded user signals throughout the simulations. For the purpose of performance evaluation only, the permutation ambiguity among the out-of-cell users is resolved using a greedy least square matching algorithm [54]. This permutation ambiguity among the out-of-cell users cannot be solved at BS without additional side information, but this indeterminacy is irrelevant in practice.

Let $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{C})\mathbf{S}^T + \mathbf{W}_p$ be the received noisy data, for $p = 1, \cdots, P$, where $\mathbf{W}_p$ are the AWGN matrices. We define the sample SNR at the input of the multiuser receiver as

$$\text{SNR} := 10\log_{10}\frac{\sum_{p=1}^{P} \parallel \mathbf{A}\mathbf{D}_p(\mathbf{C})\mathbf{S}^{\mathbf{T}} \parallel_F^2}{\sum_{p=1}^{P} \parallel \mathbf{W}_p \parallel_F^2}\text{d}B. \tag{4.11}$$

We first show that the proposed algebraic initialization significantly accelerates the convergence of least square refinement and improves the performance. In order to have a benchmark, we consider cases wherein TALS-based COMFAC [54] is also applicable, but note that the approach developed herein can work well when COMFAC fails. When both methods are applicable, our simulations show that the new approach yields better performance.

Fig. 4.1 plots BER versus average SNR, without out-of-cell interference and for $M_{in} = 4$, DE-BPSK, $K = 2, N = 50$, and $P = 4$. Results are averaged over 10^2 i.i.d. Rayleigh channels ($\mathbf{A}$ - no power control is assumed), and 10^6 realizations per Rayleigh channel. Note that total averaging is $O(10^8)$. The spreading codes are randomly drawn from a continuous distribution and fixed throughout the simulations. Fig. 4.2 depicts average BER for the in-cell users for $M_{in} = 4, M_{out} = 2, K = 4, N = 50, P = 4$, and otherwise the same simulation setup. Note that in the second experiment, both the number of antennae and spreading gain are less than the number of total active users. It is seen from those figures that, as expected, the proposed algorithm has provided better BER performance than COMFAC; in particular, such improvement

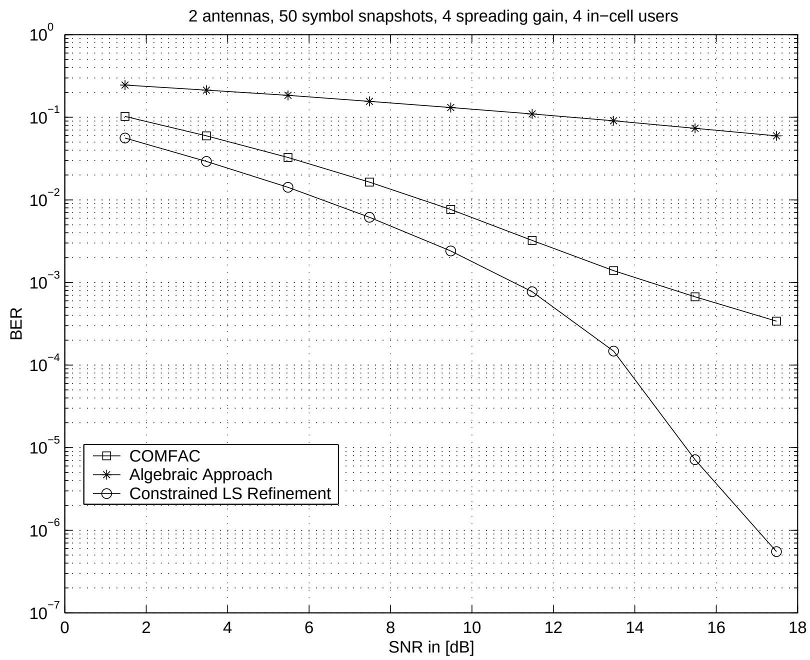

Figure 4.1: No out-of-cell user interference

is significant in the high SNR regime. In addition, the proposed algorithm has been observed to converge at least 70 percent faster (in time) than the general TALS with random initialization, and comparably with respect to the computation-efficient TALS-based COMFAC, especially in the high SNR regime.

Next, the performance of the proposed algorithm and that of the linear group-blind decorrelating detector [56] with different sample sizes is shown in Fig. 4.3. The original group-blind multiuser detector is designed for uplink CDMA with a single receive antenna, but the approach of [56] can be easily extended to handle multiple BS antennas, provided that the array steering vectors, in addition to the spreading codes, of all the in-cell users are known. Estimating steering vectors is more difficult than estimating codes, because the former vary faster due to mobility-induced fast fading. In our simulation, in contrast to the proposed algorithm, the linear group-blind decorrelating detector assumes perfect knowledge of in-cell user's steering matrix $\mathbf{A}_{in}$, i.e., we provide the linear group-blind decorrelating detector

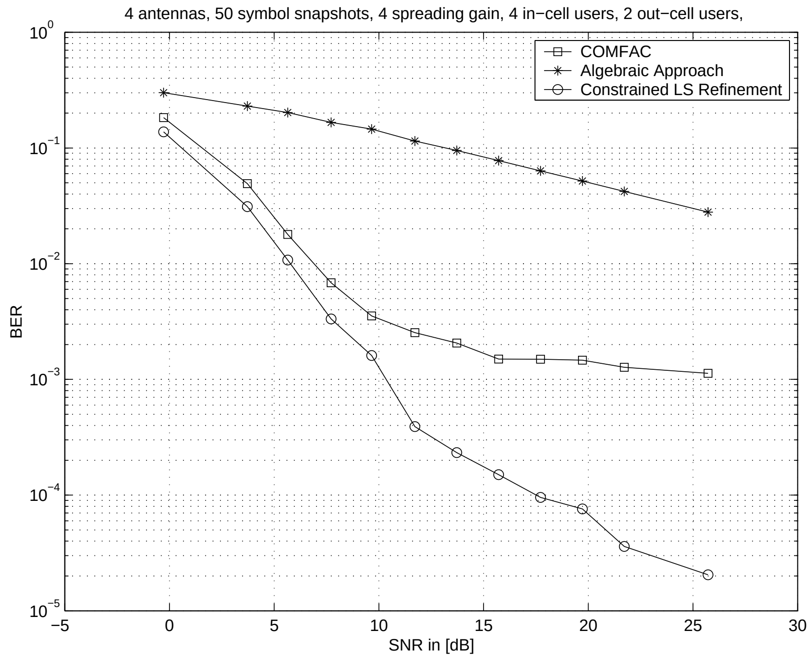

Figure 4.2: More active users than spreading gain

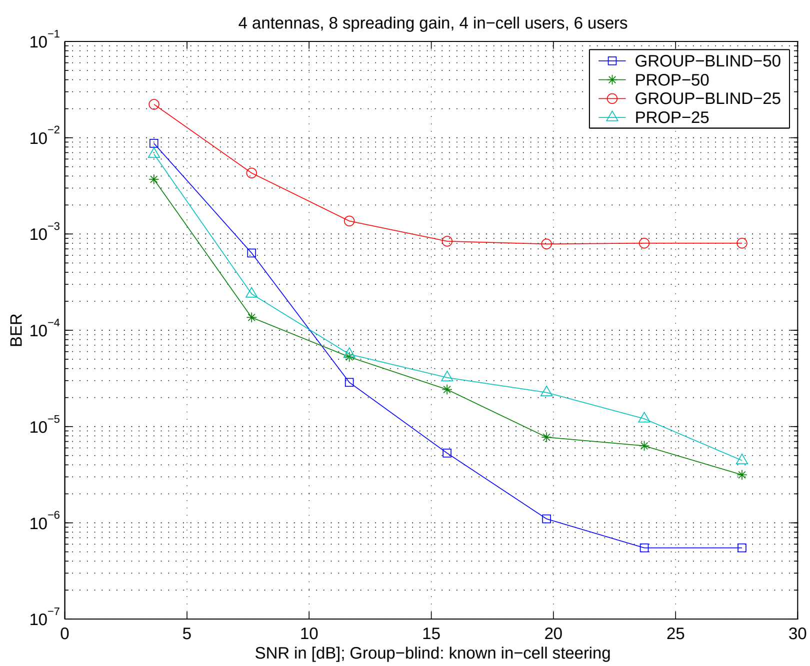

Figure 4.3: Support for small sample sizes

with perfect knowledge of $(\mathbf{C}_{in} \odot \mathbf{A}_{in})$ in (4.4). Fig. 4.3 depicts the performance of the two competing detectors for two different sample sizes, $N = 25$, $N = 50$. It is observed that the linear group-blind decorrelating detector exhibits an error floor in the high SNR regime due to using sample estimates of the correlation matrix. This yields cancellation errors which persist for any number of finite samples, even in the noiseless case. However, such error floor is acceptable when we use large sample sizes. With 50 snapshots, the linear group-blind decorrelating detector provides better BER performance than the proposed detector in the high SNR regime even though the error floor surfaces at about 24 dB. With a small sample size of $N = 25$, the proposed detector clearly outperforms the linear group-blind decorrelating detector, despite the fact that it uses less side information. In both cases, the proposed detector outperforms the linear group-blind decorrelating detector in the low SNR regime. We emphasize that the proposed algorithm performs well even for very small sample sizes (e.g., $N = 10$) in the high SNR regime, whereas the group-blind approach hits the error floor at very low SNR in this case.

Our proposed detector is also robust to strong out-of-cell interference. We have compared the user 1's BER performance of proposed approach against the usual MMSE receiver, which assumes exact knowledge of the in-cell user codes and steering vectors, but treats out-of-cell users as Gaussian interference. The soft MMSE solution for $\mathbf{S}$ is

$$\widehat{\mathbf{S}}_{in}^{T} = ((\mathbf{C}_{in} \odot \mathbf{A}_{in})^{H}(\mathbf{C}_{in} \odot \mathbf{A}_{in}) + \frac{1}{\mathrm{SNR}}\mathbf{I})^{-1}(\mathbf{C}_{in} \odot \mathbf{A}_{in})^{H}\mathbf{X}^{KP\times N}.$$

Fig. 4.4 shows that as the power of out-of-cell users increases, the performance of the MMSE receiver deteriorates significantly whereas the degradation of the proposed detector is marginal.

The proposed algorithm is capable of accurately estimating the steering matrix of all active users and code matrix of out-of-cell users. We wish to compare the MSE

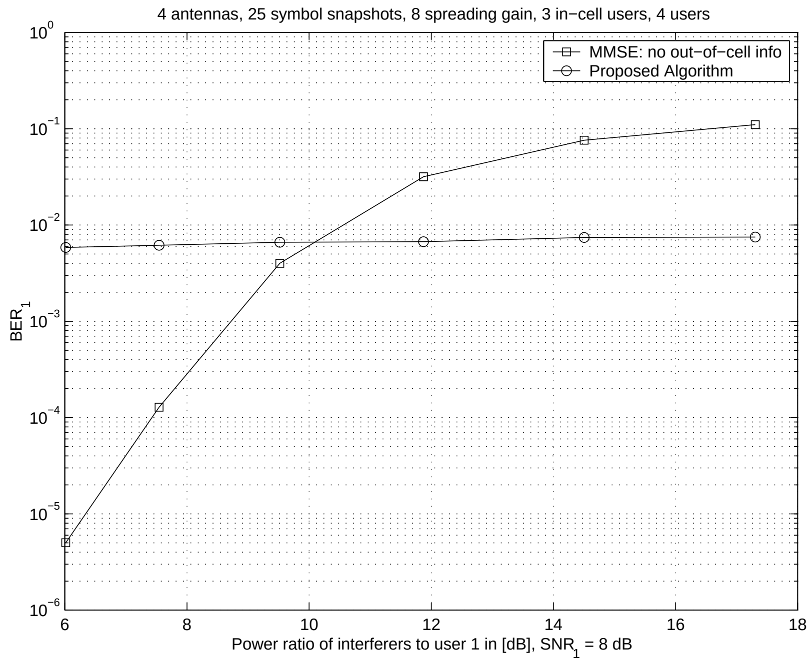

Figure 4.4: Robustness to strong out-of-cell interference

performance of proposed approach against the respective asymptotic CRB. The SNR is defined as

$$\text{SNR} := 10 \log_{10} \frac{\| \mathbf{C} \odot \mathbf{A} \|_F^2}{K P \sigma^2} dB, \tag{4.12}$$

which can be shown consistent with the definition (4.11) when we take the expectation of (4.11) with respect to $\mathbf{S}$.

Fig. 4.5 depicts simulation results comparing TALS performance to this asymptotic CRB for two different snapshots. In this simulation, $K = 4, P = 4, M = 6$, and the true parameters were used to initialize TALS. The point here is to measure how tight the asymptotic CRB is for various N; for this reason, we use the sought parameters as initialization in order to ensure the best possible scenario for TALS. It can be seen that TALS with good initialization remains very close to the CRB from medium to high SNR and relatively large sample size, $N = 64$. Note that $N = 64$ is a reasonable number of symbol snapshots in practice. When the sample size is rela-

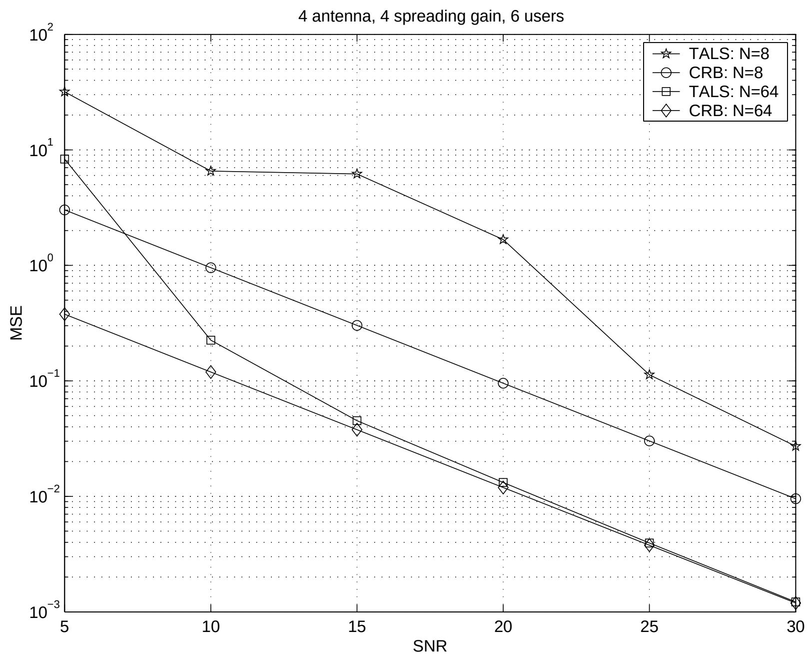

Figure 4.5: TALS performance versus Asymptotic CRB

tively small, the MSE performance of TALS is naturally worse than what predicted by the asymptotic CRB.

Fig. 4.6 presents the average MSE performance of COMFAC and the proposed algorithm against the CRB bound. We note that the performance of the proposed algorithm exceeds that of COMFAC considerably once SNR goes beyond the low SNR regime. This is because the new algebraic approach can provide fairly accurate initializations for CTALS whereas the COMFAC is forced to use random initializations in this case, wherein no two modes are full column rank. The average MSE of the proposed algorithm deviates from CRB about two to three dB. This is mainly because the initializations the algebraic approach provides are still not perfect, and the pre-specified tolerance threshold used to terminate the iterative refinement algorithm is set higher than in previous simulations, due to complexity considerations.

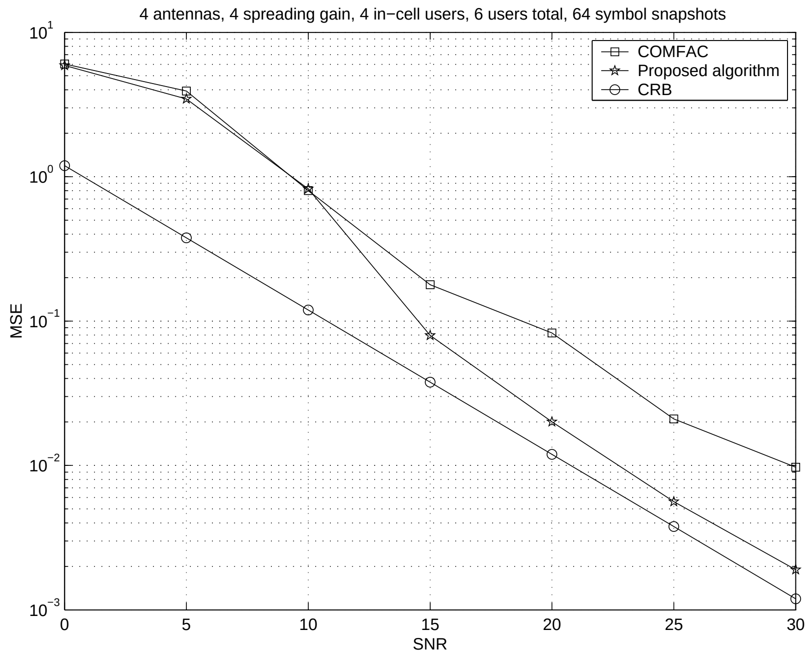

Figure 4.6: MSE performance of COMFAC and the proposed algorithm against the CRB bound

4.5 Conclusions

Out-of-cell interference in DS-CDMA systems is usually treated as noise, possibly mitigated using random cell codes. If the total number of in-cell plus out-of-cell users is smaller than the spreading gain, subspace-based suppression of out-of-cell users is possible. The assumption of more spreading than the total number of users can be quite unrealistic, even for moderately loaded cells. Completely blind reception is feasible under certain conditions (even with more users than spreading) with base station antenna arrays. We have proposed a new blind identification procedure that is capable of recovering both in-cell and out-of-cell transmissions, with sole knowledge of the in-cell user codes. The codes of the out-of-cell users and the steering vectors of all users are also recovered. The new procedure remains operational even when completely blind or subspace-based procedures fail. Interestingly, if the in-cell codes are known, then algebraic solution is possible.

Appendix 4.A Asymptotic CRB as N tends to infinity

To derive a meaningful CRB, following what has been done in [36], we assume that the first row of $\mathbf{A}$ and $\mathbf{S}$ is fixed (or normalized) to $[1 \cdots 1]_{1\times F}$ (this takes care of scale ambiguity), the first row of $\mathbf{C}_{out}$ is known and consists of distinct elements (which subsequently resolves the permutation ambiguity) and $\mathbf{C}_{in}$ is in *canonical form*. In turn, the number of unknown complex parameters is $(N+K-2)M+(P-1)M_{out}$. Let

$$\boldsymbol{\theta} := \left[\mathbf{a}_2^T;\ldots;\mathbf{a}_K^T;\mathbf{c}_{out}{}_2^T;\ldots;\mathbf{c}_{out}{}_P^T;\mathbf{s}_2^T;\ldots;\mathbf{s}_N^T;\mathbf{a}_2^H;\ldots;\mathbf{s}_N^H\right] \in \mathbb{C}^{(N+K-2)M+(P-1)M_{out}\times 1}, \tag{4.13}$$

where $\mathbf{a}_k$ denotes the kth row of $\mathbf{A}$, $\mathbf{c}_{outp}$ denotes the ith row of $\mathbf{C}_{out}$, and $\mathbf{s}_n$ denotes the nth row of $\mathbf{S}$.

It has been shown in [36] that the Fisher Information Matrix (FIM) is given by

$$\Omega(\boldsymbol{\theta}) = \mathrm{E}\left\{\left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^H \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right\} = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^* \end{bmatrix}, \tag{4.14}$$

where $f(\boldsymbol{\theta})$ is the log-likelihood function and

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ac} & \boldsymbol{\Psi}_{as} \\ \boldsymbol{\Psi}_{ac}^H & \boldsymbol{\Psi}_{cc} & \boldsymbol{\Psi}_{cs} \\ \boldsymbol{\Psi}_{as}^H & \boldsymbol{\Psi}_{cs}^H & \boldsymbol{\Psi}_{ss} \end{bmatrix}$$

with obvious notation; In addition,

$$\begin{bmatrix} \mathrm{CRB}_{aa} & \mathrm{CRB}_{ac} \\ \mathrm{CRB}_{ac}^H & \mathrm{CRB}_{cc} \end{bmatrix} = \left(\begin{bmatrix} \boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ac} \\ \boldsymbol{\Psi}_{ac}^H & \boldsymbol{\Psi}_{cc} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Psi}_{as} \\ \boldsymbol{\Psi}_{cs} \end{bmatrix} \boldsymbol{\Psi}_{ss}^{-1} \left[\boldsymbol{\Psi}_{as}^H \boldsymbol{\Psi}_{cs}^H\right]\right)^{-1}. \tag{4.15}$$

The elements of $\boldsymbol{\Psi}$ can be given[3] as follows

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{k_1,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial a_{k_2,m_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}^H_{m_1}\big(\sum_{n=1}^{N}(\mathbf{s}_n\odot\mathbf{C})^H(\mathbf{s}_n\odot\mathbf{C})\big)\mathbf{e}_{m_2}\delta_{k_1,k_2}$$
$$= \frac{1}{\sigma^2}\mathbf{e}^H_{m_1}\big(\sum_{n=1}^{N}\mathbf{s}^H_n\mathbf{s}_n\big)\diamond(\mathbf{C}^H\mathbf{C})\mathbf{e}_{m_2}\delta_{k_1,k_2},$$
$$k_1,k_2 = 2,\cdots,K, m_1,m_2 = 1,\cdots,M,$$

where we have used the following identity

$$(\mathbf{C}^H\mathbf{C})\diamond(\mathbf{D}^H\mathbf{D}) = (\mathbf{C}\odot\mathbf{D})^H(\mathbf{C}\odot\mathbf{D}),$$

and $\diamond$ stands for the Hadamard product. Similarly, we have

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial c^*_{p_1,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{p_2,m_2}}\right\} = \frac{1}{\sigma^2}\mathbf{e}^H_{m_1}(\mathbf{A}^H\mathbf{A})\diamond\big(\sum_{n=1}^{N}\mathbf{s}^H_n\mathbf{s}_n\big)\mathbf{e}_{m_2}\delta_{p_1,p_2}$$
$$p_1,p_2 = 2,\cdots,P; m_1,m_2 = M_{in}+1,\cdots,M.$$

In addition, we have

$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial s^*_{n_1,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial s_{n_2,m_2}}\right\} = \tfrac{1}{\sigma^2}\mathbf{e}^H_{m_1}(\mathbf{C}\odot\mathbf{A})^H(\mathbf{C}\odot\mathbf{A})\mathbf{e}_{m_2}\delta_{n_1,n_2}$$
$$= \tfrac{1}{\sigma^2}\mathbf{e}^H_{m_1}(\mathbf{C}^H\mathbf{C})\diamond(\mathbf{A}^H\mathbf{A})\mathbf{e}_{m_2}\delta_{n_1,n_2} \qquad (4.16)$$
$$n_1,n_2 = 2,\cdots,N; m_1,m_2 = 1,\cdots,M \qquad (4.17)$$
$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{k,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{p,m_2}}\right\} = \tfrac{1}{\sigma^2}\big(\textstyle\sum_{n=1}^{N}\mathbf{s}^*_n(m_1)\mathbf{s}_n(m_2)\big)\mathbf{c}^*_p(m_1)\mathbf{a}_k(m_2) \qquad (4.18)$$
$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{k,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial s_{n,m_2}}\right\} = \tfrac{1}{\sigma^2}\mathbf{s}^*_n(m_1)\big(\textstyle\sum_{p=1}^{P}\mathbf{c}^*_p(m_1)\mathbf{c}_p(m_2)\big)\mathbf{a}_k(m_2), \qquad (4.19)$$
$$E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial c^*_{p,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial s_{n,m_2}}\right\} = \tfrac{1}{\sigma^2}\mathbf{s}^*_n(m_1)\big(\textstyle\sum_{k=1}^{K}\mathbf{a}^*_k(m_1)\mathbf{a}_k(m_2)\big)\mathbf{c}_p(m_2). \qquad (4.20)$$

Since we have assumed that

$$E(\mathbf{s}^*_{n1}(m_1)\mathbf{s}_{n_2}(m_2)) = \delta_{n_1,n_2}\delta_{m_1,m_2},$$

[3]The forms given here can be shown to be mathematically equivalent to those in [36]. The new forms are computationally much simpler.

it follows

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{k_1,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial a_{k_2,m_2}}\right\} &= \frac{1}{\sigma^2}\mathbf{e}^H_{m_1}\Big(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mathbf{s}^H_n\mathbf{s}_n\Big)\diamond(\mathbf{C}^H\mathbf{C})\mathbf{e}_{m_2}\delta_{k_1,k_2}\\
&= \frac{1}{\sigma^2}\mathbf{e}^H_{m_1}\mathbf{I}_M\diamond(\mathbf{C}^H\mathbf{C})\mathbf{e}_{m_2}\delta_{k_1,k_2}\\
\lim_{N\to\infty}\frac{1}{N}E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial c^*_{p_1,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{p_2,m_2}}\right\} &= \frac{1}{\sigma^2}\mathbf{e}^H_{m_1}(\mathbf{A}^H\mathbf{A})\diamond\mathbf{I}_M\mathbf{e}_{m_2}\delta_{p_1,p_2}\\
\lim_{N\to\infty}\frac{1}{N}E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{k,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial c_{p,m_2}}\right\} &= \frac{1}{\sigma^2}\Big(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mathbf{s}^*_n(m_1)\mathbf{s}_n(m_2)\Big)\mathbf{c}^*_p(m_1)\mathbf{a}_k(m_2)\\
&= \frac{1}{\sigma^2}\mathbf{c}^*_p(m_1)\mathbf{a}_k(m_2)\delta_{m_1,m_2},
\end{aligned}$$

hence

$$\lim_{N\to\infty}\frac{1}{N}\begin{bmatrix}\boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ac}\\ \boldsymbol{\Psi}^H_{ac} & \boldsymbol{\Psi}_{cc}\end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix}\boldsymbol{\Psi}_{aa_{\text{limit}}} & \boldsymbol{\Psi}_{ac_{\text{limit}}}\\ \boldsymbol{\Psi}^H_{ac_{\text{limit}}} & \boldsymbol{\Psi}_{cc_{\text{limit}}}\end{bmatrix}$$

with obvious notation.

From (4.16), we know that

$$\boldsymbol{\Psi}_{ss} = \frac{1}{\sigma^2}\begin{bmatrix}(\mathbf{C}^H\mathbf{C})\diamond(\mathbf{A}^H\mathbf{A}) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & (\mathbf{C}^H\mathbf{C})\diamond(\mathbf{A}^H\mathbf{A}) & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{C}^H\mathbf{C})\diamond(\mathbf{A}^H\mathbf{A})\end{bmatrix} \in \mathbb{C}^{(N-1)M\times(N-1)M} \tag{4.21}$$

Let $\boldsymbol{H} := \big((\mathbf{C}^H\mathbf{C})\diamond(\mathbf{A}^H\mathbf{A})\big)^{-1}$; then

$$\boldsymbol{\Psi}^{-1}_{ss} = \sigma^2\begin{bmatrix}\mathbf{H} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{H} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}\end{bmatrix} \in \mathbb{C}^{(N-1)M\times(N-1)M}. \tag{4.22}$$

Recall

$$\begin{aligned}
E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial a^*_{k,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial s_{n,m_2}}\right\} &= \frac{1}{\sigma^2}\mathbf{s}^*_n(m_1)\Big(\sum_{p=1}^{P}\mathbf{c}^*_p(m_1)\mathbf{c}_p(m_2)\Big)\mathbf{a}_k(m_2),\\
E\left\{\frac{\partial f(\boldsymbol{\theta})}{\partial c^*_{p,m_1}}\frac{\partial f(\boldsymbol{\theta})}{\partial s_{n,m_2}}\right\} &= \frac{1}{\sigma^2}\mathbf{s}^*_n(m_1)\Big(\sum_{k=1}^{K}\mathbf{a}^*_k(m_1)\mathbf{a}_k(m_2)\Big)\mathbf{c}_p(m_2),
\end{aligned} \tag{4.23}$$

from which it is not difficult to see that

$$\begin{bmatrix} \mathbf{\Psi}_{as} \\ \mathbf{\Psi}_{cs} \end{bmatrix} = \frac{1}{\sigma^2}\left[\mathbf{U}_2\mathbf{R}, \mathbf{U}_3\mathbf{R}, \cdots, \mathbf{U}_N\mathbf{R}\right] \in \mathbb{C}^{\left((K-1)M+(P-1)M_{out}\right)\times(N-1)M} \tag{4.24}$$

where

$$\begin{aligned}\mathbf{U}_n =& diag([\overbrace{\mathbf{s}_n^*(1),\cdots \mathbf{s}_n^*(M)}^{1}, \cdots, \overbrace{\mathbf{s}_n^*(1),\cdots,\mathbf{s}_n^*(M)}^{K-1}, \overbrace{\mathbf{s}_n^*(M_{in}+1),\cdots,\mathbf{s}_n^*(M)}^{1}, \\ & \cdots \overbrace{\mathbf{s}_n^*(M_{in}+1),\cdots,\mathbf{s}_n^*(M)}^{P-1}]) \\ & \in \mathbb{C}^{\left((K-1)M+(P-1)M_{out}\right)\times\left((K-1)M+(P-1)M_{out}\right)}\end{aligned}$$

$$\mathbf{R} = \begin{bmatrix} \left(\left(\sum_{p=1}^{P} \mathbf{c}_p^*(m_1)\mathbf{c}_p(m_2)\right)\mathbf{a}_k(m_2)\right)_{\{(k-1)M+m_1,m_2\}} \\ \left(\left(\sum_{k=1}^{K} \mathbf{a}_k^*(m_1)\mathbf{a}_k(m_2)\right)\mathbf{c}_p(m_2)\right)_{\{(p-1)M_{out}+m_1,m_2\}} \end{bmatrix} \in \mathbb{C}^{\left((K-1)M+(P-1)M_{out}\right)\times M}$$

Let

$$\boldsymbol{G} := \begin{bmatrix} \mathbf{\Psi}_{as} \\ \mathbf{\Psi}_{cs} \end{bmatrix} \mathbf{\Psi}_{ss}^{-1}\left[\mathbf{\Psi}_{as}^H \mathbf{\Psi}_{cs}^H\right] \in \mathbb{C}^{\left((K-1)M+(P-1)M_{out}\right)\times\left((K-1)M+(P-1)M_{out}\right)}, \tag{4.26}$$

then,

$$\boldsymbol{G} = \frac{1}{\sigma^2}\sum_{n=2}^{N} \mathbf{U}_n\mathbf{R}\mathbf{H}\mathbf{R}^H\mathbf{U}_n^H. \tag{4.27}$$

With $\mathbf{Z} := \mathbf{R}\mathbf{H}\mathbf{R}^H$, we have

$$\boldsymbol{G} = \frac{1}{\sigma^2}\sum_{n=2}^{N} \mathbf{U}_n\mathbf{Z}\mathbf{U}_n^H, \tag{4.28}$$

and from

$$E(\mathbf{s}_{n1}^*(m_1)\mathbf{s}_{n_2}(m_2)) = \delta_{n_1,n_2}\delta_{m_1,m_2}, \tag{4.29}$$

we obtain

$$\lim_{N\to\infty}\frac{1}{N}\boldsymbol{G} = \frac{1}{\sigma^2}\lim_{N\to\infty}\frac{1}{N}\sum_{n=2}^{N} \mathbf{U}_n\mathbf{Z}\mathbf{U}_n^H = \frac{1}{\sigma^2}\mathbf{Z}\diamond\mathbf{Q}, \tag{4.30}$$

where

$$\mathbf{Q} = \lim_{N\to\infty}\frac{1}{N}\sum_{n=2}^{N} \text{diag}(\mathbf{U}_n)\text{diag}(\mathbf{U}_n)^H. \tag{4.31}$$

Therefore, we have

$$
\begin{aligned}
&\begin{bmatrix} \mathrm{CRB}_{aa} & \mathrm{CRB}_{ac} \\ \mathrm{CRB}_{ac}^{H} & \mathrm{CRB}_{cc} \end{bmatrix}_{limit} \\
&= \lim_{N\to\infty} \begin{bmatrix} \mathrm{CRB}_{aa} & \mathrm{CRB}_{ac} \\ \mathrm{CRB}_{ac}^{H} & \mathrm{CRB}_{cc} \end{bmatrix} \\
&= \frac{1}{N}\left(\lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} \boldsymbol{\Psi}_{aa} & \boldsymbol{\Psi}_{ac} \\ \boldsymbol{\Psi}_{ac}^{H} & \boldsymbol{\Psi}_{cc} \end{bmatrix} - \lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} \boldsymbol{\Psi}_{as} \\ \boldsymbol{\Psi}_{cs} \end{bmatrix} \boldsymbol{\Psi}_{ss}^{-1} \left[\boldsymbol{\Psi}_{as}^{H} \boldsymbol{\Psi}_{cs}^{H} \right] \right)^{-1} \\
&= \frac{\sigma^2}{N} \left(\begin{bmatrix} \boldsymbol{\Psi}_{aa_{\mathrm{limit}}} & \boldsymbol{\Psi}_{ac_{\mathrm{limit}}} \\ \boldsymbol{\Psi}_{ac_{\mathrm{limit}}}^{H} & \boldsymbol{\Psi}_{cc_{\mathrm{limit}}} \end{bmatrix} - \mathbf{Z} \diamond \mathbf{Q} \right)^{-1} .
\end{aligned}
\tag{4.32}
$$

Chapter 5

Direct Blind Receiver for SIMO and MIMO OFDM Subject to Frequency Offset

Several OFDM CFO estimators for frequency-selective fading channels have been proposed in the literature. Among them, the non-blind methods require either symbol repetition [44] or pilot symbols [47]; the blind MUSIC-like CFO estimator [38, 63, 64] based on guard null subcarriers needs sufficiently many null subcarriers to achieve satisfactory high-resolution CFO estimation, and in certain cases CFO may not be identifiable [42]. Identifiability of null-subcarrier-based CFO estimation has been studied in [42], wherein a null subcarrier hopping scheme was proposed which guarantees identifiability and improves performance. Null subcarriers are incorporated in several wireless OFDM standards, but their placement is fixed as in [38, 63, 64] and does not guarantee identifiability. Subcarrier hopping restores identifiability, but also gives up the adjacent channel interference protection afforded by fixed guard null subcarriers. Null subcarrier methods are not really blind, in the sense that null subcarriers are equivalent to zero-padding and hence suffer roughly the same rate loss as training. Another blind OFDM CFO estimator [3] exploits the cyclostationarity of OFDM signals instead of guard null subcarriers to achieve a carrier frequency acquisition and relies on second-order statistics only. This approach is appealing, but it unrealistically assumes that the time-dispersive channel impulse response is known at the receiver, and it requires more samples to estimate second-order cyclic statistics reliably.

In addition to CFO, the frequency-selective fading channel information is also unknown at the receiver in wireless OFDM systems. The literature on blind receiver design for OFDM systems in the presence of unknown frequency offset and multipath is relatively scarce. Recently, a Bayesian blind turbo receiver [41] for *coded* OFDM systems under the Markov chain Monte Carlo (MCMC) framework for Bayesian computation and utilizing the turbo principle has been proposed. A Bayesian demodulator is able to compute *approximate a posteriori* probabilities of the data symbols based on the received signals by employing MCMC techniques, and is followed by a maximum a *posteriori* (MAP) channel decoder. The Bayesian turbo receiver iteratively exchanges the extrinsic *a posteriori* probabilities of data symbols between the Bayesian demodulator and MAP decoder. It requires knowledge of the specific probability distributions of the data symbols, multipath, and CFO. Perfect data symbol recovery in the noiseless scenario is not guaranteed. The Bayesian turbo receiver in [41] does not yield an estimate of CFO or the underlying multipath channel, and it is computationally intensive.

Relying on *parallel factor* (PARAFAC) analysis, we explore herein a novel approach for data symbol detection in wireless uncoded or coded OFDM systems subject to unknown multipath and CFO. This requires the use of at least two antennas at the receiver, but in return yields important benefits besides the anticipated receive-diversity benefit: blind CFO estimation and direct symbol recovery (up to subcarrier scaling) with guaranteed identifiability under mild conditions. The key is to recognize that, after suitable preprocessing, the baseband-equivalent models of single input multiple output (SIMO) and multiple input multiple output (MIMO) OFDM systems subject to multipath fading and CFO conform to the parallel factor (PARAFAC) analysis model. It has been shown [30] that under mild conditions, PARAFAC models admit unique decomposition up to permutation and scaling. Capitalizing on this link, we develop a blind SIMO/MIMO OFDM receiver that works irrespective of CFO and yields direct data estimates up to subcarrier scaling plus high-resolution CFO estimates. The resulting receiver is typically only a few dB away from the non-blind MMSE receiver that assumes exact knowledge of channel and CFO. Given a CFO estimate from one batch of data, low-complexity adaptive (e.g., decision-directed) CFO tracking can be used to return to the low-complexity FFT receiver for subsequent data blocks. That is, the proposed PARAFAC receiver can be run periodically, and the simpler FFT receiver can be used in-between to coherently combine all antenna channels to improve performance.

The rest of this chapter is structured as follows. Section 5.1 describes our assumptions and the discrete-time baseband equivalent data model of SIMO OFDM in the presence of quasi-static block multipath fading and CFO. In Section 5.2, we link the problem of interest to PARAFAC analysis, and elaborate on identifiability. A receiver based on PARAFAC analysis is proposed in Section 5.3. The corresponding data model for MIMO OFDM and associated identifiability are outlined in Section 5.4. Section 5.5 presents simulation results, and Section 5.6 summarizes our conclusions.

5.1 System Model of SIMO-OFDM

Consider a synchronous discrete Fourier transform (DFT) based SIMO-OFDM system with $I \geq 2$ antennas at the receiver, N subcarriers, subject to frequency-selective block fading and frequency offset. As a final step for symbol recovery, we will need to take care of the scaling ambiguity that is inherent in all blind methods. A simple (but not the only) way to do this is to use differential encoding on a per-subcarrier basis. This is the reason why the proposed receiver is really semi-blind as opposed to fully-blind. Let $\mathbf{s}(k) = [s_1(k), s_2(k), \cdots, s_N(k)]$ be the kth block of differentially-encoded data to be transmitted. OFDM modulation is implemented by left multiplying $\mathbf{s}^T(k)$ by an inverse FFT matrix, and the resulting N-point time-domain signal is given by $\mathbf{F}^H\mathbf{s}^T(k)$. To guard against inter-block interference (IBI) induced by the channel's time-dispersive effect, a cyclic prefix (CP) of length L is added before transmission, where L is chosen to exceed the anticipated maximum delay spread of the I point-to-point channels. The resulting transmitted block has length $N + L$, and is given by

$$\mathbf{x}(k) = \mathbf{T}_{cp}\mathbf{F}^H\mathbf{s}^T(k),$$

where $\mathbf{T}_{cp} := [(\mathbf{I}_{L\times N})^T\mathbf{I}_N^T]^T \in \mathbb{C}^{(N+L)\times N}$. Each block $\mathbf{x}(k)$ is then converted for serial transmission through the I unknown frequency-selective block fading channels whose discrete-time equivalent form has finite impulse response $\{h_i(l)\}_{l=1}^{L_i}$ of order $L_i \leq L$ for the multipath channel between the transmitter and ith receiver.

In the presence of zero-mean additive white Gaussian noise (AWGN) and CFO, the samples at the ith receiver with proper sampling are

$$u_i(t) = e^{\sqrt{-1}\phi t}\sum_{l=0}^{L_i} h_i(l)x(t-l) + v_i(t),$$

where ϕ is the normalized CFO due to the Doppler effects and/or mismatch between the receivers and transmitter oscillators. The initial phase due to CFO is assumed to be zero (equivalently, the initial phase can be absorbed into $h_i(l)$). We also assume that the absolute value of CFO is less than or equal to half of the OFDM subcarrier spacing, i.e., $|\phi| \leq \frac{\Delta\omega}{2}$, where $\Delta\omega = \frac{2\pi}{N}$. This is a valid assumption in practice since the large frequency offset has been already compensated via an automatic frequency control [11], and what remains is the residual frequency offset. The $I \geq 2$ down-conversion chains at the receiver utilize a common local oscillator, and Doppler is essentially common for all receive antennas for long- and medium-range propagation scenarios. The result is that the same CFO appears in the baseband model of all I receive chains. Let $x(k(N+L)+t)$ be the serialized version of the kth block $\mathbf{x}(k)$ with the tth entry $[\mathbf{x}(k)]_t = x(k(N+L)+t)$, and $v_i(t)$ the corresponding AWGN. Form the $(N+L)\times 1$ block $\mathbf{u}_i(k)$ from $u_i(t)$ such that $[\mathbf{u}_i(k)]_t := u_i(k(N+L)+t)$. The CP removal can be accomplished by left multiplying $\mathbf{u}_i(k)$ by the CP removing matrix $\mathbf{T}_{rm} := [\mathbf{0}_{N\times L}\mathbf{I}_N] \in \mathbb{C}^{N\times(N+L)}$. This yields $\mathbf{y}_i(k) := \mathbf{T}_{rm}\mathbf{u}_i(k)$. After simple calculations, the receiver input for the kth block at ith antenna in the presence of carrier frequency offset is given by

$$\mathbf{y}_i(k) = \mathbf{P}\mathbf{F}^H\mathbf{H}_i\mathbf{s}^T(k)e^{j\phi((k+1)(N+L)-N)} + \mathbf{w}_i(k), \tag{5.1}$$

for $i = 1, \cdots, I$, where $\mathbf{w}_i(k) := \mathbf{T}_{rm}\mathbf{v}_i(k)$,

$$\mathbf{H}_i := \begin{bmatrix} H_i(1) & \cdots & 0 \\ \vdots & H_i(n) & \vdots \\ 0 & \cdots & H_i(N) \end{bmatrix} \in \mathbb{C}^{N\times N},$$

$H_i(n) = \sum_{l=1}^{L_i} h_i(l)e^{-\frac{j2\pi ln}{N}}$ is the channel frequency response for the nth subcarrier frequency, corresponding to the ith receiver antenna, and

$$\mathbf{P} := \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & e^{j(n-1)\phi} & \vdots \\ 0 & \cdots & e^{j(N-1)\phi} \end{bmatrix} \in \mathbb{C}^{N\times N} \tag{5.2}$$

is the carrier offset matrix.

Both channel response matrix $\mathbf{H}_i$ and CFO matrix P are assumed to remain constant over K blocks. Therefore, the received signals over K blocks at the ith antenna, can be written as

$$\begin{aligned} \mathbf{Y}_i =& \mathbf{P}\mathbf{F}^H\mathbf{H}_i(\mathbf{Q}\mathbf{S})^T + \mathbf{W}_i \\ =& \mathbf{A}\mathbf{H}_i\mathbf{B}^T + \mathbf{W}_i, i = 1, \cdots, I, \end{aligned} \tag{5.3}$$

where $\mathbf{A} := \mathbf{P}\mathbf{F}^H$, $\mathbf{B} = \mathbf{Q}\mathbf{S}$, and

$$\begin{aligned} &\mathbf{Y}_i = [\mathbf{y}_i(1), \cdots, \mathbf{y}_i(K)] \in \mathbb{C}^{N\times K} \\ &\mathbf{S} = \begin{bmatrix} \mathbf{s}(1) \\ \vdots \\ \mathbf{s}(K) \end{bmatrix} \in \mathbb{C}^{K\times N} \\ &\mathbf{Q} = diag([e^{j\phi(N+2L)}, \cdots, e^{j\phi((K+1)(N+L)-N)}]) \in \mathbb{C}^{K\times K} \\ &\mathbf{W}_i = [\mathbf{w}_i(1), \cdots, \mathbf{w}_i(K)] \in \mathbb{C}^{N\times K}. \end{aligned}$$

Since $\mathbf{F}\mathbf{P}\mathbf{F}^H \neq \mathbf{I}$, the CFO matrix P destroys the orthogonality among the subcarriers and hence induces inter-carrier interference (ICI). Therefore, the CFO ϕ, needs to be estimated and compensated before performing DFT and detecting $\mathbf{S}$.

The problem of interest in this chapter is that of estimating the signal matrix $\mathbf{S}$ and CFO ϕ from the receiver outputs $\{\mathbf{Y}_i\}_{i=1}^{I}$ subject to unknown channels $\mathbf{H}_i$.

5.2 Identifiability

Stacking all matrices $\{\mathbf{Y}_i\}$ in (5.3) one over another, the following compact model representation can be derived

$$\mathbf{Y} := \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_I \end{bmatrix} = \begin{bmatrix} \mathbf{AH}_i \\ \vdots \\ \mathbf{AH}_I \end{bmatrix} \mathbf{B}^T + \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_I \end{bmatrix} \tag{5.4}$$
$$=(\mathbf{H} \odot \mathbf{A})\mathbf{B}^T + \mathbf{W},$$

where the ith row of $\mathbf{H} \in \mathbb{C}^{I \times N}$ is given as $[H_i(1), \cdots, H_i(N)]$; $\mathbf{W} \in \mathbb{C}^{NI \times K}$ with obvious notation.

The data model in (5.4) is in fact symmetric (although this is not immediately apparent from (5.4)) and thus admits two more convenient matrix system rearrangements (cf. [53]). In particular

$$\mathbf{Z} := \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_K \end{bmatrix} = (\mathbf{B} \odot \mathbf{H})\mathbf{A}^T + \mathbf{W}^{\mathbf{Z}} \tag{5.5}$$

where $\mathbf{Y}_i(n,k) = \mathbf{Z}_k(i,n)$, and $\mathbf{W}_i(n,k) = \mathbf{W}_k^{\mathbf{Z}}(i,n)$.

The noiseless model

$$\mathbf{Y} = (\mathbf{H} \odot \mathbf{A})\mathbf{B}^T \tag{5.6}$$

is known as a PARAFAC model. The ubiquitous feature of the PARAFAC model in (5.6) is its uniqueness. Under a mild condition, the matrices $\mathbf{H}$, $\mathbf{A}$, and $\mathbf{B}$ that gives rise to the data $\mathbf{Y}$, are unique up to column permutation and scaling. Among several results regarding uniqueness of the PARAFAC model, the most popular and practical one is due to Kruskal [30] : $\mathbf{H}$, $\mathbf{A}$, and $\mathbf{B}$ that give rise to $\mathbf{Y}$ in (5.6), are unique up to permutation and scaling of columns provided $k_{\mathbf{H}} + k_{\mathbf{A}} + k_{\mathbf{B}} \geq 2N + 2$ (cf. [30, 53]).

Note that $\mathbf{A}$ in (5.3) has full k-rank ($k_{\mathbf{A}} = N$) due to the nature of the FFT matrix, regardless of CFO ϕ. Also, it can be shown that a matrix whose elements are drawn from a jointly continuous distribution has full k-rank (equal to its rank) with probability one. This means that, even if the receive antennas and channel taps are correlated, so long as there is diversity and the channels are not fully coherent, the matrix $\mathbf{H}$ will be full k-rank ($k_{\mathbf{H}} = \min(I, N)$) with probability one. finally, when the block size K is equal or larger than the number of carriers N, it can be shown that the probability of rank deficiency of $\mathbf{S}$ drawn from i.i.d. BPSK source is bounded by $O(\frac{(N-1)^2}{2^{(N-1)(K-1)}})$. For proper K and N, this probability can be fairly small (e.g. $N = K = 8$, this probability is less than 10^{-13}). Therefore, differentially encoded $\mathbf{S}$ drawn from i.i.d. BPSK source will be full k-rank ($k_{\mathbf{S}} = \min(K, N)$) with very high probability. Under these justification for the full k-rank conditions, the data model (5.6) is identifiable provided

$$\min(I, N) + \min(K, N) \geq N + 2.$$

In practice, I will always be less than N. Hence the practical condition is

$$I + \min(K, N) \geq N + 2.$$

This means that for $K \geq N$, just $I = 2$ antennas are sufficient for blind identifiability.

5.3 PARAFAC Receiver

The design of our PARAFAC receiver is based on the principle of trilinear alternating least squares (TALS), a basic algorithm to fit the PARAFAC model (5.6) on the basis of noisy observations (5.4). The idea behind TALS is simple: update one matrix (e.g. $\mathbf{A}$) using least squares (LS) conditioned on previously obtained estimates for the other two matrices (e.g. $\mathbf{H}, \mathbf{B}$), and repeat this in a circular fashion until convergence. TALS guarantees monotonic convergence and exact recovery of $\mathbf{H}, \mathbf{A}, \mathbf{B}$ from $\mathbf{Y}$ under only the model identifiability condition in Section 5.2, in the noiseless case. In the noisy case, TALS often yields the global LS (ML under our AWGN assumption) solution and remains close to the CRB, although infrequent bad runs and convergence to a local minimum cannot be excluded. The reason for the good performance is uniqueness and the strong structure of the PARAFAC model.

The generic TALS algorithm described above must be adapted for our present context, to take advantage of the fact that $\mathbf{A} := \mathbf{P}\mathbf{F}^H$ is known up to row-scaling (which contains the desired CFO information). This yields a performance improvement, as expected, but it also speeds-up convergence, because the number of free variables is significantly reduced.

The appropriately modified TALS algorithm is as follows:

1) Conditional LS Update of $\mathbf{A}$: Recall (5.5), we have

$$\mathbf{Z} = (\mathbf{B} \odot \mathbf{H})\mathbf{A}^T + \mathbf{W}^{\mathbf{Z}} = (\mathbf{B} \odot \mathbf{H})\mathbf{F}^*\mathbf{P} + \mathbf{W}^{\mathbf{Z}}$$

Let $\overline{\mathbf{Z}} := vec(\mathbf{Z})$, $\overline{\mathbf{W}}^{\mathbf{Z}} := vec(\mathbf{W}^{\mathbf{Z}})$, and $\mathbf{p}$ denote the diagonal of $\mathbf{P}$, i.e., $\mathbf{p} = [1, e^{j\phi}, \cdots, e^{j(N-1)\phi}]^T$, it can be verified that

$$\overline{\mathbf{Z}} = (\mathbf{I}_N \odot ((\mathbf{B} \odot \mathbf{H})\mathbf{F}^*))\,\mathbf{p} + \overline{\mathbf{W}}^{\mathbf{Z}},$$

therefore, the conditional LS update of $\mathbf{p}$ is given as

$$\widehat{\mathbf{p}} = \left(\mathbf{I}_N \odot ((\widehat{\mathbf{B}} \odot \widehat{\mathbf{H}})\mathbf{F}^*)\right)^{\dagger} \overline{\mathbf{Z}},$$

where $\widehat{\mathbf{B}}, \widehat{\mathbf{H}}$ denote previously obtained estimates of $\mathbf{B}$, $\mathbf{H}$. We remark that the above update ignores the Vandermonde structure of the $\mathbf{p}$ vector. This will be exploited in the end to yield a CFO estimate. After the $\widehat{\mathbf{p}}$-update, we set $\widehat{\mathbf{A}} := diag(\widehat{\mathbf{p}})\mathbf{F}^H$.

2) Conditional LS Updates of $\mathbf{B}$ and $\mathbf{H}$: The conditional LS updates for $\mathbf{B}$ and $\mathbf{H}$ are

$$\widehat{\mathbf{B}} = \left((\widehat{\mathbf{H}} \odot \widehat{\mathbf{A}})^{\dagger}\mathbf{Y}\right)^T ; \widehat{\mathbf{H}} = \left((\widehat{\mathbf{A}} \odot \widehat{\mathbf{B}})^{\dagger}\mathbf{X}\right)^T$$

where $\widehat{\mathbf{A}}$ is the previously obtained estimate of $\mathbf{A}$.

3) CFO Estimation and Resolution of Permutation Ambiguity:

Note that, in the noiseless case and after convergence of the iterative algorithm, the final $\widehat{\mathbf{p}}$ need not be the diagonal of $\mathbf{P}$. PARAFAC uniqueness only implies that $\widehat{\mathbf{A}} = diag(\widehat{\mathbf{p}})\mathbf{F}^H$ will be the same as the exact $\mathbf{A}$ up to column scaling and permutation. Column order and scaling are both fixed in the modified TALS update of $\mathbf{A}$, but note that the *unconstrained* updates of the row-scaling diagonal may effectively yield a circular shift of the columns of $\mathbf{F}^H$. E.g., if $\widehat{\mathbf{p}}$ is a Vandermonde vector of frequency $-\frac{2\pi}{N}$, all carriers will be shifted by one. Thus the peak of the periodogram of $\widehat{\mathbf{p}}$ will yield the true CFO plus an unknown multiple of the subcarrier spacing. In the noisy case, a CFO estimate $\widehat{\phi}$ is obtained from the peak of the periodogram of $\widehat{\mathbf{p}}$ modulus the frequency spacing, and then $\widehat{\mathbf{P}}$ is constructed using $\widehat{\phi}$ and (5.2). The final $\widehat{\mathbf{A}} := \widehat{\mathbf{P}}\mathbf{F}^H$.

Note that, unlike other existing TALS-like PARAFAC fitting algorithms, this modified TALS automatically column-pairs the final $\widehat{\mathbf{A}}$ to the columns of the exact $\mathbf{A}$, by enforcing the knowledge that the sought $\mathbf{A}$ matrix is a row-scaled inverse FFT matrix (and the assumption that the absolute value of CFO is less than half the subcarrier spacing). This fixes the permutation ambiguity for all three matrices. What remains is the column scaling ambiguity, i.e., the corresponding columns of two of $\mathbf{A}, \mathbf{B}, \mathbf{H}$ can be scaled arbitrarily, and the third can be counter-scaled accordingly.

Since $\mathbf{B} = \mathbf{QS}$, $\widehat{\mathbf{S}}$ cannot be recovered from $\widehat{\mathbf{B}}$ until the row-scaling (due to the diagonal matrix $\mathbf{Q}$) and column scaling (due to blind estimation) ambiguities are removed. For the former, we note that $\mathbf{Q}$ is completely determined by ϕ and the constants K, N, L. Hence we can reconstruct an estimate of $\mathbf{Q}$ from the recovered estimate of ϕ. This allows removal of the row scaling ambiguity. The column-scaling ambiguity can be dealt with using differential encoding and decoding. This yields a blind solution for all quantities of interest. Alternatively we can keep sending 1's along the 1st carrier, i.e., $\mathbf{S}(:,1) = [1, \cdots, 1]^T \in \mathbb{C}^{K \times 1}$, to decouple row-scaling removal from the CFO estimate. This gives somewhat better performance for relatively large

K, N, L, but note that it is not needed for identifiability.

5.4 MIMO-OFDM System

In the context of spatially-multiplexed MIMO-OFDM systems with $I \geq 2$ antennas at the receiver, and $J \geq 2$ antennas at the transmitter in the presence of CFO, the noiseless samples at the ith receiver filter are given by

$$\mathbf{y}_i(k) = \sum_{j=1}^{J} \mathbf{P}_j \mathbf{F}^H \mathbf{H}_{ij} \mathbf{s}_j(k) e^{\sqrt{-1}\phi_j(k(N+L)+L)}. \tag{5.7}$$

Note that the transmitted block depends on j, that is, different data streams can be transmitted through the different transmit antennas. By inserting nulls into the transmitted symbol vectors, different overlapping or non-overlapping subcarrier allocation strategies may be implemented. Again, it can been shown that data model (5.7) can be blindly identified with proper system parameters (N, I, J and K). The results depend strongly on the subcarrier allocation and multiplexing strategy utilized. For full-rate multiplexing with J independent data streams and no subcarrier partinioning among the different streams, then identifiability will hold with very high probability provided

$$N + \min(I, JN) + \min(K, JN) \geq 2JN + 2. \tag{5.8}$$

However, this condition demands small N to be practical. Judicious subcarrier partitioning and stream multiplexing allows much better results. These will be reported elsewhere.

5.5 Simulation Results

In this section, computer simulations are provided to illustrate the performance of the proposed blind PARAFAC receiver for an uncoded SIMO-OFDM system with

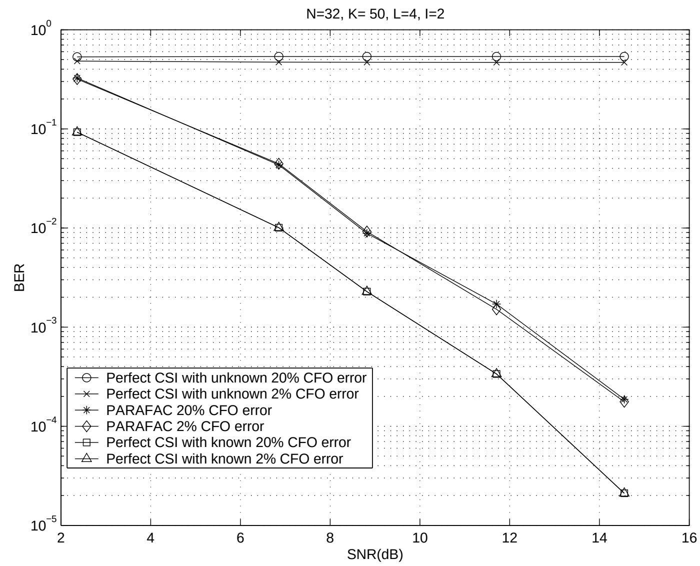

Figure 5.1: BER versus SNR

receive antenna array subject to unknown multipath and carrier frequency offset. In the simulations, we assume differentially encoded BPSK transmission, and a receiver equipped with $I = 2$ or 4 antennas. A total of 1 MHZ available bandwidth is split into $N = 32$ subcarriers. Performance assessed in terms of the average bit-error-rate (BER) versus the average signal-noise-ratio (SNR) for all carriers except the 1st carrier (used to remove the row ambiguity within symbol matrix S). The blind receiver based on reconstructing $\mathbf{Q}$ via estimated $\widehat{\phi}$ will be reported elsewhere. We define the sample SNR as

$$SNR = 10 \log_{10} \frac{\|(\mathbf{H} \odot \mathbf{A})\mathbf{B}^T\|^2}{\|\mathbf{W}\|^2} dB.$$

We have averaged the BER performance over Rayleigh frequency-selective fading channels for CFO $\phi = 0.2\Delta\omega$ or $0.02\Delta\omega$ where $\Delta\omega = \frac{2\pi}{N}$ is the subcarrier spacing. The channel order is $L = 4$, and the multipath channel is assumed constant over $K = 50$ blocks. Our results corroborate the identifiability claims in Section 5.2 and show that the proposed receiver exhibits very good performance.

In contrast to proposed PARAFAC receiver, the CFO compensated MMSE re-

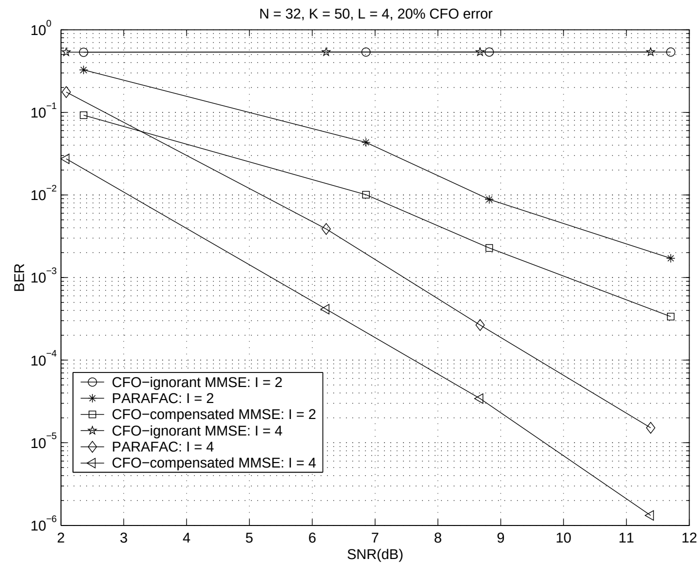

Figure 5.2: BER versus SNR

ceiver assumes perfect knowledge of multipath fading $\mathbf{H}$ and the exact carrier frequency offset ϕ while the CFO-ignorant MMSE receiver only assumes perfect knowledge of multipath fading $\mathbf{H}$ and ignores the existence of CFO ϕ. Let

$$\mathbf{G}_1 := \left((\mathbf{H}\odot\mathbf{A})^H(\mathbf{H}\odot\mathbf{A}) + \frac{1}{SNR}\mathbf{I}\right)$$

$$\mathbf{G}_2 := \left((\mathbf{H}\odot\mathbf{F}^H)^H(\mathbf{H}\odot\mathbf{F}^H) + \frac{1}{SNR}\mathbf{I}\right),$$

then, the CFO-ignorant and CFO-compensated MMSE receivers are given as follows

$$\widehat{\mathbf{B}}_{MMSE}^{CFO-compensated} = sign(Re(\mathbf{Q}^H\left(\mathbf{G}_1^{-1}(\mathbf{H}\odot\mathbf{A})\mathbf{Y}\right)^T)) \quad (5.9)$$

$$\widehat{\mathbf{B}}_{MMSE}^{CFO-ignorant} = sign(Re\left(\mathbf{G}_2^{-1}(\mathbf{H}\odot\mathbf{F}^H)\mathbf{Y}\right)^T) \quad (5.10)$$

Fig. 5.1 plots the performances of proposed PARAFAC receiver and the benchmark MMSE receivers with different carrier frequency offsets. As expected, simulations show the significant performance degradation of the non-blind MMSE receiver due to carrier frequency offsets, even with small CFO. The PARAFAC receiver is only about 2 dB away from the non-blind CFO-compensated MMSE receiver. Furthermore, the performance of the PARAFAC receiver is essentially the same for both 2% and 20% CFO.

In Fig. 5.2, we depict results for $N = 32, K = 50, L = 4$, $\phi = 0.2\Delta\omega$ with 2 and 4 antennas at the receiver. As expected, the performance of both the PARAFAC receiver and the CFO-compensated MMSE receiver improves with the number of receive antennas. On the other hand, the CFO-ignorant MMSE receiver does not improve with more receive antennas, because its performance is CFO-limited.

5.6 Conclusions

We have proposed a direct (semi-)blind receiver and CFO estimator for wireless OFDM systems with receive diversity, based on PARAFAC analysis. The key ideas have been developed for the SIMO case with a single transmit antenna and two or more receive antennas, but the concept extends to spatially multiplexed MIMO systems, possibly with subcarrier allocation, and/or receive diversity created via oversampling. The latter requires some excess bandwidth, which is a form of redundancy. For a minimum of two receive antennas, our CFO solution is the only one to date that is fully blind, in the sense that it does not require excess bandwidth, pilots, repetition, or null subcarriers. The associated hardware cost can be compensated by the performance advantage of 2-branch receive diversity, thus the overall solution is well-motivated. Extensions to spatially multiplexed MIMO OFDM systems will be fully reported elsewhere.

Chapter 6

Conclusions and Future Work

In this chapter, we will briefly summarize the contributions of this thesis and point out directions for future research.

6.1 Summary of Contributions

In this thesis, we have investigated both theory and applications of multi-way analysis. The main contributions of this thesis are as follows:

- We derived the first necessary and sufficient conditions for unique decomposition of certain CP models. These new conditions explain a curious example recently discovered in the multi-way analysis community. The methodology developed herein can be extended to general CP models, and offers the possibility to examine uniqueness of CP solution, albeit the associated necessary and sufficient uniqueness conditions quickly get of hand with increasing problem size.
- Kruskal's Permutation Lemma has been demystified. The new proof should be accessible to a much wider readership than Kruskal's original proof.
- A strong similarity between the conditions for unique decomposition of bilinear

models subject to CM constraints and certain restricted CP models has been pointed out. It is expected that this link will facilitate cross-fertilization and unification of associated uniqueness results.

- We derived the most general identifiability conditions to date for multi-dimensional harmonic retrieval in arbitrary dimensions, with important applications in wireless channel sounding. Our results have subsequently led to the development of an effective algebraic identification algorithm for 2-D harmonic retrieval [37].

- We developed a novel receiver to deal with the blind identification of out-of-cell users in DS-CDMA. This receiver not only detects the in-cell users' data symbols reliably, but also helps identify out-of-cell transmissions, the steering vectors of all active users and spreading codes of out-of-cell users. Simulations show that the proposed identification algorithm remains close to the pertinent asymptotic (symbol-independent) Cramér-Rao bound, which is also included in this thesis.

- We designed an effective blind reception scheme for SIMO/MIMO OFDM subject to unknown frequency offset and multipath.Computer simulation results show that the proposed receiver, based on the theory of multi-way array analysis, can achieve performance that is only a few dB away from the non-blind MMSE receiver, and works well for a wide range of carrier frequency offsets.

6.2 Future Work

Promising thrusts for on-going and future work include the following:

- Algebraic Decomposition Algorithms for General CP Models: A wide variety of algorithms have been developed for computing CP decompositions. These range from eigenvalue-type algebraic techniques to Alternating Least Squares (ALS) algorithms as mentioned earlier. The existing algebraic or direct fitting algorithms only work in restricted cases and ALS is mostly preferred in practice. Algebraic decomposition algorithms that are applicable in more general settings would be very useful in reducing complexity and improving the speed of ALS algorithms via better initializations. One possible lead in this direction is the algebraic solution of simultaneous diagonalization problems.

- Wireless Channel Sounding To further improve the spectral efficiency on the uplink, multiple antennas are utilized at both the BS and the MS, and such design has been incorporated in 3G systems and emerging 4G proposals. The resulting MIMO propagation channels are well modeled using a finite path parameterization, and estimated through double-directional MIMO *channel sounding*. Jointly estimating several multipath signal parameters like azimuth, elevation, delay, and Doppler, all of which can often be viewed as or transformed into frequency parameters, usually gives rise to a multi-dimensional harmonic retrieval problem. The further advance of multi-dimensional harmonic retrieval theory and algorithms with applications in MIMO wireless channel sounding, particularly so-called double-directional channel sounding for 4G should be pursued.

Bibliography

[1] J. Beltran, J. Guiteras, and R. Ferrer, "Three-way multivariate calibration procedures applied to high-performance liquid chromatography coupled with fast-scanning spectrometry (HPLC-FSFS) detection. Determination of aromatic hydrocarbons in water samples," *Analytical Chemistry*, vol. 70, pp. 1949–1955, 1998.

[2] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory", *IEEE Trans. on Circuits and Systems*, 25(9):772-781, Sept. 1978.

[3] H. Bölcskei, "Blind estimation of symbol timing and carrier frequency offset in wireless OFDM systems," *IEEE Trans. Commun.*, 49:988-999, Jun. 2001.

[4] R. Bro, "PARAFAC: tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 149–171, 1997.

[5] R. Bro, *Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications*, *Ph.D. Thesis*, University of Amsterdam, The Netherlands, 1998.

[6] C. Carathéodory, and L. Fejér, "Uber den Zusammenghang der Extemen von harmonischen Funktionen mit ihren Koeffizienten und uber den Picard-Landauschen Satz", *Rendiconti del Circolo Matematico di Palermo*, 32:218–239, 1911.

[7] J. D. Carrol and J. J. Chang, "Analysis of Individual Differences in Multidimensional Scaling via an N-way generalization of "Eckardt-Young" Decomposition", *Psychometrika*, 35:283–319, 1970.

[8] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267–283, 1944.

[9] J. B. Conway, *Functions of One Complex Variable*, 2nd Edition, Graduate Texts in Mathematics vol. 11, Springer-Verlag, 1978.

[10] E. Ebbini, “Region-adaptive motion tracking of speckle imagery”, in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2000)*, 6:2075 - 2078, June 2000, Istanbul, Turkey.

[11] S. A. Fechtel, “OFDM carrier and sampling frequency synchronization and its performance on stationary and mobile channels,” *IEEE Trans. Consummer Electron.*, 46:438-441, Aug. 2000.

[12] M. Gasca, T. Sauer, “Polynomial interpolation in several variables”, *Advances in Computational Mathematics*, 12(4):377-410, 2000.

[13] M. Haardt, and J.H. Nossek, “Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems”, *IEEE Trans. on Signal Processing*, 46(1):161–169, Jan. 1998.

[14] M. Haardt, and J.H. Nossek, “3-D unitary ESPRIT for joint 2-D angle and carrier estimation,” in *Proc. ICASSP-97*, 1:255 -258, 1997.

[15] M. Haardt, C. Brunner, and J.H. Nossek, “Efficient high-resolution 3-D channel sounding,” in *Proc. 48th IEEE Vehicular Technology Conf. (VTC ’98)*, pp. 164-168, Ottawa, Canada, May 1998.

[16] M. Haardt, C. Brunner, and J.H. Nossek, “Joint estimation of 2-D arrival angles, propagation delays, and Doppler frequencies to determine realistic directional simulation models for smart antennas,” in *Proc. IEEE Digital Signal Processing Workshop*, Bryce Canyon National Park, Utah, August 1998.

[17] R. Harshman, “Foundations of PARAFAC procedure: Models and Conditions for an ’Explanatory’ Multi-Mode Factor Analysis”, *UCLA working papers in phonetics*, 16:1–84, 1970.

[18] R. A. Harshman, “Determination and proof of minimum uniqueness conditions for PARAFAC1,” *UCLA Working Papers in Phonetics*, vol. 22, pp. 111–117, 1972.

[19] R. A. Harshman, “PARAFAC2: mathematical and technical notes,” *UCLA Working Papers in Phonetics*, vol. 22, pp. 33–44, 1972.

[20] R. A. Harshman, and M. E. Lundy, “The PARAFAC model for three-way factor analysis and multidimensional scaling,” in *Research Methods for Multimode Data*

Analysis, H. G. Law, C. W. Jr. Snyder, J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.

[21] Y. Hua, "High Resolution Imaging of Continuously Moving Object Using Stepped Frequency Radar," *Signal Processing*, 35:33–40, Jan. 1994.

[22] T. Jiang, N. D. Sidiropoulos, and J.M.F. ten Berge, "Almost Sure Identifiability of Multi-dimensional Harmonic Retrieval," *ICASSP 2001*, Salt Lake City, Utah, May 7-11, 2001.

[23] T. Jiang and N. D. Sidropoulos "Blind Identification of Out-of-cell Users in DS-CDMA: An Algebraic Approach," *ICASSP 2002*, Orlando, FL, May. 13 - 17, 2002.

[24] T. Jiang and N. D. Sidropoulos "A Direct Semi-Blind Receiver for SIMO and MIMO OFDM Systems Subject to Frequency Offset," in *Proc. SPAWC 2003*, Rome, Italy, Jun. 15 - 18, 2003.

[25] T. Jiang, N. D. Sidiropoulos, and J. M. F. ten Berge, "Almost Sure Identifiability of Multidimensional Harmonic Retrieval," *IEEE Transactions on Signal Processing*, 49(9): 1849-1859, Sep. 2001

[26] T. Jiang and N. D. Sidiropoulos, "Blind Identification of Out-of-cell Users in DS-CDMA," *EURASIP Journal on Applied Signal Processing*, special issue on Advances in Smart Antennas, (invited submission, May 2003)

[27] T. Jiang and N. D. Sidiropoulos, "A Direct Semi-Blind Receiver for SIMO and MIMO OFDM Subject to Frequency Offset," *IEEE Transactions on Signal Processing*, (in preparation, 2003)

[28] T. Jiang and N. D. Sidiropoulos, "Kruskal's Permutation Lemma and the Identification of CANDECOMP/PARAFAC and Bilinear Models with Constant Modulus Constraints," *IEEE Transactions on Signal Processing*, (submitted, May 2003)

[29] W. P. Krjnen, *The Analysis of three-way arrays by constrained PARAFAC methods*, Leiden, DSWO Press, 1993.

[30] J. B. Kruskal, "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics," *Linear Algebra and Its Applications*, 18:95–138, 1977.

[31] J. B. Kruskal, "Rank, decomposition, and uniqueness for 3-way and N-way arrays," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, Eds.

[32] J. K. Lee, R. T. Ross, S. Thampi, and S. E. Leurgans, "Resolution of the properties of hydrogen-bonded tyrosine using a trilinear model of fluorescence," *Journal of Physical Chemistry*, vol. 96, pp. 9158-6192, 1992.

[33] A. Leshem, N. Petrochilos, and A. van der Veen, "Finite Sample Identifiability of Multiple Constant Modulus Sources ," *IEEE Sensor Array and Multichannel Signal Processing Workshop Proceedings*, 408–412, 2002.

[34] J. Li, P. Stoica, and D. Zheng, "An efficient algorithm for two-dimensional frequency estimation", *Multidimensional Systems and Signal Processing*, 7(2):151–178, April 1996.

[35] X. Liu and N. Sidiropoulos, "PARAFAC methods for blind beamforming: multilinear ALS performance and CRB," in *Proc ICASSP 2000*, Istanbul, Turkey, Jun. 5–9, 2000, pp. 3128–3131.

[36] X. Liu and N. D. Sidiropoulos, "Cramer-Rao Lower Bounds for Low-rank Decomposition of Multidimensional Arrays," *IEEE Trans. on Signal Processing*, 49:2074–2086, 2001.

[37] X. Liu, N. D. Sidiropoulos, and A. Swami, "High Resolution Localization and Tracking of Multiple Frequency Hopped Signals," *IEEE Trans. on Signal Processing*, 50:891–901, Apr. 2002.

[38] H. Liu and Tureli, "A high-efficiency carrier estimator for OFDM communications," *IEEE Commun. Lett.*, 2:104-106, Apr. 1998.

[39] H. Liu and G. Xu, "A subspace method for signature waveform estimation in synchronous CDMA systems," *IEEE Trans. on Commun.*, pp. 1346-1354, 1996.

[40] R. Lorentz "Multivariate Birkhoff Interpolation", *Lecture Notes in Mathematics*, Springer-Verlag, 1992.

[41] B. Lu and X. D. Wang, "Bayesain blind turbo receiver for coded OFDM systems with frequency offset and frequency-selective fading," *IEEE J. Select. Areas Commun.*, 19:2516-2527, Dec. 2001.

[42] X. L. Ma, C. Tepedelenlioglu, G. B. Giannakis, and S. Barbarossa, "Non-data-aided carrier offset estimators for OFDM with null subcarriers: identifiability, algorithms, and performance," *IEEE J. Select. Areas Commun.*, 19:2504-2515, Dec. 2001.

[43] C. D. Meyer, "Matrix Analysis and Applied Linear Algebra," *Society for Industrial and Applied Mathematics*, 2000.

[44] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, 42:2908-2914, Oct. 1994.

[45] V.F. Pisarenko, "The retrieval of harmonics from a covariance function", *Geophys. J. Roy. Astron. Soc.*, 33:347–366, 1973.

[46] T. Sauer , Y. Xu "On Multivariate Hermite Interpolation", *Advances in Computational Mathematics*, 4:207-259, 1995.

[47] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, 45:1613-1621, Dec. 1997.

[48] P. H. Schonemann, "An algebraic solution for a class of subjective metrics models," *Psychometrika*, vol. 37, pp. 441, 1972.

[49] V. Shtrom, "CDMA vs. OFDM in Broadband Wireless Access - The Fundamental Characteristics of OFDM Make It Ideally Suited for Broadband Data," *Broadband Wireless Online.*, http://www.shorecliffcommunications.com/magazine/index.asp, Vol 5, No 3, July/Aug. 2002.

[50] N.D. Sidiropoulos, "Generalizing Carathéodory's Uniqueness of Harmonic Parameterization to N Dimensions", *IEEE Trans. Information Theory*, 47:1687-1690, May 2001.

[51] N.D. Sidiropoulos, and R. Bro, "On the Uniqueness of Multilinear Decomposition of N-way Arrays", *J. Chemometrics*, special cross-disciplinary issue on multi-way analysis, 14(3):229-239, May 2000.

[52] N.D. Sidiropoulos, and R. Bro, "On Communication Diversity for Blind Identifiability and the Uniqueness of Low-Rank Decomposition of N-way Arrays", in *Proc. ICASSP2000*, June 5-9, 2000, Istanbul, Turkey.

[53] N.D. Sidiropoulos, R. Bro, and G.B. Giannakis, "Parallel Factor Analysis in Sensor Array Processing", *IEEE Trans. Signal Processing*, 48(8):2377–2388, Aug. 2000.

[54] N.D. Sidiropoulos, G.B. Giannakis, and R. Bro, "Blind PARAFAC Receivers for DS-CDMA Systems", *IEEE Trans. Signal Processing*, 48(3):810–823, Mar. 2000.

[55] N.D. Sidiropoulos, and X. Liu, "Identifiability Results for Blind Beamforming in Incoherent Multipath with Small Delay Spread", *IEEE Trans. Signal Processing*, 49(1):228-236, Jan. 2001.

[56] P. Spasojević, X. Wang, and A. Høst-Madsen, "Nonlinear Group-blind Multiuser Detection," *IEEE Trans. on Commun.*, vol. 49, pp. 1631-1641, Sept 2001.

[57] D. Starer and A. Nehorai, "Passive Localization of Near-Field Sources by Path Following", *IEEE Trans. Signal Processing*, 42(3):677-680, March 1994.

[58] P. Stoica, and R. Moses, *Introduction to Spectral analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 1997.

[59] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, 37(5):720-741, May 1989.

[60] P. Stoica, and T. Söderström, "Parameter identifiability problem in signal processing", *IEE Proc., Pt F–Radar and Sonar Navig*, vol. 141:133-136, 1994.

[61] S. Talwar, M. Viberg, and A. Paulraj, "Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array - Part I: Algorithms," *IEEE Trans. Signal Processing.*, 44:1184–1196, May. 1996.

[62] J. M. F. ten Berge and N. D. Sidiropoulos, "On Uniqueness in CANDECOMP/PARAFAC," *Psychometrika*, 67, 2002.

[63] U. Tureli, H. Liu, and M. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. Commun.*, 48:1459-1461, Sept. 2000.

[64] U. Tureli, D. Kivanc, and H. Liu "Experimental and analystical studies on a high-resolution OFDM carrier frequencyoffset estimator," *IEEE Trans. Veh. Technol.*, 50:629-643, Mar. 2001.

[65] A. van der Veen, "Asymptotic properties of the algebraic constant modulus algorithm," *IEEE Trans. Signal Processing.*, 49:1796–1807, Aug. 2001.

[66] M.C. Vanderveen, A-J. van der Veen, and A.J. Paulraj, "Estimation of Multipath Parameters in Wireless Communications", *IEEE Trans. Signal Processing*, 46(3):682–690, Mar. 1998.

[67] H. Yang, and Y. Hua, "On rank of block Hankel Matrix for 2-D frequency detection and estimation", *IEEE Trans. Signal Processing* , vol. 44(4):1046-1048, Apr. 1996.

[68] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D Angle Estimation with Rectangular Arrays in Element Space or Beamspace via Unitary ESPRIT", *IEEE Trans. Signal Processing*, 44(2):316-328, Feb. 1996.

[69] M. D. Zoltowski, and C. P. Mathews, "Real-time Frequency and 2-D Angle Estimation with Sub-Nyquist Temporal Sampling", *IEEE Trans. Signal Processing*, 42(10):2781-2794, Oct. 1994.