

Research Reports from the Communications Research Laboratory  
at Ilmenau University of Technology

*João Paulo Carvalho Lustosa da Costa*

# Parameter Estimation Techniques for Multi-Dimensional Array Signal Processing



Fakultät für Elektrotechnik und Informationstechnik  
der Technischen Universität Ilmenau

# PARAMETER ESTIMATION TECHNIQUES FOR MULTI-DIMENSIONAL ARRAY SIGNAL PROCESSING

*João Paulo Carvalho Lustosa da Costa, M.Sc.,  
22.05.1981, Fortaleza, Ceará, Brasilien*

Dissertation zur Erlangung des  
akademischen Grades Doktor-Ingenieur (Dr.-Ing.)

Anfertigung im: Fachgebiet Nachrichtentechnik  
Institut für Informationstechnik  
Fakultät für Elektrotechnik und Informationstechnik

Gutachter: Univ.-Prof. Dr.-Ing. Martin Haardt  
Dr. Rodrigo de Lamare, University of York  
Univ.-Prof. Dr.-Ing. Gerald Schuller

Vorgelegt am: 17.12.09

Verteidigt am: 01.03.10



## ACKNOWLEDGEMENTS

It is a matter of honor and justice to acknowledge the inestimable support of many diligent people during the development of this thesis.

First I would like to thank God, who gave me strength and insights to cope with my PhD course.

Also I leave my immense gratitude to the Brazilian and German people represented respectively by the Brazilian Army and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Deutscher Akademischer Austausch Dienst (DAAD). Their taxes financed my research during my whole stay in Germany.

An invaluable character is my PhD tutor Univ.-Prof. Dr.-Ing. Martin Haardt, who directed me to the field of the multilinear algebra, particularly applied to multi-dimensional array signal processing. This long period working together with him made me a much better researcher mainly in terms of critical viewing, perfectionism, and patience. Moreover, I would like to thank him also for accepting me as a member of his high-level team at the Communications Research Laboratory (CRL), and also for all the several challenging tasks that he trusted to me during these years, such as research assistant of the NORTEL project, reviewer of international conferences, tutor of an advanced seminar of Adaptive & Array Signal Processing (AASP), assistant and lecturer of seminars during the subjects Mobile Communications and AASP.

I express my thanks also to Dr.-Ing. Giovanni Del Galdo and Florian Roemer, who were essential during my PhD. The first one supported me with the preparation of my research project accepted by CNPq and DAAD, as well as with important survival and scientific hints. In general, his guidance at the beginning of my PhD was decisive. The second one never hesitated to share his knowledge and experience obtained through the many years working at the CRL. Additionally, his unconditional helpfulness and psychological support were simply crucial, mainly to handle hard bureaucratic problems in German language and also to discuss about multi-dimensional problems.

Another very essential character in my PhD is Prof. Dr. José Antonio Apolinário Jr. from the Instituto Militar de Engenharia (IME). Without his support, the nine month extension provided by the Brazilian Army would not be possible. My thanks also go to Prof. Dr. Maria Thereza Miranda Rocco Giraldi from IME for accepting being my supervisor for the Brazilian Army.

The other members of CRL deserve equally to be referred here not only for their friendship, but also for their suggestions given during the meetings once a week. Therefore, I extend my thanks to Stanislaw Gorlow, Martin Weis, Cristian Kaes, Ulrike Korger, Anja Grosch, Dr.-Ing.

---

Martin Fuchs-Lautensack, Marko Hennhoefer, Dominik Schulz, Liane Grobe, Bin Song, Nuan Song, Dr.-Ing. Mike Wolf, Jianhui Li, Marko Milojevic, Veljko Stankovic, Jianshu Zhang, and Mrs. Christina Patotschka. Here I leave special thanks to Wolfgang Erdtmann, a key member of the Lab. Due to his technical support, cooperativeness and efficiency, I and the other members of our lab can develop more plenary our research.

My talented sister Nádia Carvalho Lustosa da Costa, who accompanied me during these years in Germany and proofread my CNPq reports in Portuguese, also deserves my special gratitude. I am eternally thankful to my parents Eugênio Pacelli Lustosa da Costa and Rosa Maria Carvalho Lustosa da Costa, whom I can count and trust unconditionally.

During my six month German course supported by DAAD in Leipzig, I got to know several friends that were also PhD students and researchers of other fields and countries, and that supported me many times. Therefore, my sincere gratitude for the precious support of Carlos Henrique Cardonha, Dr. iur. Juliano Zaiden Benvindo, Daniela Azoubel, Miguel Montero-Baker, Napoleon Molina, Brand Arenari, Dr. Lucia Luque Nadal, Ahmad Hammdy, and Reichiele Vanessa Vervloet de Carvalho.

To be licensed by the Brazilian Army, the support of other members of the Brazilian Army was required. Therefore, I would like to sincerely thank general João Carlos Pedroza Rêgo, colonel R/1 Sussumo Ohashi Suzukawa, first lieutenant Augusto José Batista da Silva, first lieutenant Alexandre César Nass, colonel R/1 Macarino Bento Garcia de Freitas, first lieutenant Danieli Dyba Kellner, and first sergeant José Ferreira Silva. Also during this period, the encouragement and the assistance of Eliégia Freitas Rodrigues was extremely important.

For the hospitality and also for the permanent support during the completion of my mission in Germany, I also express my sincere gratitude to colonel Marcio Roland Heise, colonel Klauss Erich Klein, lieutenant Ilmar Rainoldo Schuck, lieutenant Claudino Rachor, lieutenant Lauro Liberto Wachholz, and Amely Duetthorn of the Brazilian Army Attaché at the Brazilian Embassy in Berlin, as well as to my uncle João Eudes Brownsville. Here my thanks also go to my uncle and colonel Fernando Sávio Parente de Carvalho, and also to my brother and captain Giovanni Pacelli Carvalho Lustosa da Costa due to their constant support and example.

I would like to thank the members of the Brazilian Army, who assisted me in several bureaucratic problems before and during my stay in Germany, specifically, the members of the Seção de Redes Telefônicas (SRTel) do Quartel General do Exército (QGEx) subordinate to Centro Integrado de Telemática do Exército (CITEx), Departamento de Ciência Tecnologia (DCT), Gabinete do Comandante do Exército (GAB CMT EX), Estado-Maior do Exército (EME), and Departamento Geral de Pessoal (DGP).

I would like to thank also Prof. Dr. Ricardo Zelenovsky and Prof. Dr. Geovany Araujo Borges from Universidade de Brasília (UnB) for their assistance. I also would like to recognize the support of my uncle and newspaper columnist Francisco José Lustosa da Costa, who included my achievements and of our whole family in his daily column. Similarly, I am grateful to my uncle and writer José Elcias Lustosa da Costa for his motivating support. For partially supporting the publication of this thesis, I would like to thank my grandmother

Maria Dolores Lustosa da Costa.

My inestimable gratitude goes also to Dr. Rodrigo de Lamare for always being available and for providing intelligent suggestions about my PhD thesis, as well as for accepting being the reviewer of this thesis. In the same way, I am very thankful to Univ.-Prof. Dr.-Ing. Gerald Schuller for accepting being my reviewer and for his important suggestions, which improved the readability as well as the presentation of the thesis. In addition, I am grateful to Dr.-Ing. Karl Schran, who kindly accepted to be a member of my Rigorosum. Also Mrs. Heike Rimmrott of the Faculty of Electrical Engineering and Information Technology at the Ilmenau University of Technology deserves my gratitude for her courtesy and kindness, and for providing an excellent service for the students.

I am also grateful to the Brazilian community and its aggregates of other nationalities in Ilmenau for their cooperativeness and friendship. In particular, my thanks go to Andrea Iffland, Wolfgang Bolduan, Traute Angélica Hiltel, Carmen Ruiz Ruiz, Yahia Cheringen, Guilherme Zandona, Liele Silveira Costa, Sliman S. Mukarker, Denny Horn, Ricardo Nuber, Tiago Scholz, Rafael dos Santos Alves Figueiredo Brasil, Mariana Pimenta Alves, Claysson Bruno Santos Vimieiro, Priscila Barbara Marques Heleno, Sara Del Vecchio, Israel da Costa, Alexandre Tsuno, and William E. Blunk.

Finally, my affectionate and sincere thanks go to my beloved girlfriend Lola Sabirova, whose comprehension, care and love allowed me to improve my life significantly in all dimensions.





# ABSTRACT

In the literature, one-dimensional array signal processing techniques are extensively used in a variety of applications including radar, mobile communications, sonar, and seismology. One class of array signal processing techniques is applied to estimate parameters such as the number of signal components, also known as model order, the direction of arrival, the direction of departure, the time of direction of arrival and the Doppler frequency.

By taking into account only one dimension, the problem is seen from just one perspective, i.e., one projection. Consequently, parameters cannot be estimated properly for certain scenarios. To handle that, multi-dimensional array signal processing, which considers several dimensions, is studied. These dimensions can correspond to time, frequency, or polarization, but also spatial dimensions such as one- or two-dimensional arrays at transmitter and receiver. With multi-dimensional array signal processing, it is possible to estimate parameters using all the dimensions jointly, even if they are not resolvable for each dimension separately. Moreover, by considering all dimensions jointly, an improved accuracy, reliability and robustness can be achieved. Another important advantage of using tensors is the identifiability, since with tensors the typical rank can be much higher than using matrices. Here we focus particularly on the development of techniques for the estimation of the model order and the associated spatial frequencies.

The estimation of the model order, also known as the number of principal components, has been investigated in several science fields, and usually model order selection schemes are proposed only for specific scenarios in the literature. Therefore, as a first important contribution, we propose the one-dimensional model order selection scheme called Modified Exponential Fitting Test (M-EFT), which outperforms all the other schemes for scenarios involving white Gaussian noise. Additionally, we propose improved versions of the Akaike's Information Criterion (AIC) and Minimum Description Length (MDL).

As shown here the multi-dimensional structure of the data can be taken into account in order to improve the estimation of the model order. As an example of such improvement, we propose the  $R$ -dimensional Exponential Fitting Test ( $R$ -D EFT) for multi-dimensional applications, where the noise is additive white Gaussian. The  $R$ -D EFT successfully outperforms the M-EFT confirming that even the technique with best performance can be improved by taking into account the multi-dimensional structure of the data. In addition, we also extend our modified versions of AIC and MDL to their respective multi-dimensional versions  $R$ -D AIC and  $R$ -D MDL. For scenarios with colored noise, we propose the multi-dimensional model order selection technique called closed-form PARAFAC based model order selection (CFP-MOS) scheme.

---

Once the model order is correctly obtained, the next step is the extraction of the desired parameters from the main components, in our case the estimation of the spatial frequencies. For this task, we have proposed the closed-form PARAFAC based parameter estimation (CFP-PE) scheme. Besides the advantage of being a closed-form scheme, the CFP-PE is more robust than other schemes in the literature with respect to the violation of the narrowband assumption and arrays with positioning errors. This is due to the fact that the CFP-PE decouples the dimensions, and the error of one dimension does not affect the other dimensions. Due to the fabrication tolerances, the array elements in a uniform rectangular array may not be exactly equi-spaced, which results in slight positioning errors violating the assumed data model. Therefore, in practical applications, the CFP-PE is very appealing.

For colored noise applications, subspace prewhitening schemes are necessary to avoid a degradation in the estimation of the spatial frequencies. Therefore, in case of a special noise correlation model, we propose the deterministic prewhitening scheme, which takes into account the structure of the colored noise in order to increase the Signal-to-Noise Ratio (SNR) inversely proportional to the noise correlation level. For high correlation levels, the deterministic prewhitening is significantly superior to the stochastic prewhitening in the literature.

For certain biomedical applications, like electroencephalography (EEG), as well as for certain communications applications with multiple antennas, i.e., Multiple Input Multiple Output (MIMO) systems, the noise correlation possesses a specific multi-dimensional structure that can be exploited to improve prewhitening schemes further. In this sense, a multi-dimensional prewhitening scheme is required for such problems, and we propose the Sequential Generalized Singular Value Decomposition (S-GSVD). When the knowledge of the noise statistics information is not available, we propose the Iterative S-GSVD, which is an excellent alternative, since its estimation is comparable to the S-GSVD with knowledge of the noise statistics information for scenarios with high and intermediate SNR regimes.

## ZUSAMMENFASSUNG

Methoden der Antennenarraysignalverarbeitung stellen ein sehr wichtiges und bedeutsames Forschungsgebiet dar. Insbesondere spielen hochauflösende Parameterschätzverfahren eine entscheidende Rolle in Anwendungsgebieten wie RADAR, SONAR, Mobilkommunikation, Biosignalverarbeitung oder Seismologie. Dabei muss einerseits die Modellordnung, also die Anzahl der dominanten Komponenten im gemessenen Signal, bestimmt werden. Andererseits sind häufig auch Parameter wie räumliche Winkel der einfallenden sowie der abgestrahlten Wellen, Signallaufzeit oder Dopplerverschiebung zu schätzen.

Im Allgemeinen sind die zugrunde liegenden Signale mehrdimensional. Betrachtet man diese Dimensionen nicht gemeinsam sondern separat, steht jeweils nur eine Projektion der Daten in diese Dimension zur Verfügung. Das führt in vielen Szenarien zu einer unzureichenden Schätzgenauigkeit, so dass die Parameter nicht mehr korrekt erfasst werden können. Deshalb sind mehrdimensionale Arraysignalverarbeitungsalgorithmen, die alle natürlichen Dimensionen des Signals gemeinsam behandeln und dabei dessen Struktur ausnutzen, ein bedeutsames Werkzeug und Gegenstand der aktuellen Forschung. Diese Dimensionen können beispielsweise korrespondieren zu Zeit, Frequenz oder Polarisierung, aber auch räumlichen Dimensionen wie etwa bei ein- oder zweidimensionalen Antennenarrays am Sender und am Empfänger. Mehrdimensionale Parameterschätzverfahren sind in der Lage, all diese Parameter gemeinsam zu schätzen, sogar in solchen Fällen, wo diese in den einzelnen Dimensionen für sich nicht auflösbar sind. Darüber hinaus sind durch das Ausnutzen der mehrdimensionalen Struktur wesentliche Verbesserungen der Schätzgenauigkeit, der Zuverlässigkeit, sowie der Robustheit der Algorithmen möglich. Schließlich besteht ein weiterer wichtiger Vorteil in der verbesserten Identifizierbarkeit. Da im mehrdimensionalen Fall der typische Rang wesentlich höher werden kann, sind im Allgemeinen mehr Quellen auflösbar. In dieser Arbeit konzentrieren wir uns hauptsächlich auf zwei Arten von Parametern: Die Modellordnung sowie die zugehörigen räumlichen Frequenzen.

Die Schätzung der Modellordnung, auch bekannt als die Anzahl der Hauptkomponenten, wurde in unterschiedlichen wissenschaftlichen Gebieten eingehend untersucht. Deshalb finden sich in der Literatur für spezifische Szenarien zugeschnittene Verfahren zur Modellordnungsschätzung. Ein erster wichtiger Beitrag dieser Arbeit ist daher ein generisches Schätzverfahren für die Modellordnung, der so genannte *Modified Exponential Fitting Test* (M-EFT). Dieser übertrifft alle anderen bekannten Schätzverfahren in dem Fall, dass ausschließlich additives Gaußverteiltes weißes Rauschen das gewünschte Signal stört. Darüber hinaus schlagen wir Verbesserungen der bekanntesten Modellordnungsschätzer vor, dem Akaike's Information Criterion (AIC) sowie der Minimum Description Length (MDL).

---

Im mehrdimensionalen Fall lässt sich die Modellordnungsschätzung erheblich verbessern, wenn man dafür die mehrdimensionale Struktur explizit ausnutzt. Für Szenarien mit additivem Gaußverteilterm weißen Rauschen schlagen wir dafür den *R-D Exponential Fitting Test* (R-D EFT) vor. Wir zeigen numerisch, dass der R-D-EFT-Algorithmus den M-EFT-Algorithmus in Schätzgenauigkeit übertrifft. Dies bestätigt, dass selbst das beste bekannte Verfahren verbessert werden kann, wenn man die mehrdimensionale Struktur der Daten explizit ausnutzt. Im Anschluss werden auch die verbesserten Versionen des AIC- und des MDL-Verfahrens auf die mehrdimensionalen Verfahren *R-D AIC* und *R-D MDL* erweitert. Da *R-D EFT* weißes Rauschen voraussetzt untersuchen wir schließlich noch den Fall des gefärbten Rauschens. Hier schlagen wir das *closed-form PARAFAC-based model order selection scheme* (CFP-MOS) vor und zeigen numerisch, dass es bestehende Verfahren übertrifft.

Sobald die Modellordnung bestimmt wurde, können im nächsten Schritt weitere gewünschte Parameter aus den dominanten Komponenten bestimmt werden. Für diese Aufgabe schlagen wir das *closed-form PARAFAC based parameter estimation* (CFP-PE) *scheme* vor. Das Verfahren ist nicht nur geschlossen sondern insbesondere robuster als andere Verfahren gegen Modellierungsfehler. Diese verbesserte Robustheit ergibt sich, da CFP-PE die verschiedenen Dimensionen entkoppelt voneinander betrachtet, so dass derartige Fehler in einer Dimension die anderen Dimensionen unberührt lässt. Solche Modellierungsfehler geschehen beispielsweise, wenn die Phasenzentren der Elemente der Antennenarrays durch Toleranzen in der Herstellung nicht exakt positioniert sind oder auch dadurch, dass die Bandbreite im Datenmodell infinitesimal klein angenommen wird. Daher ist CFP-PE für praktische Anwendungen besonders gut geeignet.

Im Fall von farbigem Rauschen leidet die Schätzgenauigkeit aller Parameterschätzverfahren, wodurch die Anwendung von *Prewhitening*-Techniken erforderlich wird. Der Grund für diese Verschlechterung ist, dass sich die Rauschleistung des gefärbten Rauschens stärker in den dominanten Komponenten, also im Signalraum, konzentriert. Für ein spezifisches Rauschkorrelationsmodell schlagen wir, als Gegenstück zu existierenden stochastischen Prewhitening-Verfahren, ein deterministisches Prewhitening vor, welches die spezielle Struktur des gefärbten Rauschens explizit ausnutzt. Damit wird eine Verbesserung des Signal-zu-Rausch-Verhältnisses (SNR) erreicht, welches umgekehrt proportional zur Rauschkorrelation wächst. Dies impliziert, dass das deterministische Prewhitening für sehr stark korreliertes Rauschen ein wesentlich besseres Ergebnis als das stochastische Prewhitening erzielt.

Für bestimmte biomedizinische Signale wie Elektroenzephalogramme (EEG) sowie für bestimmte Kommunikations-Anwendungen mit Mehrantennen-Systemen besitzt das Rauschen eine ganz spezifische mehrdimensionale Struktur. Diese kann ausgenutzt werden um das Prewhitening weiter zu verbessern, wodurch jedoch ein mehrdimensionales Prewhitening-Verfahren erforderlich wird. Unser Vorschlag dazu ist die *Sequential Generalized Singular Value Decomposition* (S-GSVD). Wie gezeigt wird, ist die S-GSVD der *Generalized Higher Order SVD* (HOSVD) deutlich überlegen. Für den Fall, dass die Rauschstatistik nicht bekannt ist, schlagen wir die iterative S-GSVD vor. Wir demonstrieren numerisch, dass für mittleres bis hohes SNR die erreichbare Genauigkeit der iterativen S-GSVD vergleichbar ist

---

mit der S-GSVD mit bekannter Rauschstatistik.



# CONTENTS

<b>Acknowledgements</b> . . . . .	i
<b>Abstract</b> . . . . .	v
<b>Zusammenfassung</b> . . . . .	vii
<b>Contents</b> . . . . .	xi
<b>1. Introduction</b> . . . . .	1
1.1 Overview and Contributions . . . . .	3
1.2 Observation . . . . .	6
<b>2. Multi-dimensional and Matrix-based Model Order Selection</b> . . . . .	7
2.1 Matrix-based Model Order Selection . . . . .	11
2.1.1 Goal and Problem . . . . .	12
2.1.1.1 Eigenvalue based Estimation of the Model Order . . . . .	15
2.1.2 Previous Approaches: Review of the State-of-the-art Matrix-based Model Order Estimation Techniques . . . . .	17
2.1.2.1 Subspace based Approaches . . . . .	22
2.1.2.2 Data based Approach via Measured Noise Power . . . . .	24
2.1.3 Proposed Approaches . . . . .	25
2.1.3.1 1-D AIC and 1-D MDL . . . . .	25
2.1.3.2 Modified Exponential Fitting Test (M-EFT) . . . . .	26
2.1.3.3 Applying Forward-Backward Averaging (FBA) . . . . .	29
2.1.4 Evaluation of Approaches . . . . .	31
2.1.5 Main Results of the Section . . . . .	35
2.2 Multi-dimensional Model Order Selection . . . . .	45
2.2.1 Goal and Problem . . . . .	46
2.2.2 Previous Approach: CORCONDIA . . . . .	49
2.2.3 Proposed Approaches . . . . .	51
2.2.3.1 $R$ -dimensional Exponential Fitting Test ( $R$ -D EFT) . . . . .	52
2.2.3.2 $R$ -D AIC and $R$ -D MDL . . . . .	55
2.2.3.3 Exponentiation of the Global Eigenvalues . . . . .	56
2.2.3.4 Threshold CORCONDIA (T-CORCONDIA) . . . . .	56

---

2.2.3.5	Closed-Form PARAFAC based Model Order Selection (CFP-MOS) . . . . .	59
2.2.3.6	Applying Forward-Backward Averaging (FBA) . . . . .	62
2.2.4	Evaluation of Approaches . . . . .	64
2.2.4.1	MOS for Special Array Sizes . . . . .	68
2.2.5	Main Results of the Section . . . . .	69
2.3	Main Results of the Chapter . . . . .	71
<b>3.</b>	<b><i>R</i>-D Parameter Estimation via Closed-Form PARAFAC . . . . .</b>	<b>77</b>
3.1	Goal and Problem . . . . .	79
3.2	Previous Approaches: MALS based and <i>R</i> -D ESPRIT-type Approaches . . . .	82
3.3	Proposed Approach: Closed-Form PARAFAC based Approach . . . . .	84
3.3.1	Merging Dimensions . . . . .	84
3.3.2	Least Squares Khatri-Rao Factorization (LSKRF) . . . . .	85
3.3.3	Estimation of Spatial Frequencies via Peak Search (PS) . . . . .	86
3.3.4	Estimation of Spatial Frequencies via Shift Invariance (SI) . . . . .	86
3.4	Evaluation of Approaches . . . . .	87
3.5	Main Results of the Chapter . . . . .	90
<b>4.</b>	<b>Multi-dimensional and Matrix-based Subspace Prewhitening . . . . .</b>	<b>97</b>
4.1	Matrix-based Prewhitening . . . . .	99
4.1.1	Goal and Problem . . . . .	101
4.1.2	Previous Approach: Stochastic Matrix-based Prewhitening . . . . .	101
4.1.2.1	Computation of the Stochastic Prewhitening Matrix . . . . .	101
4.1.2.2	Stochastic Prewhitening . . . . .	102
4.1.3	Proposed Approach: Deterministic Matrix-based Prewhitening . . . . .	102
4.1.4	Estimation of the Correlation Coefficient . . . . .	106
4.1.5	Evaluation of Approaches . . . . .	107
4.1.6	Main Results of the Section . . . . .	110
4.2	Multi-dimensional Prewhitening . . . . .	112
4.2.1	Goal and Problem . . . . .	112
4.2.2	Previous Approaches: Stochastic Prewhitening and Generalized HOSVD	112
4.2.2.1	Generalized HOSVD of [97] . . . . .	112
4.2.2.2	Prewhitening matrix estimation . . . . .	113
4.2.2.3	Matrix prewhitening . . . . .	114
4.2.3	Proposed Approaches . . . . .	114
4.2.3.1	Prewhitening Correlation Factors Estimation . . . . .	114
4.2.3.2	Tensor Prewhitening Scheme: <i>n</i> -mode Products using Matrix Inversions . . . . .	114
4.2.3.3	Tensor Prewhitening Scheme: S-GSVD . . . . .	115
4.2.4	Evaluation of Approaches . . . . .	117
4.2.5	Main Results of the Section . . . . .	127

---



4.3	Iterative Multi-dimensional Preshwhitening . . . . .	128
4.3.1	Goal and Problem . . . . .	128
4.3.2	Previous Approach . . . . .	128
4.3.3	Proposed Approach: Iterative S-GSVD (I-S-GSVD) . . . . .	128
4.3.4	Evaluation of Approaches . . . . .	129
4.3.5	Main Results of the Section . . . . .	130
4.4	Main Results of the Chapter . . . . .	135
<b>5.</b>	<b>Concluding remarks . . . . .</b>	<b>137</b>
5.1	Outlook and Future Works . . . . .	138
<b>Appendix A. Probabilities to evaluate the model order selection schemes . . . . .</b>		<b>143</b>
<b>Appendix B. 1-D Information Theoretic Criteria - Explanation . . . . .</b>		<b>147</b>
<b>Appendix C. Properties of Eigenvalues with a Wishart Distribution . . . . .</b>		<b>149</b>
<b>Appendix D. Modified Exponential Fitting Test II - Derivation . . . . .</b>		<b>153</b>
D.1	Review of the EFT . . . . .	154
D.2	Derivation of the M-EFT II . . . . .	155
<b>Appendix E. Tensor Calculus . . . . .</b>		<b>159</b>
E.1	$r$ -mode vectors and $r$ -mode unfoldings . . . . .	160
E.2	Tensor operators . . . . .	161
<b>Appendix F. Important Definitions and Relations for Tensors . . . . .</b>		<b>165</b>
<b>Appendix G. Generalized SVD (GSVD) . . . . .</b>		<b>171</b>
<b>Appendix H. Applications Considering Kronecker Colored Noise . . . . .</b>		<b>173</b>
<b>Appendix I. Multi-dimensional Colored Noise . . . . .</b>		<b>177</b>
<b>Appendix J. Error Analysis for the Estimation of Parameters of MIMO Channels . . . . .</b>		<b>181</b>
<b>List of Figures . . . . .</b>		<b>187</b>
<b>List of Tables . . . . .</b>		<b>197</b>
<b>Glossary of Acronyms, Symbols and Notation . . . . .</b>		<b>199</b>
<b>Bibliography . . . . .</b>		<b>201</b>
<b>Theses . . . . .</b>		<b>209</b>
<b>Erklärung . . . . .</b>		<b>211</b>

---



# 1. INTRODUCTION

Nowadays the amount of subscribers connected to mobile networks exceeds three billions, and according to the Wireless World Research Forum (WWRF) it is expected to have seven billions subscribers for 2017 [46]. Moreover, high data rate applications are foreseen for the next generation (3D) mobile Internet with 3D graphics and animations [46]. As a consequence, the wireless communication systems should evolve in order to fit to the future demands.

In order to achieve higher data rates, several solutions have been investigated, such as using higher frequency bands, decreasing the cell size, and applying Multiple Input Multiple Output (MIMO) systems. Since the path loss is inversely proportional to the square of the frequency, by using higher frequency bands the area of coverage of the communication system reduces.<sup>1</sup> Decreasing the cell size has already been extremely exploited, and a hard limit is close to be achieved. For instance, by having very small cells, a very high co-channel interference level is obtained, and for indoor applications, such an effect is even more pronounced [101]. Moreover, there is an economic impact, since the density of base stations in a certain area increases. Consequently, MIMO systems show to be the most promising solution for the future wireless communications systems.

MIMO systems are basically communication systems composed of multiple antennas at the transmitter and multiple antennas at the receiver together with advanced signal processing algorithms. This whole set of hardware and software increases the spectral efficiency with a low cost compared to other solutions.

In order to carry out deployment planning, algorithm design, and system simulations, a profound knowledge of the underlying physical MIMO channel is necessary [31]. In the literature, the most common model for the physical MIMO channel is called multipath components (MPCs), where each MPC describes a discrete link between the stations characterized by physical parameters, like path loss, phase, delay, direction of departure (DoD), and direction of arrival (DoA) [25]. These unique links between DoAs and DoDs are referred as double-directional radio channel [25, 94]. For instance, the Wireless World Initiative New Radio (WINNER)<sup>2</sup> channel model (WIM) is based on the MPCs concept [73]. Moreover, the International Telecommunication Union (ITU)<sup>3</sup> has adopted a slightly modified WINNER channel model for evaluation of IMT-Advanced radio interface technologies [73, 8]. To provide more realistic channel models, the physical parameters should be extracted from measurements

---

<sup>1</sup> Note that by using MIMO in higher frequency bands and considering a rich scattering environment, the diversity gain and the array gain can compensate the path loss [102].

<sup>2</sup> <http://projects.celtic-initiative.org/winner+/>

<sup>3</sup> The file can be downloaded from the site <http://www.itu.int/publ/R-REP-M.2135/en>.

---

given a known scenario. Therefore, the development of robust algorithms to extract these parameters of the MIMO channel is fundamental.

Here we focus on proposing advanced signal processing algorithms to estimate the model order and the associated spatial frequencies considering multi-dimensional array signal processing applications.

In the MIMO system context, the model order is also referred to as the number of multipath components or specular paths of a MIMO channel, while the spatial frequencies are parameters that can be mapped to the direction of departure (DOD) at the transmitter (TX), the direction of arrival (DOA) at the receiver (RX), the time delay of arrival (TDOA), and the Doppler shift. Usually a data model composed of undamped complex exponentials is assumed as a function of the spatial frequencies. Moreover, we also investigate multi-dimensional prewhitening schemes to improve the estimation of spatial frequencies in the presence of colored Gaussian noise.

By reviewing the literature, some important aspects are clearly not solved yet.

- Most model order selection (MOS) schemes do not take into account a limited number of snapshots, and also the case that the number of sensors is larger than the number of snapshots. Such scenarios are common in multi-dimensional data, since representing the multi-dimensional data into a matrix form, the dimensions with sensors are all stacked together. Moreover, since the performance of the MOS schemes depends on the array size, it is not possible to affirm, based on the literature, which of the state-of-the-art MOS technique achieves a best estimation for different SNR regimes in the presence of white or colored Gaussian noise.
- The multi-dimensional model order selection (MOS) scheme, which takes into account the multi-dimensional structure of the data in the estimation of the model order, called CORE CONSistency DIAGnostics (CORCONDIA) [15], has a model order estimation comparable to MOS schemes that are one-dimensional. Therefore, there is clear lack of multi-dimensional techniques with a significant improved performance compared to the one-dimensional MOS schemes.
- Once the model order, i.e., the number of MPCs, is estimated, the estimation of the spatial frequencies can be performed. For this task, closed-form schemes as *R*-D Standard Tensor-ESPRIT (*R*-D STE) or *R*-D Unitary Tensor-ESPRIT (*R*-D UTE) [44] are very appealing, since they are not iterative and their performance is close to the Cramér-Rao lower bound (CRLB) [44]. However, these multi-dimensional parameter estimation schemes fail for arrays with positioning errors and for arrays with more general structure, i.e., outer product based arrays (OPAs) or mixed arrays composed of OPAs and arbitrary arrays. Therefore, there is a necessity of closed-form multi-dimensional schemes that provides a robust estimation of spatial frequencies in such scenarios.

Since in several signal processing applications, the noise is colored, prewhitening schemes should be additionally applied in order to improve the estimation of parameters. Usually

stochastic prewhitening schemes [45, 47, 89] based on matrix inversion, Generalized Singular Value Decomposition (GSVD) or Generalized Eigenvalue Decomposition (GEVD) are applied.

- However, for the previous approaches, which are named as stochastic prewhitening schemes, the noise correlation structure is not taken into account in order to improve substantially the prewhitening. Moreover, prewhitening the data taking into account the structure of the colored noise also requires the estimation of the correlation coefficient.
- In some MIMO and EEG applications, the noise has a multi-dimensional Konecker structure [52, 76], i.e., the noise can be represented as the Kronecker product of two correlation matrices, one related to the time and other related to the space. Therefore, a prewhitening scheme for such a problem should also take into account this multi-dimensional structure. Moreover, the statistics of the multi-dimensional structure of the colored noise should be extracted in order to apply such a prewhitening scheme.
- Since also in several multi-dimensional signal processing applications, the colored noise is only present, when the signal is present. A multi-dimensional prewhitening scheme able to improve the estimation of parameters, even when noise samples are not available, is fundamental.

In order to solve these pointed out open problems, we start describing an overview of our contributions, where we investigate these problems, as well as we propose new solutions.

## 1.1 Overview and Contributions

In this section, we summarize the main contributions of each chapter.

### Chapter 2: Multi-dimensional and Matrix-based Model Order Selection

Here the model order selection problem for the case of one-dimensional and multi-dimensional data model is presented in a didactic fashion. The Chapter 2 is divided into two main sections. In Section 2.1 only the matrix-based model order selection schemes are investigated and proposed, while in Section 2.2 several tensor-based model order selection approaches are classified and proposed.

At the beginning of Subsection 2.1.2 a detailed bibliographic review of the state-of-the-art matrix-based model order selection approaches is provided. In Subsection 2.1.3, we show how to obtain an improved performance for model order selection schemes in scenarios where the number of sensors is larger than the number of snapshots by proposing the 1-dimensional Akaike's Information Criterion (1-D AIC), the 1-dimensional minimum description length (1-D MDL) and the modified exponential fitting test (M-EFT), which are improved versions of AIC [98], MDL [98] and EFT [42, 79], respectively. Details about the conception of our 1-D AIC and 1-D MDL can be found in the Appendix B, while the derivation of the M-EFT is shown in Appendices C and D. Also in Appendix D, we have extended the M-EFT to the case of real-valued noise, which is encountered in sound applications.

---

Since the M-EFT outperforms all the other schemes in the literature in scenarios with white Gaussian noise, we have proposed a multi-dimensional extension of the M-EFT called  $R$ -dimensional exponential fitting test ( $R$ -D EFT) in Subsubsection 2.2.3.1. In general, the M-EFT is only outperformed by the  $R$ -D EFT, since the latter takes into account the multi-dimensional structure of the data. Due to the fact that the techniques in Section 2.2 are based on tensor calculus, a brief review of basic concepts of tensor calculus is provided in Appendix E.

Similarly to the  $R$ -dimensional extension of M-EFT in Subsubsection 2.2.3.1, we have also extended 1-D AIC and 1-D MDL in Subsubsection 2.2.3.2 by applying the concept of global eigenvalues. Therefore, a description of the proposed global eigenvalues is also done in Subsubsection 2.2.3.1. Moreover, inspired by the application of second order statistics of the eigenvalues in [49], we have improved even more the  $R$ -D AIC and  $R$ -D MDL via the exponentiation of the global eigenvalues in Subsubsection 2.2.3.3.

We compare the proposed multi-dimensional MOS schemes to the well-known multi-dimensional model order selection scheme called CORE CONSistency Analysis (CORCONDIA). In order to perform such comparison, we extend CORCONDIA into a non-subjective approach called Threshold-CORCONDIA (T-CORCONDIA) in Subsection 2.2.3.4.

Another multi-dimensional MOS scheme to estimate the model order in the presence of colored noise is proposed in Subsubsection 2.2.3.5 by taking into account the multiple estimates provided by the closed-form PARAFAC [84]. In contrast to the  $R$ -D EFT, the proposed technique called closed-form PARAFAC based model order selection (CFP-MOS) scheme is applicable to colored noise applications. The proposed CFP-MOS scheme achieves a performance close to the proposed  $R$ -D AIC,  $R$ -D MDL and  $R$ -EFT in the presence of white Gaussian noise, and, in the presence of colored Gaussian noise with severe correlation, a performance similar to RADOI [80], the state-of-the-art matrix-based model order selection technique designed for scenarios with colored noise.

By assuming a data model with centro-symmetric arrays, the model order estimation can be improved via forward backward averaging (FBA) [104]. Therefore, in Subsubsection 2.2.3.6, we check the improvement obtained in the model order selection techniques by applying it.

In some applications one dimension of the array can be much greater than the other dimensions. The influence of such characteristic in the performance of M-EFT and  $R$ -D EFT is verified in Subsubsection 2.2.4.1.

### Chapter 3: $R$ -D Parameter Estimation via Closed-Form PARAFAC

In this chapter, we propose a new scheme for a robust estimation of spatial frequencies by applying the closed-form PARAFAC, which is called closed-form PARAFAC based parameter estimation (CFP-PE). As benchmark, comparisons with  $R$ -D ESPRIT-type algorithms [44] and Multilinear Alternating Least Squares (MALS) based approach [66] are performed. The former is a closed-form approach restricted to arrays with a shift invariance structure, while the latter provides the PARAFAC decomposition of the multi-dimensional data iteratively.

---

Particularly for scenarios with modeling errors, such as arrays have positioning errors and also in scenarios generated by *IlmProp*<sup>4</sup> [30, 32], where the narrowband assumption is violated, the CFP-PE gives a significant improvement in the estimation of spatial frequencies compared to the MALS based and ESPRIT-type schemes in the literature.

In Subsection 3.3.1, we show how to increase the maximum model order possible for the CFP-PE by merging dimensions. Moreover, in Subsection 3.3.2, we show how to separate the merged dimensions via the Least Squares Khatri-Rao Factorization (LSKRF). In order to obtain the spatial frequencies, we integrate the CFP to a shift invariance scheme in Subsection 3.3.4 and to a peak search technique in Subsection 3.3.3.

In Section 3.4, simulations are presented to show which combination is more suitable for different scenarios.

## Chapter 4: Multi-dimensional and Matrix-based Subspace Prewhitening Schemes

Sometimes the estimation of parameters is performed in the presence of colored noise. Therefore, in order to avoid the degradation of the techniques in such a scenario we propose matrix-based and multi-dimensional subspace based prewhitening schemes in this chapter.

In Subsubsection 4.1.2.1, by exploiting the structure of the colored noise, we propose the matrix-based deterministic prewhitening scheme, whose the prewhitening matrix has a certain structure and is a function of the noise correlation coefficient. Additionally, we propose schemes to estimate the noise correlation coefficient, which is required for the deterministic prewhitening scheme. In comparison to the stochastic prewhitening schemes [47, 89, 45] in the literature, our scheme presents a considerable improvement mainly for scenarios with severely correlated noise.

Since for certain MIMO and EEG applications, the colored noise has a Kronecker structure, first we show how such structure is obtained in Appendix H. Assuming this type of noise, we propose the Sequential Generalized Singular Value Decomposition (S-GSVD) in Subsubsection 4.2.3.3, which has a low computational complexity and an improved estimation accuracy compared to the stochastic prewhitening schemes [47, 89, 45]. Moreover, we show how to compute the correlation factors of the colored noise in Appendix I, which are necessary for the S-GSVD.

Finally, taking into account signal processing applications, where the noise samples can only be collected in the presence of signal components, we propose the Iterative S-GSVD (I-S-GSVD) in Subsection 4.3.3. The I-S-GSVD inherits from the S-GSVD the low computational complexity, and its performance without the noise statistics information for high and intermediate SNR regimes is quite close to the performance of the S-GSVD with the noise statistics information.

---

<sup>4</sup> *IlmProp* is a flexible geometry-based multi-user MIMO channel modeling tool [30, 32].

## 1.2 Observation

Note that, in the table of contents, the chapters and the sections of this thesis follow the pattern: goal, problem, previous approaches, proposed approaches, and evaluation of approaches. By using such a structure, we intend to help the reader to navigate faster to some point of interest and also to provide the reader a very clear distinction between our proposed contributions and the contributions reviewed in the literature.

---



## 2. MULTI-DIMENSIONAL AND MATRIX-BASED MODEL ORDER SELECTION

This chapter deals with multi-dimensional model order selection, which is a combination of two main subjects or research fields: model order selection and tensor calculus. As we present later, research has been done in these fields for more than one hundred years. Therefore, we describe a brief historic overview of remarkable developments of both subjects until converging to the state of the art of both topics, which includes our contributions [2, 1, 7, 6].

The model order selection (MOS) problem together with the estimation of the principal components is also referred to in the literature as principal component analysis (PCA). In [9] and in [18], PCA was first addressed for applications in chemistry and in physics, respectively. However, just in the 1930s and the 1940s due to [51], a considerable attention was given to multivariate methods for application in psychometrics.

Nowadays the scientific fields where model order selection is necessary are vastly spread, and we list a few examples: stock markets [77], chemistry [9, 67], physics [18], pharmacology [74, 36], psychometrics [51], electroencephalography [100], resonance imaging [64], computer vision [37] and signal processing [2, 1, 7]. Therefore, other terms used in the literature as synonyms for model order selection are signal enumeration [41], estimation of the number of PARAFAC components [15, 2, 1], and clustering enumeration [49].

Until 1985 most of the model order selection (MOS) techniques were characterized as being applied for a matrix data model and as being based on a threshold or a hypothesis [11]. In general, these MOS approaches are not robust, since they depend on parameters that frequently vary, e.g., the Signal-to-Noise Ratio (SNR). Therefore, a remarkable contribution in [98] was the expression for two Information Theoretic Criteria called Akaike's Information Criterion (AIC) and Minimum Description Length (MDL). Due to the good performance, robustness, and simplicity of the proposed expressions in [98], until today they are referred and used as benchmarks.

In Section 2.1 of this chapter, the state-of-the-art matrix-based model order selection (MOS) techniques are described in detail. In Subsection 2.1.2, a detailed bibliographic review is shown. Our contributions related to matrix-based model order selection techniques are explained in Subsection 2.1.3, where we propose modified versions, called one-dimensional Akaike's Information Criterion (1-D AIC), one-dimensional minimum description length (1-D MDL), and Modified Exponential Fitting Test (M-EFT). Comparisons of the matrix-based

MOS techniques proposed by other authors in the literature and by us are performed in Subsection 2.1.4.

Our proposed matrix-based MOS schemes are particularly advantageous for scenarios where the number of sensors is greater than the number of snapshots, which is a common condition in multi-dimensional problems. As a result of several simulation campaigns, our proposed M-EFT has outperformed all the state-of-the-art matrix-based model order selection schemes in the presence of white Gaussian noise [7]. In the M-EFT, the noise eigenvalue profile, which follows a Wishart distribution, is approximated by an exponential curve. Due to this close approximation, an improved performance is obtained.

We emphasize that without our proposed modified versions of the matrix-based model order selection techniques (1-D AIC, 1-D MDL, and M-EFT), some of the multi-dimensional extensions proposed by us in Section 2.2 would not be possible.

Based on a description of a matrix-based model order selection technique used in Multiple Input Multiple Output (MIMO) channel measurements in [81], we formulate a practical MOS approach, which is similar to [81], in Subsection 2.1.2.2, and we compare it to other matrix-based model order selection approaches in Subsection 2.1.4.

Moore’s law, which was first published in 1965 [71], predicts that every two years the number of transistors to be placed in an integrated circuit is approximately doubled.<sup>1</sup> Therefore, with the rising development of faster processors and of memories with higher capacities, it has become common to process considerable amounts of data even in real-time applications [62], which years ago would be prohibitive.

Usually the larger the processed data becomes, the greater is the amount of information to be exploited, which in several cases may present some structure. By exploiting such a structure of the data, it is possible to improve the estimation of desired parameters. For instance, in communication systems, the data composed of different dimensions, e.g., code, space, time, and frequency, is assumed to have a structure, which is usually exploited to estimate parameters, such as Direction-of-Arrival (DOA), Direction-of-Departure (DOD), and Time Delay of Arrival (TDOA). Consequently, tensor-based techniques, which take into account such a multi-dimensional structure, provide significant improvements in the estimation of MIMO channel parameters [44, 6].

In order to exploit the multi-dimensional structure of the data, we apply the tensor calculus, whose some basic concepts are presented in Appendix E. The application of tensor calculus has only more recently been investigated in some scientific fields, e.g., telecommunications, while for others, e.g., psychometrics [48, 17], it has been studied for several years. Already in [48, 17], a very important proposed contribution is the decomposition of the multi-dimensional data into factors by using the parallel factor analysis (PARAFAC). One requirement for the PARAFAC decomposition is also the estimation of the number of PARAFAC components, also called multi-dimensional model order.

---

<sup>1</sup> Note that memories, processors and sensors are composed of transistors. Therefore, several devices are affected by Moore’s law.

Some of the earliest multi-dimensional model order selection schemes in the literature are iterative approaches called LOSS function and Relative Fitness (RELFIT) [15]. Both LOSS and RELFIT are outperformed by another iterative technique called CORE CONSistency DIAGnostics (CORCONDIA) in [15]. One particularity of CORCONDIA is the fact that it depends on a subjective interpretation of the core consistency curve to estimate the rank of the tensor. Therefore, in Subsubsection 2.2.3.4, we show our proposed T-CORCONDIA (Threshold-CORCONDIA) [1], which is a non-subjective form of CORCONDIA [15] based on the threshold coefficients.

In contrast to T-CORCONDIA, whose *Probability of correct Detection (PoD)* is comparable to matrix-based MOS schemes [1], in this chapter we have proposed the  $R$ -dimensional model order selection ( $R$ -D MOS) techniques, whose PoD is much superior to the PoD of T-CORCONDIA and of the matrix-based MOS schemes [1]. During the whole chapter, the *Probability of correct Detection (PoD)* is used to compare the performance between the different model order selection techniques. The PoD is defined as the probability that the correct model order is estimated and is computed as the number of correct estimations divided by the total number of performed realizations for a given scenario. Therefore, with the PoD information, the probability that a certain model order selection technique has to estimate the correct model order in a certain scenario is known. In literature, there are also other ways to measure the efficiency of a model order selection technique by assuming a certain application. For example, in RADAR applications, the overestimation of the model order may be preferable rather than underestimation [80]. However, in order to be objective in our evaluation, we do not include other comparison tools, since we consider in this chapter a long list of model order selection schemes and scenarios, and since we avoid to specify applications. More information about the PoD and other types of probabilities is described in Appendix A.

In Section 2.1, we review several matrix-based model order selection techniques and we propose modified versions of MOS schemes for scenarios where the number of sensors is greater than the number of snapshots. Such scenario is common in multi-dimensional applications, since one-dimensional is usually smaller than the all other dimensions stacked together. In Section 2.2, we present a detailed description of the  $R$ -D MOS techniques proposed by us, and also their classification as global eigenvalue based,  $R$ -D subspace based, or tensor-data based method.

We list below some open questions in the literature solved by us in this chapter.

- 1) Which state-of-the-art model order selection (MOS) technique in the literature presents the best performance considering white Gaussian noise? And which one considering colored Gaussian noise?
- 2) How robust is the state-of-the-art MOS techniques for different array sizes? For example, in case of a small number of snapshots, which approach outperforms the others? Also in case that the number of sensors is greater than the number of snapshots, should the MOS techniques be modified? How significant is the improvement in terms of the

Probability of correct Detection (PoD) applying these modifications?

- 3) How close is the performance of the subspace based MOS techniques compared to the state-of-the-art eigenvalue based MOS techniques?
- 4) Is the performance of the state-of-the-art MOS techniques comparable to the practical MOS approach used in MIMO channel measurements?
- 5) Can the model order selection techniques be extended in order to take into account the multi-dimensional structure of the data? How significant is the improvement obtained?
- 6) Are these extensions possible taking into account the eigenvalues, the subspace or even the multi-dimensional structure of the data?
- 7) Are these new multi-dimensional techniques comparable to CORCONDIA [15]?
- 8) The eigenvalues raised to the power of different values and forward backward averaging (FBA) [104] provide some improvement? How significant is this improvement?
- 9) Does this multi-dimensional gain in the performance depend on the array size?

## 2.1 Matrix-based Model Order Selection

In many scientific publications [84, 44, 91, 69, 26], the model order is always assumed as being known, and usually, attention is only given to the estimation of other parameters, e.g., spatial frequencies. However, for several signal processing techniques, the model order selection is a first required step. Although it seems to be a very simple task to estimate a natural number, the authors of [49] state that the problem of determining the number of principal components, i.e., the model order, is even a more difficult problem than to estimate the principal components.

After the expressions of AIC and MDL were proposed in [98], many other Information Theoretic Criteria (ITC) were investigated assuming different constraints. For instance, the penalty function for AIC is calculated to make the cost function an unbiased estimate of the Kullback-Liebler distance between the modeled density and the estimated density [98]. A few examples of other ITC with penalty functions obtained from different constraints are the Bayesian Information Criterion (BIC) [92], Criterion for Prediction (CP) [68], Shortest Data Description (SDD) [82], Surface Selection Criterion (SSC) [12], Consistent AIC (CAIC) [14], Geometric BIC (GBIC) [21], Geometric AIC (GAIC) [55], and Geometric MDL (GMDL) [56].

A more general ITC approach is the Efficient Detection Criterion (EDC) in [106], since it is shown that the penalty function should obey only two properties, which are reviewed in Subsection 2.1.2. Therefore, based on these properties, each designer can build his own penalty function.

For different data models, noise types or even array sizes, a certain Information Theoretic Criterion may give a better performance than the other. However, note that it is not the scope of this chapter to verify which scenario is more appropriate for different penalty functions. Therefore, we restrict here our comparisons and extensions only to the following ITC approaches: AIC, MDL, and EDC, which are reviewed in Subsection 2.1.2.

As we have mentioned, AIC, MDL, and all similar expressions with different penalty functions are classified as Information Theoretic Criteria (ITC) [98, 54]. However, the model order estimation is a problem which can be solved not only via ITC, but also in many different and creative ways. In [54] the matrix-based model order selection methods are classified into the following categories: information theoretic criterion, eigenvector, data-based, threshold, and root-finding. Also according to [54], another way to classify the MOS schemes is verifying if they are pure detection or combined detection-estimation schemes.

Similarly to [54], we propose in Table 2.1 a slightly modified classification for the matrix-based model order selection approaches. In our proposed classification, ITC is a subgroup of the techniques based on the eigenvalues. Examples of non-ITC eigenvalue based matrix-based model order selection techniques are Stein's Unbiased Risk Estimator (SURE) [95], Nadakuditi Edelman Model Order selection (NEMO) [72], Exponential Fitting Test (EFT) [42, 79] and RADOI [80]. For nonlinear problems, the Stein's Unbiased Risk Estimator (SURE) [95] provides an unbiased estimation, while RADOI [80] is suitable for scenarios with the presence

	Detection	Combined Detection-Estimation
Eigenvalue based	AIC [98], MDL [98], EDC [106], SURE [95], NEMO [72], EFT [42, 79], and RADOI [80] in Subsection 2.1.2 M-EFT [2] in Subsubsection 2.1.3.2	
Subspace based	ESTER[13] and SAMOS [75] in Subsubsection 2.1.2.1	
Data based		UMCS in Subsubsection 2.1.2.2

Tab. 2.1: Classification of matrix-based model order selection techniques

of colored noise. NEMO and EFT are model order selection (MOS) techniques proposed for scenarios with a small number of snapshots and are only applicable in scenarios with the presence of additive white Gaussian noise. In order to handle scenarios where the number of sensors is greater than the number of snapshots, we derive our proposed Modified Exponential Fitting Test (M-EFT) [2] in Subsubsection 2.1.3.2.

Specifically for harmonic retrieval applications, the Vandermonde structure of the array steering vectors, which spans the same subspace of the eigenvectors, can be used to estimate the model order. ESTimation ERror (ESTER) [13] and Subspace-based Automatic Model Order Selection (SAMOS) [75] are examples of subspace based techniques. Other examples of subspace based MOS techniques are the Variance of Transformed Rotational Submatrix (VTRS) [54] and GEneralized Eigenvalues using signal Subspace Eigenvectors (GEESE) [78, 93]. The latter is based on the ESPRIT eigenvalues, which are computed via the shift invariance of the subspace. Since GEESE is a subjective approach, and since ESTER outperforms VTRS, we do not include them in this chapter.

Finally, in Subsection 2.1.2.2, we present a matrix-based model order selection technique based on ESPRIT [88], whose the main algorithm is similar the one used with SAGE [34] by users of MIMO channel sounders (UMCS) [81]. Limitations of this practical MOS approach are shown, as well as simulations comparing it to other matrix-based MOS schemes.

### 2.1.1 Goal and Problem

Since we have now concluded a short review and classification of the state-of-the-art matrix-based model order selection techniques, we start to show specifically what the model order selection problem is. In [23], the author defines independent component analysis (ICA) of a random vector as the search for a linear transformation that minimizes the statistical dependence between its components. However, to estimate these components, it is necessary first to know how many main components there are, i.e., the model order. Therefore, we define model order as the minimum number of components, i.e., principal components, that still characterizes certain properties of the data. For example, for AIC [98], the estimation of the model order is based on the Kullback-Leibler distance between the modeled and estimated density, while for ESTER [13], the estimation of the model order is based on the shift

invariance property.

The estimation of the number of principal components, i.e., the model order, is a first step in a blind source separation problem, also known as cocktail-party problem. To illustrate a typical matrix-based model order problem for a blind source separation application, let us consider the following noiseless data model

$$\mathbf{X}_0 = \sum_{i=1}^d \mathbf{a}_i \cdot \mathbf{s}_i^T, \quad (2.1)$$

where  $\mathbf{X}_0 \in \mathbb{C}^{M \times N}$  contains the noiseless samples,  $\mathbf{a}_i \in \mathbb{C}^{M \times 1}$  and  $\mathbf{s}_i \in \mathbb{C}^{N \times 1}$  are the mixing vector and the source symbols of the  $i$ -th source, respectively. Since  $\mathbf{X}_0$  in (2.1) is the sum of rank-one matrices, and since we consider that none of the  $d$  rank-one matrices can be represented as a linear combination of the others, then  $\mathbf{X}_0$  has rank  $d$ . Therefore, without the presence of noise, the model order selection problem is very simple to solve.

Instead of using vectors, the data model in (2.1) can be represent as the product of the matrices  $\mathbf{A}$  and  $\mathbf{S}$  defined as

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_d], \quad (2.2)$$

and

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1^T \\ \mathbf{s}_2^T \\ \vdots \\ \mathbf{s}_d^T \end{bmatrix}. \quad (2.3)$$

Since, in practice,  $\mathbf{X}$  is contaminated by noise, the following more realistic data model is given

$$\begin{aligned} \mathbf{X} &= \mathbf{A} \cdot \mathbf{S} + \mathbf{N}, \\ &= \mathbf{X}_0 + \mathbf{N}, \end{aligned} \quad (2.4)$$

where  $\mathbf{X} \in \mathbb{C}^{M \times N}$  contains the measurements, and the noise samples are collected in the matrix  $\mathbf{N} \in \mathbb{C}^{M \times N}$ . Here the elements of  $\mathbf{N}$  are considered independent and identically-distributed (i.i.d.) zero mean circularly symmetric (ZMCS) complex Gaussian  $n_{i,j}$  with variance  $\sigma_n^2$ , where  $i$  and  $j$  indicates the respective column and row of the element. Moreover, we assume that  $d \leq \min(M, N)$  (overdetermined case).

Note that since there is noise present in (2.4),  $\mathbf{X}$  has rank  $\min(M, N)$  with probability one, which is usually different from the model order  $d$ . Therefore, in the model order selection problem, it is our goal to estimate  $d$  from the noisy measurements.

---

Here we assume that the symbols of the different sources are uncorrelated. Then

$$\mathbf{R}_{ss} = \mathbb{E}\{\mathbf{S} \cdot \mathbf{S}^H\} \quad (2.5)$$

$$= \sigma_s^2 \cdot \mathbf{I}_d, \quad (2.6)$$

where  $\mathbb{E}\{\cdot\}$  denotes the expected value operator and  $\mathbf{I}_d$  is an identity matrix of size  $d \times d$ . Depending on the level of correlation of the source symbols, a degradation in the estimation of the applied model order selection scheme is observed in practice.

Since, in channel modeling applications, it is also common scenarios where the sources are correlated, we replace  $\mathbf{S}$  by  $\mathbf{S}^{(c)}$ , such that

$$\mathbf{S}^{(c)} = \mathbf{L}_s \cdot \mathbf{S}, \quad (2.7)$$

where the correlation factor  $\mathbf{L}_s$  is such that  $\mathbf{R}_{ss} = \mathbb{E}\left\{\mathbf{S}^{(c)} \left(\mathbf{S}^{(c)}\right)^H\right\} = \sigma_s^2 \cdot \mathbf{L}_s \cdot \mathbf{L}_s^H$ .

In (2.4), we assume the noise as white Gaussian. However, since in several applications colored noise is present, we also consider the colored Gaussian noise case, which is represented by a colored ZMCS complex Gaussian noise matrix  $\mathbf{N}^{(c)}$  defined as

$$\mathbf{N}^{(c)} = \mathbf{L} \cdot \mathbf{N}, \quad (2.8)$$

where the correlation factor  $\mathbf{L}$  is such that  $\mathbf{R}_{nn} = \mathbb{E}\left\{\mathbf{N}^{(c)} \left(\mathbf{N}^{(c)}\right)^H\right\} = \sigma_n^2 \cdot \mathbf{L} \cdot \mathbf{L}^H$ . Note that  $\mathbf{N}$  in (2.4) is the specific case of  $\mathbf{N}^{(c)}$  in (2.8), where  $\mathbf{L} = \mathbf{I}_M$ , which means that the noise is uncorrelated.

In order to estimate the model order  $d$  from the noisy data, one important technique is the eigenvalue decomposition (EVD) in (2.4).<sup>2</sup> To apply the EVD, it is necessary to estimate the covariance matrix. The true covariance matrix  $\mathbf{R}_{xx}$  satisfies

$$\begin{aligned} \mathbf{R}_{xx} &= \mathbb{E}\{\mathbf{X} \cdot \mathbf{X}^H\} \in \mathbb{C}^{M \times M} \\ &= \mathbf{A} \cdot \mathbf{R}_{ss} \cdot \mathbf{A}^H + \mathbf{R}_{nn}, \end{aligned} \quad (2.9)$$

where the noise covariance matrix  $\mathbf{R}_{nn} = \mathbb{E}\{\mathbf{N} \cdot \mathbf{N}^H\} = \sigma_n^2 \cdot \mathbf{I}_M$ .

Since in practice only  $N$  samples of  $\mathbf{X}$  may be available, the sample covariance matrix  $\hat{\mathbf{R}}_{xx}$  can be computed as

$$\hat{\mathbf{R}}_{xx} = \frac{1}{N} \cdot \mathbf{X} \cdot \mathbf{X}^H, \quad (2.10)$$

where  $\hat{\mathbf{R}}_{xx} \in \mathbb{C}^{M \times M}$  is a Hermitian matrix, i.e.,  $\hat{\mathbf{R}}_{xx} = \hat{\mathbf{R}}_{xx}^H$ . Therefore, we can compute its EVD as

$$\hat{\mathbf{R}}_{xx} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H, \quad (2.11)$$

where  $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_M] \in \mathbb{C}^{M \times M}$  is a unitary matrix, which contains the eigenvectors,

---

<sup>2</sup> Instead of applying the EVD to  $\mathbf{R}_{xx}$ , the singular value decomposition (SVD) can also be applied to  $\mathbf{X}$ .



and  $\mathbf{\Lambda} \in \mathbb{C}^{M \times M}$  is a real-valued diagonal matrix, which contains the sorted eigenvalues  $\lambda_i$ , such that  $\lambda_1 > \lambda_2 > \dots > \lambda_M$ . Let us denote  $\mathbf{U}^{[s]} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_P] \in \mathbb{C}^{M \times P}$  as being the truncated matrix composed of  $P$  eigenvectors of  $\mathbf{U}$  corresponding to the  $P$  largest eigenvalues of  $\mathbf{\Lambda}$ . In case that  $P = d$ , the dominant eigenvectors  $\mathbf{U}^{[s]} \in \mathbb{C}^{M \times d}$  and the column space of  $\mathbf{A}$  span the same subspace. Therefore, to find  $d$  is a crucial step, and without  $d$  it is not possible to select the principal components in order to estimate  $\mathbf{A}$  and  $\mathbf{S}$  correctly. Independently of using the EVD, ICA [33] or another type of decomposition,  $d$  must be estimated.

Hence, using the mathematical representation already introduced, we can state the matrix-based model order selection problem as to estimate the model order  $d$  of  $\mathbf{X}$  in (2.4).

Next in Subsubsection 2.1.1.1, the profile of the noise eigenvalues, which is a crucial aspect for various model order selection schemes, is analyzed.

#### 2.1.1.1 Eigenvalue based Estimation of the Model Order

Several matrix-based model order selection techniques estimate the model order via the eigenvalues  $\lambda_i$  of the sample covariance matrix  $\widehat{\mathbf{R}}_{xx}$  in (2.11). Again the eigenvalues of  $\widehat{\mathbf{R}}_{xx}$  are given by

$$\{\lambda_1, \dots, \lambda_M\} = \text{diag}(\mathbf{\Lambda}), \quad (2.12)$$

where  $\mathbf{\Lambda}$  is defined in (2.11) and  $\text{diag}(\cdot)$  is an operator that extracts the main diagonal of the matrix  $\mathbf{\Lambda}$ .

In the absence of noise, an example of the eigenvalues profile is presented in Fig. 2.1. In this case, the model order is easily estimated by counting the number of eigenvalues that are different from zero.

Now fixing the SNR to 0 dB, and considering that  $N \rightarrow \infty$ , we have the following covariance matrix

$$\mathbf{R}_{xx} = \lim_{N \rightarrow \infty} \widehat{\mathbf{R}}_{xx} = \mathbf{A} \mathbf{R}_{ss} \mathbf{A}^H + \sigma_n^2 \cdot \mathbf{I}_M \in \mathbb{C}^{M \times M}. \quad (2.13)$$

An example of the eigenvalues profile of  $\mathbf{R}_{xx}$  in (2.13) is presented in Fig. 2.2. Since in (2.4), the noise elements  $n_{i,j}$  are i.i.d., then  $\mathbb{E}\{\mathbf{N} \cdot \mathbf{N}^H\} = \sigma_n^2 \cdot \mathbf{I}_M$ . Note that (2.13) is assumed in the derivation of AIC [98] and MDL [98].

Obviously, in practical applications the assumption that  $N \rightarrow \infty$  is also not valid, since only a limited number of samples  $N$  is available. Therefore, for a small  $N$ , an example of the eigenvalues profile is given in Fig. 2.3. In this case, the data is contaminated by white Gaussian noise, and, for this reason, the profile of the noise eigenvalues follows a Wishart distribution [72, 42, 79].

For scenarios with a small number of samples, i.e., where the noise eigenvalues follow a Wishart distribution exemplified in Fig. 2.3, the Nadakuditi Edelman Model Order selection (NEMO) scheme [72], the Exponential Fitting Test (EFT) [42, 79] and the Modified Exponential Fitting Test (M-EFT) in Subsubsection 2.1.3.2 have been proposed. In the case of

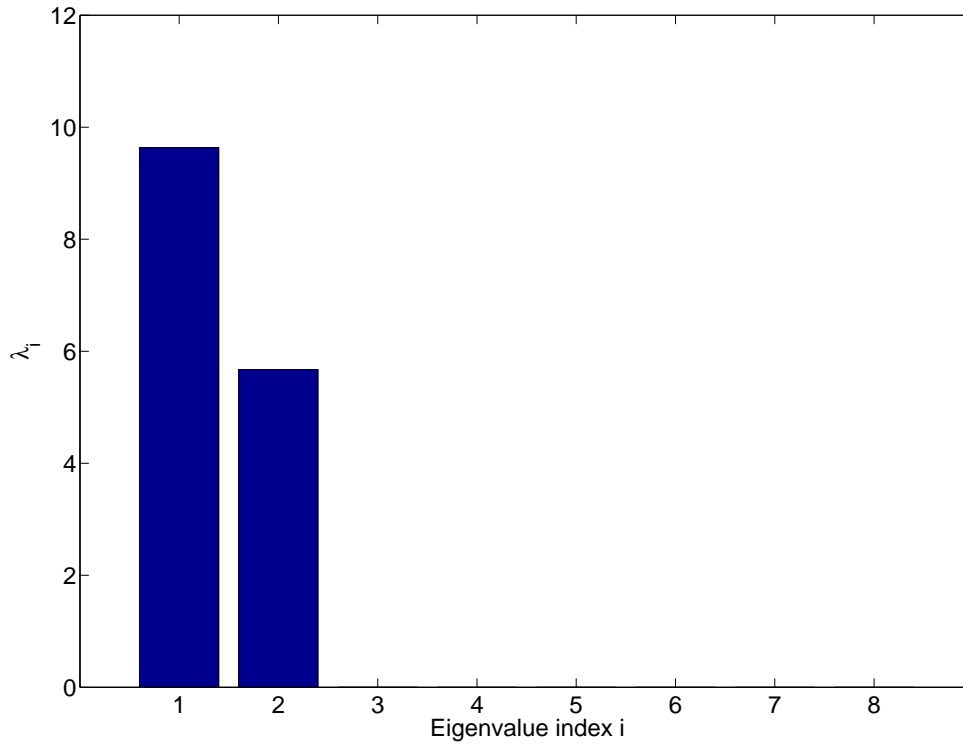


Fig. 2.1: Eigenvalues profile in case that  $\text{SNR} \rightarrow \infty$  and  $N \rightarrow \infty$ , and considering  $M = 8$  and  $d = 2$ .

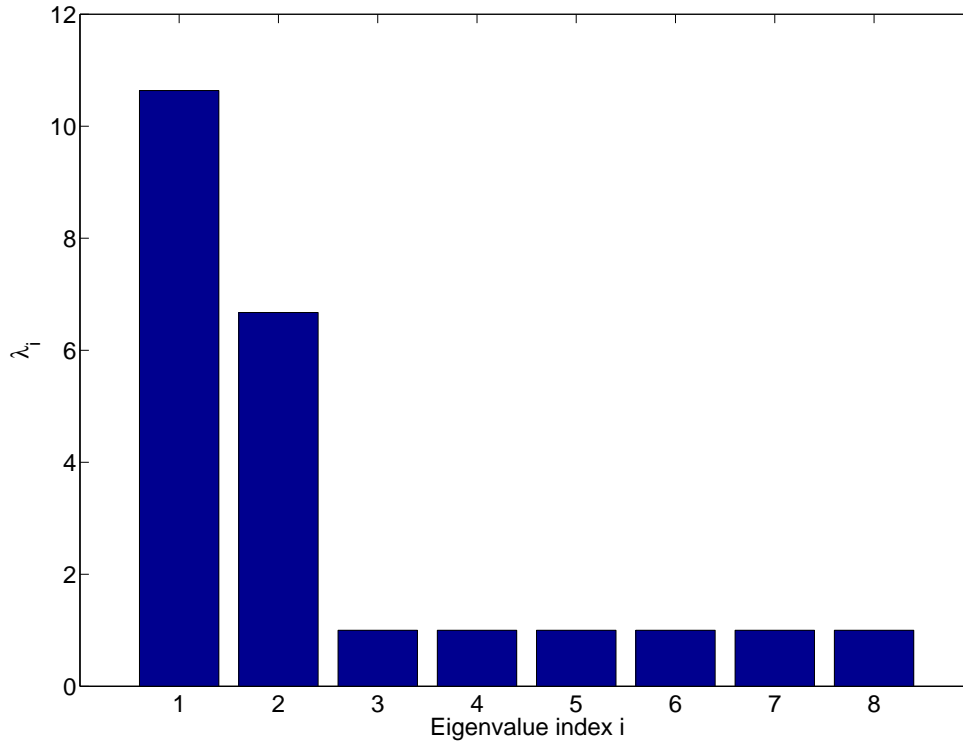


Fig. 2.2: Eigenvalues profile in case that  $N \rightarrow \infty$ , and considering  $\text{SNR} = 0$  dB,  $M = 8$  and  $d = 2$ .

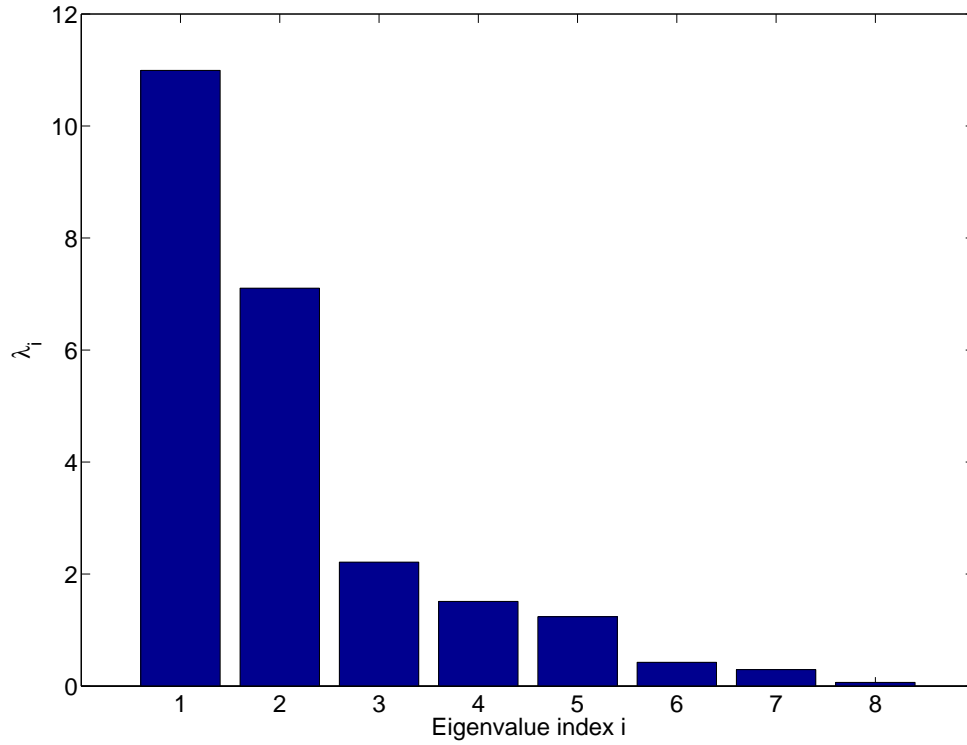


Fig. 2.3: Eigenvalues profile considering SNR = 0 dB,  $M = 8$ ,  $N = 10$  and  $d = 2$ .

EFT and M-EFT, the Wishart distribution<sup>3</sup> is approximated by an exponential profile.

Next the state-of-the-art matrix-based model order estimation techniques are briefly reviewed in Subsection 2.1.2. In Subsection 2.1.3, we present 1-D AIC and 1-D MDL, and the Modified Exponential Fitting Test (M-EFT), which are extensions of AIC, MDL and EFT for any values of  $M$  and  $N$ . Comparisons between all these matrix-based model order selection techniques are depicted in Subsection 2.1.4.

### 2.1.2 Previous Approaches: Review of the State-of-the-art Matrix-based Model Order Estimation Techniques

In this subsection, we first present the state-of-the-art eigenvalue based model order selection techniques found in the literature, which are AIC [98], MDL [98], EDC [106], NEMO [72], SURE [95], RADOI [80], and EFT [42, 79]. These techniques in their original publications have already outperformed several other model order selection techniques, which are not included here. In addition, in Subsubsection 2.1.2.1, we present the state-of-the-art subspace based matrix-based model order selection techniques, while in Subsubsection 2.1.2.2 we show a practical model order selection scheme used in MIMO channel measurements.

First we show the model order estimation via AIC, MDL, and EDC. In (2.14), we present

---

<sup>3</sup> Note that the one-dimensional Wishart distribution is a  $\chi^2$  distribution.

Approach	Penalty function $p(P, M, N)$
AIC [98]	$P \cdot (2 \cdot M - P)$
MDL [98]	$\frac{1}{2} \cdot P \cdot (2 \cdot M - P) \cdot \log(N)$
EDC [106, 13]	$\frac{1}{2} \cdot P \cdot (2 \cdot M - P) \cdot \sqrt{N \cdot \log(\log N)}$

Tab. 2.2: Penalty functions for Information Theoretic Criteria

a general expression used to estimate the model order by the ITC approaches. It is given by

$$\begin{aligned} \widehat{d} &= \arg \min_P J(P) \quad \text{where} \\ J(P) &= -N(M - P) \log \left( \frac{g(P)}{a(P)} \right) + p(P, M), \end{aligned} \quad (2.14)$$

$P$  is a candidate value for estimated model order  $\widehat{d}$ ,  $\log(\cdot)$  is the natural logarithm function,  $g(P) = \left( \prod_{i=\alpha-P}^{\alpha} \lambda_i \right)^{\frac{1}{P}}$  and  $a(P) = \frac{1}{P} \cdot \sum_{i=\alpha-P}^{\alpha} \lambda_i$  are the geometric and the arithmetic mean of the  $P$  smallest eigenvalues  $\lambda_i$  of (2.12), respectively. We define  $\alpha = \min(M, N)$ . Here,  $p(P, M)$  is the penalty function, and it varies for each ITC approach. In Table 2.2, we present the penalty functions of AIC, MDL, and EDC.

In [106], it is shown that the penalty function for EDC should obey two properties, which are  $\lim_{N \rightarrow +\infty} \frac{p(P, M, N)}{f_p \cdot N} = 0$  and  $\lim_{N \rightarrow +\infty} \frac{p(P, M, N)}{f_p \cdot \log(\log(N))} = +\infty$ .  $f_p$  denotes the number of free parameters and how to compute them is shown in Appendix B. Note that we use in Table 2.2 the same penalty function used in [13] for the EDC technique.

In contrast to the previous techniques, Nadakuditi Edelman Model Order selection (NEMO) scheme [72] and Exponential Fitting Test (EFT) [42, 79] are proposed for scenarios, where only a small number of snapshots is available.

In the NEMO scheme [72], the cost function to be minimized is

$$\begin{aligned} \widehat{d} &= \arg \min_P \text{NEMO}(P) \quad \text{where} \\ \text{NEMO}(P) &= \frac{\psi}{4} \left[ \frac{N}{M} \right] t_P^2 + 2(P + 1) \quad , \\ t_P &= \left[ (M - P) \frac{\sum_{i=P+1}^M \lambda_i^2}{\left( \sum_{i=P+1}^M \lambda_i \right)^2} - \left( 1 + \frac{M}{N} \right) \right] N - \left( \frac{2}{\psi} - 1 \right) \frac{M}{N} \quad , \end{aligned} \quad (2.15)$$

$\psi = 1$  if  $\mathbf{X} \in \mathbb{R}^{M \times N}$  and  $\psi = 2$  if  $\mathbf{X} \in \mathbb{C}^{M \times N}$ . In NEMO, we have that  $0 \leq P < \min(M, N)$ .

In the EFT [42, 79], the Wishart profile of the noise eigenvalues is approximated by an exponential profile. Therefore, this facilitates the prediction of the noise eigenvalues. In Fig. 2.4, we exemplify how EFT estimates the model order by comparing the eigenvalue  $\lambda_{M-P}$  calculated from the measurements with the predicted eigenvalue  $\widehat{\lambda}_{M-P}$ , where  $P$  is our candidate value for the number of noise eigenvalues. Therefore, our estimate of the model order  $\widehat{d} = M - P$ . The threshold  $\eta_P$ , which is compared to the relative distance between  $\lambda_{M-P}$

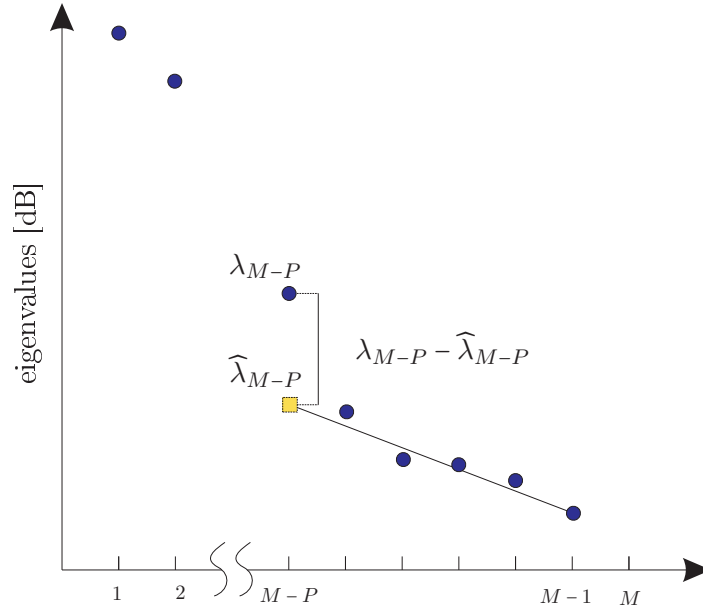


Fig. 2.4: Typical profile of the eigenvalues. The last  $P - 1$  eigenvalues are used to estimate the  $(M - P)$ -th eigenvalue, denoted by a yellow square. The EFT method considers the gap between the measured eigenvalue and the predicted one.

and  $\hat{\lambda}_{M-P}$ , is a function of the design parameter  $P_{fa}$ , which is the probability of false alarm. Usually a very good Probability of correct Detection (PoD) is obtained for very small values of  $P_{fa}$ , e.g.,  $P_{fa} = 10^{-3}$ , or even smaller.

The EFT is divided into two parts: one preprocessing step to compute the threshold coefficients  $\eta_P$  and the estimation of the model order. Although the calculation of the threshold coefficients is computationally expensive, it is a task only performed once for a given scenario where  $M$ ,  $N$ , and the probability of false alarm  $P_{fa}$  do not vary. In this sense, since  $M$ ,  $N$ , and  $P_{fa}$  are fixed parameters, the threshold coefficients  $\eta_P$  can be computed previously. Therefore, the estimation of the model order via EFT has a very low computational complexity, which is comparable to AIC and MDL [79]. Next we show the expressions to compute the predicted eigenvalues, as well as the expressions to calculate the threshold coefficients  $\eta_P$  and to estimate the model order  $d$ .

The derivation of EFT is based on three equations, which are related to the exponential assumption, the sum of Wishart eigenvalues, and the variance of the Wishart eigenvalues. By using these three equations, it is possible to derive the following correction equations [79]

$$\hat{\lambda}_{M-P} = (P+1) \cdot \frac{1 - q(P+1, N)}{1 - q(P+1, N)^{P+1}} \hat{\sigma}^2 \quad (2.16)$$

$$\hat{\sigma}^2 = \frac{1}{P+1} \sum_{i=0}^P \lambda_{M-i}, \quad (2.17)$$

where  $q(M, N)$  is given by

$$q(M, N) = \exp \left\{ -\sqrt{\frac{30}{M^2 + 2}} - \sqrt{\frac{900}{(M^2 + 2)^2} - \frac{720 \cdot M}{N(M^4 + M^2 - 2)}} \right\}, \quad (2.18)$$

where  $q$  is the rate of the exponential profile. Instead of presenting here the derivation of EFT, we describe in Appendix D a more general derivation considering any value of  $M$  and  $N$ . This general derivation is necessary for our proposed M-EFT in Subsubsection 2.1.3.2.

Note that  $\lambda_{M-P}$  in (2.17) is included in the estimation of  $\widehat{\lambda}_{M-P}$  in (2.16). Therefore, (2.16) is a correction equation rather than a prediction equation. As shown in Subsubsection 2.1.3.2, one of the modifications in the M-EFT is the use of a prediction equation instead of a correction equation.

In order to estimate the threshold coefficients  $\eta_P$ , we have to consider  $N_l$  different realizations of a white Gaussian noise matrix  $\mathbf{N} \in \mathbb{C}^{M \times N}$ . Therefore, to obtain the Probability of False Alarm  $P_{fa}$  as a function of the threshold coefficient  $\eta_P$ , we assume the following hypotheses

$$\begin{aligned} H_{P+1} : \lambda_{M-P} \text{ is a noise EV, } & \frac{\lambda_{M-P} - \widehat{\lambda}_{M-P}}{\widehat{\lambda}_{M-P}} \leq \eta_P \\ \bar{H}_{P+1} : \lambda_{M-P} \text{ is a signal EV, } & \frac{\lambda_{M-P} - \widehat{\lambda}_{M-P}}{\widehat{\lambda}_{M-P}} > \eta_P, \end{aligned} \quad (2.19)$$

where the range for  $\eta_P$  is also a design parameter. We define  $N_{fa}$  as the number of times that  $\bar{H}_{P+1}$  is observed for all  $N_l$  noise realizations. Therefore,  $P_{fa} = \frac{N_{fa}}{N_l}$ , and for each value of the predefined range of  $\eta_P$ , a certain value of  $P_{fa}$  is computed. An example of the range of  $\eta_P$  is from 0 to 10 with steps of 0.1. For instance, a curve of the Probability of False Alarm  $P_{fa}$  versus the threshold coefficient  $\eta_P$  by applying (2.19) is presented in Fig. 2.5. For instance, in Fig. 2.5, by choosing a  $P_{fa} = 10^{-2}$ , then we get  $\eta_1 = 0.277$ .

Once all  $\eta_P$  are found for a certain  $M$ ,  $N$ , and  $P_{fa}$  via (2.19), the model order can be estimated by the following cost function

$$\begin{aligned} \widehat{d} &= M - \min(\mathcal{P}) \quad \text{where} \\ \mathcal{P} &\in \mathcal{P}, \quad \text{if } \frac{\lambda_{M-P} - \widehat{\lambda}_{M-P}}{\widehat{\lambda}_{M-P}} > \eta_P, \end{aligned}$$

where  $\widehat{\lambda}_{M-P}$  is computed according to (2.16) and  $\mathcal{P}$  is the set of all values of  $P$ , such that the relative distance  $\frac{\lambda_{M-P} - \widehat{\lambda}_{M-P}}{\widehat{\lambda}_{M-P}}$  is greater than  $\eta_P$ .

In [79], the EFT outperforms the AIC, MDL, MDL version B (MDLB) [99], and the predictive description length (PDL) criterion [96]. According to [79], since MDLB and PDL are maximum likelihood (ML) based approaches, they require a much higher computational complexity than AIC, MDL and EFT.

Another recent MOS technique in the literature is the Stein's Unbiased Risk Estimator (SURE) [95], which provides an unbiased estimation. In this approach, the risk  $\widehat{R}(P)$  must

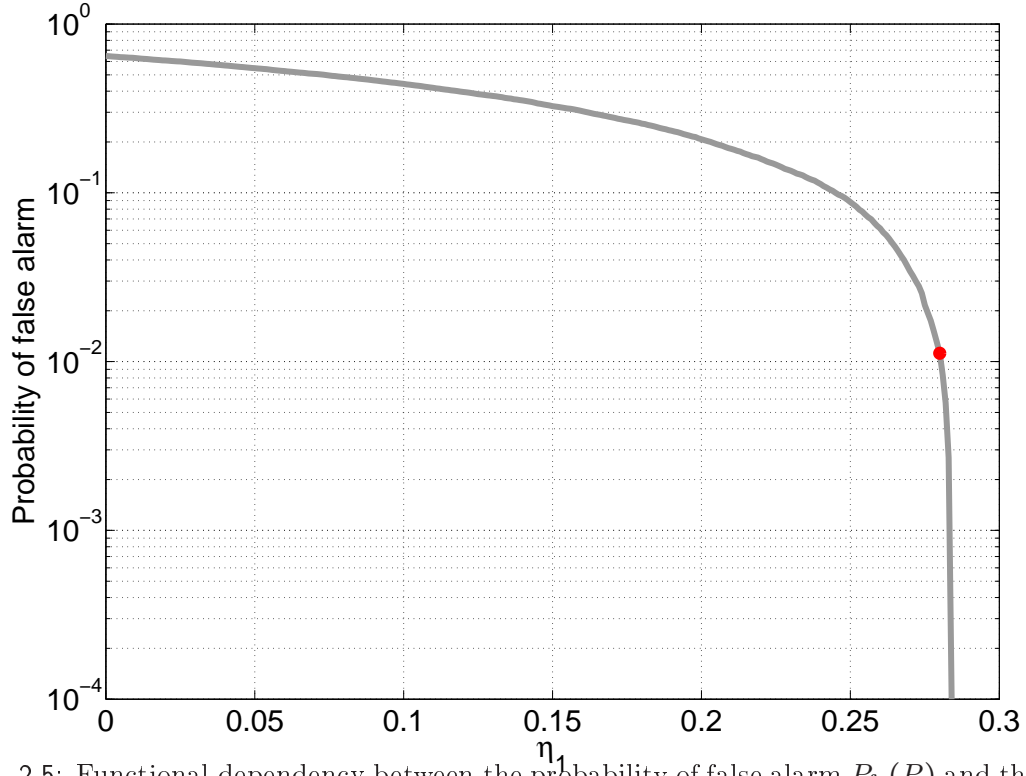


Fig. 2.5: Functional dependency between the probability of false alarm  $P_{fa}(P)$  and the threshold  $\eta_P$ . The plot corresponds to  $M = 5$ ,  $N = 6$ , and  $P = 1$ .

be minimized according to the following expression

$$\hat{d} = \arg \min_P \hat{R}(P) \quad \text{where} \quad (2.20)$$

$$\begin{aligned} \hat{R}(P) = & (M - P)\hat{\sigma}_P^2 + 2\sigma^2 P + (\hat{\sigma}_P^4 - 2\hat{\sigma}_P^2\sigma^2 + \frac{4\hat{\sigma}_P^2\sigma^2}{N}) \sum_{i=1}^P \frac{1}{\lambda_i} \\ & + \frac{4\sigma^2}{N} \sum_{i=1}^P \sum_{j=P+1}^M \frac{\lambda_i - \hat{\sigma}_P^2}{\lambda_i - \lambda_j} + \frac{2\sigma^2}{N} P(P-1) \\ & - \frac{2\sigma^2}{N} (M-1) \sum_{i=1}^P (1 - \frac{\hat{\sigma}_P^2}{\lambda_i}) \end{aligned}$$

where  $\hat{\sigma}_P^2 = \frac{1}{M-P} \sum_{i=P+1}^M \lambda_i$  and the noise variance  $\sigma^2$  is estimated via random matrix theory (RMT) using the algorithm also proposed in [95]. In the estimation of  $\sigma^2$  via RMT, a crude estimation of  $d$  is performed, and for this step, it is necessary to assume that a certain percentage, e.g., 25 %, of the smallest eigenvalues is only due to the noise.

In [95], SURE outperforms the Laplace [70] and the BIC [57] methods in terms of the PoD.

The previous MOS approaches fail in the presence of colored noise, and for these scenarios the empirical MOS scheme called RADOI [80] has been proposed. The RADOI scheme solves

the following optimization problem

$$\widehat{d} = \arg \min_P \text{RADOI}(P) \quad \text{where} \quad (2.21)$$

$$\text{RADOI}(P) = \lambda_{P+1} \cdot \left( \sum_{i=2}^M \lambda_i \right)^{-1} - \xi_P \cdot \left( \sum_{i=1}^{M-1} \xi_i \right)^{-1}, \quad (2.22)$$

where  $\xi_P = 1 - \frac{\alpha_R(\lambda_P - \widehat{\sigma}_P^2)}{\widehat{\sigma}_P^2}$ ,  $\widehat{\sigma}_P^2 = \frac{1}{M-P} \cdot \sum_{i=P+1}^M \lambda_i$ , and  $\alpha_R$  is given by

$$\alpha_R = \left[ \arg \max_P \frac{(\lambda_P - \widehat{\sigma}_P^2)}{\widehat{\sigma}_P^2} \right]^{-1}, \quad (2.23)$$

where  $\alpha_R$  is the inverse of the index  $P$  corresponding to the greatest normalized difference between  $\lambda_P$  and  $\widehat{\sigma}_P^2$ .

In [80], RADOI outperforms the Gerschgoerin disk estimator (GDE) criterion [103] in the presence of colored noise, while its performance in the presence of white noise is close to the GDE criterion.

### 2.1.2.1 Subspace based Approaches

In this subsection, we present a different way to estimate the model order, not based on the eigenvalues as the previous, but based on the subspace. Here we introduce two subspace based techniques called ESTimation ERror (ESTER) [13] and Subspace-based Automatic Model Order Selection (SAMOS) [75]. The subspace based approaches are restricted to shift invariance data model, which includes the case where the matrix  $\mathbf{A}$  in (2.4) has a Vandermonde structure.

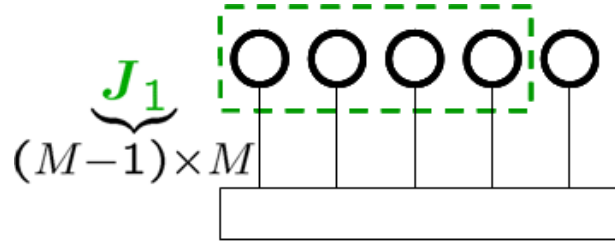
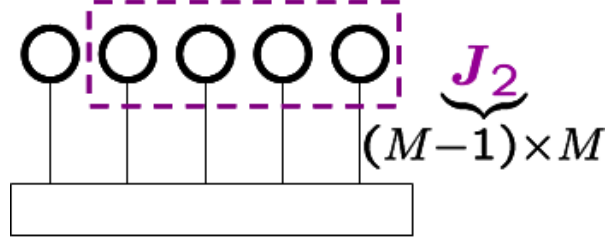
Comparing to the data model in (2.4) using (2.7) in Subsection 2.1.1, here we assume that the vectors of  $\mathbf{A}$  have a Vandermonde structure, and that they can be written as a function of the spatial frequency  $\mu_i$  as follows

$$\mathbf{A} = [\mathbf{a}(\mu_1), \mathbf{a}(\mu_2), \dots, \mathbf{a}(\mu_d)]. \quad (2.24)$$

Note that, in case of that Vandermonde structure, the matrix  $\mathbf{A}$  has a certain element  $a_i(m) = e^{j\mu_i(m-1)}$ , where  $i = 1, \dots, d$ , and  $m = 1, \dots, M$ .

In ESTER and SAMOS, only the eigenvectors are applied to estimate the model order, and particularly, in the presence of colored Gaussian noise a high probability of correct detection (PoD) is reached if compared to some classical eigenvalues based approaches [13]. It is important to emphasize that although the performance of SAMOS is slightly superior to ESTER, the candidate value  $P$  for  $\widehat{d}$  in SAMOS is restricted to  $1 \leq P < \min(\lfloor \frac{M-1}{2} \rfloor, N)$ , while in ESTER,  $P$  is restricted to  $1 \leq P < \min(M, N)$ .  $\lfloor \cdot \rfloor$  is the floor operator, which selects the closest integer smaller than the value inside the brackets.



Fig. 2.6: Selection of the first  $M - 1$  sensors multiplying by the matrix  $\mathbf{J}_1$ Fig. 2.7: Selection of the last  $M - 1$  sensors multiplying by the matrix  $\mathbf{J}_2$ 

The model order estimation via the ESTER's residual error is given by

$$\begin{aligned} \hat{d} &= \arg \max_P \frac{1}{E_E(P)^2} \quad \text{where} \\ E_E(P)^2 &= \|\mathbf{J}_1 \cdot \mathbf{U}_P \cdot \hat{\Psi} - \mathbf{J}_2 \cdot \mathbf{U}_P\|_2^2 \quad \text{where} \\ \hat{\Psi} &= (\mathbf{J}_2 \cdot \mathbf{U}_P)^+ \cdot \mathbf{J}_1 \cdot \mathbf{U}_P \end{aligned} \quad (2.25)$$

where  $\mathbf{J}_1 \in \mathbb{R}^{M-1 \times M}$  and  $\mathbf{J}_2 \in \mathbb{R}^{M-1 \times M}$  are the selection matrices for the first  $M - 1$  sensors and for the last  $M - 1$  sensors respectively, and  $\mathbf{U}_P$  represents the  $P$  first left eigenvectors of  $\hat{\mathbf{R}}_{xx}$  in (2.11). In general, the matrices  $\mathbf{J}_1 \in \mathbb{R}^{M^{(\text{sel})} \times M}$  and  $\mathbf{J}_2 \in \mathbb{R}^{M^{(\text{sel})} \times M}$  can be represented mathematically in the following way

$$\begin{aligned} \mathbf{J}_1 &= [\mathbf{I}_{M^{(\text{sel})}} \quad \mathbf{0}_{M^{(\text{sel})} \times 1}] \\ \mathbf{J}_2 &= [\mathbf{0}_{M^{(\text{sel})} \times 1} \quad \mathbf{I}_{M^{(\text{sel})}}], \end{aligned} \quad (2.26)$$

where  $M^{(\text{sel})} = M - 1$  for maximum overlap. As illustrated in Fig. 2.6, we select the first  $M - 1$  sensors, while in Fig. 2.7, the last  $M - 1$  sensors are selected.

Due to the structure of the matrix  $\mathbf{A}$  in (2.24), the shift invariance equation is satisfied by  $\mathbf{A}$  shown as

$$\mathbf{J}_1 \cdot \mathbf{A} \cdot \Phi = \mathbf{J}_2 \cdot \mathbf{A}, \quad (2.27)$$

where  $\text{diag}(\Phi) = \{e^{j\mu_1}, \dots, e^{j\mu_d}\}$ . Therefore, ESTER exploits the fact (2.27) is not satisfied for  $P \neq d$  and in the presence of noise.

In SAMOS [75], a certain matrix  $\mathbf{U}_P^{\text{tb}} = [\mathbf{J}_2 \mathbf{U}_P \ \mathbf{J}_1 \mathbf{U}_P] \in \mathbb{C}^{(M-1) \times 2 \cdot P}$  is defined. Without noise  $\mathbf{U}_P^{\text{tb}}$  is a rank- $d$  matrix, since  $\mathbf{J}_2 \mathbf{U}_P$  and  $\mathbf{J}_1 \mathbf{U}_P$  span the same subspace. Therefore, the last  $d$  singular values  $\lambda_{d+1}^{\text{tb}}, \dots, \lambda_{2d}^{\text{tb}}$  are equal to zero.

In the case that  $P > d$ , i.e., overestimation of the model order, the additional columns of  $\mathbf{J}_2 \mathbf{U}_P$  and  $\mathbf{J}_1 \mathbf{U}_P$  are not the same. Therefore, including columns when  $P > d$ , the rank of  $\mathbf{U}_P^{\text{tb}}$  is increased by two [75]. In case that  $P < d$ , i.e., underestimation of the model order, the columns of  $\mathbf{U}_P$  contain contributions of all  $d$  components. Therefore, the model order is equal to  $\min(2 \cdot P, d)$  [75]. Hence, the model order selection expression for SAMOS is given by

$$\begin{aligned} \hat{d} &= \arg \max_P \frac{1}{E_S(P)} \quad \text{where} \\ E_S(P) &= \frac{1}{P} \sum_{i=P+1}^{2P} \lambda_i^{\text{tb}} \quad \text{where} \end{aligned} \quad (2.28)$$

where  $\lambda_i^{\text{tb}}$  denotes the  $i$ -th singular value of  $\mathbf{U}_P^{\text{tb}}$ .

For real time application, the subspace based techniques are particularly appealing, since the subspace tracking is computationally much less expensive than computing an EVD. Moreover, in [13], it was shown that ESTER outperforms EDC [106] in the presence of colored noise, while in [75], it was shown that SAMOS outperforms ESTER.

### 2.1.2.2 Data based Approach via Measured Noise Power

In order to estimate the model order, users of MIMO channel sounders (UMCS) apply a model order scheme based on the stability of the noise power for different time instants [81]. In this subsection, we present a general insight of this model order selection scheme, although we know that different implementations are possible. Moreover, we compare this scheme to the previous matrix-based model order techniques in Subsection 2.1.4.

The model order selection via measured noise power (MOS-MNP) is divided into two steps. In the first step, noise samples  $\mathbf{N}'$  are collected with the absence of signals, and the noise power  $\sigma'^2 = \left\| \frac{\mathbf{N}' \cdot \mathbf{N}'^H}{N} \right\|_{\text{F}}^2$  is estimated. In the second step, the noise plus signal samples  $\mathbf{X}$  are collected. Then, the highest power components of  $\mathbf{X}$  are gradually removed, and the power of the remaining components  $\hat{\sigma}^2$  is computed. When the condition  $\hat{\sigma}^2 \leq \sigma'^2$  is satisfied, the number  $P$  of the strongest components obtained is our estimate of the model order  $d$ . The operator  $\|\cdot\|_{\text{F}}^2$  denotes the Frobenius norm.

Next we summarize an algorithm describing the second step of the model order selection via measured noise power (MOS-MNP) scheme, and as usual our candidate for the model order is represented by  $P$ .

- 1)  $P = 1$ ;

- 2) Given noise plus signal samples  $\mathbf{X}$ , estimate the spatial frequencies  $\hat{\mu}_i^{(r)}$  for  $r = 1, \dots, R$ , and for  $i = 1, \dots, P$ . In this step, any direction-of-arrival approach can be applied, e.g., MUSIC [91], ESPRIT [88], EM [69], or SAGE [34].
- 3) Given  $\hat{\mu}_i^{(r)}$ , compute  $\hat{\mathbf{A}}$  considering the Vandermonde structure. Moreover, compute  $\hat{\mathbf{S}}$ , since  $\hat{\mathbf{S}} = \hat{\mathbf{A}}^+ \cdot \mathbf{X}$ . A particular structure of  $\mathbf{S}$  can be taken into account in order to improve the estimation of  $\hat{\mathbf{S}}$ .
- 4) Given  $\hat{\mathbf{S}}$  and  $\hat{\mathbf{A}}$ , then compute  $\hat{\mathbf{X}}$ .
- 5) Given  $\hat{\mathbf{X}}$ ,  $\hat{\mathbf{N}} = \mathbf{X} - \hat{\mathbf{X}}$  is computed. Then,  $\hat{\sigma}^2 = \left\| \frac{\hat{\mathbf{N}} \cdot \hat{\mathbf{N}}^H}{N} \right\|_F^2$  is computed.
- 6) If  $\hat{\sigma} \leq \alpha_{PW} \cdot \sigma'$  is true, then  $\hat{d} = P$  and the algorithm should stop. Otherwise, increment  $P$  and return to 2).

Note that  $\alpha_{PW}$  is a very important design parameter in the MOS-MNP scheme as we show in the simulations in Subsection 2.1.4.

### 2.1.3 Proposed Approaches

In this subsection, we present our extensions of AIC, MDL, and EFT called respectively 1-D AIC, 1-D MDL, and M-EFT for scenarios where  $M > N$ .

#### 2.1.3.1 1-D AIC and 1-D MDL

In order to apply AIC and MDL for the cases where  $M > N$ , practitioners have proposed a new version of AIC and MDL by restricting  $P$  to integer values  $0 \leq P < \min(M, N)$  [72]. Since only by restricting  $P$ , the performance obtained using AIC and MDL is degraded, we have proposed more suitable versions in [2, 7] called 1-dimensional AIC (1-D AIC) and 1-dimensional MDL (1-D MDL), where the case  $M > N$  is taken into account in the derivation.

In [2], we have modified AIC, MDL and EDC for the case that  $M > N$ , and we called them 1-D AIC, 1-D MDL, and 1-D EDC, respectively. Similarly to (2.14), we write our proposed 1-D ITC techniques in the following general form

$$\begin{aligned} \hat{d} &= \arg \min_P J(P) \quad \text{where} \\ J(P) &= -N(\alpha - P) \log \left( \frac{g(P)}{a(P)} \right) + p(P, N, \alpha), \end{aligned} \tag{2.29}$$

where  $\hat{d}$  represents an estimate of the model order  $d$ ,  $P$  is a candidate value for  $\hat{d}$ ,  $g(P)$  and  $a(P)$  are the geometric mean and arithmetic mean of the  $P$  smallest eigenvalues of (2.12), respectively. The penalty functions for 1-D AIC, 1-D MDL and 1-D EDC are given in Table 2.3. According to [2]  $\alpha = \min[M, N]$ , while according to [72], we should use  $\alpha = M$ , and  $0 \leq P \leq \min[M, N]$ . An explanation for the expression in (2.29) of the 1-D Information Theoretic Criteria can be found in Appendix B.

Approach	Penalty function $p(P, N, \alpha)$
1-D AIC	$P \cdot (2 \cdot \alpha - P)$
1-D MDL	$\frac{1}{2} \cdot P \cdot (2 \cdot \alpha - P) \cdot \log(N)$
1-D EDC	$\frac{1}{2} \cdot P \cdot (2 \cdot \alpha - P) \cdot \sqrt{N \cdot \log \cdot (\log N)}$

Tab. 2.3: Penalty functions for 1-D Information Theoretic Criteria

### 2.1.3.2 Modified Exponential Fitting Test (M-EFT)

In this subsection, we present our proposed Modified Exponential Fitting Test (M-EFT), which differs from EFT in a few but crucial aspects. First we generalize the two fundamental equations (2.18) and (2.16) of EFT for any  $M$  and  $N$ , which is shown in detail in Appendix C. Consequently, instead of (2.18), a new derivation of the rate  $q$  for the exponential profile generalized for any  $M$  and  $N$  is proposed. The derivation is shown in Appendix D. As we mentioned previously, another important modification in M-EFT is that, instead of the correction form in (2.16), we apply the prediction form.

According to our derivation in Appendix D, we propose a new equation for  $q$

$$q(\alpha, \beta) = \exp \left\{ -\sqrt{\frac{30}{\alpha^2 + 2}} - \sqrt{\frac{900}{(\alpha^2 + 2)^2} - \frac{720\alpha}{\beta(\alpha^4 + \alpha^2 - 2)}} \right\}, \quad (2.30)$$

where  $\alpha = \min(M, N)$  and  $\beta = \max(M, N)$ . In order to be even more precise in the computation of  $q$ , the following polynomial can be solved

$$(C - 1) \cdot q^{\alpha+1} + (C + 1) \cdot q^\alpha - (C + 1) \cdot q + 1 - C = 0. \quad (2.31)$$

Although from (2.31)  $\alpha + 1$  solutions are possible, we select only the  $q$  that belongs to the interval  $(0, 1)$ . For  $M \leq N$ , (2.30) is equal to (2.18), which means that the PoD of EFT and the PoD of M-EFT for  $M < N$  are the same. Consequently, M-EFT automatically inherits from EFT the property that it outperforms the other matrix-based MOS techniques in the literature for  $M \leq N$  in the presence of white Gaussian noise as shown in Subsection 2.1.4.

In order to check the correctness of our first modification related to the estimation of  $q$  by using (2.30) or (2.31), we compare the probability density functions (PDF) of the relative value of the  $i$ -th eigenvalue of a noise matrix of size  $350 \times 10$ , defined as  $\frac{\widehat{\lambda}_i}{\mathbb{E}\{\lambda_i\}}$ , obtained via EFT and via M-EFT in Figs. 2.8 and 2.9. Note that  $\mathbb{E}\{\lambda_i\}$  is computed via Monte Carlo simulations using trials over the noise matrix.

In Figs. 2.8, 2.9, 2.10 and 2.11, M-EFT\* stands for M-EFT using the fixed  $q(M, N)$  instead of having  $q(P + 1, N)$  in (2.16). The ideal PDF is obtained by computing  $\frac{\lambda_i}{\mathbb{E}\{\lambda_i\}}$  for several realizations. Therefore, in the Ideal PDF, the actual value of  $\lambda_i$  is used for each iteration, instead of  $\widehat{\lambda}_i$  via prediction or correction. In all figures, the PDF obtained in M-EFT and M-EFT\* are quite close.

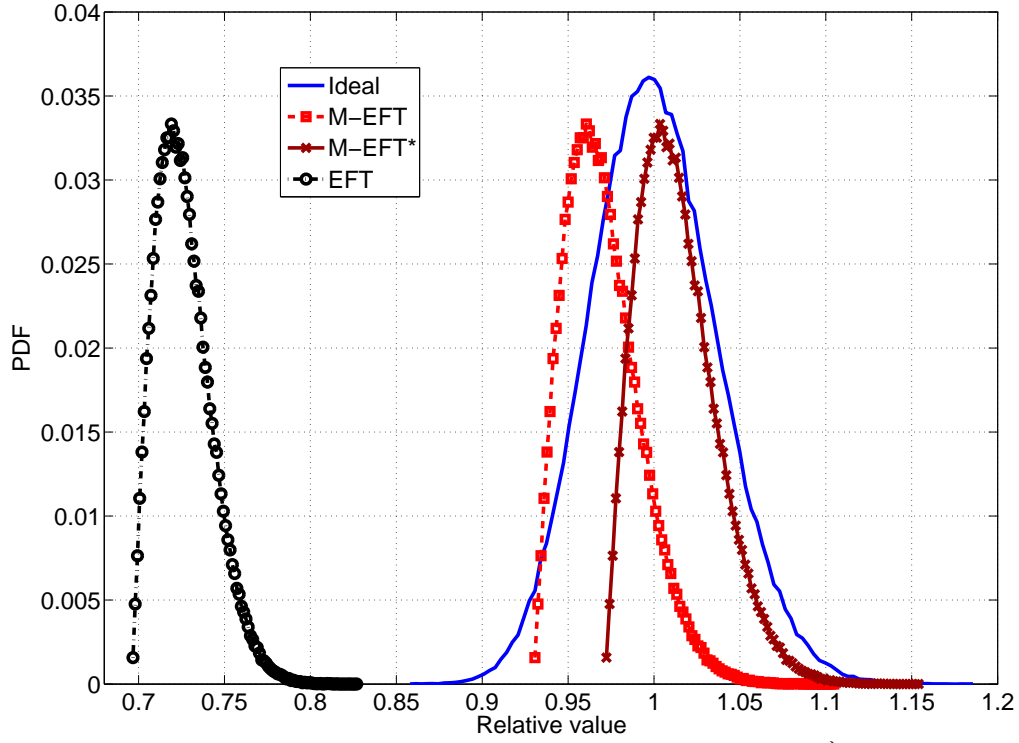


Fig. 2.8: Probability Density Function (PDF) of the relative value  $\frac{\lambda_2}{\mathbb{E}\{\lambda_2\}}$  for an array of size  $M_1 = 350$  in an only white noise scenario. The number of snapshots  $N$  is set to 10. In the EFT and M-EFT, the correction version is used.

In Fig. 2.8, for the PDF of the relative value of the second eigenvalue, there is already some significant difference between the PDF of EFT based eigenvalues and the PDF of the M-EFT based eigenvalues. Actually the PDF of the EFT based eigenvalues should be centered close to 1, as the ideal PDF and also the PDF of the M-EFT based eigenvalues, however, the PDF of the EFT based eigenvalues is centered around 0.72. In Fig. 2.9, for the PDF of the relative value of the 10<sup>th</sup> eigenvalue, the M-EFT fits even better to the ideal curve. On the other hand, the EFT, which is now centered around 0.38, becomes even worse compared to its PDF in Fig. 2.8. In order to make a fair comparison, both M-EFT and EFT use the correction form of  $\widehat{\lambda}_i$  according to (2.16).

Our next proposed modification for the M-EFT is to replace the correction expression in (2.17) by the prediction expression in (2.33).

$$\widehat{\lambda}_{M-P} = (P+1) \cdot \frac{1 - q(P+1, N)}{1 - q(P+1, N)^{P+1}} \widehat{\sigma}^2 \quad (2.32)$$

$$\widehat{\sigma}^2 = \frac{1}{P} \sum_{i=0}^{P-1} \lambda_{M-i}. \quad (2.33)$$

Note that  $\widehat{\lambda}_{M-P}$  in (2.32) is estimated without using  $\lambda_{M-P}$  in (2.33), which characterizes the prediction.

In Figs. 2.10 and 2.11, instead of applying the correction version of M-EFT and EFT as

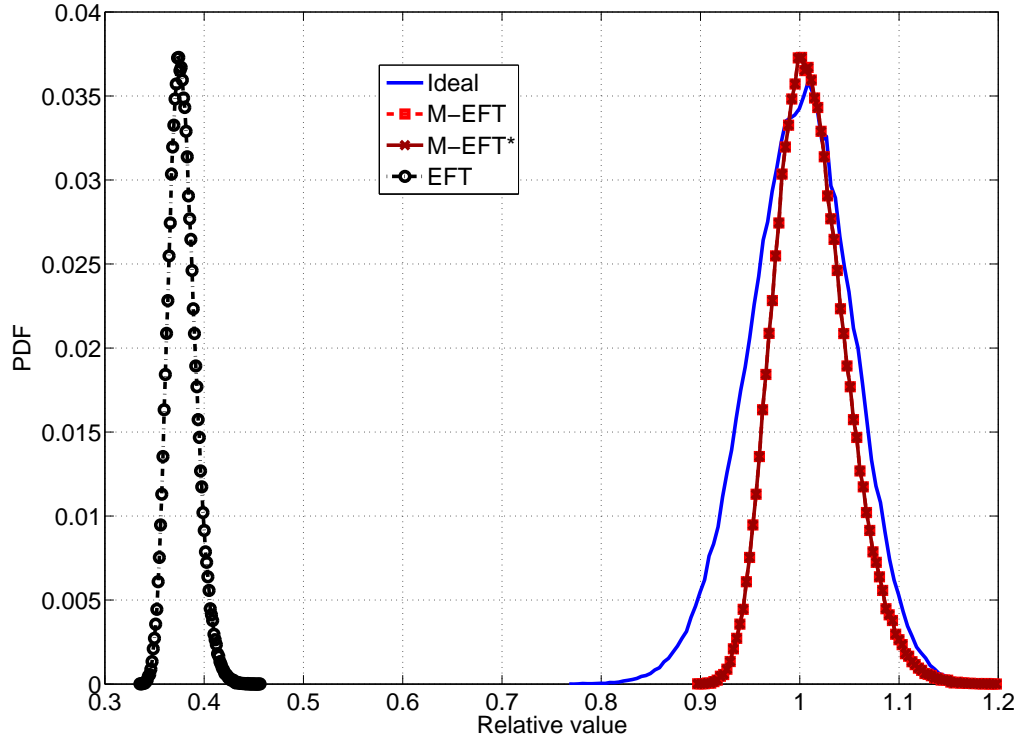


Fig. 2.9: Probability Density Function (PDF) of the relative value  $\frac{\lambda_{10}}{E\{\lambda_{10}\}}$  for an array of size  $M_1 = 350$  in an only white noise scenario. The number of snapshots  $N$  is set to 10. In the EFT and M-EFT, the correction version is used.

in (2.17), we apply the prediction version based on (2.33). Obviously without the information of the  $P+1$  eigenvalue, the estimation via the prediction version is worse for both approaches compared to the one via the correction version. Therefore, although we point out a drawback of the prediction version here, later we present an advantage as compared to the correction version.

Now we compare the M-EFT and the EFT only with respect to the second modification and without being influenced by the first one. To this end, we consider a scenario where  $M \leq N$  in Fig. 2.12. We set  $d = 2$ . Therefore, the greater is the distance between the actual  $\lambda_2$  and the estimated one  $\hat{\lambda}_2$ , the better is the approach. In Fig. 2.12,  $\hat{\lambda}_2 - \lambda_2$  via the M-EFT is greater than  $\hat{\lambda}_2 - \lambda_2$  via the EFT. Due to this reasoning, we include this second modification in the M-EFT.

In Section 2.2, we show that one way to outperform the M-EFT is taking into account the multi-dimensional structure of the data by using the  $R$ -dimensional EFT, only if such multi-dimensional structure is available. However, the correct  $R$ -D EFT is only obtained based on the M-EFT approach, and the  $R$ -D EFT based on the EFT gives us a PoD even worse than the EFT. Such a behavior is expected, since the error in the prediction caused by the EFT is amplified in its  $R$ -dimensional version.

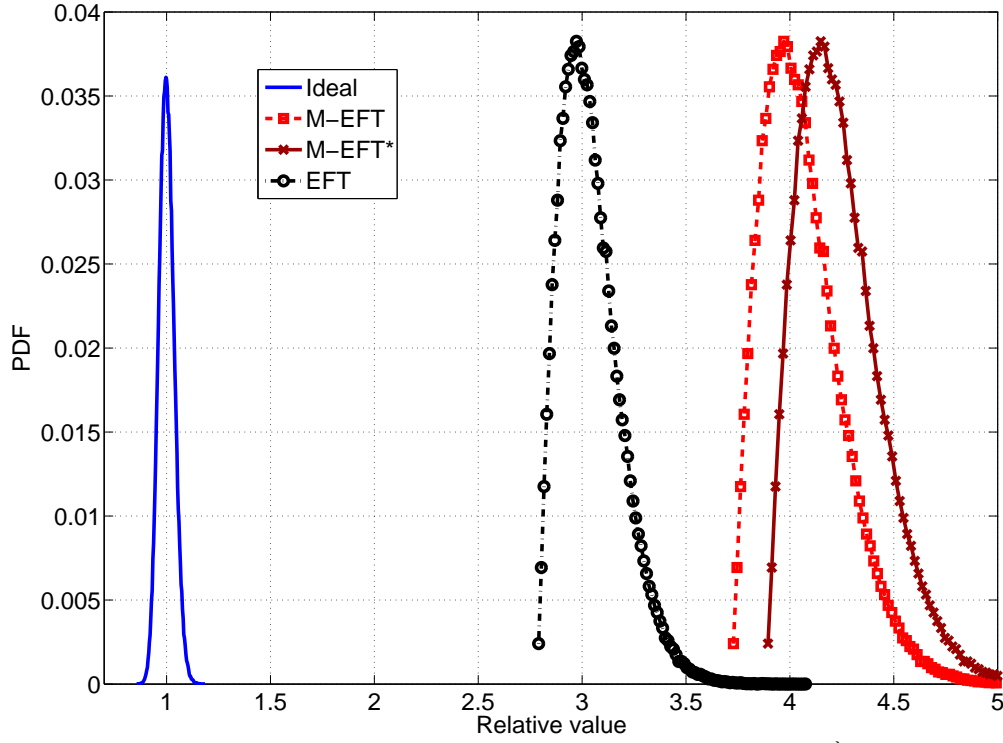


Fig. 2.10: Probability Density Function (PDF) of the relative value  $\frac{\lambda_2}{E\{\lambda_2\}}$  for an array of size  $M_1 = 350$  in an only white noise scenario. The number of snapshots  $N$  is set to 10. In the EFT and M-EFT, the prediction version is used.

### 2.1.3.3 Applying Forward-Backward Averaging (FBA)

A frequently encountered special case is the presence of *undamped* exponentials, i.e.,  $\mu_i = 0 \ \forall i = 1, 2, \dots, d$ . In this case we can enhance the performance further by incorporating forward-backward averaging (FBA) as a preprocessing step [104]. Thereby, we virtually double the number of available snapshots without sacrificing array aperture. Here a data model similar to the one in Subsubsection 2.1.2.1 is assumed.

In [2], for scenarios with centro-symmetric arrays, we have proposed to incorporate FBA [104] to all model order selection schemes and we have shown a significant improvement. Here additionally we also consider scenarios, where the correlation of the sources increases, which are scenarios where the gain of the FBA are even greater.

In the matrix case, forward-backward averaging [104] may be incorporated by replacing the measurement matrix  $\mathbf{X} \in \mathbb{C}^{M \times N}$  by a modified matrix  $\mathbf{Z} \in \mathbb{C}^{M \times 2N}$  given by [43]

$$\mathbf{Z} = [\mathbf{X} \quad \mathbf{\Pi}_M \mathbf{X}^* \mathbf{\Pi}_N], \quad (2.34)$$

where  $\mathbf{\Pi}_n$  represents the  $n \times n$  exchange matrix having ones on its anti-diagonal and zeros

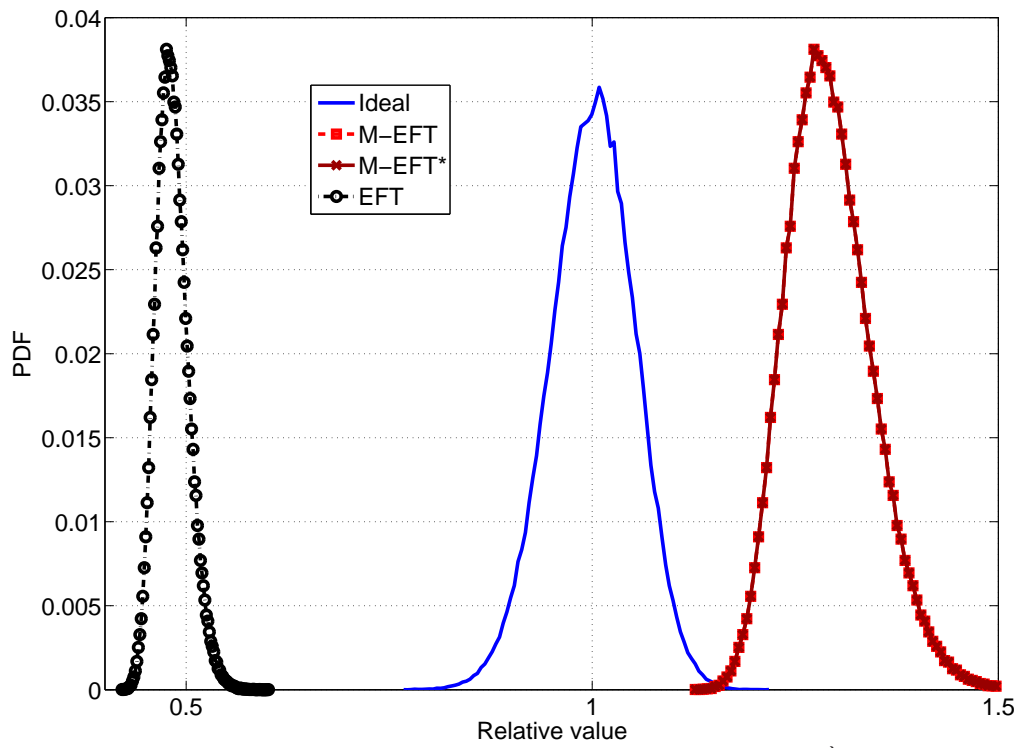


Fig. 2.11: Probability Density Function (PDF) of the relative value  $\frac{\lambda_{10}}{E\{\lambda_{10}\}}$  for an array of size  $M_1 = 350$  in an only white noise scenario. The number of snapshots  $N$  is set to 10. In the EFT and M-EFT, the prediction version is used.

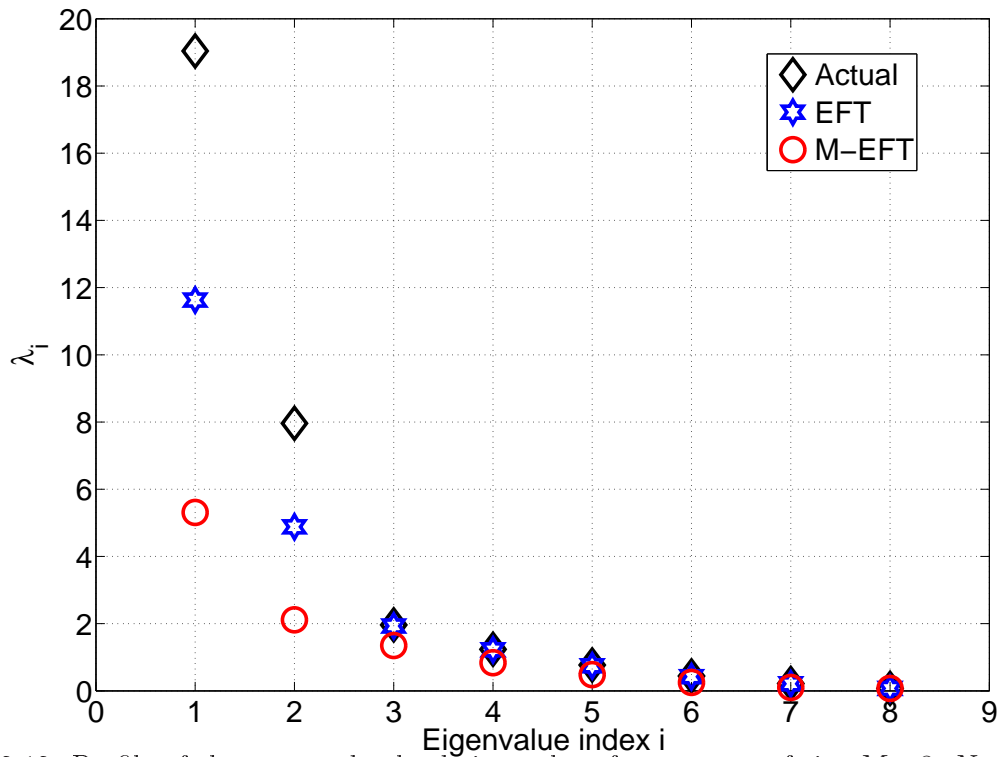


Fig. 2.12: Profile of the expected-valued eigenvalues for an array of size  $M = 8$ ,  $N = 10$ ,  $d = 2$  and SNR = 25 dB.



elsewhere, as shown as follows

$$\mathbf{\Pi}_n = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}. \quad (2.35)$$

In Subsection 2.1.4, we show simulations incorporating the FBA to the model order selection schemes considering correlated sources.

### 2.1.4 Evaluation of Approaches

In this subsection, we evaluate the performance of all matrix-based model order selection techniques presented previously in terms of probability of correct detection (PoD) versus SNR via Monte Carlo simulations for different scenarios.

In Fig. 2.13, 1-D AIC and 1-D MDL outperform AIC and MDL respectively with a gain of 0.5 dB. The gap between 1-D AIC/MDL and AIC/MDL is a function of  $M$  and  $N$ . Therefore, increasing  $M$ , the gain of 1-D AIC/MDL over AIC/MDL increases. Note that 1-D AIC/MDL are fundamental for the multi-dimensional extensions of AIC/MDL. Surprisingly, 1-D EDC and EDC present exactly the same performance for  $M > N$ .

In Figs. 2.14 and 2.15, we compare the state-of-the-art eigenvalue based model order estimation techniques. For the EFT, we set  $P_{fa} = 10^{-3}$ . We consider an array size such that  $N > M$  in Fig. 2.14 and also a small number of snapshots. Due to the small number of snapshots, the PoD of AIC and MDL are very degraded. In Fig. 2.14, the EFT and the SURE present the best performance, while the EFT and the NEMO have the highest PoD in Fig. 2.15.

As we have mentioned previously, the Probability of correct Detection (PoD) of the ITC schemes depends on the array size. For example, in Fig. 2.14, where  $M = 8$  and  $N = 10$ , EDC outperforms MDL, which outperforms AIC. On the other hand, in Fig. 2.15, where  $M = 100$  and  $N = 10$ , AIC outperforms MDL, which outperforms EDC. As we have also mentioned finding the best penalty function for different scenarios is not in the scope of this thesis.

Note that the EFT is only slightly outperformed for  $M > N$  in Fig. 2.15. Therefore, improving the EFT for scenarios where  $M > N$  means obtaining a MOS scheme that outperforms the other techniques in the literature for all values of  $M$  and  $N$ . This is another important motivation for the Modified Exponential Fitting Test (M-EFT).

In Fig. 2.16, where  $M > N$ , M-EFT outperforms all the state-of-the-art matrix-based model order selection approaches in the literature, and the gap between the M-EFT, which is the best one, and the second best technique is approximately 5 dB. Therefore, M-EFT has the best PoD among the considered state-of-the-art matrix-based MOS techniques for scenarios with the presence of white Gaussian noise.

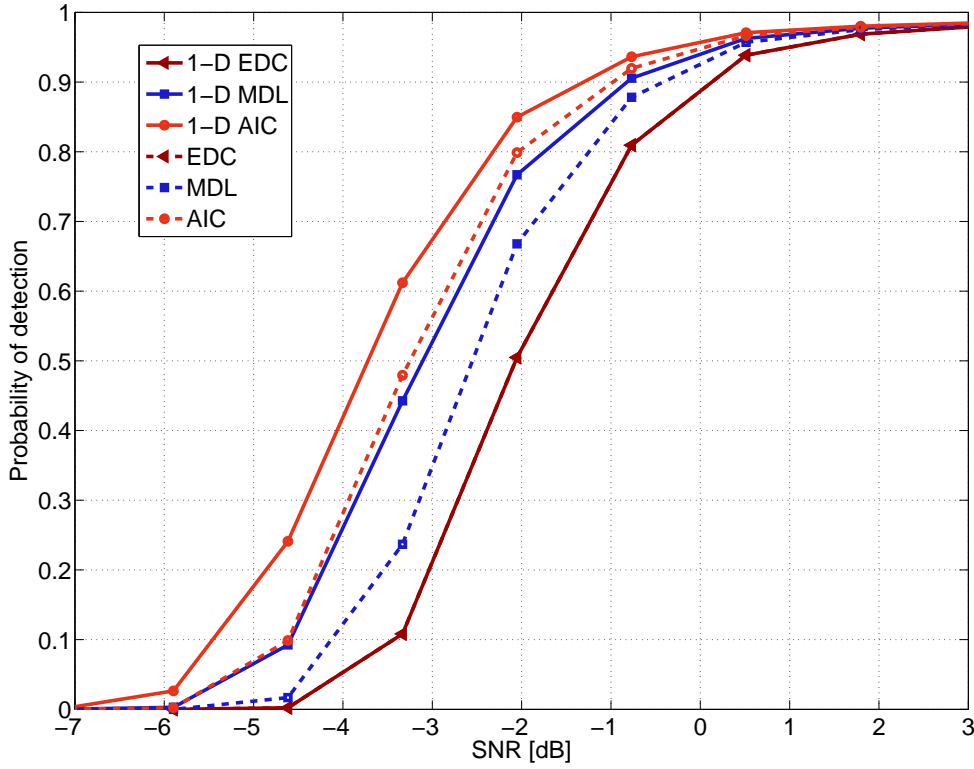


Fig. 2.13: Probability of correct Detection (PoD) vs. SNR considering  $M = 100$ ,  $N = 10$  and  $d = 3$ . Comparison between the proposed 1-D ITC and ITC.

In Figs. 2.17, 2.18 and 2.19, ESTER and SAMOS are outperformed by several model order selection schemes. Therefore, in case of white Gaussian noise, the subspace based matrix model order selection schemes should not be applied.

In (2.8), we have introduced the colored noise model. Now for the simulations we consider the following structure for the noise correlation matrix  $\mathbf{R}_{nn} \in \mathbb{C}^{M \times M}$

$$\mathbf{R}_{nn} = \sigma_n^2 \cdot \mathbf{L} \cdot \mathbf{L}^H = \sigma_n^2 \cdot \begin{bmatrix} 1 & \rho^* & \dots & (\rho^*)^{M-1} \\ \rho & 1 & \dots & (\rho^*)^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{M-1} & \rho^{M-2} & \dots & 1 \end{bmatrix}, \quad (2.36)$$

where  $\rho$  is the correlation coefficient.

In Figs. 2.20 and 2.21, the model order is estimated in the presence of colored Gaussian noise, whose correlation matrix in (2.36) has a correlation coefficient  $\rho = 0.9$ . As expected, the performance of most matrix-based model order selection techniques has been very degraded, except RADOI, SAMOS and ESTER, which have been proposed for scenarios with colored noise. Therefore, in Figs. 2.20 and 2.21, only these three techniques still estimate the model order with a reasonable PoD.

Note that in Fig. 2.21,  $d = 4$  is greater than the maximum limit for the candidate value  $P$  for SAMOS. Therefore, only RADOI and ESTER can estimate the model order.

In Fig. 2.22, UCMS stands for users of MIMO channel sounders scheme, which denotes

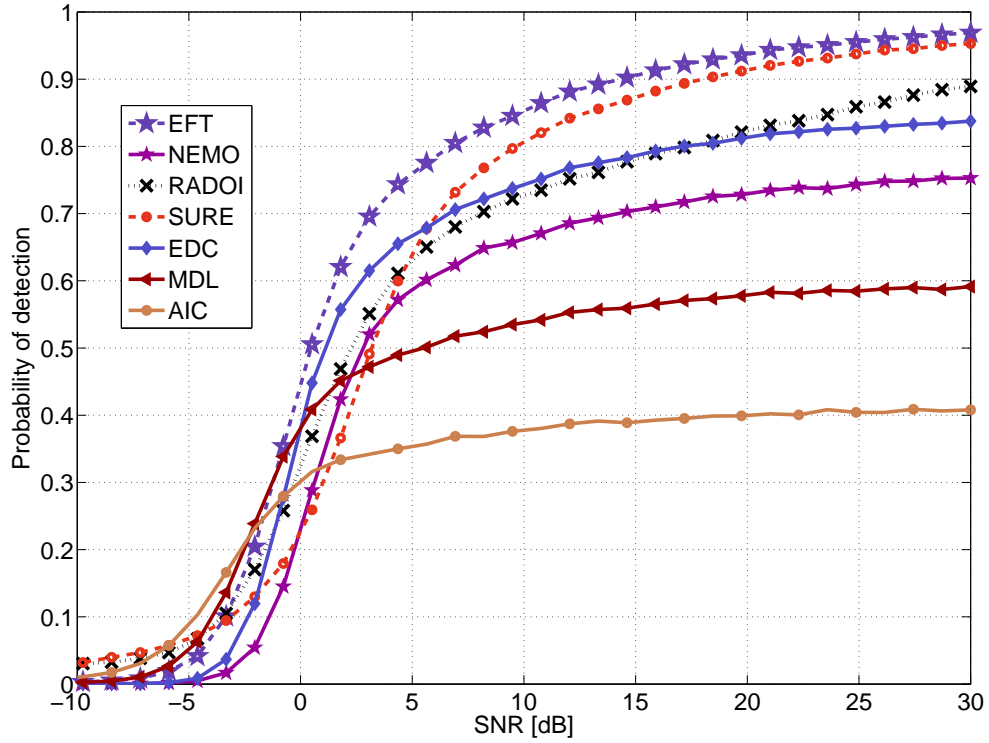


Fig. 2.14: Probability of correct Detection (PoD) vs. SNR considering  $M = 8$ ,  $N = 10$  and  $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques found in the literature.

the MOS-MNP approach from Subsubsection 2.1.2.2. The suffix Low, Med and Hi indicates the respective variations of the design parameter  $\alpha_{PW} = 1$ ,  $\alpha_{PW} = 1.5$ , and  $\alpha_{PW} = 2$ , where  $\alpha_{PW}$  is also defined in Subsubsection 2.1.2.2. Moreover, the suffix Mag and Vand denotes that information about  $\mathbf{S}$  is applied to improve the estimation. Mag means that the magnitude of the elements of  $\mathbf{S}$  should be equal to one, while Vand means that  $\mathbf{S}$  has a Vandermonde structure.

Comparing all the curves in Fig. 2.22, the two with highest Probability of correct Detection (PoD) are UCMS Med and UCMS Mag Med. In general, for  $\alpha_{PW} = 1$ , and for  $\alpha_{PW} = 2$ , a degraded performance is obtained. For instance, in Fig. 2.22 for SNR = 50 dB, UCMS Med with  $\alpha_{PW} = 1.5$  has a PoD = 0.98, while UCMS Hi with  $\alpha_{PW} = 2$  has a PoD = 0, and UCMS Low with  $\alpha_{PW} = 1$  has a PoD = 0.75. Therefore, small variations in  $\alpha_{PW}$  imply a severe degradation of the PoD. Hence, one first disadvantage of MOS-MNP is the instability due to  $\alpha_{PW}$ . Note that practitioners usually set  $\alpha_{PW} = 1$ , which in this example would give a considerable degradation in the model order estimation.

In Fig. 2.23, we select the two best curves of Fig. 2.22 to compare to the best state-of-the-art model order selection schemes. In general, the two selected MOS-MNP approaches are not comparable to the M-EFT, EFT, and 1-D MDL. However, they are comparable in this scenario to the 1-D AIC and the NEMO. Therefore, their performance is worse or similar than the state-of-the-art model order selection techniques in the literature.

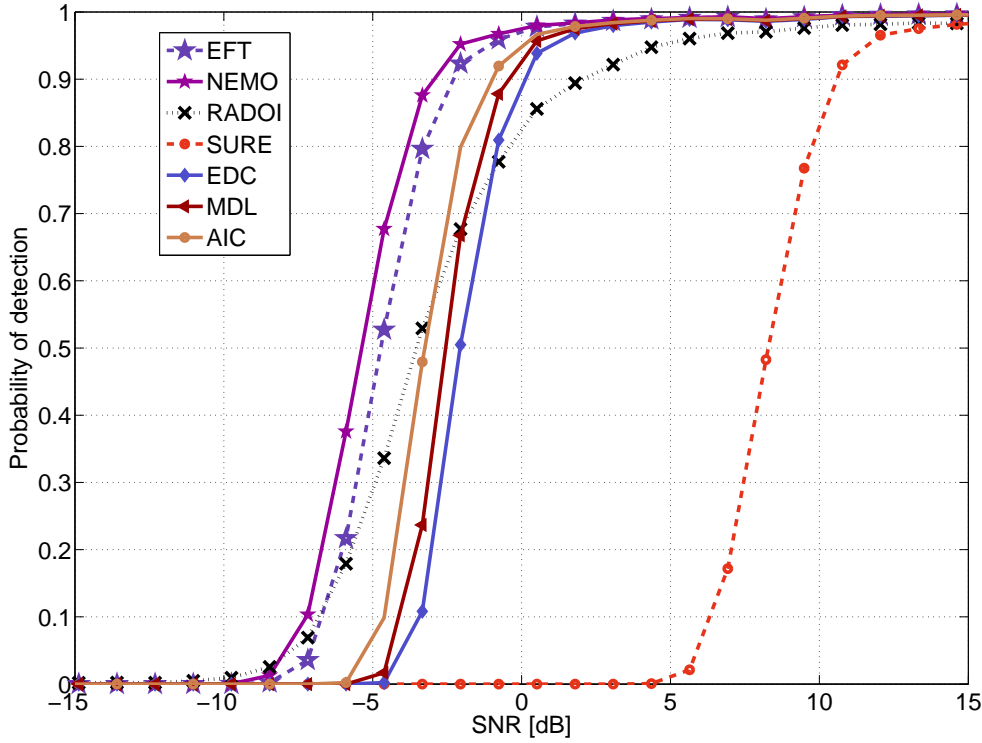


Fig. 2.15: Probability of correct Detection (PoD) vs. SNR considering  $M = 100$ ,  $N = 10$  and  $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques found in the literature.

Note that  $\alpha_{PW}$  is a parameter that depends on how stable the noise power is in time. If the noise power from the first step to the second step varies 50 %, according to Fig. 2.22, the model order estimation is drastically degraded. Hence, MOS-MNP is very instable. Moreover, since in the best choice of  $\alpha_{PW}$ , the PoD with MOS-MNP is close to the model order selection techniques found in the literature, it is certainly a much better option to use any state-of-the-art matrix-based model order selection present in the literature than using MOS-MNP.

In (2.7), we have introduced the model for correlated sources. Now for the simulations we consider the following structure for the noise correlation matrix  $\mathbf{R}_{ss} \in \mathbb{C}^{M \times M}$

$$\mathbf{R}_{ss} = \sigma_s^2 \cdot \mathbf{L} \cdot \mathbf{L}^H = \sigma_s^2 \cdot \begin{bmatrix} 1 & \rho_s^* & \dots & (\rho_s^*)^{M-1} \\ \rho_s & 1 & \dots & (\rho_s^*)^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_s^{M-1} & \rho_s^{M-2} & \dots & 1 \end{bmatrix}, \quad (2.37)$$

where  $\rho_s$  is the correlation coefficient.

In Figs. 2.24 and 2.25, we depict the PoD versus the  $\rho_s$  for  $\rho_s = 0, \dots, 0.999$ . In Fig. 2.24, the model order selection schemes are applying without FBA, while in Fig. 2.25 FBA is applied. By applying FBA, an improved gain is obtained in all techniques, except for the EFT.

In Figs. 2.26 and 2.27, we depict the Pfa versus the  $\rho_s$  for  $\rho_s = 0, \dots, 0.999$ . In Fig. 2.26, the model order selection schemes are applying without FBA, while in Fig. 2.27 FBA is applied.

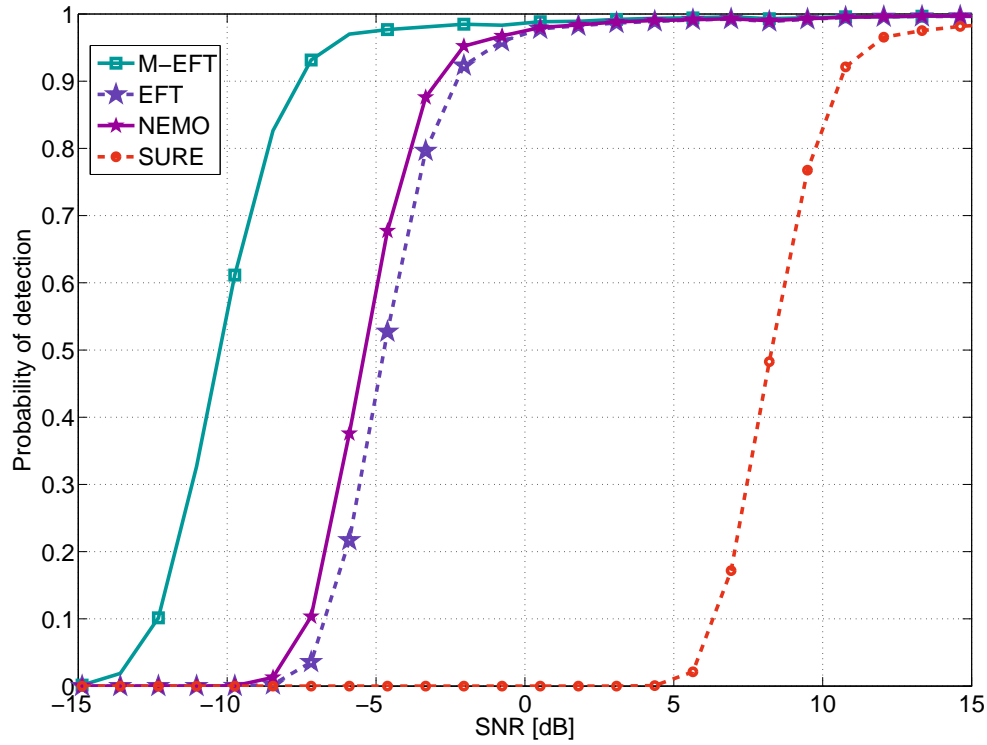


Fig. 2.16: Probability of correct Detection (PoD) vs. SNR considering  $M = 100$ ,  $N = 10$  and  $d = 3$ . Comparison between the proposed M-EFT and the state-of-the-art eigenvalue based matrix-based model order selection techniques in the literature.

In Figs. 2.28 and 2.29, we depict the Pfa versus the  $\rho_s$  for  $\rho_s = 0, \dots, 0.999$ . In Fig. 2.28, the model order selection schemes are applying without FBA, while in Fig. 2.29 FBA is applied.

In Figs. 2.30 and 2.31, we depict the PoD versus Pfa for  $\rho_s = 0, \dots, 0.999$ . In Fig. 2.30, the model order selection schemes are applying without FBA, while in Fig. 2.31 FBA is applied.

### 2.1.5 Main Results of the Section

In this section, we have proposed the 1-D AIC, 1-D MDL and the M-EFT, which are extensions of the AIC, MDL, and EFT, for the case that  $M > N$ . Moreover, we have performed comparisons of the state-of-the-art matrix-based model order selection techniques, from which we can summarize the following main results

- 1) In case of white Gaussian noise scenarios, the M-EFT achieves the highest PoD, and particularly for  $M \gg N$ , the PoD of M-EFT is far superior than any other matrix-based model order selection technique already proposed.
- 2) In presence of colored noise, the best approach is RADOI. Note that in this case obviously the value of the correlation coefficient  $\rho$  also plays a very important role. One drawback of the RADOI, the SAMOS and the ESTER, which are applied for colored noise, is

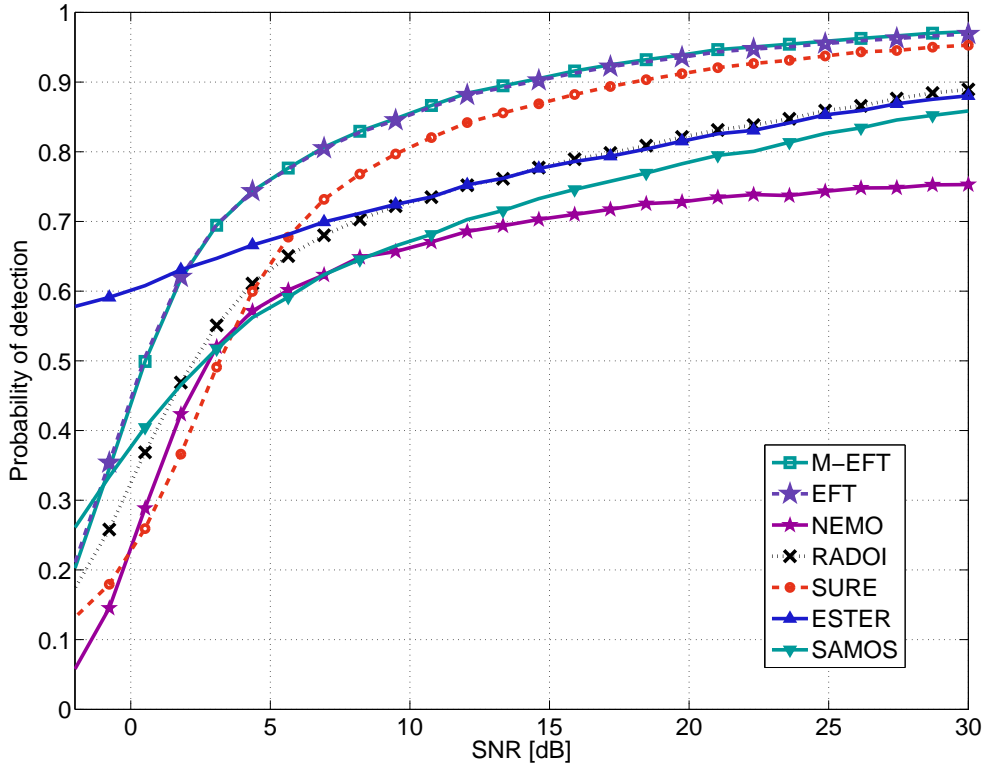


Fig. 2.17: Probability of correct Detection (PoD) vs. SNR considering  $M = 8$ ,  $N = 10$  and  $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques and the subspace based ones.

that their performance is very degraded compared to several matrix-based model order selection techniques for  $M > N$  in the presence of white noise.

- 3) For scenarios where  $M > N$ , the ITC should be modified just as we have proposed the 1-D AIC and the 1-D MDL. Our proposed modifications are fundamental for the extension of AIC and MDL to the multi-dimensional version.

In the next section, we compare only the best state-of-the-art matrix-based model order selection techniques to the multi-dimensional model order selection schemes in order to reduce the number of curves to be plotted.

Note that the model order estimation in the presence of colored noise can be improved by applying prewhitening schemes presented in Chapter 4. However, in many real-time applications samples with the absence of signals may not be available. Therefore, to estimate the noise statistics and the signal parameters jointly as proposed in Section 4.3, it is first necessary to estimate the model order.

In Table 2.4, we summarize the scenarios to apply the different techniques shown in this section. Note that for SURE, there is a  $\eta$ , which is the percentage of eigenvalues which are not assume as noise eigenvalues. Theoretically SURE and RADOI should work for  $d = 0$ , however this is not observed according to the simulations in Appendix A.

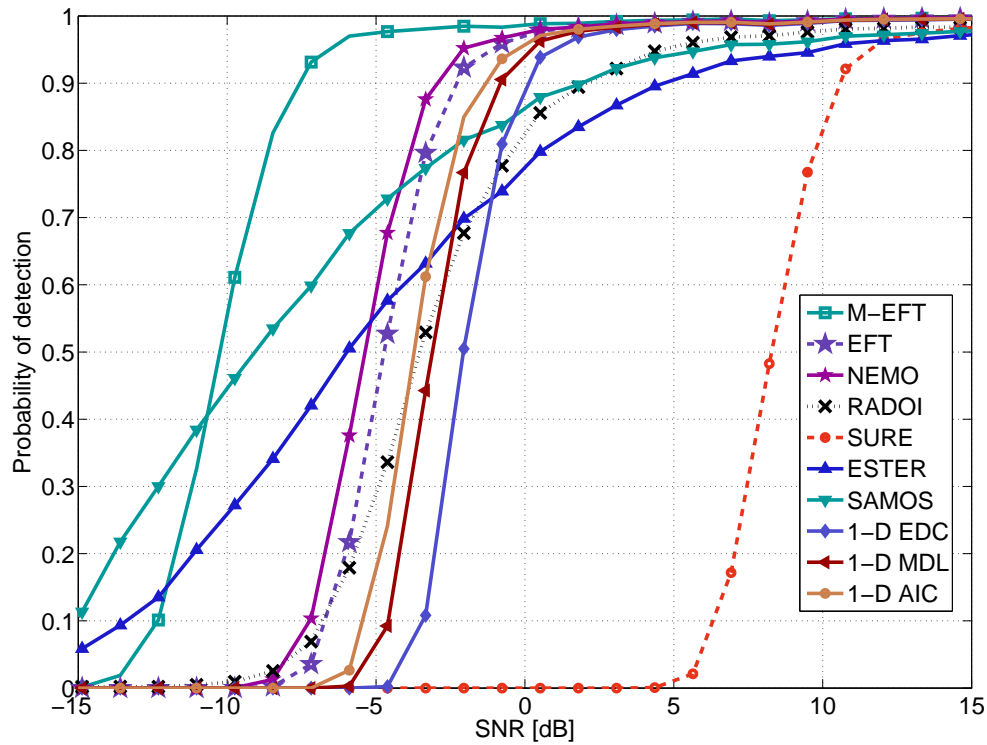


Fig. 2.18: Probability of correct Detection (PoD) vs. SNR considering  $M = 100$ ,  $N = 10$  and  $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques and the subspace based ones.

Scheme	Minimum $d$	Maximum $d$	Noise	Performance
AIC [98]	0	$\min(M, N) - 1$	Wht	Good only if $N \gg M$
MDL [98]	0	$\min(M, N) - 1$	Wht	Good only if $N \gg M$
EDC [106]	0	$\min(M, N) - 1$	Wht and clr	Good only if $N \gg M$
1-D AIC [7]	0	$\min(M, N) - 1$	Wht	Better than AIC if $M > N$
1-D MDL [7]	0	$\min(M, N) - 1$	Wht	Better than MDL if $M > N$
1-D EDC	0	$\min(M, N) - 1$	Wht and clr	Same as EDC
NEMO [72]	0	$\min(M, N) - 1$	Wht	Better than AIC and MDL
SURE [95]	0	$\eta \cdot \min(M, N) - 1$	Wht and clr	Better than MDL
ESTER [13]	1	$\min(M, N)$	Wht and clr	Better than EDC
SAMOS [75]	1	$\min(\lfloor \frac{M-1}{2} \rfloor, N)$	Wht and clr	Better than ESTER
RADOI [80]	0	$\min(M, N) - 1$	Wht and clr	Better than SAMOS
EFT [42, 79]	0	$\min(M, N) - 1$	Wht	Best for $M < N$
M-EFT [2, 7]	0	$\min(M, N) - 1$	Wht	Best for $\forall M$ and $N$

Tab. 2.4: Summarized table comparing characteristics of the matrix-based model order selection schemes.

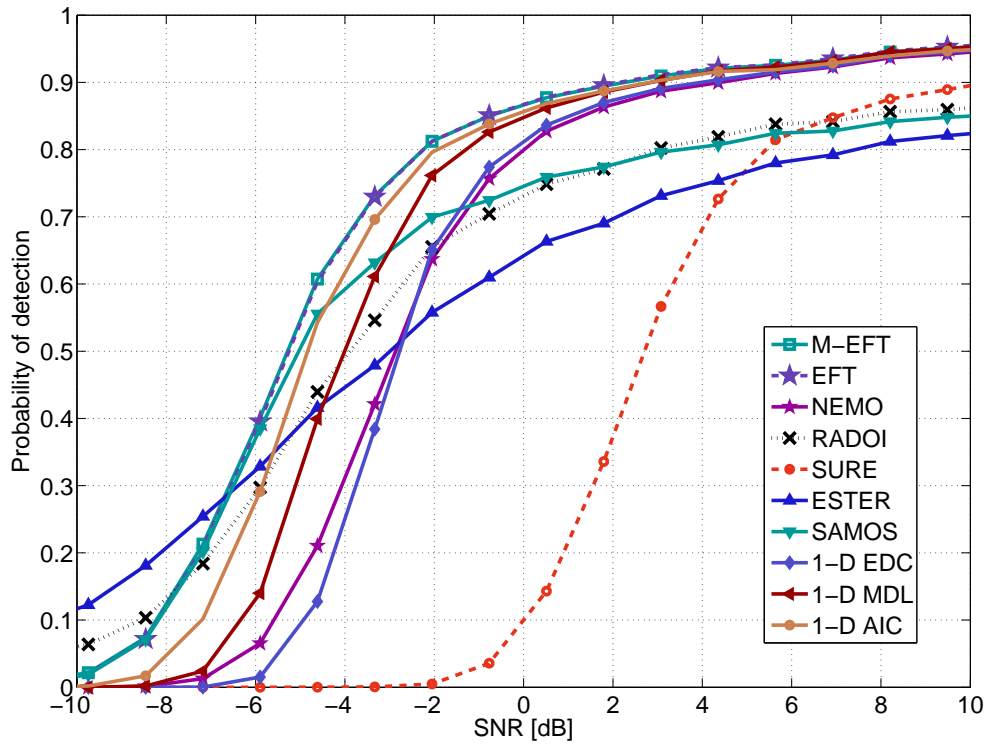


Fig. 2.19: Probability of correct Detection vs. SNR for an array of size  $M = 15$ . The number of snapshots  $N$  is set to 10 and the number of sources  $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques and the subspace based ones.



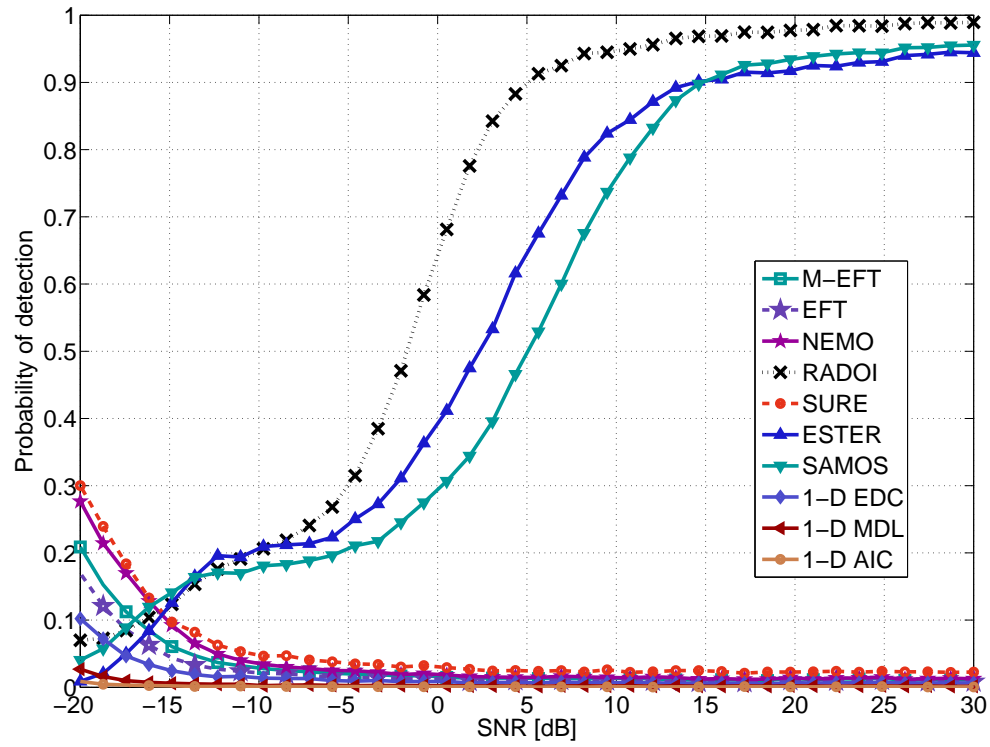


Fig. 2.20: Probability of correct Detection (PoD) vs. SNR, considering  $\rho = 0.9$ ,  $M = 8$ ,  $N = 10$  and  $d = 2$ .

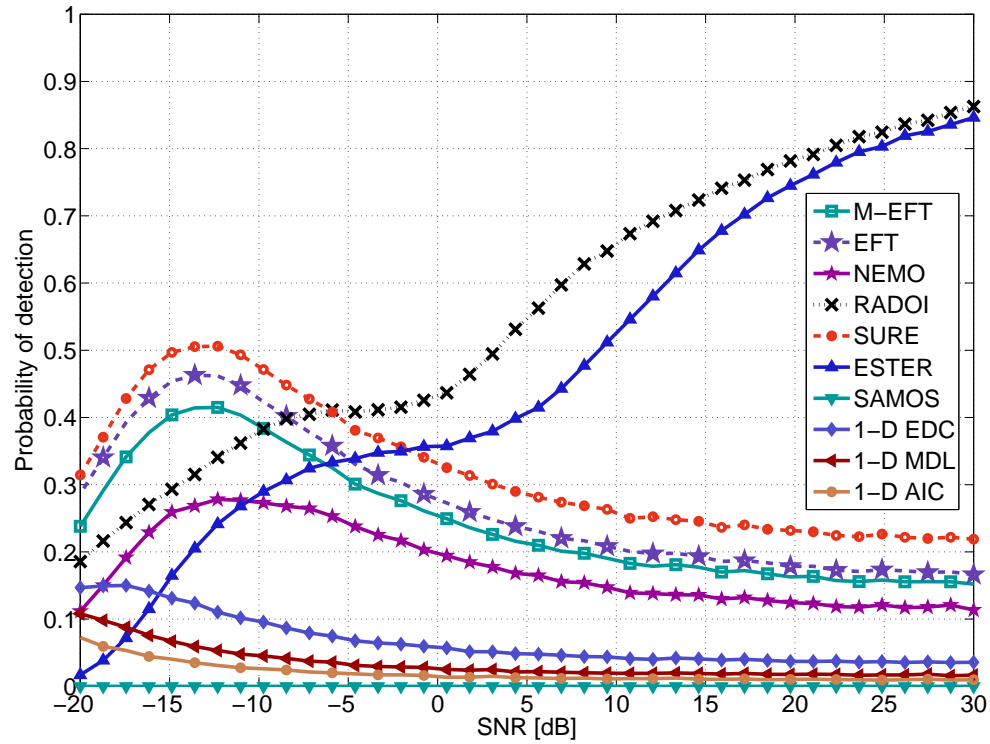


Fig. 2.21: Probability of correct Detection (PoD) vs. SNR, considering  $\rho = 0.9$ ,  $M = 8$ ,  $N = 10$  and  $d = 4$ .

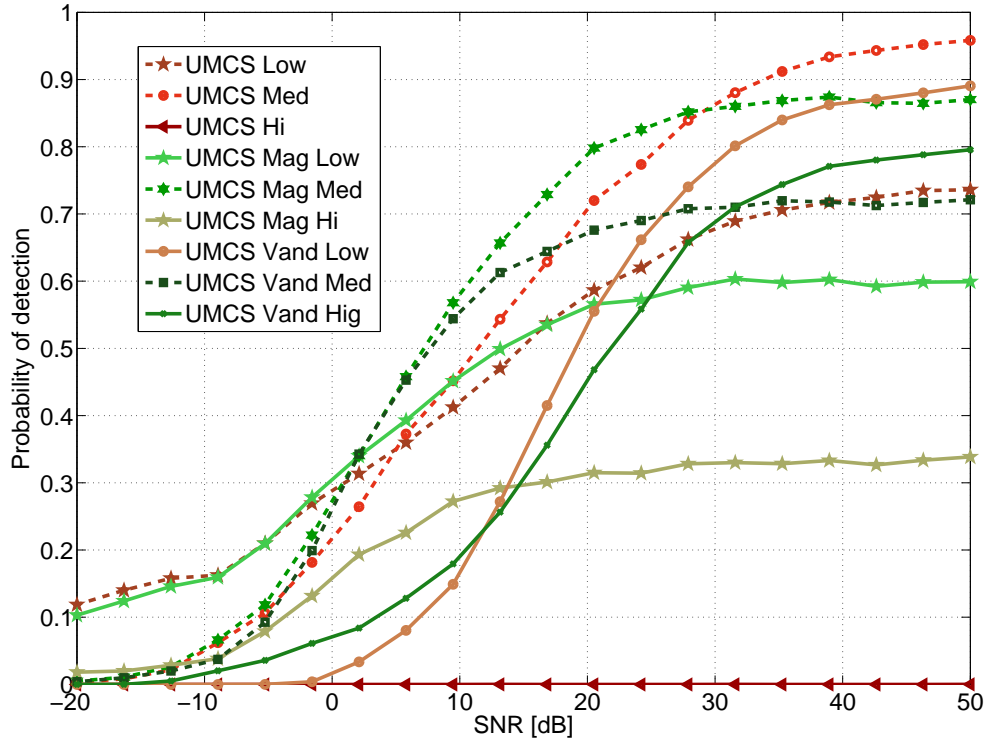


Fig. 2.22: Probability of correct Detection vs. SNR for an array of size  $M_1 = 10$ . The number of snapshots  $N$  is set to 20 and the number of sources  $d = 3$ .

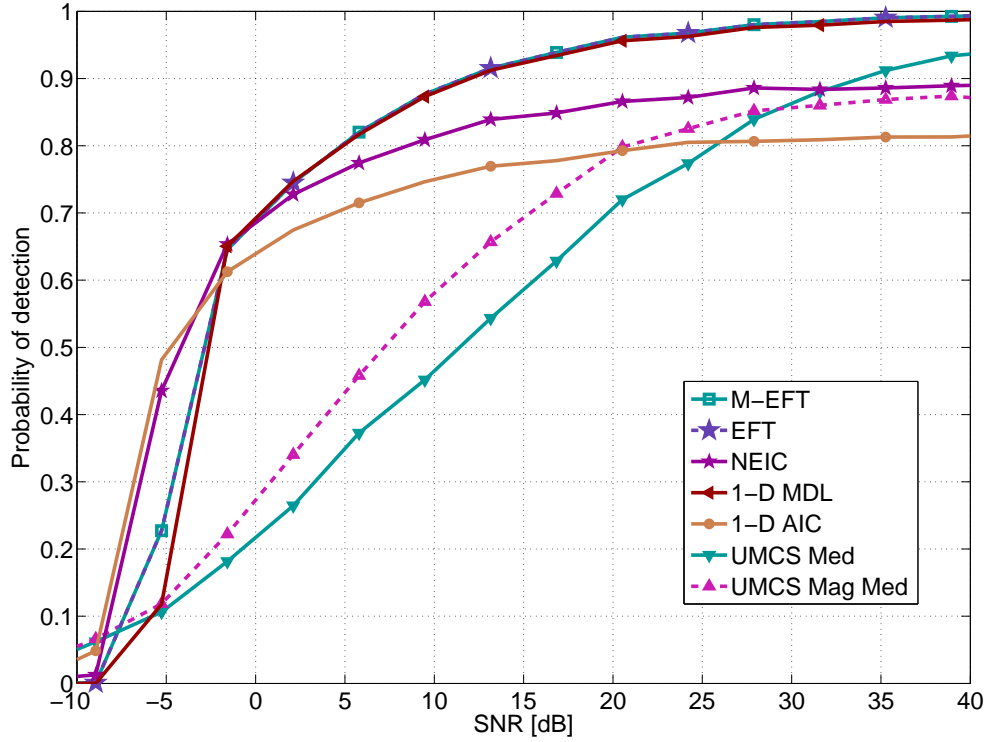


Fig. 2.23: Probability of correct Detection vs. SNR for an array of size  $M_1 = 10$ . The number of snapshots  $N$  is set to 20 and the number of sources  $d = 3$ .

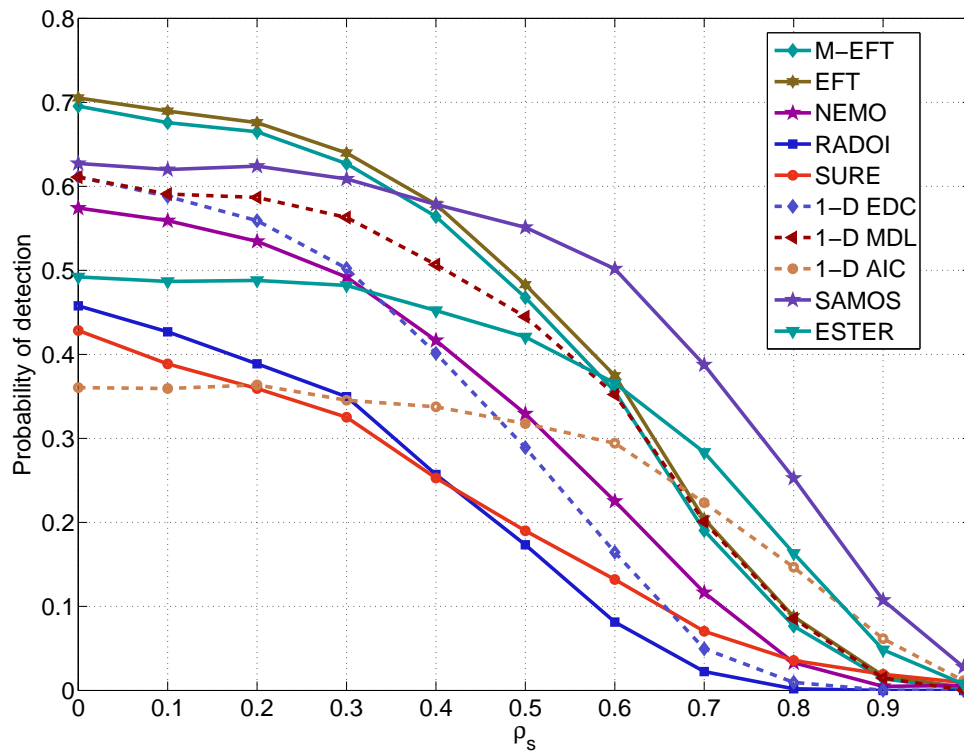


Fig. 2.24: PoD versus  $\rho_s$  for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

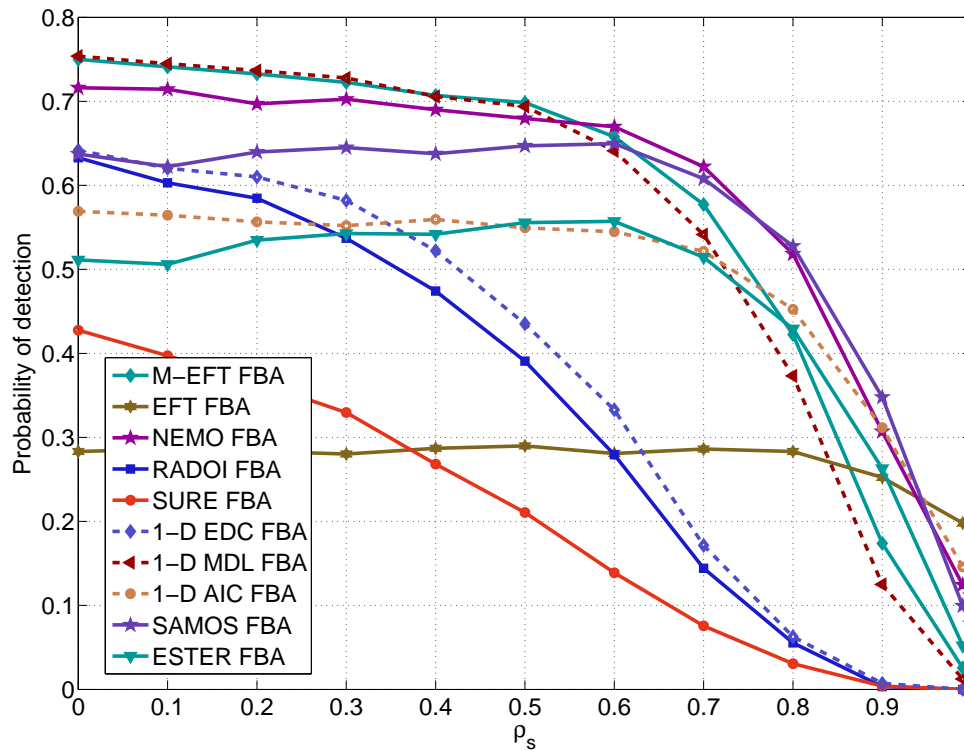


Fig. 2.25: PoD versus  $\rho_s$  for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

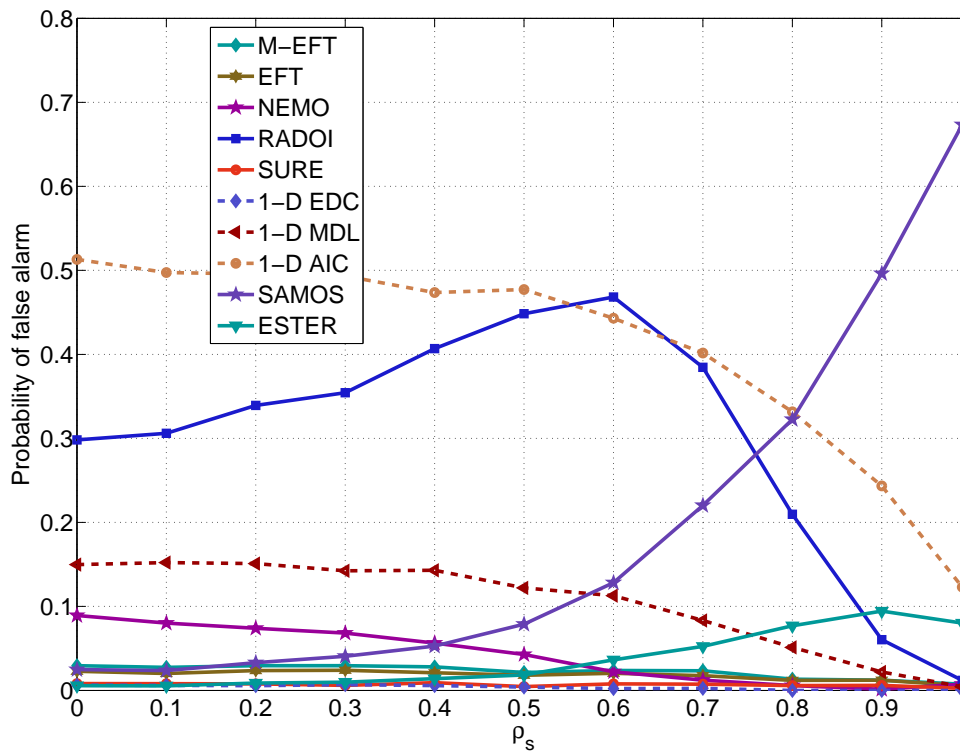


Fig. 2.26: Pfa versus  $\rho_s$  for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

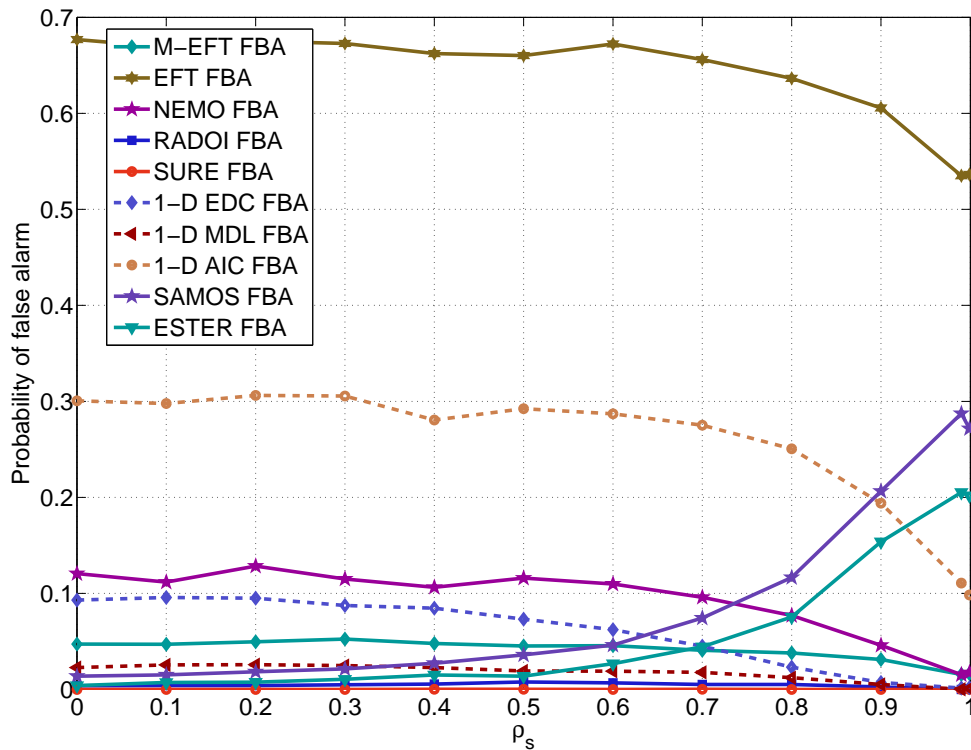


Fig. 2.27: Pfa versus  $\rho_s$  for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

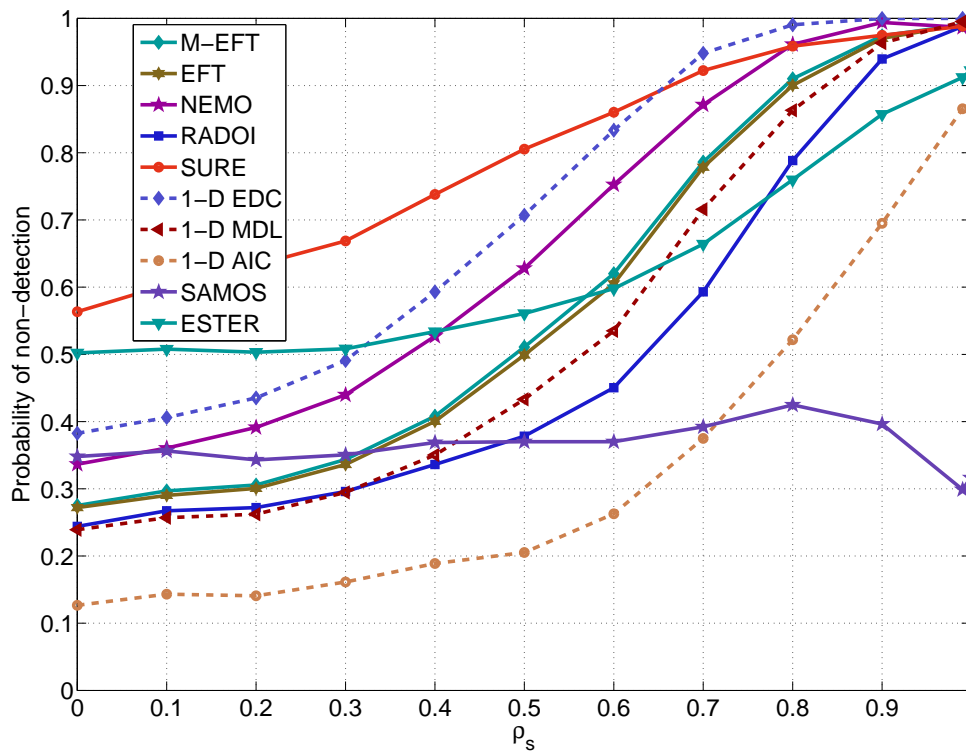


Fig. 2.28: Pnd versus  $\rho_s$  for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

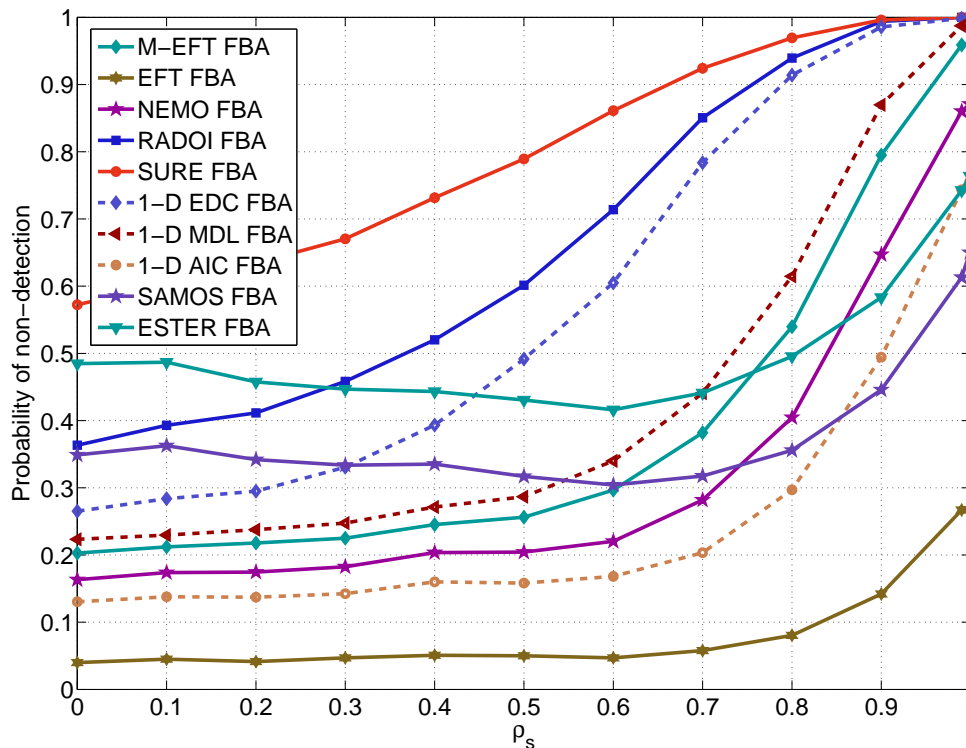


Fig. 2.29: Pnd versus  $\rho_s$  for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

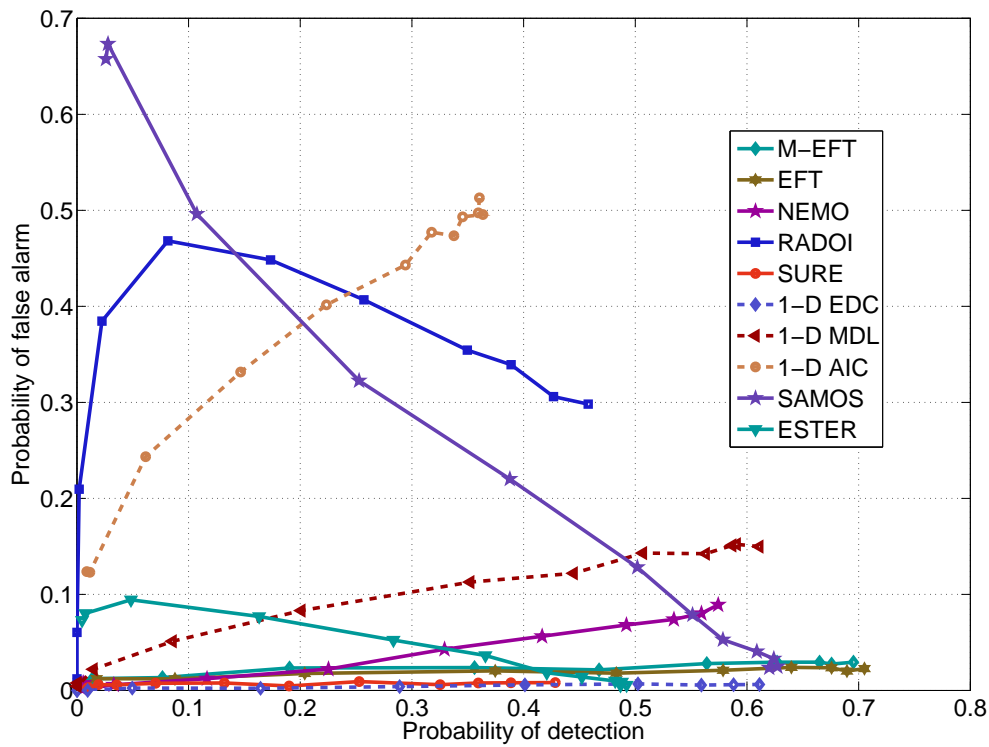


Fig. 2.30: PoD versus Pfa for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

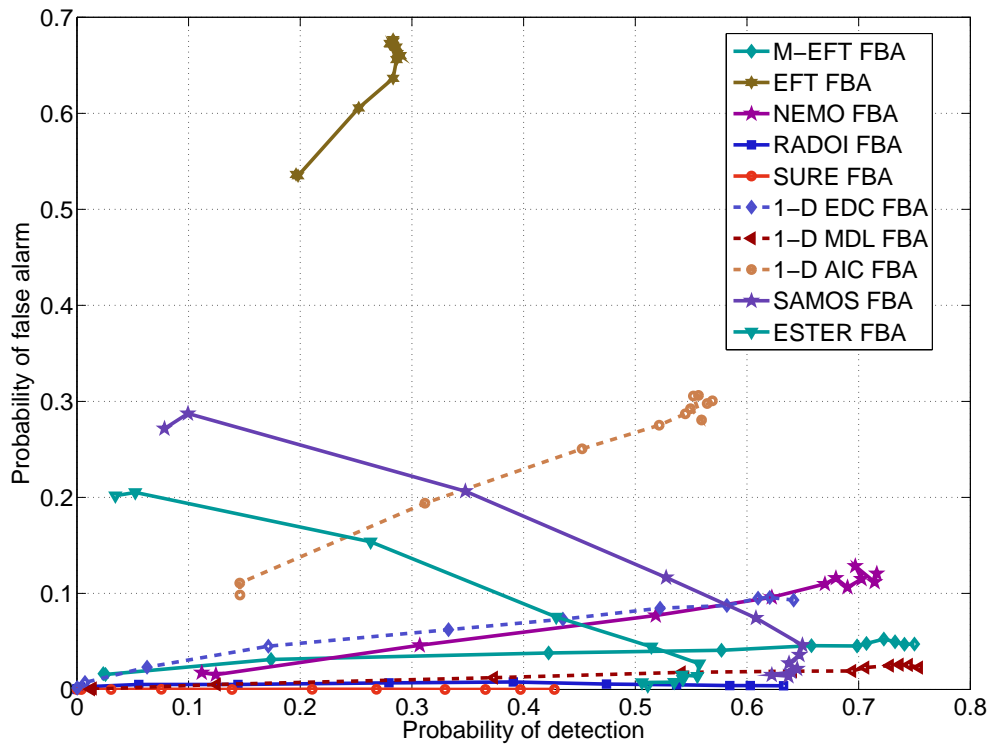


Fig. 2.31: PoD versus Pfa for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We fix the number of sources  $d = 3$ .

## 2.2 Multi-dimensional Model Order Selection

As presented previously the multi-dimensional model order selection is the combination of two important fields: model order selection and tensor calculus. Therefore, in this section, we propose to use tensor calculus in order to upgrade matrix-based MOS schemes. Moreover, we also propose a multi-dimensional model order selection scheme based on the closed-form PARAFAC algorithm [84], which is only applicable to multi-dimensional data. In general, the PoD of our proposed multi-dimensional MOS schemes is much better than the PoD of schemes already proposed in the literature.

First let us show a short bibliographic review and performance analysis of the state-of-the-art multi-dimensional model order selection techniques proposed by the other authors and the one-dimensional model order selection schemes in Section 2.1. One of the most well-known multi-dimensional model order selection schemes is the CORE CONSistency DI-Agnostics (CORCONDIA) [15]. We have shown in [1] that its objective version Threshold-CORCONDIA (T-CORCONDIA) has a similar performance than 1-D AIC/MDL. Similar to T-CORCONDIA with PoD  $\approx 78$  % [1], other multi-dimensional schemes with limited performance<sup>4</sup>, i.e., are DIFFIT [61] with PoD  $\approx 80$  % [58], and Fast DIFFIT [59] also with PoD  $\approx 80$  %. The convex hull based method [61, 19] has a PoD slightly superior than DIFFIT [58]. However, also similarly to T-CORCONDIA [1] such PoDs are only reached for intermediate or high SNR regimes.

Note, for instance, that the M-EFT reaches a PoD  $\approx 99.99$  % depending on the  $P_{\text{fa}}$ , and even for low SNR as shown in Fig. 2.16, the performance of the M-EFT is quite good, i.e., SNR =  $-5$  dB gives a PoD  $\approx 98$  %. Based on our results obtained in Section 2.1, and comparing the multi-dimensional model order selection schemes to the matrix-based model order selection schemes, we can conclude that the M-EFT proposed by us in [2] for white Gaussian noise and RADOI proposed in [80] for colored Gaussian noise achieve by means of simulations the best PoD for their respective types of noise. In Subsection 2.2.4, we use these results, and comparisons between the multi-dimensional model order selection schemes and the matrix-based model order selection schemes are provided.

Hence, our strategy is to incorporate the tensor calculus to the M-EFT approach in order to improve even more the model order estimation in case of white Gaussian noise. This multi-dimensional version of the M-EFT is called *R*-Dimensional Exponential Fitting Test (*R*-D EFT), and it is based on the concept of global eigenvalues also proposed by us. Moreover, we have also applied the concept of global eigenvalues to extend the 1-D AIC and 1-D MDL to their multi-dimensional forms denoted as *R*-D AIC and *R*-D MDL.

In Section 2.1, the matrix-based model order selection techniques are classified according to Table 2.1. Similarly to Table 2.1, we also propose a classification for the multi-dimensional model order selection techniques, which is shown in Table 2.5. In the multi-dimensional case,

---

<sup>4</sup> Limited in the sense that for high and intermediate SNR regimes a PoD  $\approx 100$  % is not achieved. Moreover, for these approaches and considering low SNR regimes, a PoD  $\approx 0$  % is obtained.

	Detection	Combined Detection-Estimation
Global Eigenvalue based	$R$ -D AIC, $R$ -D MDL, and $R$ -D EFT	
Tensor Data-based		T-CORCONDIA, CFP-MOS
$R$ -D Subspace based	CFP-MOS	

Tab. 2.5: Classification of multi-dimensional model order selection techniques

instead of using eigenvalues, we apply the global eigenvalues. Therefore, the multi-dimensional techniques can be also classified as global eigenvalue based. Moreover, instead of subspace based, in the multi-dimensional case, each dimension has a subspace. Therefore, they are classified as  $R$ -D Subspace based. Finally, instead of being data-based as in Table 2.1, we have a multi-dimensional data with lattice, which implies that the multi-dimensional MOS schemes should be Tensor Data-based. Note that since the CFP-MOS belongs to both Tensor data-based and  $R$ -D Subspace based schemes, since it uses the structure and the subspaces of the tensor.

### 2.2.1 Goal and Problem

Now we introduce the applied multi-dimensional data model. In order to facilitate the understanding of the tensor notation and operators, the reader is referred to Appendix E, where basic concepts of tensor calculus are presented. To validate the general applicability of our proposed schemes, we adopt the PARAFAC data model below

$$x_0(m_1, m_2, \dots, m_R) = \sum_{n=1}^d f_n^{(1)}(m_1) \cdot f_n^{(2)}(m_2) \dots f_n^{(R)}(m_R) \quad (2.38)$$

where  $f_n^{(r)}(m_r)$  is the  $m_r$ -th element of the  $n$ -th factor of the  $r$ -th mode for  $m_r = 1, \dots, M_r$  and  $r = 1, 2, \dots, R$ .

By defining the vectors  $\mathbf{f}_n^{(r)} = \begin{bmatrix} f_n^{(r)}(1) & f_n^{(r)}(m_r) & \dots & f_n^{(r)}(M_r) \end{bmatrix}^T$  and by using the outer product operator  $\circ$ , another possible representation of (2.38) is given by

$$\mathbf{x}_0 = \sum_{n=1}^d \mathbf{f}_n^{(1)} \circ \mathbf{f}_n^{(2)} \circ \dots \circ \mathbf{f}_n^{(R)}, \quad (2.39)$$

where  $\mathbf{x}_0 \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  is composed of the sum of  $d$  rank one tensors. Therefore, the tensor rank of  $\mathbf{x}_0$  coincides with the model order  $d$ .

For applications with the PARAFAC decomposition, it is important to estimate the factors of the tensor  $\mathbf{x}_0$ , which are defined as  $\mathbf{F}^{(r)} = \begin{bmatrix} \mathbf{f}_1^{(r)} & \dots & \mathbf{f}_d^{(r)} \end{bmatrix} \in \mathbb{C}^{M_r \times d}$ , and we assume that each  $\mathbf{F}^{(r)}$  has full rank. This definition of the factor matrices allows us to rewrite (2.39)



according to the notation proposed in [85]

$$\mathcal{X}_0 = \mathcal{I}_{R,d} \times_1 \mathbf{F}^{(1)} \times_2 \mathbf{F}^{(2)} \dots \times_R \mathbf{F}^{(R)}, \quad (2.40)$$

where  $\times_r$  is the  $r$ -mode product defined in (E.5) in Appendix E, and the tensor  $\mathcal{I}_{R,d}$  represents the  $R$ -dimensional identity tensor of size  $d \times d \dots \times d$ , whose elements are equal to one when the indices  $i_1 = i_2 \dots = i_R$  and zero otherwise.

In practice, the data is contaminated by noise, which we represent by the following data model

$$\mathcal{X} = \mathcal{I}_{R,d} \times_1 \mathbf{F}^{(1)} \times_2 \mathbf{F}^{(2)} \dots \times_R \mathbf{F}^{(R)} + \mathcal{N}, \quad (2.41)$$

where  $\mathcal{N} \in \mathbb{C}^{M_1 \times M_2 \dots \times M_R}$  is the additive noise tensor, whose elements are i.i.d. ZMCS Gaussian random variables. Thereby, the tensor rank is different from  $d$  and usually it assumes extremely large values as shown in [24]. Hence, the problem we are solving can therefore be stated in the following fashion: Given a noisy measurement tensor  $\mathcal{X}$ , we desire to estimate the model order  $d$ . Note that according to [24], the typical rank of  $\mathcal{X}$  is much bigger than any of the dimensions  $M_r$  for  $r = 1, \dots, R$ .

The objective of the PARAFAC decomposition is to compute the estimated factors  $\widehat{\mathbf{F}}^{(r)}$ , such that

$$\mathcal{X} \approx \mathcal{I}_{R,d} \times_1 \widehat{\mathbf{F}}^{(1)} \times_2 \widehat{\mathbf{F}}^{(2)} \dots \times_R \widehat{\mathbf{F}}^{(R)}. \quad (2.42)$$

Since  $\widehat{\mathbf{F}}^{(r)} \in \mathbb{C}^{M_r \times d}$ , one requirement to apply the PARAFAC decomposition is to estimate  $d$ .

Similarly to the previous section, we evaluate the performance of the model order selection scheme in the presence of colored noise, which is given by replacing the white Gaussian white noise tensor  $\mathcal{N}$  by the colored Gaussian noise tensor  $\mathcal{N}^{(c)}$  in (2.41).

Particularly, for multi-dimensional data, the colored noise with a Kronecker structure is present in several applications. For example, in EEG applications [52], the noise is correlated in both space and time dimensions, and it has been shown that a model of the noise combining these two correlation matrices using the Kronecker product can fit noise measurements. Moreover, for MIMO systems the noise covariance matrix is often assumed to be the Kronecker product of the temporal and spatial correlation matrices [76]. In Appendix H, we show why this type of noise is assumed for these applications.

The multi-dimensional colored noise, which is assumed to have a Kronecker correlation structure, can be written as

$$[\mathcal{N}^{(c)}]_{(R+1)} = [\mathcal{N}]_{(R+1)} \cdot (\mathbf{L}_1 \otimes \mathbf{L}_2 \otimes \dots \otimes \mathbf{L}_R)^T, \quad (2.43)$$

where  $\otimes$  represents the Kronecker product as defined in (F.1) in Appendix F. We can also

rewrite (2.43) by using the  $n$ -mode products in the following fashion

$$\mathcal{N}^{(c)} = \mathcal{N} \times_1 \mathbf{L}_1 \times_2 \mathbf{L}_2 \dots \times_R \mathbf{L}_R, \quad (2.44)$$

where  $\mathcal{N} \in \mathbb{C}^{M_1 \times M_2 \dots \times N}$  is a tensor with uncorrelated ZMCSCG (zero-mean circularly symmetric complex Gaussian) elements with variance  $\sigma_n^2$ , and  $\mathbf{L}_i \in \mathbb{C}^{M_i \times M_i}$  is the correlation factor of the  $i$ -th dimension of the colored noise tensor. The noise covariance matrix in the  $i$ -th mode is defined as

$$\mathbb{E} \left\{ \left[ \mathcal{N}^{(c)} \right]_{(i)} \cdot \left[ \mathcal{N}^{(c)} \right]_{(i)}^H \right\} = \alpha \cdot \mathbf{W}_i = \alpha \cdot \mathbf{L}_i \cdot \mathbf{L}_i^H, \quad (2.45)$$

where  $\alpha$  is a normalization constant, such that  $\text{tr}(\mathbf{L}_i \cdot \mathbf{L}_i^H) = M_i$ . The equivalence between (2.43), (2.44), and (2.45) is shown in [4] and also in Appendix I.

To simplify the notation, let us define  $M = \prod_{r=1}^R M_r$ . For the  $r$ -mode unfolding we compute the sample covariance matrix as

$$\hat{\mathbf{R}}_{xx}^{(r)} = \frac{M_r}{M} [\mathcal{X}]_{(r)} \cdot [\mathcal{X}]_{(r)}^H \in \mathbb{C}^{M_r \times M_r}. \quad (2.46)$$

The eigenvalues of these  $r$ -mode sample covariance matrices play a major role in the model order estimation step. Let us denote the  $i$ -th eigenvalue of the sample covariance matrix of the  $r$ -mode unfolding as  $\lambda_i^{(r)}$ . Notice that  $\hat{\mathbf{R}}_{xx}^{(r)}$  possesses  $M_r$  eigenvalues, which we order in such a way that  $\lambda_1^{(r)} \geq \lambda_2^{(r)} \geq \dots \geq \lambda_{M_r}^{(r)}$ . The eigenvalues may be computed from the HOSVD of the measurement tensor

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_R \mathbf{U}_R \quad (2.47)$$

as

$$\text{diag}(\lambda_1^{(r)}, \lambda_2^{(r)}, \dots, \lambda_{M_r}^{(r)}) = \frac{M_r}{M} [\mathcal{S}]_{(r)} \cdot [\mathcal{S}]_{(r)}^H. \quad (2.48)$$

Note that the eigenvalues  $\lambda_i^{(r)}$  are related to the  $r$ -mode singular values  $\sigma_i^{(r)}$  of  $\mathcal{X}$  through  $\lambda_i^{(r)} = \frac{M_r}{M} (\sigma_i^{(r)})^2$ . The  $r$ -mode singular values  $\sigma_i^{(r)}$  can be also computed via the SVD of the  $r$ -mode unfolding of  $\mathcal{X}$  as follows

$$[\mathcal{X}]_{(r)} = \mathbf{U}_r \cdot \mathbf{\Sigma}_r \cdot \mathbf{V}_r^H, \quad (2.49)$$

where  $\mathbf{U}_r \in \mathbb{C}^{M_r \times M_r}$  and  $\mathbf{V}_r \in \mathbb{C}^{\frac{M}{M_r} \times \frac{M}{M_r}}$  are unitary matrices, and  $\mathbf{\Sigma}_r \in \mathbb{C}^{M_r \times \frac{M}{M_r}}$  is a diagonal matrix, which contains the singular values  $\sigma_i^{(r)}$  in the main diagonal.

In Subsection 2.2.3, (2.48) is applied to the global eigenvalue based model order selection techniques. On the other hand, (2.47) is also the basis for the  $R$ -D Subspace based model order selection technique presented in Subsubsection 2.2.3.5. Both types of multi-dimensional model order selection techniques are classified in Table 2.5.

In Subsubsection 2.2.3.1 and in Subsubsection 2.2.3.2, we extend M-EFT, 1-D AIC, and 1-D MDL to their correspondent multi-dimensional forms  $R$ -D EFT,  $R$ -D AIC, and  $R$ -D MDL by applying the global eigenvalues concept. Moreover, in [49] second order statistics of the

eigenvalues are empirically used in order to obtain a higher PoD. Therefore, in Subsubsection 2.2.3.3, we also apply second order statistics of the square global eigenvalues, and we improve even more the PoD of  $R$ -D AIC, and  $R$ -D MDL as examples.

In Table 2.5, we classify Threshold-CORE CONSistency DIAGnostics (T-CORCONDIA) and closed-form PARAFAC model order selection (CFP-MOS) as Tensor Data-based model order selection techniques. In Subsubsection 2.2.3.4, we describe the T-CORCONDIA, which is an important benchmark of the multi-dimensional model order selection techniques. In Subsubsection 2.2.3.5, we present CFP-MOS, which is a multi-dimensional model order selection technique that presents PoD close to  $R$ -D EFT in the presence of white Gaussian noise and is also applicable in the presence of colored Gaussian noise.

In case of centro-hermitian arrays, one important way to improve the estimation is incorporating forward-backward averaging (FBA) [104], which we present in Subsection 2.2.3.6. Moreover, for case of tensor with special sizes, we provide practical results in Subsection 2.2.4.1.

Almost this whole section is composed of contributions proposed by us in the field of multi-dimensional model order selection.

## 2.2.2 Previous Approach: CORCONDIA

The CORE CONSistency DIAGnostics (CORCONDIA) [15] is a well-known multi-dimensional model order selection technique, which is applied in several scientific fields. Therefore, in this chapter, we intend to use it, as well as the best matrix-based model order selection schemes from the previous section, as a benchmark for our proposed multi-dimensional MOS techniques. In this subsection, we review CORCONDIA, which is fundamental for our non-subjective form of CORCONDIA called Threshold-CORCONDIA (T-CORCONDIA) in Subsubsection 2.2.3.4.

For the sake of simplicity, consider the ALS (Alternating Least Squares) solution [60] for the special case  $R = 3$ . Then we have the following iterative equations for the PARAFAC decomposition of a tensor

$$\begin{aligned}\widehat{\mathbf{F}}_{t+1}^{(3)} &= [\mathcal{X}]_{(3)} \cdot \left[ \left( \widehat{\mathbf{F}}_t^{(1)} \diamond \widehat{\mathbf{F}}_t^{(2)} \right)^+ \right]^T \\ \widehat{\mathbf{F}}_{t+1}^{(2)} &= [\mathcal{X}]_{(2)} \cdot \left[ \left( \widehat{\mathbf{F}}_t^{(3)} \diamond \widehat{\mathbf{F}}_t^{(1)} \right)^+ \right]^T \\ \widehat{\mathbf{F}}_{t+1}^{(1)} &= [\mathcal{X}]_{(1)} \cdot \left[ \left( \widehat{\mathbf{F}}_t^{(2)} \diamond \widehat{\mathbf{F}}_t^{(3)} \right)^+ \right]^T,\end{aligned}\tag{2.50}$$

where  $\diamond$  stands for the Khatri-Rao product also known as the column-wise Kronecker product, which is defined in (F) in Appendix F.2 and  $t$  denotes the  $t$ -th iteration. The factor matrices are initialized via the Direct Trilinear Decomposition (DTLD) [90], which is an improved version of the Generalized Rank Annihilation Method (GRAM). In the DTLD and GRAM, a two slab-analysis is performed in order to solve the problem in a closed-form way. Therefore, only crude estimations of the factor matrices are obtained by applying these

---

schemes. However, these estimates are used to initialize the iterative algorithm in (3.10). To compute iteratively the factors as shown (2.50), the PARAFAC algorithm<sup>5</sup> in MATLAB is applied. In the simulations, the ALS algorithm is restricted to a maximum number of iterations equal to 2500, and the threshold of the reconstruction error is equal to 10<sup>-6</sup>.

We can rewrite the last equation from the ALS equations (2.50) using another representation

$$[\widehat{\mathcal{X}}]_{(1)} = \widehat{\mathbf{F}}^{(1)} \cdot \mathbf{T} \cdot \left( \widehat{\mathbf{F}}^{(2)} \otimes \widehat{\mathbf{F}}^{(3)} \right)^T \quad (2.51)$$

where  $\mathbf{T} \in \mathbb{C}^{d \times d^2}$  is defined as  $\mathbf{T} = [\mathcal{I}_{3,d}]_{(1)}$ , which can be seen by computing the 1-mode unfolding of  $\mathcal{X}$  in (2.42). More information about (2.51) can be found in Appendix F. Moreover, from the third equation of (2.50), (2.51) is obtained through the following relation

$$\widehat{\mathbf{F}}^{(2)} \diamond \widehat{\mathbf{F}}^{(3)} = \left[ \left( \widehat{\mathbf{F}}^{(2)} \otimes \widehat{\mathbf{F}}^{(3)} \right) \right] \cdot [\mathcal{I}_{3,d}]_{(1)}^T, \quad (2.52)$$

where this equality is proven in (F.5) in Appendix F.

Once the matrices  $\mathbf{F}^{(1)}$ ,  $\mathbf{F}^{(2)}$  and  $\mathbf{F}^{(3)}$  are known for a given  $d$  using ALS, the following matrix  $\mathbf{G}$  can be computed via

$$\text{vec}(\mathbf{G}) = \left( \widehat{\mathbf{F}}^{(3)} \otimes \widehat{\mathbf{F}}^{(2)} \otimes \widehat{\mathbf{F}}^{(1)} \right)^+ \text{vec}([\mathcal{X}]), \quad (2.53)$$

where (2.53) is derived in Appendix F and in the presence of noise, (2.53) becomes an approximation. If the PARAFAC model is exactly fulfilled, then  $\mathbf{G}$  is equal to  $\mathbf{T}$ . Otherwise, the closeness of  $\mathbf{G}$  to  $\mathbf{T}$  provides a measure of how well the PARAFAC model fits the observations. In the case of a perfect fit, i.e.,  $\mathbf{G} = \mathbf{T}$ .

Defining  $[\mathcal{G}]_{(1)} = \mathbf{G}$ , then the following CORE CONSistency cost function can be obtained as defined in [15]

$$CC(P) = 100 \left( 1 - \frac{\|\mathcal{G}_{3,P} - \mathcal{I}_{3,P}\|_{\text{H}}^2}{\|\mathcal{G}_{3,P}\|_{\text{H}}^2} \right) \quad (2.54)$$

where  $\|\cdot\|_{\text{H}}^2$  is the higher order norm of the tensor, which can be computed via (E.4) in Appendix E, and  $P$  is defined as a candidate value for  $\widehat{d}$ , which is the estimate of the model order  $d$ .

An example of the CORCONDIA profile can be seen in Figure 2.32, where we have a break distance of approximately 99.6 %. In practice, for different values of the SNR, the break distance varies. Therefore, in order to avoid a subjective analysis of the CORCON profile with respect to the break distance, it becomes of great necessity to define a threshold value.

---

<sup>5</sup> The N-way toolbox v3.10 from <http://www.models.life.ku.dk/source/nwaytoolbox/index.asp> is used.

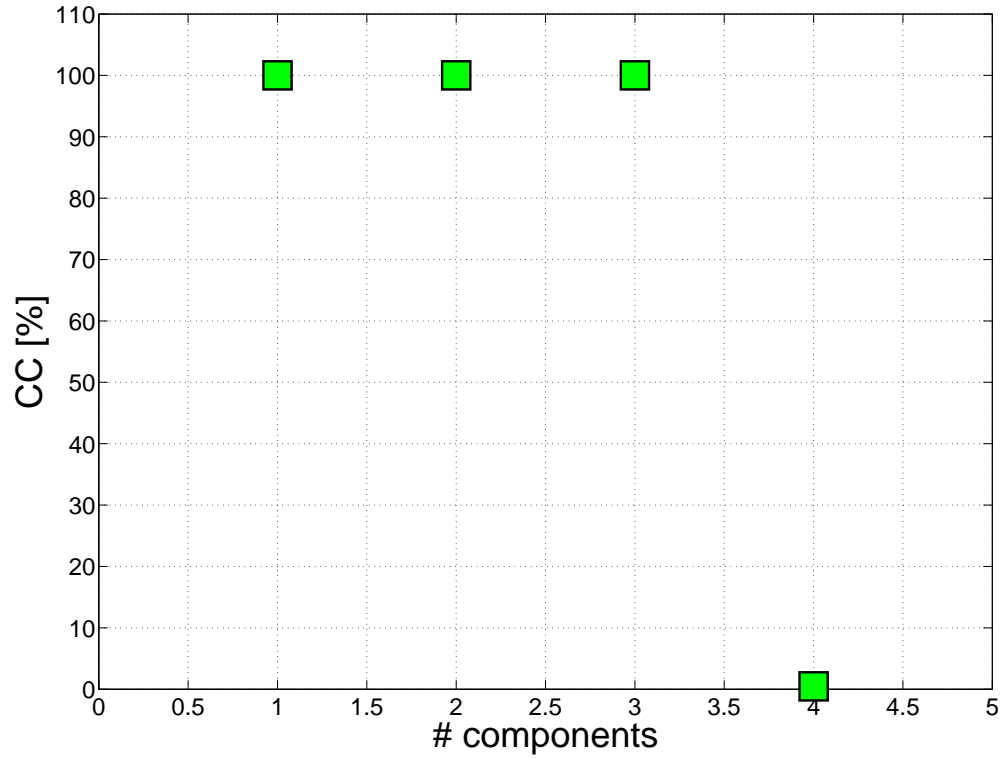


Fig. 2.32: CCore CONSistency in % vs. the number of components considering a scenario with a data model of  $M_1 = 7$ ,  $M_2 = 7$ ,  $M_3 = 7$ ,  $M_4 = 7$  and  $d = 3$  components and with SNR = 50 dB. The break point distance between the components 3 and 4 is approximately 99.6 %.

In the next subsection, we propose the Threshold-CORCONDIA (T-CORCONDIA), which does not require the subjective analysis, since it is based on the threshold coefficients. We also present the proposed multi-dimensional model order selection schemes for white and colored Gaussian noise.

### 2.2.3 Proposed Approaches

In this subsection, multi-dimensional model order selection schemes are proposed based on the global eigenvalues,  $R$ -D subspace, or tensor-based data model. First we show the proposed definition of the global eigenvalues together with the presentation of the proposed  $R$ -D EFT in Subsubsection 2.2.3.1. Examples of other extendable schemes are  $R$ -D AIC and  $R$ -D MDL, which are described in Subsubsection 2.2.3.2. The application of the exponentiation of the global eigenvalues is also empirically investigated in Subsubsection 2.2.3.3.

Besides the global eigenvalue based schemes, we also propose two tensor data-based multi-dimensional model order selection schemes. The first one in Subsubsection 2.2.3.4 is the Threshold-CORCONDIA (T-CORCONDIA), a non-subjective form of CORCONDIA re-

viewed in Subsection 2.2.2. In Subsubsection 2.2.3.5, a closed-form PARAFAC based model order selection scheme is also proposed for white and also colored noise scenarios.

### 2.2.3.1 $R$ -dimensional Exponential Fitting Test ( $R$ -D EFT)

The global eigenvalues are based on the  $r$ -mode eigenvalues represented by  $\lambda_i^{(r)}$  for  $r = 1, \dots, R$  and for  $i = 1, \dots, M_r$ . To obtain the  $r$ -mode eigenvalues, there are two ways. The first way shown in (2.46) is possible via the EVD of each  $r$ -mode sample covariance matrix, and the second way in (2.48) is given via an HOSVD.

According to [42, 79], the noise eigenvalues that exhibit a Wishart profile can have their profile approximated by an exponential curve. Therefore, by applying the exponential approximation for every  $r$ -mode, we obtain that

$$\mathbb{E} \left\{ \lambda_i^{(r)} \right\} = \mathbb{E} \left\{ \lambda_1^{(r)} \right\} \cdot q(\alpha_r, \beta_r)^{i-1}, \quad (2.55)$$

where  $\alpha_r = \min \left\{ M_r, \frac{M}{M_r} \right\}$ ,  $\beta_r = \max \left\{ M_r, \frac{M}{M_r} \right\}$ ,  $i = 1, 2, \dots, M_r$  and  $r = 1, 2, \dots, R$ . The rate of the exponential profile  $q(\alpha_r, \beta_r)$  is defined in (2.30) and is the same used for the M-EFT.

For the sake of simplicity, let us first assume that  $M_1 = M_2 = \dots = M_R$ . Then we can define *global eigenvalues* as being [2]

$$\lambda_i^{(\text{G})} = \lambda_i^{(1)} \cdot \lambda_i^{(2)} \dots \lambda_i^{(R)}. \quad (2.56)$$

Therefore, it is straightforward that the noise global eigenvalues also follow an exponential profile, since

$$\mathbb{E} \left\{ \lambda_i^{(\text{G})} \right\} = \mathbb{E} \left\{ \lambda_1^{(\text{G})} \right\} \cdot \left( q(\alpha_1, \beta_1) \cdot \dots \cdot q(\alpha_R, \beta_R) \right)^{i-1}, \quad (2.57)$$

where  $i = 1, \dots, M_R$ .

The product across modes increases the threshold between the predicted and the actual eigenvalue. In Fig. 2.2.3.1, we compare the threshold between the actual eigenvalue and the predicted eigenvalue in the  $r$ -th mode to the threshold between the actual global eigenvalue and the predicted global eigenvalue. Here we consider that  $\mathcal{X}_0$  is a rank one tensor, and noise is added according to (2.41). Then, in this case,  $d = 1$ . For the first threshold, we have  $\lambda_i^{(r)} - \widehat{\lambda}_i^{(r)} = 2.4 \cdot 10^2$ , while for the second one, we have  $\lambda_1^{(\text{G})} - \widehat{\lambda}_1^{(\text{G})} = 2.4 \cdot 10^{12}$ . Therefore, the break in the profile is easier to detect via global eigenvalues than using only one mode eigenvalues.

Without loss of generality, let us consider the case in which  $M_1 \geq M_2 \geq \dots \geq M_{R+1}$ . We start by estimating  $\widehat{d}$  with a certain eigenvalue based model order selection method considering the first unfolding only, which in the example in Fig. 2.34 has a size  $M_1 = 13$ . If  $\widehat{d} < M_2$ , we could have taken advantage of the second mode as well. Therefore, we compute the global eigenvalues  $\lambda_i^{(\text{G})}$  as in (2.56) for  $1 \leq i \leq M_2$ , thus discarding the  $M_1 - M_2$  last eigenvalues of the first mode. We can obtain a new estimate  $\widehat{d}$ . As illustrated in Fig. 2.35, we utilize only

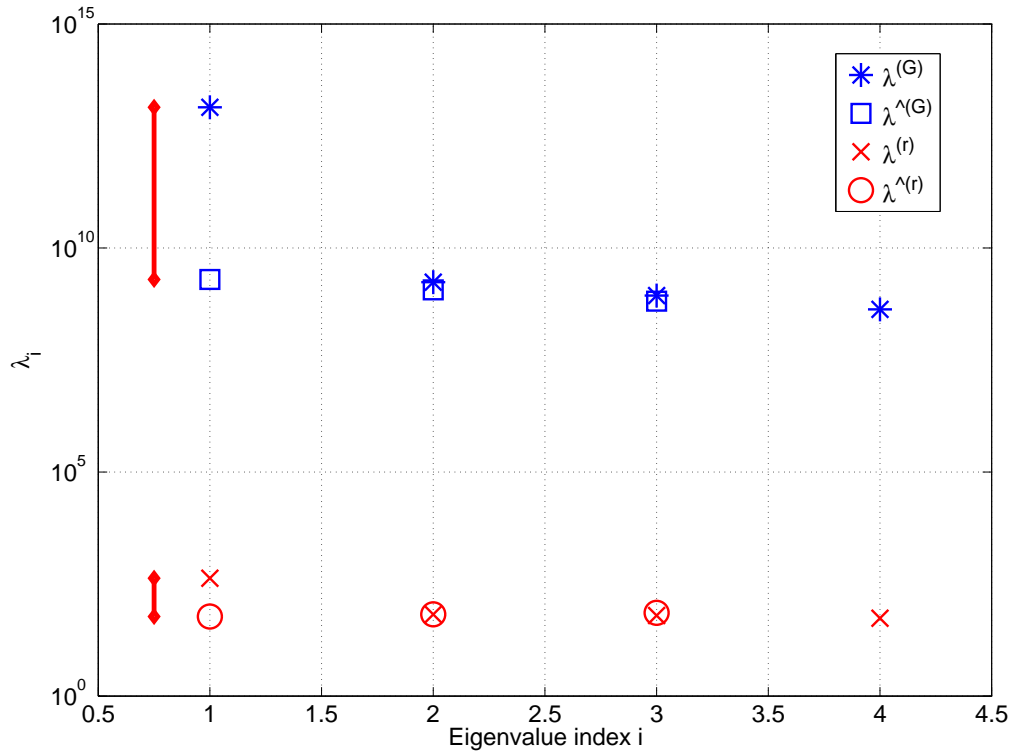


Fig. 2.33: Comparison between the global eigenvalues profile and the  $R$ -mode eigenvalues profile for a scenario with array size  $M_1 = 4, M_2 = 4, M_3 = 4, M_4 = 4, M_5, d = 1$  and  $\text{SNR} = 0$  dB.

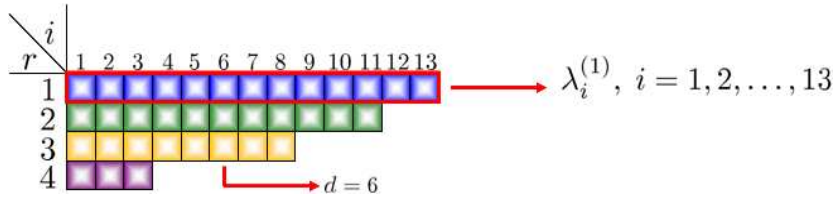


Fig. 2.34: Adaptive definition of the global eigenvalues - 1<sup>st</sup> eigenvalue set

the first  $M_2$  highest eigenvalues of the first and of the second modes to estimate the model order. If  $\widehat{d} < M_3$  we could continue in the same fashion, by computing the global eigenvalues considering the first 3 modes. In the example in Fig. 2.36, since the model order is equal to 6, which is greater than  $M_4$ , the adaptive definition algorithm of the global eigenvalues stops using the three first modes. Clearly, the full potential of the proposed method can be achieved when all modes are used to compute the global eigenvalues. This happens when  $\widehat{d} < M_{R+1}$ , so that  $\lambda_i^{(G)}$  can be computed for  $1 \leq i \leq M_{R+1}$ .

Note that by using the global eigenvalues, the assumptions of M-EFT, that the noise eigenvalues can be approximated by an exponential profile, and the assumptions of AIC and MDL, that the noise eigenvalues are constant, still hold. Moreover, the maximum model order possible to estimate is equal to  $\max_r M_r$ , for  $r = 1, \dots, R$ .

Next the M-EFT is extended to its multi-dimensional version  $R$ -D EFT, whereas in Sub-

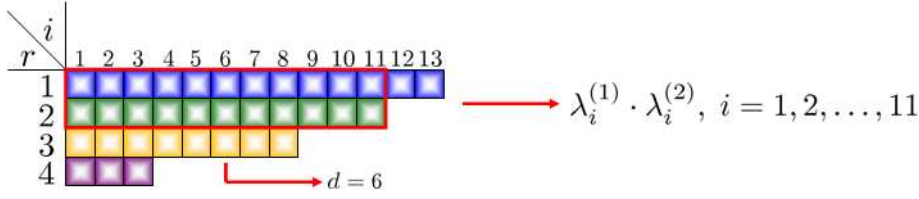


Fig. 2.35: Adaptive definition of the global eigenvalues - 1<sup>st</sup> and 2<sup>nd</sup> eigenvalue sets

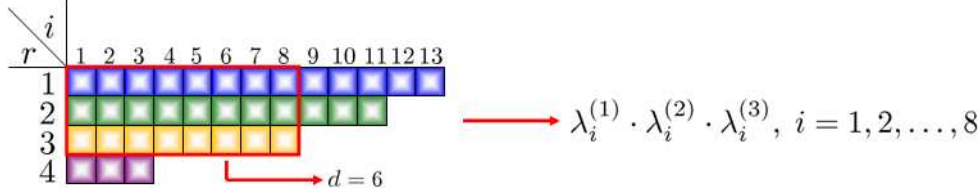


Fig. 2.36: Adaptive definition of the global eigenvalues - 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> eigenvalue sets

subsection 2.2.3.2, 1-D AIC and 1-D MDL are extended to  $R$ -D AIC and  $R$ -D MDL.

The  $R$ -D EFT is an extended version of the M-EFT operating on the  $\lambda_i^{(G)}$ . Therefore,

- 1) It exploits the fact that the noise global eigenvalues still exhibit an exponential profile;
- 2) The increase of the threshold between the actual signal global eigenvalue and the predicted noise global eigenvalue leads to a significant improvements in the performance;
- 3) It is able to adapt to arrays of arbitrary size and dimension through the adaptive definition of the global eigenvalues.

To derive the proposed multi-dimensional extension of the M-EFT algorithm, namely the  $R$ -D EFT, we start by looking at an  $(R)$ -dimensional noise-only case. For the  $R$ -D EFT, it is our intention to predict the noise global eigenvalues defined in (2.57). Each  $r$ -mode eigenvalue can be estimated via

$$\widehat{\lambda}_{M-P}^{(r)} = (P+1) \cdot \frac{1 - q\left(P+1, \frac{M}{M_r}\right)}{1 - q\left(P+1, \frac{M}{M_r}\right)^{P+1}} \left(\widehat{\sigma}^{(r)}\right)^2 \quad (2.58)$$

$$\left(\widehat{\sigma}^{(r)}\right)^2 = \frac{1}{P} \sum_{i=0}^{P-1} \lambda_{M-i}^{(r)}. \quad (2.59)$$

Equations (2.58) and (2.59) are the same expressions as in the case of the M-EFT in (2.32) and (2.33), however, in contrast to the M-EFT, here they are applied to each  $r$ -mode eigenvalue.

Let us apply the definition of the global eigenvalues according to (2.56)

$$\widehat{\lambda}_i^{(G)} = \widehat{\lambda}_i^{(1)} \cdot \widehat{\lambda}_i^{(2)} \dots \widehat{\lambda}_i^{(R)}, \quad (2.60)$$



where in (2.57) the approximation by an exponential profile is assumed. Therefore,

$$\widehat{\lambda}_i^{(G)} = \widehat{\lambda}_{\alpha^{(G)}}^{(G)} \cdot \left( q(P+1, \frac{M}{M_1}) \cdot \dots \cdot q(P+1, \frac{M}{M_R}) \right)^{i-1}, \quad (2.61)$$

where  $\alpha^{(G)}$  is the minimum  $\alpha_r$  for all the  $r$ -modes considered in the adaptive definition of the global eigenvalue. In (2.59)  $\widehat{\lambda}_i^{(G)}$  is a function of only the last global eigenvalue  $\widehat{\lambda}_{\alpha^{(G)}}^{(G)}$ , which is the smallest global eigenvalue and is assumed a noise eigenvalue, and of the rates  $q(P+1, \frac{M}{M_r})$  for all the  $r$ -modes considered in the adaptive definition. Instead of using directly (2.61), we use  $\widehat{\lambda}_{M-P}^{(r)}$  according to (2.59) for all the  $r$ -modes considered in the adaptive definition. Therefore, the previous eigenvalues that were already estimated as noise eigenvalues are taken into account in the prediction step.

Similarly to the M-EFT, by using the predicted global eigenvalue expression (2.60) considering white Gaussian noise samples, we compute the global threshold coefficients  $\eta_P^{(G)}$  via the hypotheses for the tensor case

$$\begin{aligned} H_{P+1}: \lambda_{M-P}^{(G)} \text{ is a noise EV, } & \frac{\lambda_{M-P}^{(G)} - \widehat{\lambda}_{M-P}^{(G)}}{\widehat{\lambda}_{M-P}^{(G)}} \leq \eta_P^{(G)} \\ \bar{H}_{P+1}: \lambda_{M-P}^{(G)} \text{ is a signal EV, } & \frac{\lambda_{M-P}^{(G)} - \widehat{\lambda}_{M-P}^{(G)}}{\widehat{\lambda}_{M-P}^{(G)}} > \eta_P^{(G)}. \end{aligned} \quad (2.62)$$

Once all  $\eta_P^{(G)}$  are found for a certain higher order array of sizes  $M_1, M_2, \dots, M_R$ , and for a certain  $P_{fa}$ , then the model order can be estimated by applying the following cost function

$$\begin{aligned} \widehat{d} &= \alpha^{(G)} - \min(\mathcal{P}) \quad \text{where} \\ P \in \mathcal{P}, \quad & \text{if } \frac{\lambda_{M-P}^{(G)} - \widehat{\lambda}_{M-P}^{(G)}}{\widehat{\lambda}_{M-P}^{(G)}} > \eta_P^{(G)}, \end{aligned}$$

where  $\alpha^{(G)}$  is the total number of adaptively defined global eigenvalues.

### 2.2.3.2 $R$ -D AIC and $R$ -D MDL

In AIC and MDL, it is assumed that the noise eigenvalues are all equal. Therefore, once this assumption is valid for all  $r$ -mode eigenvalues, it is straightforward that it is also valid for our global eigenvalue definition. Moreover, since we have shown that 1-D AIC and 1-D MDL are more general and superior in terms of performance than AIC and MDL, respectively, we extend 1-D AIC and 1-D MDL to the multi-dimensional form by using the global eigenvalues. Note that PoD of 1-D AIC and 1-D MDL is only greater than PoD of AIC and MDL for cases where  $\frac{M}{M_r} > M_r$ , which are certainly more common in multi-dimensional data.

The corresponding  $R$ -dimensional versions of 1-D AIC and 1-D MDL are obtained by first replacing the eigenvalues of  $\widehat{\mathbf{R}}_{xx}$  by the global eigenvalues  $\lambda_i^{(G)}$  defined in (2.56). Additionally, in order to compute the number of free parameters for the 1-D AIC and 1-D MDL methods

Approach	Penalty function $p(P, N, \alpha^{(G)})$
$R$ -D AIC	$P \cdot (2 \cdot \alpha^{(G)} - P)$
$R$ -D MDL	$\frac{1}{2} \cdot P \cdot (2 \cdot \alpha^{(G)} - P) \cdot \log(N)$

Tab. 2.6: Penalty functions for  $R$ -D Information Theoretic Criteria

and their  $R$ -D extensions, we propose to set the parameter  $N = \max_r M_r$  and  $\alpha^{(G)}$  is the total number of adaptively defined global eigenvalues similarly as we propose in [2]. Therefore, the optimization problem for the  $R$ -D AIC and  $R$ -D MDL is given by

$$\begin{aligned} \widehat{d} &= \arg \min_P J^{(G)}(P) \quad \text{where} \\ J^{(G)}(P) &= -N(\alpha^{(G)} - P) \log \left( \frac{g^{(G)}(P)}{a^{(G)}(P)} \right) + p(P, N, \alpha^{(G)}), \end{aligned} \quad (2.63)$$

where  $\widehat{d}$  represents an estimate of the model order  $d$ ,  $g^{(G)}(P)$  and  $a^{(G)}(P)$  are respectively the geometric and arithmetic means of the global eigenvalues. The penalty functions for  $R$ -D AIC and  $R$ -D MDL are given in Table 2.6.

Note that the  $R$ -dimensional extension described in this subsection can be applied to *any* model order selection scheme that is based on the profile of eigenvalues, i.e., also to the 1-D MDL and the 1-D AIC methods.

### 2.2.3.3 Exponentiation of the Global Eigenvalues

In this section, inspired by the empirical application of second order statistics of the eigenvalues in [49], we propose to apply the exponentiation of the global eigenvalues in order to improve the PoD.

We define the higher order of the global eigenvalues proposed in (2.56) as

$$\lambda_i^{(\text{PW})} = \left( \lambda_i^{(G)} \right)^u, \quad (2.64)$$

where  $u > 1$ .

### 2.2.3.4 Threshold CORCONDIA (T-CORCONDIA)

In this subsection, we present T-CORCONDIA proposed by us in [1], which is the non-subjective form of CORCONDIA [15] reviewed in Subsubsection 2.2.2. In contrast to the CORCONDIA, we can compare the T-CORCONDIA to the other model order selection schemes in an objective way in terms of PoD by introducing threshold coefficients. After finding the threshold coefficients considering some criterion, we compare the T-CORCONDIA to the multi-dimensional model order selection techniques in order to check the efficiency of CORCONDIA.

Let us denote the threshold distance from the component  $P$  to the component  $P + 1$  as being  $\Delta(P + 1)$  and define the following hypotheses:

$$\begin{aligned} H_{P+1} : \Delta(P + 1) \text{ is NC } , (CC(P + 1) - CC(P)) &\geq \Delta(P + 1) \\ \bar{H}_{P+1} : \Delta(P + 1) \text{ is PC } , (CC(P + 1) - CC(P)) &< \Delta(P + 1) \end{aligned} \quad (2.65)$$

where PC means Principal Component and NC means Noise Component. Note that CORCONDIA as well as T-CORCONDIA assumes that there is at least one principal component in the data.

In general, all threshold coefficients  $\Delta(P)$  can be adjusted individually. To find a set of coefficients that is optimal we need to minimize a cost function that depends on many variables (i.e., perform a multi-dimensional search). Therefore, the computational complexity of this method which we call T-CORCONDIA Var is relatively high. As a simpler alternative we propose another method which we call T-CORCONDIA Fix where we set all threshold coefficients to one value  $\Delta$  and obtain an optimal  $\Delta$  by minimizing a cost-function that depends only on one variable (i.e., a one-dimensional search). Applying Monte Carlo simulations for the hypotheses from (2.65), we obtain a curve of the Probability of correct Detection (PoD) v.s.  $\Delta$  ( $\text{PoD}_\Delta$ ) for the T-CORCONDIA Fix. Note that  $\text{PoD}_\Delta$  represents the probability of correct Detection averaged over the SNR. An example of a SNR range considered to compute this average is from  $-20$  dB to  $20$  dB. The vector of equal thresholds,  $\Delta_{\text{fix}}$ , can be estimated by using:

$$\Delta_{\text{fix}} = \arg \max_{\Delta} \text{PoD}_\Delta(\Delta) \quad (2.66)$$

An example of this estimation is presented in Figure 2.37, where the maximum Probability of correct Detection (PoD) of 52 % is obtained when  $\Delta \approx 1.6$  % . One drawback of the T-CORCONDIA Fix is the fact that the Probability of correct Detection (PoD) v.s. SNR ( $\text{PoD}_{\text{SNR}}$ ) for each  $\hat{d}$  may be different. In other words, for some values of  $\hat{d}$  the estimation is better than for other values. Varying the threshold coefficients over  $\hat{d}$  allows us to mitigate this drawback in T-CORCONDIA Var. In order to achieve more balanced thresholds we therefore define a cost function for  $\Delta$  that reflects any imbalance in the threshold coefficients. To this end, let  $\text{PoD}_{\text{SNR}}(\Delta(\hat{d}))$  be the Monte Carlo estimate of the probability of correct Detection (PoD) curve versus the SNR for a certain model order which depends on the threshold coefficient  $\Delta(\hat{d})$ . Then, an appropriate cost function for T-CORCONDIA Var is given by

$$\begin{aligned} \Delta_{\text{var}} &= \arg \min_{\Delta} J_{\text{var}}(\Delta) \quad \text{where} \\ J_{\text{var}}(\Delta) &= \sum_{P=1}^{\hat{d}_{\text{max}}-1} \left| E\{\text{PoD}_{\text{SNR}}(\Delta(P + 1))\} - E\{\text{PoD}_{\text{SNR}}(\Delta(P))\} \right| \end{aligned} \quad (2.67)$$


---

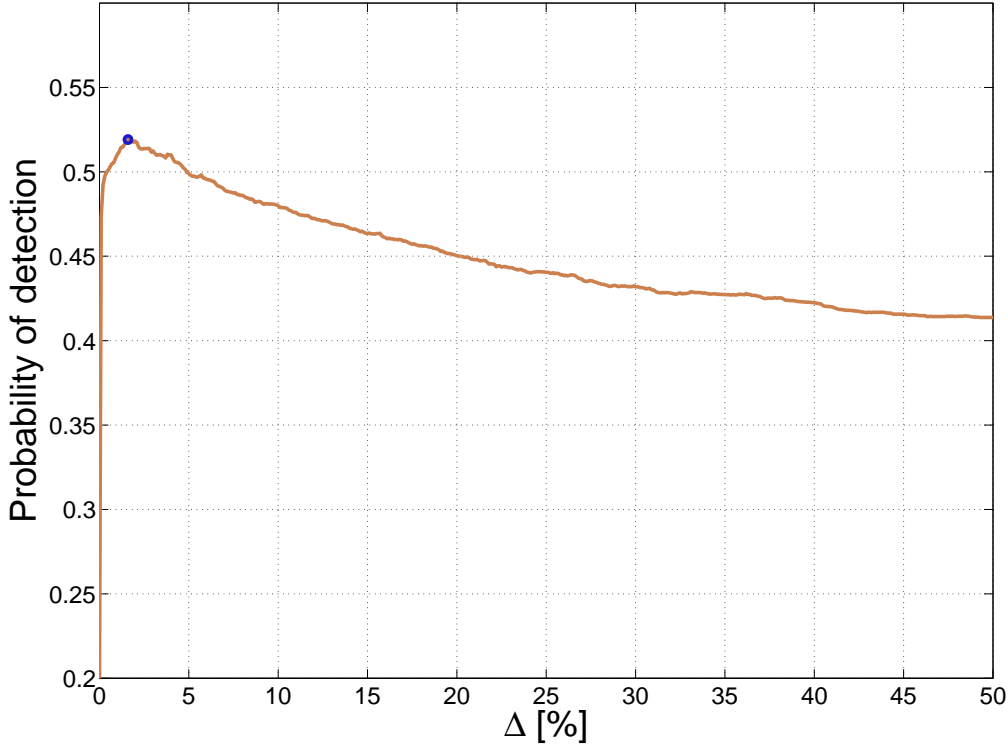


Fig. 2.37: Probability of correct Detection v.s.  $\Delta$  ( $\text{PoD}_\Delta$ ) for the threshold computations of the T-CORCONDIA Fix. Varying  $\Delta$  from 0 to 50 with steps of 0.1. The curve is estimated using the four first  $\Delta$  thresholds, i.e.,  $d$  is varied from 1 to 5. In this example scenario, the data has a size of  $M_1 = 7$ ,  $M_2 = 7$ ,  $M_3 = 7$  and  $M_4 = 7$ . The SNR varies from  $-20$  dB to  $50$  dB.

where  $\hat{d}_{\max}$  is defined as being the maximum candidate value of  $P$  and  $\Delta_{\text{var}}$  is the vector with the threshold coefficients for each value of  $P$ .

Note that  $\hat{d}_{\max}$  is not limited by the number of sensors or observations (as for all eigenvalue-based approaches). However, the computational complexity required to find the threshold coefficients increases dramatically with  $\hat{d}_{\max}$ . Therefore, the available computing resources impose a practical limit on  $\hat{d}_{\max}$ . For example, for the scenario in Fig. 2.37,  $\hat{d}_{\max}$  was set to 4. Testing all possible thresholds between 0 and 50 % in steps of 0.1 % the cost function had to be evaluated  $501^4$  times.

One advantage of the CORCONDIA and T-CORCONDIA in comparison to the global eigenvalue based and  $R$ -D subspace based multi-dimensional model order selection schemes is the possibility to estimate the model order even for data with typical rank of the tensor [23], i.e.,  $d \gg M_r$  for  $r = 1, \dots, R$ .

In Subsection 2.2.4, simulations comparing the PoD of T-CORCONDIA to the  $R$ -D EFT,  $R$ -D AIC, and  $R$ -D MDL are performed.

### 2.2.3.5 Closed-Form PARAFAC based Model Order Selection (CFP-MOS)

In this subsection, we present the Closed-form PARAFAC based model order selection (CFP-MOS) technique proposed by us in [6]. The major motivation of CFP-MOS is the fact that  $R$ -D AIC,  $R$ -D MDL and  $R$ -D EFT are applicable only in the presence of white Gaussian noise. Therefore, it is very appealing to apply CFP-MOS, since it has a performance close to  $R$ -D EFT in the presence of white Gaussian noise, and that at the same time it is also applicable in the presence of colored Gaussian noise.

According to [84], the estimation of the factors  $\mathbf{F}^{(r)}$  via the PARAFAC decomposition is transformed into a set of simultaneous diagonalization problems based on the relation between the truncated HOSVD [29]-based low-rank approximation of  $\mathcal{X}$

$$\begin{aligned}\mathcal{X} &\approx \mathcal{S}^{[s]} \times_1 \mathbf{U}_1^{[s]} \dots \times_R \mathbf{U}_R^{[s]} \\ &\approx \mathcal{S}^{[s]} \bigtimes_{r=1}^R \mathbf{U}_r^{[s]},\end{aligned}\quad (2.68)$$

and the PARAFAC decomposition of  $\mathcal{X}$

$$\begin{aligned}\mathcal{X} &\approx \mathcal{I}_{R,d} \times_1 \widehat{\mathbf{F}}^{(1)} \dots \times_R \widehat{\mathbf{F}}^{(R)} \\ &\approx \mathcal{I}_{R,d} \bigtimes_{r=1}^R \widehat{\mathbf{F}}^{(r)},\end{aligned}\quad (2.69)$$

where  $\mathcal{S}^{[s]} \in \mathbb{C}^{p_1 \times p_2 \times \dots \times p_R}$ ,  $\mathbf{U}_r^{[s]} \in \mathbb{C}^{M_r \times p_r}$ ,  $p_r = \min(M_r, d)$ , and  $\widehat{\mathbf{F}}^{(r)} = \mathbf{U}_r^{[s]} \cdot \mathbf{T}_r$  for a nonsingular transformation matrix  $\mathbf{T}_r \in \mathbb{C}^{d \times d}$  for all modes  $r \in \mathcal{R}$ , where  $\mathcal{R} = \{r | M_r \geq d, r = 1, \dots, R\}$  denotes the set of non-degenerate modes. As shown in (2.68) and in (2.69), the operator  $\bigtimes_{r=1}^R$  denotes a compact representation of  $R$   $r$ -mode products between a tensor and  $R$  matrices.

The closed-form PARAFAC (CFP) [84] decomposition constructs two simultaneous diagonalization problems for every tuple  $(k, \ell)$ , such that  $k, \ell \in \mathcal{R}$ , and  $k < \ell$ . In order to reference each simultaneous matrix diagonalization (SMD) problem, we define the enumerator function  $e(k, \ell, i)$  that assigns the triple  $(k, \ell, i)$  to a sequence of consecutive integer numbers in the range  $1, 2, \dots, T$ . Here  $i = 1, 2$  refers to the two simultaneous matrix diagonalizations (SMD) for our specific  $k$  and  $\ell$ . Consequently,  $\text{SMD}(e(k, \ell, 1), P)$  represents the first SMD for a given  $k$  and  $\ell$ , which is associated to the simultaneous diagonalization of the matrices  $\mathbf{S}_{k, \ell, (n)}^{\text{rhs}}$  by  $\mathbf{T}_k$ . Initially we consider that the candidate value of the model order  $P = d$ , which the mode order. Similarly,  $\text{SMD}(e(k, \ell, 2), P)$  corresponds to the second SMD for a given  $k$  and  $\ell$  referring to the simultaneous diagonalizations of  $\mathbf{S}_{k, \ell, (n)}^{\text{lhs}}$  by  $\mathbf{T}_\ell$ .  $\mathbf{S}_{k, \ell, (n)}^{\text{rhs}}$  and  $\mathbf{S}_{k, \ell, (n)}^{\text{lhs}}$  are defined in [84]. Note that each  $\text{SMD}(e(k, \ell, i), P)$  yields an estimate of all factors  $\mathbf{F}^{(r)}$  [84, 100], where  $r = 1, \dots, R$ . Consequently, for each factor  $\mathbf{F}^{(r)}$  there are  $T$  estimates.

For instance, consider a 4-D tensor, where the third mode is degenerate, i.e.,  $M_3 < d$ . Then, the set  $\mathcal{R}$  is given by  $\{1, 2, 4\}$ , and the possible  $(k, \ell)$ -tuples are  $(1, 2)$ ,  $(1, 4)$ , and  $(2, 4)$ . Consequently, the six possible SMDs are enumerated via  $e(k, \ell, i)$  as follows:  $e(1, 2, 1) = 1$ ,

$e(1, 2, 2) = 2$ ,  $e(1, 4, 1) = 3$ ,  $e(1, 4, 2) = 4$ ,  $e(2, 4, 1) = 5$ , and  $e(2, 4, 2) = 6$ . In general, the total number of SMD problems  $T$  is equal to  $\#(\mathcal{R}) \cdot [\#(\mathcal{R}) + 1]$ .

There are different heuristics to select the best estimates of each factor  $\mathbf{F}^{(r)}$  as shown in [84]. We define the function to compute the residuals (RESID) of the simultaneous matrix diagonalizations (SMD) as  $\text{RESID}(\text{SMD}(\cdot))$ . For instance, we apply it for  $e(k, \ell, 1)$

$$\text{RESID}(\text{SMD}(e(k, \ell, 1), P)) = \sum_{n=1}^{N_{\max}} \left\| \text{off} \left( \mathbf{T}_k^{-1} \cdot \mathbf{S}_{k, \ell, (n)}^{\text{rhs}} \cdot \mathbf{T}_k \right) \right\|_{\text{F}}^2, \quad (2.70)$$

and for  $e(k, \ell, 2)$

$$\text{RESID}(\text{SMD}(e(k, \ell, 2), P)) = \sum_{n=1}^{N_{\max}} \left\| \text{off} \left( \mathbf{T}_{\ell}^{-1} \cdot \mathbf{S}_{k, \ell, (n)}^{\text{lhs}} \cdot \mathbf{T}_{\ell} \right) \right\|_{\text{F}}^2, \quad (2.71)$$

where  $N_{\max} = \prod_{r=1}^R M_r \cdot N / (M_k \cdot M_{\ell})$ .

Since each residual is a positive real-valued number, we can order the SMDs by the magnitude of the corresponding residual. For the sake of simplicity, we represent the ordered sequence of SMDs corresponding to  $e(k, \ell, i)$  by a single index  $e^{(t)}$  for  $t = 1, 2, \dots, T$ , such that  $\text{RESID}(\text{SMD}(e^{(t)}, P)) \leq \text{RESID}(\text{SMD}(e^{(t+1)}, P))$ . Since in practice  $d$  is not known,  $P$  denotes a candidate value for  $\widehat{d}$ , which is our estimate of the model order  $d$ . Our task is to select  $P$  from the interval  $\widehat{d}_{\min} \leq P \leq \widehat{d}_{\max}$ , where  $\widehat{d}_{\min}$  is a lower bound and  $\widehat{d}_{\max}$  is an upper bound for our candidate values. For instance,  $\widehat{d}_{\min}$  equal to 1 is used, and  $\widehat{d}_{\max}$  is chosen such that no dimension is degenerate [84], i.e.,  $d > M_r$  for  $r = 1, \dots, R$ . We define  $\text{RESID}(\text{SMD}(e^{(t)}, P))$  as being the  $t$ -th lowest residual of the SMD considering the number of components per factor equal to  $P$ . Based on the definition of  $\text{RESID}(\text{SMD}(e^{(t)}, P))$ , one first direct way to estimate the model order  $d$  can be done by using the following properties

- 1) If there is no noise and  $P < d$ , then  $\text{RESID}(\text{SMD}(e^{(t)}, P)) > \text{RESID}(\text{SMD}(e^{(t)}, d))$ , since the matrices generated are composed of mixed components as shown in [13].
- 2) If the noise is present and  $P > d$ , then  $\text{RESID}(\text{SMD}(e^{(t)}, P)) > \text{RESID}(\text{SMD}(e^{(t)}, d))$ , since the matrices generated with the noise components are not diagonalizable commuting matrices. Therefore, the simultaneous diagonalizations are not valid anymore.

Based on these properties, a first model order selection scheme can be proposed

$$\widehat{d} = \arg \min_P \text{RESID}(\text{SMD}(e^{(1)}, P)). \quad (2.72)$$

However, the model order selection scheme in (2.72) gives us a Probability of correct Detection (PoD) inferior to the some MOS techniques found in the literature. Therefore, in order to improve the PoD of (2.72), we propose to exploit the redundant information provided only by the closed-form PARAFAC (CFP) [84].

Let  $\mathbf{F}_{e^{(t)}, P}^{(r)}$  denote the ordered sequence of estimates for  $\mathbf{F}^{(r)}$  assuming that the model order is  $P$ . For the correct model order and in the absence of noise, the subspaces of  $\mathbf{F}_{e^{(t)}, P}^{(r)}$

should not depend on  $t$ . Consequently, a measure for the reliability of the estimate is given by comparing the angle between the vectors  $\hat{\mathbf{f}}_{v,e(t),P}^{(r)}$  for different  $t$ , where  $\hat{\mathbf{f}}_{v,e(t),P}^{(r)}$  corresponds to the estimate of the  $v$ -th column of  $\mathbf{F}_{e(t),P}^{(r)}$ . Hence, this gives rise to an expression to estimate the model order using CFP-MOS

$$\begin{aligned} \hat{d} &= \arg \min_P \text{RMSE}(P) \quad \text{where} \\ \text{RMSE}(P) &= \Delta(P) \cdot \sqrt{\sum_{t=2}^{T_{\text{lim}}} \sum_{r=1}^R \sum_{v=1}^P \angle \left( \hat{\mathbf{f}}_{v,e(t),P}^{(r)}, \hat{\mathbf{f}}_{v,e(1),P}^{(r)} \right)}, \end{aligned} \quad (2.73)$$

where the operator  $\angle$  gives the angle between two vectors and  $T_{\text{lim}}$  represents the total number of simultaneous matrix diagonalizations taken into account.  $T_{\text{lim}}$ , a design parameter of the CFP-MOS algorithm, can be chosen between 2 and  $T$ . Similarly to the T-CORCONDIA in [1], the CFP-MOS requires weights  $\Delta(P)$ , otherwise the Probabilities of correct Detection (PoD) for different values of  $d$  have a significant gap from each other. Therefore, in order to have a fair estimation for all candidates  $P$ , we introduce the weights  $\Delta(P)$ , which are calibrated in a scenario with white Gaussian noise, where the number of sources  $d$  varies. For the calibration of weights, we use the probability of correct detection (PoD) of the  $R$ -D EFT [2, 1] as a reference, since the  $R$ -D EFT achieves the best PoD in the literature even in the low SNR regime. Consequently, we propose the following expression to obtain the calibrated weights  $\Delta_{\text{var}}$

$$\begin{aligned} \Delta_{\text{var}} &= \arg \min_{\Delta} J_{\text{var}}(\Delta) \quad \text{where} \\ J_{\text{var}}(\Delta) &= \sum_{P=d_{\min}}^{d_{\max}} \left| \mathbb{E}\{\text{PoD}_{\text{SNR}}^{\text{CFP-MOS}}(\Delta(P))\} - \mathbb{E}\{\text{PoD}_{\text{SNR}}^{R\text{-D EFT}}(P)\} \right| \end{aligned} \quad (2.74)$$

where  $\mathbb{E}\{\text{PoD}_{\text{SNR}}^{R\text{-D EFT}}(P)\}$  returns the averaged probability of correct detection over a certain predefined SNR range using the  $R$ -D EFT for a given scenario assuming  $P$  as the model order,  $d_{\max}$  is defined as being the maximum candidate value of  $P$ , and  $\Delta_{\text{var}}$  is the vector with the threshold coefficients for each value of  $P$ . Note that the elements of the vector of weights  $\Delta$  vary according to a certain defined range and interval and that the averaged PoD of the CFP-MOS is compared to the averaged PoD of the  $R$ -D EFT. When the cost function is minimized, then we have the desired  $\Delta_{\text{var}}$ .

Up to this point, the CFP-MOS is applicable to scenarios without any specific structure in the factor matrices. In case that the vectors  $\mathbf{f}_{v,e(t),P}^{(r)}$  have a Vandermonde structure, we can propose another expression. Again let  $\hat{\mathbf{F}}_{e(t),P}^{(r)}$  be the estimate for the  $r$ -th factor obtained from  $\text{SMD}(e(t), P)$ . Using the Vandermonde structure of each factor we can estimate the spatial frequencies  $\mu_{v,e(t),P}^{(r)}$  corresponding to the  $v$ -th column of  $\hat{\mathbf{F}}_{e(t),P}^{(r)}$ . Similarly to previous, for the correct model order and in the absence of noise, the estimated spatial frequencies should not depend on  $t$ . Consequently, a measure for the reliability of the estimate is given

by comparing the estimates for different  $t$ . Hence, this gives rise to the new cost function

$$\begin{aligned}\widehat{d} &= \arg \min_P \text{RMSE}(P) \quad \text{where} \\ \text{RMSE}(P) &= \Delta(P) \cdot \sqrt{\sum_{t=2}^{T_{\text{lim}}} \sum_{r=1}^R \sum_{v=1}^P \left( \widehat{\mu}_{v,e^{(t)},P}^{(r)} - \widehat{\mu}_{v,e^{(1)},P}^{(r)} \right)^2}.\end{aligned}\tag{2.75}$$

Similarly to cost function in (2.73), in order to have a fair estimation for all candidates  $P$ , we introduce the weights  $\Delta(P)$ , which are calculated considering in a similar fashion to the T-CORCONDIA Var [1] in Subsubsection 2.2.3.4 by considering data contaminated by white Gaussian noise.

In Subsection 2.2.4, we compare the proposed CFP-MOS technique to the following approaches:  $R$ -D EFT [2] proposed in Subsubsection 2.2.3.1,  $R$ -D AIC [2] and  $R$ -D MDL [2] proposed in Subsubsection 2.2.3.2, M-EFT [2] proposed in Subsubsection 2.1.3.2, EFT [42, 79] reviewed in Subsubsection 2.1.1.1, RADOI [80] reviewed in Subsubsection 2.1.1.1, AIC [98, 7] reviewed in Subsubsection 2.1.1.1, and MDL [98, 7] reviewed in Subsubsection 2.1.1.1. According to [7], for white noise, the M-EFT has the best performance compared to the state-of-the-art matrix-based model order selection techniques in the literature. For the case of colored noise, the RADOI technique has the best performance according to [80].<sup>6</sup>

### 2.2.3.6 Applying Forward-Backward Averaging (FBA)

Similarly to Subsubsection 2.1.3.3, for scenarios with centro-symmetric arrays, we can incorporate FBA [104] to all model order selection schemes even with a multi-dimensional data model. First let us present modifications in the data model, which should be considered in order to apply the FBA. Comparing the data model of (2.40) to the data model to be introduced in this subsection, we summarize two main differences. The first one is the size of  $\mathcal{X}_0$ , which has  $R+1$  dimensions instead of the  $R$  dimensions as in (2.40). Therefore, the noiseless data tensor is given by

$$\mathcal{X}_0 = \mathcal{I}_{R+1,d} \times_1 \mathbf{F}^{(1)} \times_2 \mathbf{F}^{(2)} \dots \times_R \mathbf{F}^{(R)} \times_{R+1} \mathbf{F}^{(R+1)} \in \mathbb{C}^{M_r \times d}.\tag{2.76}$$

This additional  $R+1$  dimension is due to the fact that the  $R+1$  factor represents the source symbols matrix  $\mathbf{F}^{(R+1)} = \mathbf{S}^T$ . The second difference is the restriction of the factor matrices  $\mathbf{F}^{(r)}$  for  $r = 1, \dots, R$  of the tensor  $\mathcal{X}_0$  in (2.76) to a matrix, where each vector is a function of a certain spatial frequency  $\mu_i^{(r)}$  related to the  $r$ -th dimension and the  $i$ -th source. In many applications, these vectors have a Vandermonde structure. For the sake of notation, the factor matrices for  $r = 1, \dots, R$  are represented by  $\mathbf{A}^{(r)}$ , and it can be written as a function of  $\mu_i^{(r)}$

---

<sup>6</sup> RADOI can be extended to  $R$ -D RADOI by replacing the eigenvalues of the sample covariance matrix by the global eigenvalues [2] similar to the definition of  $R$ -D AIC and  $R$ -D MDL. However, in our simulations we found that RADOI and  $R$ -D RADOI have the same performance. Therefore, we only include RADOI.

---





and  $p = 2 \cdot n + 1$ . On the other hand, if  $p$  is even, then  $\mathbf{Q}_p$  is given as

$$\mathbf{Q}_p = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} \mathbf{I}_n & j \cdot \mathbf{I}_n \\ \mathbf{\Pi}_n & -j \cdot \mathbf{\Pi}_n \end{bmatrix}, \quad (2.81)$$

and  $p = 2 \cdot n$ . The matrix  $\mathbf{\Pi}_n$  is defined in (2.35).

### 2.2.4 Evaluation of Approaches

In this subsection, we evaluate the performance, in terms of the probability of correct detection (PoD), of all multi-dimensional model order selection techniques presented previously via Monte Carlo simulations considering different scenarios.

In Fig. 2.38, *R-D AIC SQR* and *R-D MDL SQR* stands for *R-D AIC* and *R-D MDL* using (2.64), where (2.56) is replaced by (2.63) with  $u = 2$ . Similarly, *R-D AIC CUB* and *R-D MDL CUB* stands for *R-D AIC* and *R-D MDL* using (2.64), where (2.56) is replaced by (2.63) with  $u = 3$ . In general, the *R-D MOS* selection schemes outperform significantly the matrix-based MOS schemes as shown in [2]. Comparing *R-D AIC SQR* and *R-D AIC*, a difference of more than 2.5 dB is obtained with this small change. By increasing more  $u$  as in *R-D AIC CUB* and *R-D MDL CUB*, we obtain a  $PoD \approx 90\%$  and a  $PoD \approx 80\%$ , respectively. In general, we have also performed several simulations for different value of  $u$ , and the best value is for  $u = 2$ .

Note that *R-D AIC SQR* has the same PoD as *R-D EFT* in Fig. 2.38. Therefore, the application of the exponentiation of the global eigenvalues improves significantly the PoD compared to the *R-D* model order selection approaches. One reason for that is the fact that the exponentiation also increases the threshold distance similarly to the global eigenvalues as shown in Fig. 2.2.3.1.

In Figures 2.39 and 2.40 we observe the performance of the classical methods and the extended methods for a scenario with the following dimensions  $M_1 = 7$ ,  $M_2 = 7$ ,  $M_3 = 7$  and  $M_4 = 7$ . The methods described as M-EFT, AIC, and MDL are the simplified one-dimensional case of the *R-D* methods, in which we consider only one unfolding for  $r = 4$ .

All the methods based on eigenvalues and global eigenvalues outperform T-CORCONDIA Fix, with a vector of equal thresholds, and T-CORCONDIA Var, with a vector of variable thresholds. In addition, if we compare the curves of T-CORCONDIA Var in Figs. 2.39 and 2.40, we can realize that they have a similar profile, e.g., the peak of the former is approximately 77 % and peak of the latter is roughly 74 %. While the peaks from the T-CORCONDIA Fix have a drastical variation from 99 % to 77 % comparing the two figures. Therefore we observe that the fixed coefficient  $\Delta$  resulted in an acceptable performance for  $d = 3$  but was not the optimal choice for  $d = 4$ . The reason that T-CORCONDIA has peaks of PoD may be related to the fact that CORCONDIA is proposed for structured noise according to [15].

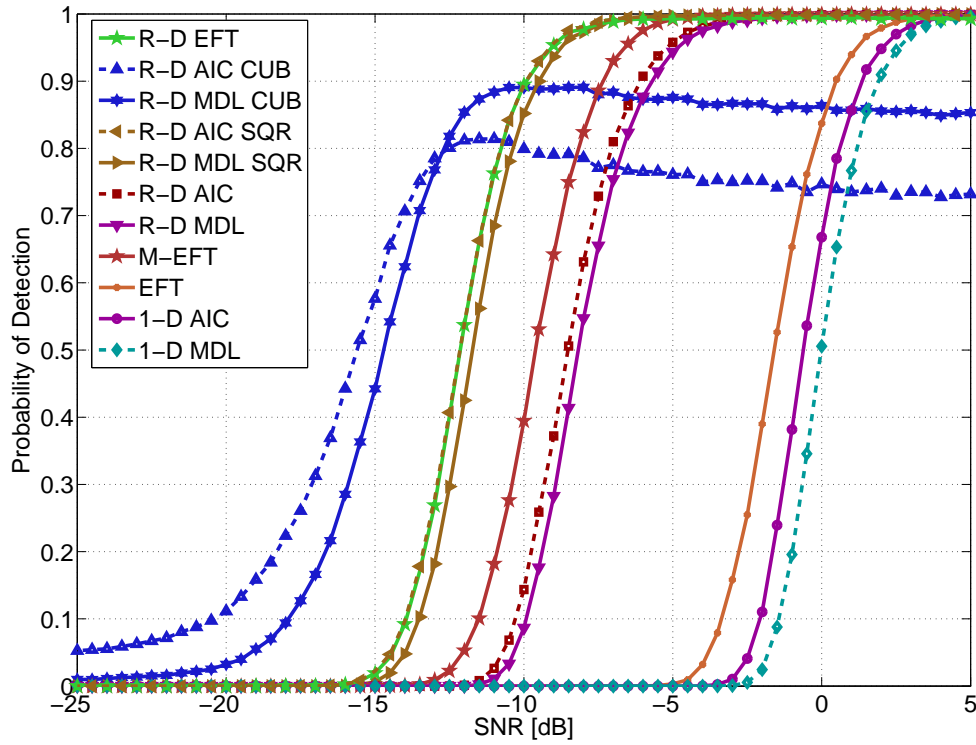


Fig. 2.38: Probability of correct Detection (PoD) vs. SNR for an array of size  $M_1 = 5$ ,  $M_2 = 7$ ,  $M_3 = 9$  and  $M_4 = 10$ . The number of sources  $d$  is set to 3. *R-D* MOS versus matrix-based MOS.

Comparing the two versions of the CORCONDIA and the HOSVD-based approaches, we can notice that the computational complexity is much lower in the *R-D* methods. Moreover, the HOSVD-based approaches outperform the iterative approaches, since none of them are close to the 100 % Probability of correct Detection (PoD). The techniques based on global eigenvalues, *R-D* EFT, *R-D* AIC, and *R-D* MDL maintain a good performance even for lower SNR scenarios, and the *R-D* EFT shows the best performance if we compare all the techniques.

In Figs. 2.41 and 2.42, we compare our proposed approach to all mentioned techniques for the case that white noise is present. To compare the performance of CFP-MOS for various values of the design parameter  $T_{\text{lim}}$ , we select  $T_{\text{lim}} = 2$  for the legend CFP 2f and  $T_{\text{lim}} = 4$  for CFP 4f. In Fig. 2.41, the model order  $d$  is equal to 2, while in Fig. 2.42,  $d = 3$ . In these two scenarios, the proposed CFP-MOS has a performance very close to *R-D* EFT, which has the best performance.

In this part of the simulations, we assume the noise correlation structure of equation (2.45), where  $\mathbf{W}_i$  of the  $i$ -th factor for  $M_i = 3$  is given by

$$\mathbf{W}_i = \begin{bmatrix} 1 & \rho_i^* & (\rho_i^*)^2 \\ \rho_i & 1 & \rho_i^* \\ \rho_i^2 & \rho_i & 1 \end{bmatrix}, \quad (2.82)$$

where  $\rho_i$  is the correlation coefficient. Note that also other types of correlation models different from (2.82) can be used.

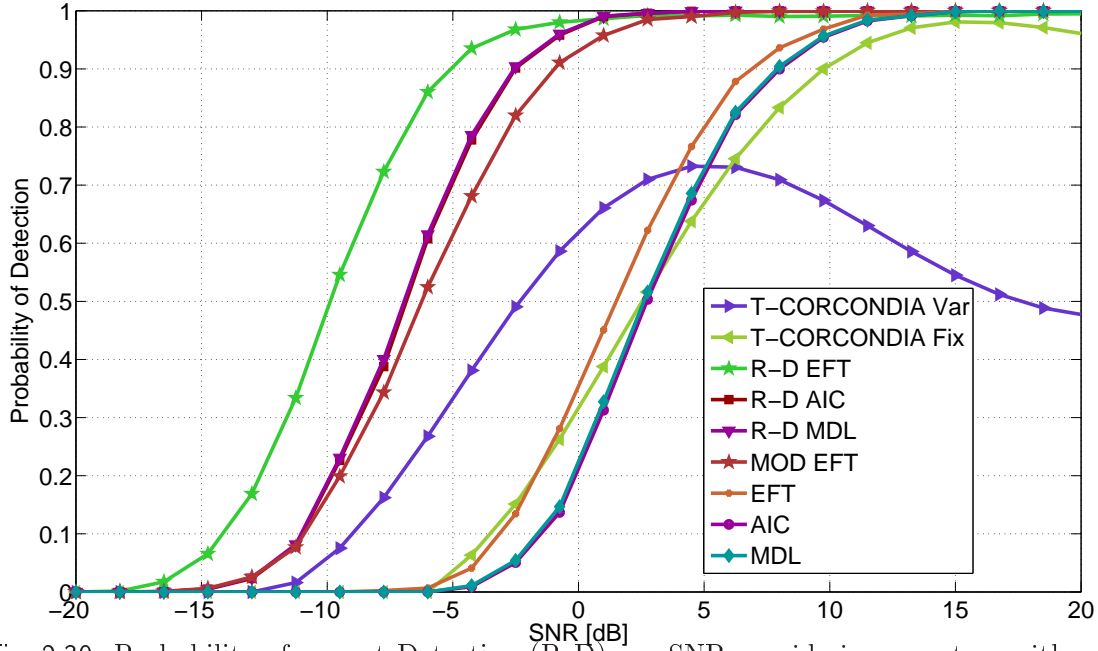


Fig. 2.39: Probability of correct Detection (PoD) vs. SNR considering a system with a data model of  $M_1 = 7$ ,  $M_2 = 7$ ,  $M_3 = 7$ ,  $M_4 = 7$  and  $d = 3$  sources.

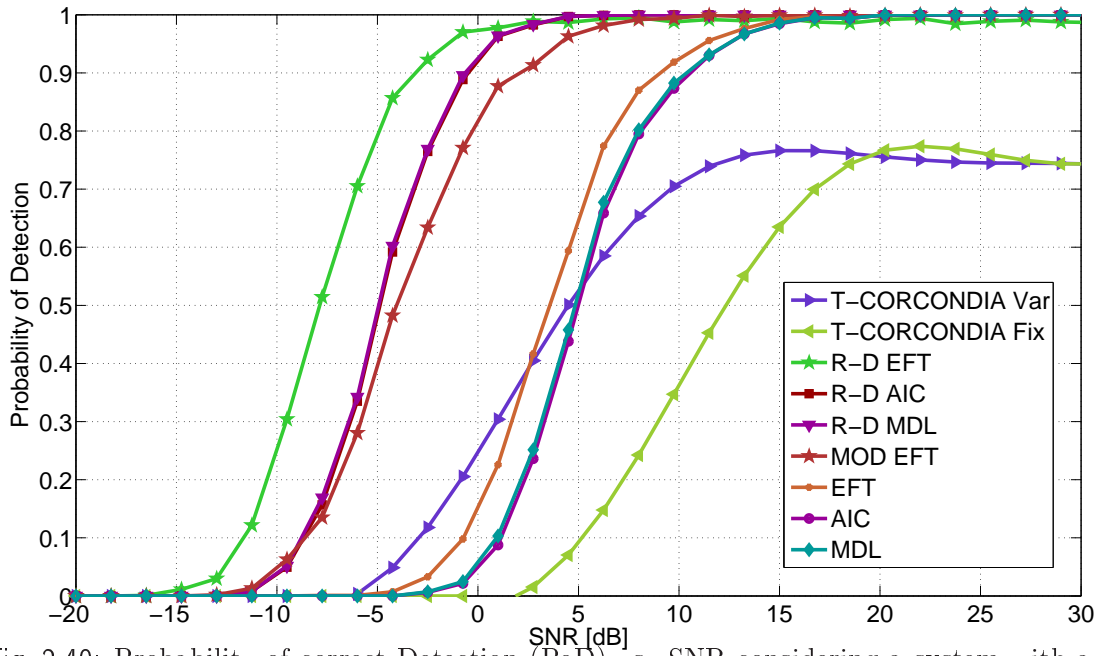


Fig. 2.40: Probability of correct Detection (PoD) vs. SNR considering a system with a data model of  $M_1 = 7$ ,  $M_2 = 7$ ,  $M_3 = 7$ ,  $M_4 = 7$  and  $d = 4$  sources.

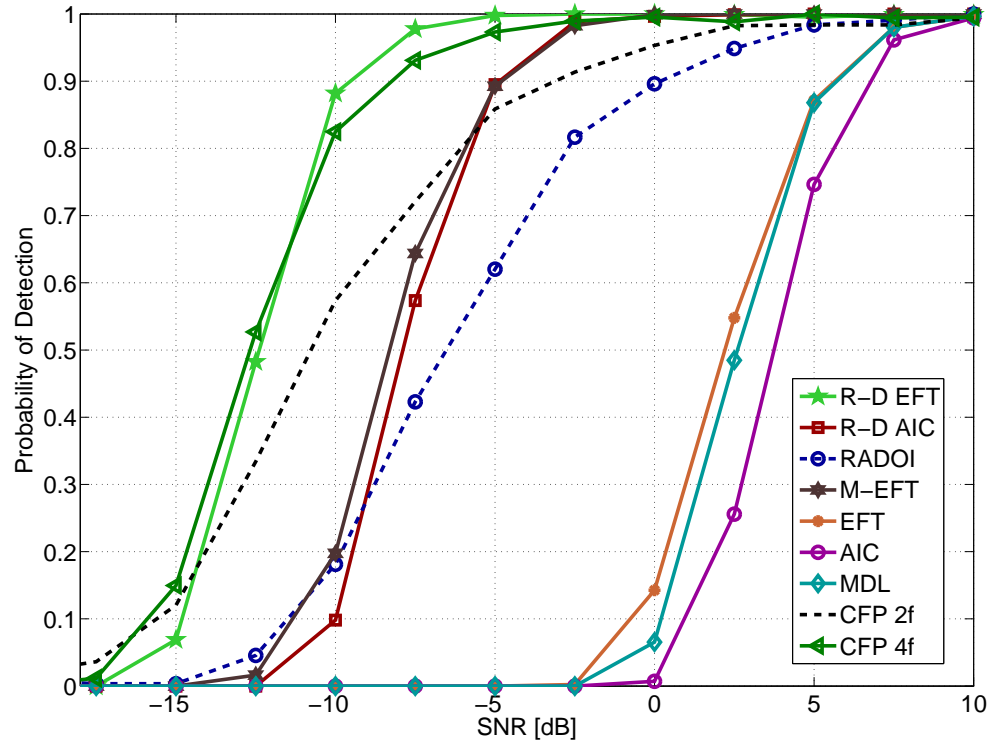


Fig. 2.41: Probability of correct Detection (PoD) versus SNR. In the simulated scenario,  $R = 5$ ,  $M_1 = 5$ ,  $M_2 = 5$ ,  $M_3 = 5$ ,  $M_4 = 5$ ,  $M_5 = 5$ , and  $N = 5$  presence of white noise. We fixed  $d = 2$ .

In Figs. 2.43 and 2.44, the noise is colored with a very high correlation, and the factors  $\mathbf{L}_i$  are computed based on (2.45) and (2.82) as a function of  $\rho_i$ . As expected for this scenario, the *R-D EFT*, *R-D AIC*, and *R-D MDL* completely fail. In case of colored noise with high correlation, the noise power is much more concentrated in the signal components. Therefore, the smaller are the values of  $d$ , the worse is the PoD. The behaviour of the CFP-MOS, AIC, MDL, and EFT are consistent with this effect. The PoD of AIC, MDL, and EFT increases from 0.85, 0.7, and 0.7 in Fig. 2.43, to 0.9, 0.85, and 0.85 in Fig. 2.44, respectively. CFP-MOS 4f has a PoD = 0.98 for SNR = 20 dB in Fig. 2.43, while a PoD = 0.98 for SNR = 15 dB in Fig. 2.44.

In contrast to CFP-MOS, AIC, MDL, and EFT, the PoD of RADOI degrades from Fig. 2.43 to Fig. 2.44. In Fig. 2.43, RADOI has a better performance than the CFP-MOS version, while in Fig. 2.44, CFP-MOS outperforms RADOI. Note that PoD for RADOI becomes constant for  $\text{SNR} \leq 3$  dB, which corresponds to a biased estimation. Therefore, for severely colored noise scenarios, the model order selection using CFP-MOS is more stable than the other approaches.

Note that the Kronecker noise or interference defined in (2.43) corresponds to the most critical type of noise for the proposed CFP based model order selection (CFP-MOS), since the CFP-MOS takes into account the Kronecker structure, which both signal and noise have in this case.

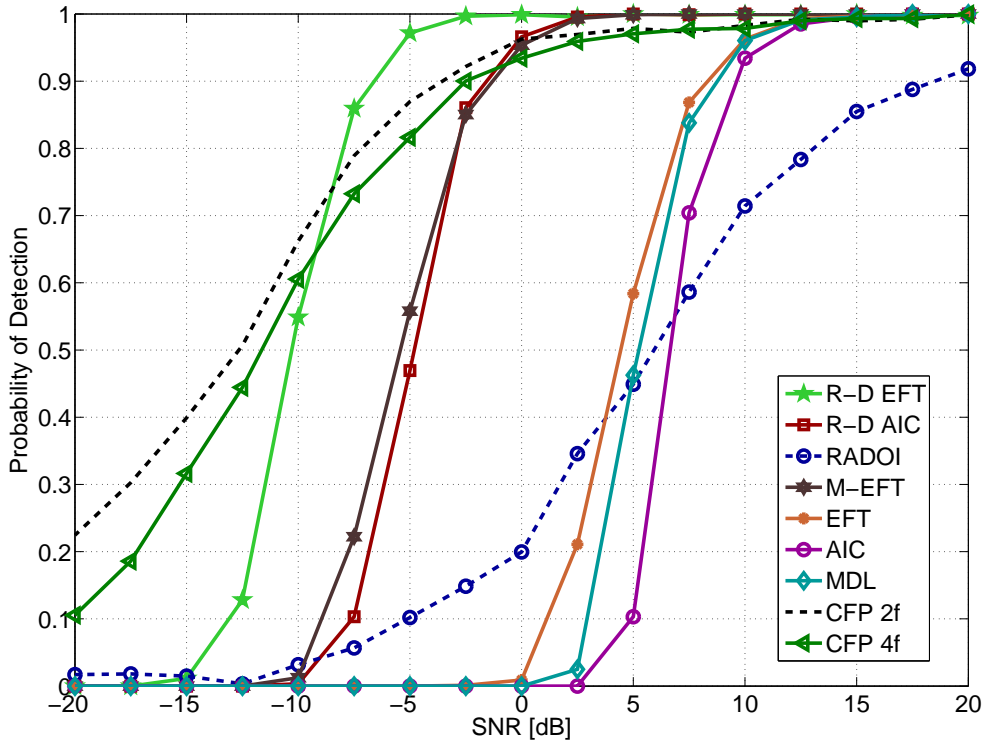


Fig. 2.42: Probability of correct Detection (PoD) versus SNR. In the simulated scenario,  $R = 5$ ,  $M_1 = 5$ ,  $M_2 = 5$ ,  $M_3 = 5$ ,  $M_4 = 5$ ,  $M_5 = 5$ , and  $N = 5$  presence of white noise. We fixed  $d = 3$ .

In Fig. 2.45, no FBA is applied in all model order selection techniques, while in Fig. 2.46 FBA is applied in all of them according to Subsubsection 2.2.3.6. In general, an improvement of approximately 3 dB is obtained when FBA is applied.

In Fig. 2.46,  $d = 3$ . Therefore, by using the adaptive definition of the global eigenvalues from Subsubsection 2.2.3.1, we can estimate the model order considering 4 modes. By increasing the number of sources to 5 in Fig. 2.47, the adaptive definition of the global eigenvalues is computed considering the second, third and fourth modes, which are related to  $M_2$ ,  $M_3$  and  $N$ .

By increasing the number of sources even more such that only one mode can be applied, the curves of the  $R$ -D EFT,  $R$ -D AIC and  $R$ -D MDL superpose the curves of M-EFT, 1-D AIC, and 1-D MDL as shown in Fig. 2.48.

#### 2.2.4.1 MOS for Special Array Sizes

As shown previously, the  $R$ -D EFT should be used instead of the M-EFT in case of multi-dimensional data model due to its improved estimation of the model order. However, the computational complexity of the  $R$ -D EFT is greater than the M-EFT, since for the former  $R + 1$  HOSVDs should be computed, and for the latter only one SVD should be computed.

In this subsection, we consider scenarios, where the M-EFT and the  $R$ -D have similar

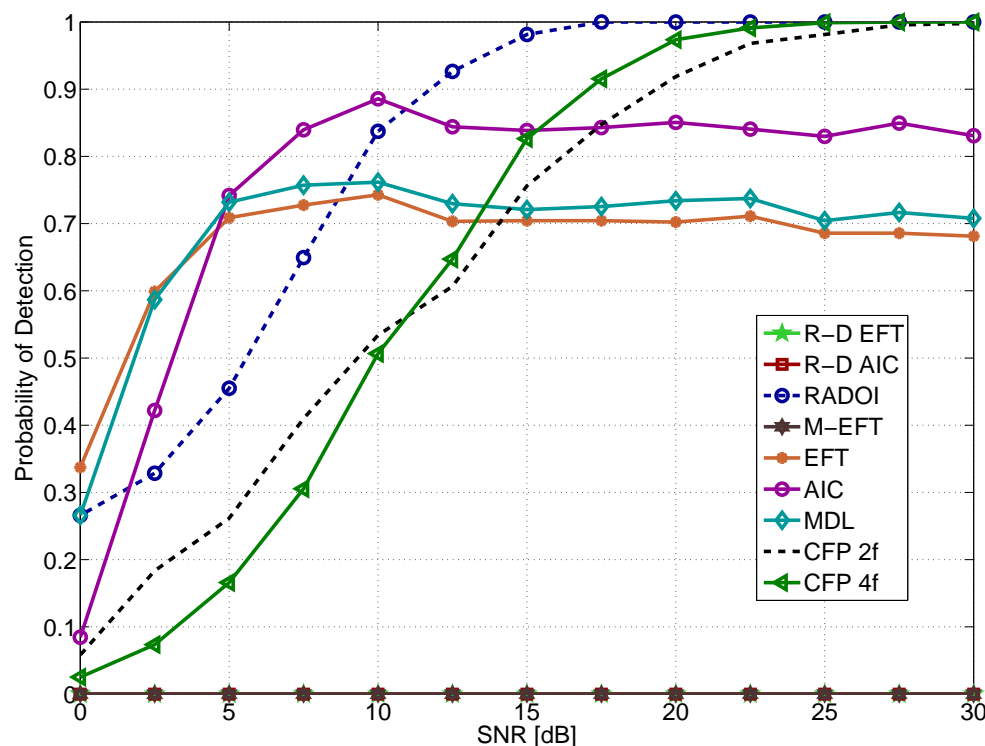


Fig. 2.43: Probability of correct Detection (PoD) versus SNR. In the simulated scenario,  $R = 5$ ,  $M_1 = 5$ ,  $M_2 = 5$ ,  $M_3 = 5$ ,  $M_4 = 5$ ,  $M_5 = 5$ , and  $N = 5$  presence of colored noise, where  $\rho_1 = 0.9$ ,  $\rho_2 = 0.95$ ,  $\rho_3 = 0.85$ , and  $\rho_4 = 0.8$ . We fixed  $d = 2$ .

performance, and, since the M-EFT has a smaller computational complexity, the M-EFT can be used instead of the  $R$ -D EFT. Let us consider a scenario, where the difference between the sizes of each dimension of the multi-dimensional data may be significant. In other words,  $M_r \gg M_i$ , where  $i = 1, \dots, r-1, r+1, \dots, R+1$ . In this case, we propose to compute the M-EFT only to the  $r$ -th unfolding, and the performance is very close to the one obtained by  $R$ -D EFT. However, we emphasize that to stay in the safe side, i.e., to be sure that the best estimation of the model order is achieved, the  $R$ -D EFT can be preferentially applied.

To exemplify the previous problem, we consider a scenario where  $M_1 = 4 \cdot M_2$ , and  $M_1 = 2 \cdot N$  in Figure 2.49. Here the M-EFT is applied for each of the three possible unfoldings, and the PoD using the M-EFT in the first unfolding, called M-EFT 1, is better than the PoD of the third unfolding, called M-EFT 3, which is also better than the PoD of the second unfolding, called M-EFT 2. In general, if the number of sensors in one dimension is at least two times bigger than the number of sensors in the second highest dimension, then we apply the M-EFT on the greatest dimension instead of applying the  $R$ -D EFT [2] to the whole tensor.

## 2.2.5 Main Results of the Section

In this section, we have proposed multi-dimensional model order selection techniques. After comparing all techniques proposed by us, and also techniques found in the literature, we have

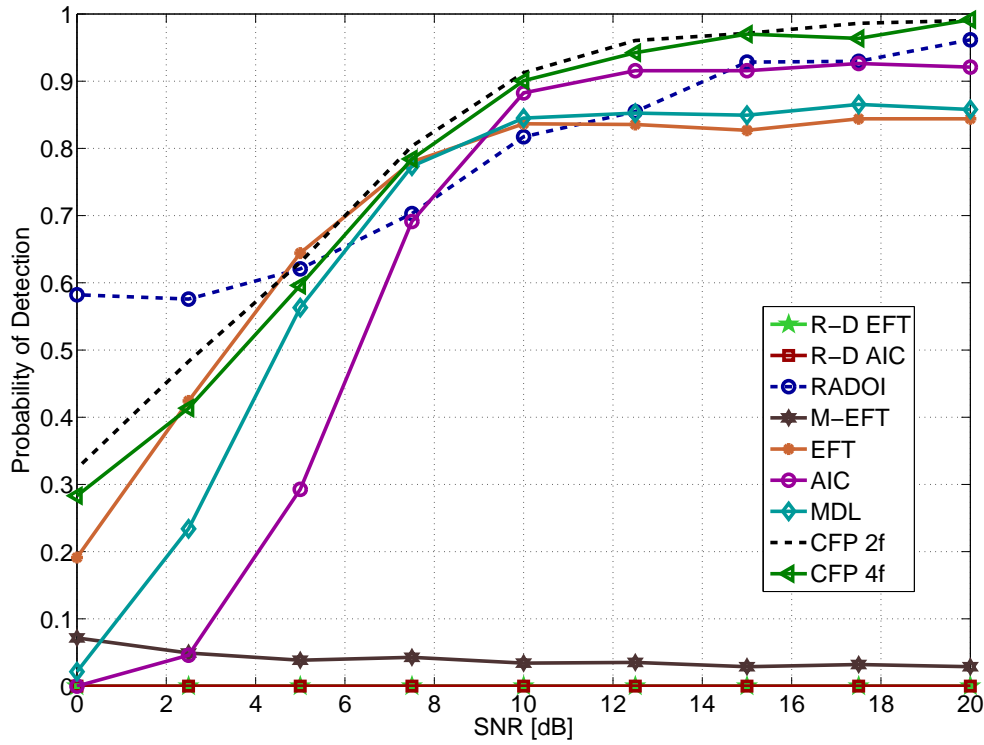


Fig. 2.44: Probability of correct Detection (PoD) versus SNR. In the simulated scenario,  $R = 5$ ,  $M_1 = 5$ ,  $M_2 = 5$ ,  $M_3 = 5$ ,  $M_4 = 5$ ,  $M_5 = 5$ , and  $N = 5$  presence of colored noise, where  $\rho_1 = 0.9$ ,  $\rho_2 = 0.95$ ,  $\rho_3 = 0.85$ , and  $\rho_4 = 0.8$ . We fixed  $d = 3$ .

achieved the following results considering a multi-dimensional data model.

- 1) In case of white Gaussian noise scenarios, our  $R$ -D EFT outperforms the other techniques presented in the literature.
- 2) In the presence of colored noise, the CFP-MOS is the best technique, since it has a performance close to the  $R$ -D EFT in case of no correlation, and a performance more stable than RADOI, in case of severely correlated noise.
- 3) For researchers, which prefer to use ITC techniques, we have also proposed multi-dimensional extensions of AIC and MDL, called  $R$ -D AIC and  $R$ -D MDL, respectively.

In Table 2.7, we summarize the scenarios to apply the different techniques shown in this section. Also in Table 2.7, wht stands for white noise and clr stands for colored noise. Note that the PoD of the CFP-MOS is close to one of the  $R$ -D EFT for white noise, which means that it has a multi-dimensional gain. Moreover, since the CFP-MOS is suitable for white and colored noise applications, we consider it the best scheme.



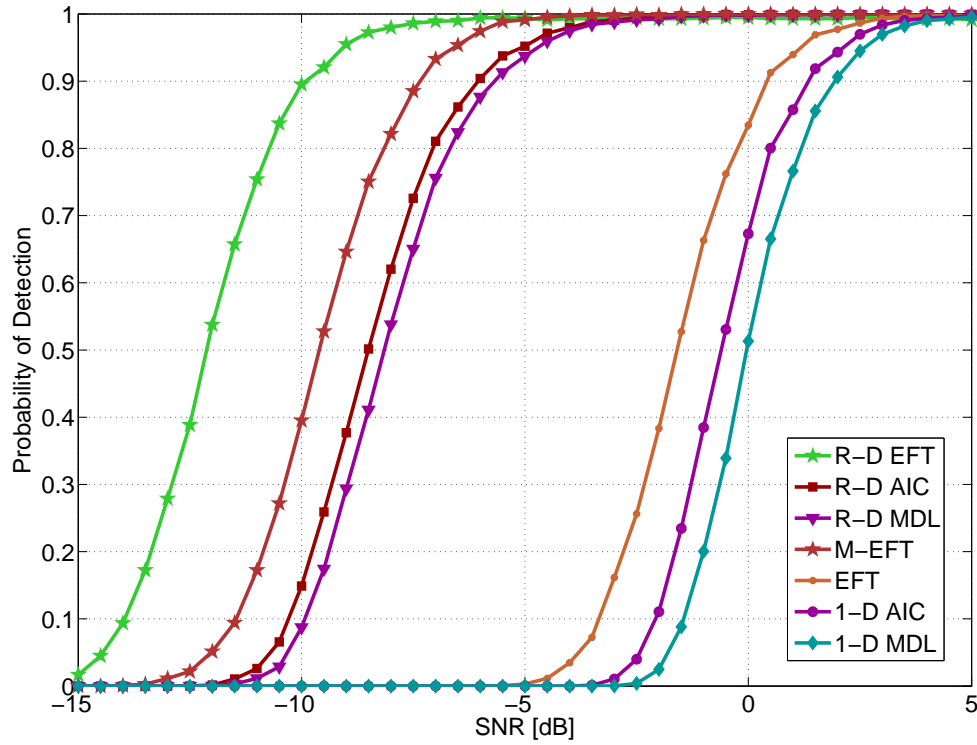


Fig. 2.45: Probability of correct Detection (PoD) vs. SNR for an array of size  $M_1 = 5$ ,  $M_2 = 7$ , and  $M_3 = 9$ . The number of snapshots  $N$  is set to 10 and the number of sources  $d = 3$ . No FBA is applied.

Scheme	Minimum $d$	Maximum $d$	Noise	Performance
T-CORCONDIA [15, 1]	1	Typical rank [24]	Wht and clr	Comparable to 1-D AIC
$R$ -D AIC [2, 1]	0	$\max_r \left[ \min \left( M_r, \frac{M}{M_r} \right) - 1 \right]$	Wht	Superior to 1-D AIC
$R$ -D MDL [2, 1]	0	$\max_r \left[ \min \left( M_r, \frac{M}{M_r} \right) - 1 \right]$	Wht	Superior to 1-D MDL
$R$ -D EFT [2, 1]	0	$\max_r \left[ \min \left( M_r, \frac{M}{M_r} \right) - 1 \right]$	Wht	Best
CFP-MOS [6]	1	$\max_r (M_r)$	Wht and clr	Best

Tab. 2.7: Summarized table comparing characteristics of the multi-dimensional model order selection schemes.

## 2.3 Main Results of the Chapter

In this chapter, a deep review of the several model order selection schemes in the literature is done and several model order selection schemes are also proposed by us. As an important result, we divide our problem in four cases, and we propose for each case the application of a certain model order selection scheme.

The first case is using the matrix-based data model contaminated by white Gaussian noise, and for this problem our proposed approach called M-EFT, which is shown in Subsubsection 2.1.3.2, is the best approach, since it outperforms all the other model order selection schemes for all possible array sizes.

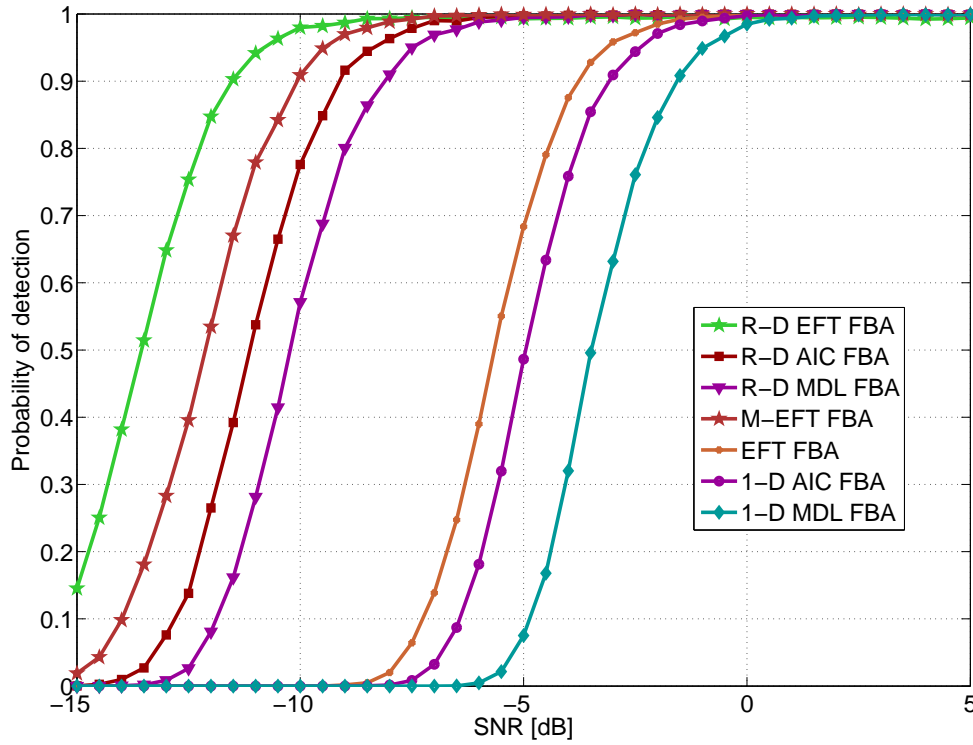


Fig. 2.46: Probability of correct Detection (PoD) vs. SNR for an array of size  $M_1 = 5$ ,  $M_2 = 7$ , and  $M_3 = 9$ . The number of snapshots  $N$  is set to 10 and the number of sources  $d = 3$ . FBA is applied.

The second case considers also matrix-based data model, but contaminated by severe colored noise, and for this problem the scheme called RADOI [80] outperforms all the other model order selection schemes as shown in the simulations in Subsection 2.1.4.

For the third case, where multi-dimensional data model in the presence of white Gaussian noise is considered, the *R-D EFT* scheme, which is described in Subsubsection 2.2.3.1 and is also proposed by us, should be definitely applied.

Finally, for scenarios with multi-dimensional data model in the presence of colored Gaussian noise, the CFP-MOS scheme, also proposed by us and presented in Subsubsection 2.2.3.5, is the best method.

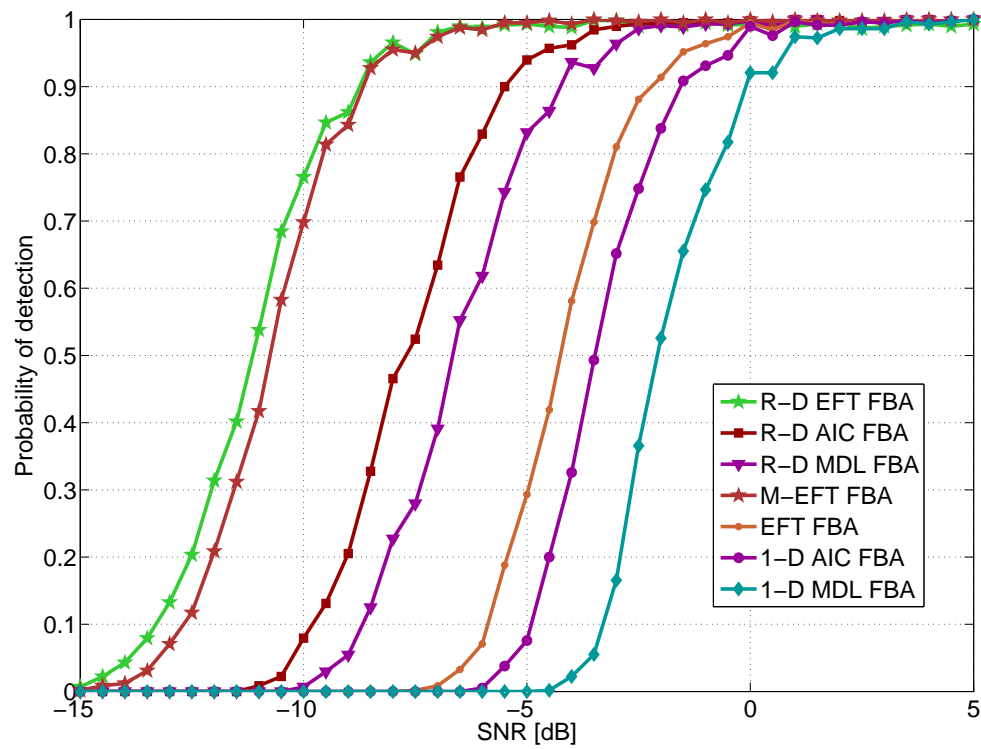


Fig. 2.47: Probability of correct Detection (PoD) vs. SNR for an array of size  $M_1 = 5$ ,  $M_2 = 7$ , and  $M_3 = 9$ . The number of snapshots  $N$  is set to 10 and the number of sources  $d = 5$ . FBA is applied.

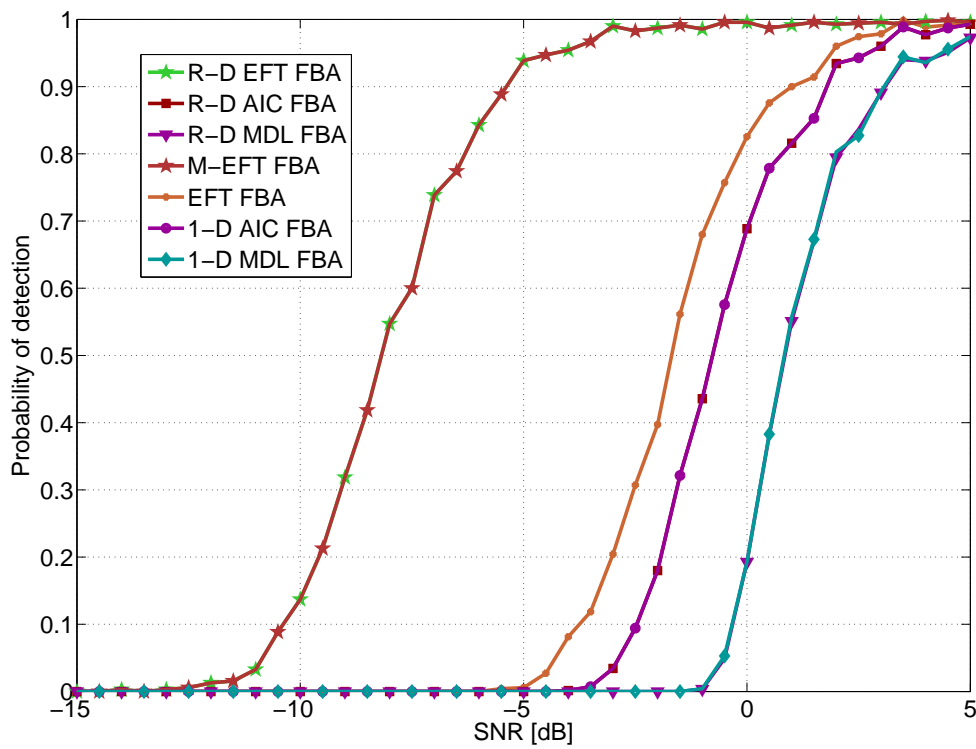


Fig. 2.48: Probability of correct Detection (PoD) vs. SNR for an array of size  $M_1 = 5$ ,  $M_2 = 7$ , and  $M_3 = 9$ . The number of snapshots  $N$  is set to 10 and the number of sources  $d = 9$ . FBA is applied.

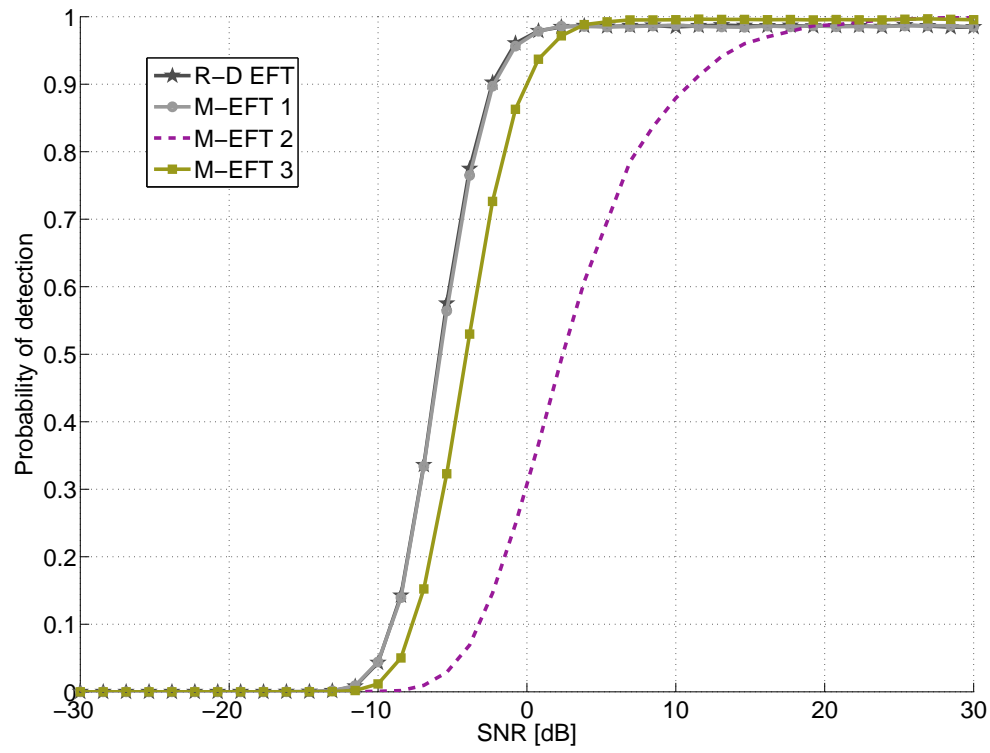


Fig. 2.49: The case of having the tensor with one dimension much higher than the other dimensions. The one-dimensional solution is the best one. Probability of correct Detection (PoD) vs. SNR for an array of size of  $M_1 = 20$  and  $M_2 = 5$ . The number of snapshots  $N$  is set to 10 and the number of sources  $d = 4$ .



### 3. $R$ -D PARAMETER ESTIMATION VIA CLOSED-FORM PARAFAC

High-resolution parameter estimation from  $R$ -dimensional signals is a task required for a variety of applications, such as estimating the multi-dimensional parameters of the dominant multipath components from MIMO channel measurements, radar, sonar, seismology, medical imaging and other medical applications, like the Intelligent Hearing Aid [27, 28]. For this task, many techniques are found in the literature, and particularly the closed-form techniques such as *Standard Tensor-ESPRIT (STE)* [83, 44] are appealing, since they are not iterative and their performance is very close to the Cramér-Rao lower bound (CRLB) [83, 44].

As shown in [66], iterative techniques for the PARAFAC decomposition such as Trilinear Alternating Least Squares (TALS) and Quadrilinear Alternating Least Squares (QALS) can also be applied to estimate spatial frequencies. By applying the PARAFAC decomposition to the multi-dimensional data,  $R$  desired factors each with  $d$  components are returned. Therefore, the multi-dimensional maximization problem with a high computational complexity is transformed into  $R \cdot d$  one-dimensional peak search problems with a very small computational complexity. Note that TALS and QALS are specific versions of the *Multilinear Alternating Least Squares (MALS)* algorithm.

Originally, the PARAFAC decomposition has been first applied to psychometrics in [48, 17]. Later many other fields have also applied it, for example, chemistry [67] and telecommunications [66]. Since the PARAFAC decomposition is very general, i.e., it only assumes the Kronecker structure of the data, some robustness is expected in its estimation. Such robustness of the PARAFAC decomposition is related to the decoupling of the dimensions, which avoids that the error in one dimension affects the other dimensions. The PARAFAC decomposition allows us to estimate parameters in arrays where the shift-invariance structure is not present, and, consequently, the STE cannot be applied. Other important advantage is the robustness exploited by us in [6] for multi-dimensional parameter estimation in arrays with positioning errors. In general, by the application of the PARAFAC decomposition, the multi-dimensional data is decomposed into rank-one tensors with automatic pairing, which relieves us from the computationally expensive multi-dimensional searches.

According to our work in [6], the *closed-form PARAFAC based parameter estimation (CFP-PE)* scheme, which is non-iterative and based on the HOSVD [29], achieves a performance even better than MALS [66] and  $R$ -D STE [83, 44] for scenarios with modeling errors. Therefore, instead of the iterative MALS, the closed-form PARAFAC (CFP) decomposition [84, 100]

should be preferentially used for high-resolution parameter estimation.<sup>1</sup>

In this chapter, we propose the closed-form PARAFAC based parameter estimation (CFP-PE) scheme [6]. Next we list some of the other important contributions of this chapter. The first one is the application of the closed-form PARAFAC (CFP) combined with different other techniques, e.g., shift invariance (SI), least squares Khatri-Rao factorization (LSKRF), and peak search (PS) in order to estimate the spatial frequencies. Advantages and drawbacks for each possible combination are investigated. The second one is the increase of the maximum model order applicable for the CFP by merging dimensions. Finally the comparison of the CFP-PE to all the estimated spatial frequencies via  $R$ -D STE,  $R$ -D UTE, and MALS is performed. We show by means of simulations that our proposed scheme is more general and robust.

As presented in Subsection 3.4, compared to the ESPRIT based approaches, the CFP presents two important advantages. The first one is the application to arrays that do not satisfy the shift invariance equation<sup>2</sup>, and the second one is the robustness by estimating spatial frequencies with arrays with positioning errors. Additionally, the parameter estimation using CFP outperforms significantly MALS for *IlmProp* [30, 31] scenarios, where the narrowband assumption is violated.

Some open problems solved by us in this chapter are presented below.

- 1) Since MALS and CFP are different approaches to decompose the multi-dimensional data into PARAFAC factors, we show if there is some advantage in using the former or the latter in case of applications for parameter estimation.
- 2) For scenarios with modeling errors, we present that the CFP based approaches have a superior performance to the  $R$ -D STE.
- 3) Since CFP has a limited maximum rank for the decomposition of  $\mathcal{X}$ , we propose a way to increase it by merging dimensions.
- 4) We show for which circumstances it is more advantageous to apply the Peak Search (PS) or the Shift Invariance (SI) combined with the CFP, and also how to combine the CFP with the Least Squares Kathri-Rao Factorization (LSKR) for the parameter estimation problem.
- 5) We investigate the effect of the violation of the narrowband assumption for the estimation of spatial frequencies using the MALS based approach, CFP based approach, and  $R$ -D STE.

During the whole chapter, we consider the multi-dimensional parameter estimation data model presented in detail in Section 3.1. Moreover, we assume in this chapter that the number of

---

<sup>1</sup> For nonnegative tensor factorizations, which is not the scope of our work, the reader is referred to [22].

<sup>2</sup> Concepts of the shift invariance property are reviewed in Subsubsection 2.1.2.1.



PARAFAC components  $d$  is known, and in order to estimate the model order, the reader is referred to the Chapter 2.

We divide this chapter as follows. Our goal and problem are introduced in Section 3.1, while the previous approaches are reviewed in Section 3.2. In Section 3.3, we describe our proposed closed-form PARAFAC based Parameter Estimation (CFP-PE) scheme. The simulations are presented in Section 3.4, and the main results are summarized in Section 3.5.

### 3.1 Goal and Problem

In general, the observations may be modeled as a superposition of  $d$  damped exponentials sampled on an  $R$ -dimensional grid of size  $M_1 \times M_2 \times \dots \times M_R$  at  $N$  subsequent time instants. The measurement samples are given by

$$x_{m_1, m_2, \dots, m_R, n} = \sum_{i=1}^d s_i(n) \cdot \prod_{r=1}^R e^{(m_r-1) \cdot (\zeta_i^{(r)} + j \cdot \mu_i^{(r)})} + n_{m_1, m_2, \dots, m_R, n}, \quad (3.1)$$

where  $m_r = 1, 2, \dots, M_r$  for  $r = 1, 2, \dots, R$ ,  $n = 1, 2, \dots, N$ ,  $s_i(n)$  denotes the complex amplitude of the  $i$ -th exponential at time instant  $n$ ,  $\mu_i^{(r)}$  symbolizes the spatial frequency of the  $i$ -th exponential in the  $r$ -th mode,  $\zeta_i^{(r)}$  represents the corresponding damping factor, and  $n_{m_1, m_2, \dots, m_R, n}$  models the additive noise component inherent in the measurement process. Here we assume that  $\zeta_i^{(r)} = 0$ . In the context of array signal processing, each of the complex exponentials represents one planar wavefront and the complex amplitudes  $s_i(n)$  are the symbols.

In the classical matrix representation, we have the following data model

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} + \mathbf{N}, \quad (3.2)$$

where  $\mathbf{X} \in \mathbb{C}^{M \times N}$  contains the stacked measurements, the matrix  $\mathbf{S} \in \mathbb{C}^{d \times N}$  contains the symbols  $s_i(n)$ , the noise samples are collected in the matrix  $\mathbf{N} \in \mathbb{C}^{M \times N}$  and the array steering matrix  $\mathbf{A}$  [43] is defined in the following way

$$\mathbf{A} = \mathbf{A}^{(1)} \diamond \mathbf{A}^{(2)} \dots \diamond \mathbf{A}^{(R)} \in \mathbb{C}^{M \times d}, \quad (3.3)$$

where  $\diamond$  stands for the Khatri-Rao product already shown in (2.50 and defined in (F.2) in Appendix F,  $M = \prod_{r=1}^R M_r$ , and

$$\mathbf{A}^{(r)} = \left[ \mathbf{a}^{(r)}(\mu_1^{(r)}), \mathbf{a}^{(r)}(\mu_2^{(r)}), \dots, \mathbf{a}^{(r)}(\mu_d^{(r)}) \right] \in \mathbb{C}^{M_r \times d}. \quad (3.4)$$

The vector  $\mathbf{a}^{(r)}(\mu_i^{(r)}) \in \mathbb{C}^{M_r \times 1}$  denotes the array response in the  $r$ -th dimension for the  $i$ -th source. Here, all the spatial dimensions are stacked into column vectors.

The array in (3.3) is called an outer product based array (OPA), since it can be written

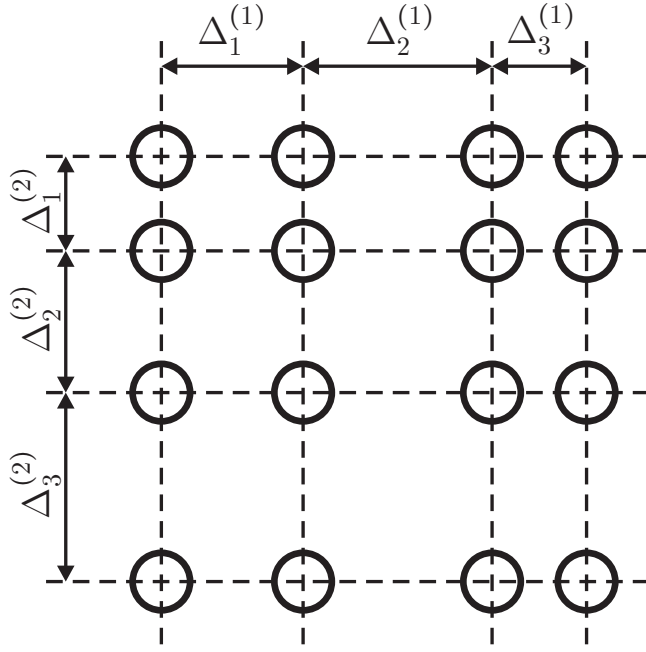


Fig. 3.1: An example of a 2-D outer product based array (OPA) of size  $4 \times 4$  is illustrated. All the distances  $\Delta_i^{(r)}$  for  $i = 1, 2, 3$  and for  $r = 1, 2$  may assume different values.

as

$$\mathcal{A} = \sum_{i=1}^d \mathbf{a}^{(1)}(\mu_i^{(1)}) \circ \mathbf{a}^{(2)}(\mu_i^{(2)}) \circ \dots \circ \mathbf{a}^{(R)}(\mu_i^{(R)}) \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}, \quad (3.5)$$

where  $\circ$  is the outer product operator. Note that (3.1) implies that (3.4) has a Vandermonde structure. In Chapter 3, only in some scenarios the Vandermonde structure is assumed, while in Chapter 4, this structure is assumed in the whole chapter.

For instance, a 2-D OPA can be formed as an outer product of two uniform linear arrays (ULAs), which gives rise to a uniform rectangular array (URA). On the other hand, the outer product of non-uniform linear arrays (NULAs) is also an OPA, which we refer to as non-uniform rectangular array (NURA). We exemplify a 2-D OPA formed by NULAs in Fig. 3.1. OPAs with more dimensions are typically present if a sensor array is used to measure a signal at consecutive time instants and/or on a regular grid of frequency bins.

In case of ULAs/URAs, the corresponding arrays steering vectors have a Vandermonde structure, while NULAs/NURAs can be divided into arrays with shift invariant array steering vectors (where ESPRIT-type methods are applicable), and those without shift invariance. The arrays that are non-outer product arrays are called here arbitrary arrays. Note that in Chapter 3 we also study the impact of the calibration errors, which means that the actual sensor positions are not perfectly known.

The CFP based approach is indicated for Outer Product based Arrays (OPAs). Moreover, our proposed technique is also applicable for mixed arrays, which have part of the dimensions

with Arbitrary Arrays (AAs) and the others with OPAs. Note that physical dimensions that do not possess an outer product structure are represented by a single dimension in our model. For example, 2-D array without outer product structure is represented by one array steering matrix  $\mathbf{A}^{(r)} \in \mathbb{C}^{M_r \times d}$ , like a 1-D array.

The matrix representation in (3.2) does not capture the structure inherent in the lattice that is used to sample the data due to the stacking operation. We therefore replace the measurement matrix  $\mathbf{X}$  by a measurement tensor  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times N}$ . Similarly to (3.2),  $\mathcal{X}$  can be modeled as

$$\mathcal{X} = \mathcal{A} \times_{R+1} \mathbf{S}^T + \mathcal{N}. \quad (3.6)$$

Here the matrix  $\mathbf{S}$  is the same as in (3.2), the tensor  $\mathcal{N}$  contains the noise samples, and the tensor  $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$  is termed the array steering tensor. It can be computed from the array response vectors  $\mathbf{a}^{(r)}(\mu_i^{(r)})$  through the outer product operator. Note that the matrix and tensor data model in (3.2) and in (3.6) respectively are connected through the relations  $\mathbf{A}^T = [\mathcal{A}]_{(R+1)}$  and  $\mathbf{X}^T = [\mathcal{X}]_{(R+1)}$ .

The array steering tensor in (3.6) can be written in terms of the PARAFAC representation

$$\mathcal{A} = \mathcal{I}_{R+1,d} \times_1 \mathbf{A}^{(1)} \dots \times_R \mathbf{A}^{(R)}, \quad (3.7)$$

where  $\mathcal{I}_{R+1,d} \in \mathbb{R}^{d \times d \times \dots \times d}$  is the identity tensor, which is equal to one if all indices are equal and zero otherwise. Consequently, in the absence of noise, (3.6) may be rewritten as

$$\mathcal{X}_0 = \mathcal{I}_{R+1,d} \times_1 \mathbf{A}^{(1)} \dots \times_R \mathbf{A}^{(R)} \times_{R+1} \mathbf{S}^T, \quad (3.8)$$

where  $\mathbf{s}_i^T$  is the  $i$ -th row of  $\mathbf{S}$ . Note that in the absence of noise, the tensor  $\mathcal{X}$  has rank  $d$ .<sup>3</sup> Therefore, all the  $n$ -ranks of  $\mathcal{X}$  are at most  $d$ . We assume the model order  $d$  to be known, although in practice  $d$  should be estimated for noisy data. For particularly efficient model order estimation schemes for the PARAFAC data model, the reader is referred to Chapter 2.

Using another fashion to express (3.6), the following representation of the data model in the presence of noise is possible

$$\mathcal{X} = \mathcal{I}_{R+1,d} \times_{r=1}^R \mathbf{A}^{(r)} \times_{R+1} \mathbf{S}^T + \mathcal{N}. \quad (3.9)$$

Therefore, the parameter estimation problem can be stated as: given the measurement tensor  $\mathcal{X}$  and the model order  $d$ , we desire to estimate all the spatial frequencies  $\mu_i^{(r)}$ , which in total are  $R \cdot d$  variables.  $\mathbf{S}$  is also unknown, and once all  $\mu_i^{(r)}$  are estimated, this information can be used to estimate  $\widehat{\mathbf{S}}$ . Note the operator  $\times_{r=1}^R$  as defined in (2.68) denotes a compact representation of  $R$   $r$ -mode products between a tensor and  $R$  matrices.

---

<sup>3</sup> The rank cannot be larger than  $d$ , but it might be smaller. This occurs only in degenerate cases which are not relevant for our discussion, e.g., two coherent sources at exactly the same position.

---

The data model considered in this chapter is not so general as in Subsection 2.2.1 due to the assumption of the complex exponentials, but also is not so specific as in Subsubsection 2.2.3.6 due to the possibility of an OPA or a mixed array.

### 3.2 Previous Approaches: MALS based and $R$ -D ESPRIT-type Approaches

In Subsection 2.2.2, we have already introduced the TALS in order to estimate the model order via the CORCONDIA [15]. For the sake of simplicity, consider the ALS (Alternating Least Squares) solution [60] for the special case  $R = 3$ . Then we have the following iterative equations for the PARAFAC decomposition of a tensor:

$$\begin{aligned}\widehat{\mathbf{S}}_{t+1}^T &= [\mathbf{x}]_{(3)} \cdot \left[ \left( \widehat{\mathbf{A}}_t^{(1)} \diamond \widehat{\mathbf{A}}_t^{(2)} \right)^+ \right]^T \\ \widehat{\mathbf{A}}_{t+1}^{(2)} &= [\mathbf{x}]_{(2)} \cdot \left[ \left( \widehat{\mathbf{S}}_t^T \diamond \widehat{\mathbf{A}}_t^{(1)} \right)^+ \right]^T \\ \widehat{\mathbf{A}}_{t+1}^{(1)} &= [\mathbf{x}]_{(1)} \cdot \left[ \left( \widehat{\mathbf{A}}_t^{(2)} \diamond \widehat{\mathbf{S}}_t^T \right)^+ \right]^T,\end{aligned}\tag{3.10}$$

where  $t$  denotes the  $t$ -th iteration. More information about the estimation of the factors such as stop criteria and initialization is provided in (2.50) in Subsection 2.2.2. After the estimation of the factors using (3.10), one can apply the schemes in Subsections 3.3.4 and 3.3.3 to obtain the spatial frequencies.

A well-known closed-form solution used in the literature to obtain the spatial frequencies is the  $R$ -D standard Tensor-ESPRIT ( $R$ -D STE) algorithm proposed in [83, 44]. We show below a summary to apply it for our problem just as in [83, 44] and for more details the reader is referred to [83, 44].

- 0) Arrange the measurement samples into a tensor  $\mathbf{x} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times N}$
- 1) Compute an HOSVD-based low-rank approximation of  $\mathbf{x}$  (see the Appendix E for more information)

$$\mathbf{x} \approx \mathbf{s}^{[s]} \times_{r=1}^{R+1} \mathbf{U}_r^{[s]}\tag{3.11}$$

From (3.11), the signal subspace tensor  $\mathbf{u}^{[s]} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$

$$\mathbf{u}^{[s]} = \mathbf{s}^{[s]} \times_{r=1}^R \mathbf{U}_r^{[s]}.\tag{3.12}$$

Here we compute the low-rank approximation by truncating the HOSVD of  $\mathbf{x}$ .

- 2) The shift invariance equations given by

$$\mathbf{u}^{[s]} \times_r \mathbf{J}_1^{(r)} \times_{R+1} \mathbf{\Psi}^{(r)} \approx \mathbf{u}^{[s]} \times_r \mathbf{J}_2^{(r)}\tag{3.13}$$

can be solved by least square methods, i.e., Least Squares (LS), Structured Least Squares (SLS), Tensor-Structured Structured Least Squares (TS-SLS) or their  $R$ -D extensions also shown in [83, 44].

- 3) Computation of the eigenvalues of the matrices  $\Psi^{(r)}$  jointly for  $r = 1, 2, \dots, R$ . This can be accomplished by joint diagonalization algorithms [35].
- 4) The spatial frequencies are related to the eigenvalues  $\lambda_i^{(r)}$  of the matrices  $\Psi^{(r)}$  through

$$\mu_i^{(r)} = \arg\left(\lambda_i^{(r)}\right), \quad i = 1, 2, \dots, d. \quad (3.14)$$

Another approach used by us in the simulations is the  $R$ -D Unitary Tensor-ESPRIT ( $R$ -D UTE) proposed in [83, 44]. We show below a summary to apply it for our problem just as in [83, 44] and for more details the reader is referred to [83, 44].

- 0a) Arrange the measurement samples into a tensor  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times N}$
- 0b) Apply FBA and also the real-valued transformation of the measurement tensor  $\varphi(\mathcal{Z}) \in \mathbb{R}^{M_1 \times M_2 \times \dots \times M_R \times 2 \cdot N}$  as shown in Subsubsection 2.2.3.6.
- 1) Compute an HOSVD-based low-rank approximation of  $\varphi(\mathcal{Z})$  (see the Appendix E for more information)

$$\varphi(\mathcal{Z}) \approx \mathcal{S}_\varphi^{[s]} \times_{r=1}^{R+1} \mathbf{E}_r^{[s]} \quad (3.15)$$

From (3.15), the signal subspace tensor  $\mathcal{E}^{[s]} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$

$$\mathcal{E}^{[s]} = \mathcal{S}_\varphi^{[s]} \times_{r=1}^R \mathbf{E}_r^{[s]}. \quad (3.16)$$

Here we compute the low-rank approximation by truncating the HOSVD of  $\mathcal{X}$ .

- 2) The shift invariance equations given by

$$\mathcal{E}^{[s]} \times_r \mathbf{K}_1^{(r)} \times_{R+1} \Upsilon^{(r)} \approx \mathcal{E}^{[s]} \times_r \mathbf{K}_2^{(r)} \quad (3.17)$$

can be solved by least square methods, i.e., LS, SLS, TS-SLS, or their  $R$ -D extensions also shown in [83, 44]. The matrices  $\mathbf{K}_1^{(r)}$  and  $\mathbf{K}_2^{(r)}$  are defined in the following way

$$\begin{aligned} \mathbf{K}_1^{(r)} &= 2 \cdot \text{Re} \left\{ \mathbf{Q}_{M_r^{(\text{sel})}}^H \cdot J_2^{(r)} \cdot \mathbf{Q}_{M_r}^H \right\}, \\ \mathbf{K}_2^{(r)} &= 2 \cdot \text{Im} \left\{ \mathbf{Q}_{M_r^{(\text{sel})}}^H \cdot J_2^{(r)} \cdot \mathbf{Q}_{M_r}^H \right\}. \end{aligned} \quad (3.18)$$

- 3) Computation of the eigenvalues of the matrices  $\Upsilon^{(r)}$  jointly for  $r = 1, 2, \dots, R$ . This can be accomplished by joint diagonalization algorithms [35].
-

4) The spatial frequencies are related to the eigenvalues  $\omega_i^{(r)}$  of the matrices  $\Upsilon^{(r)}$  through

$$\mu_i^{(r)} = 2 \cdot \arg\left(\omega_i^{(r)}\right), \quad i = 1, 2, \dots, d. \quad (3.19)$$

Note that the difference between the *R*-D Standard ESPRIT (*R*-D SE) and the *R*-D Standard Tensor-ESPRIT (*R*-D STE) is that in the former the low-rank HOSVD-based approximation is not applied, instead a low-rank SVD-based approximation is applied in the  $R + 1$ -mode unfolding. Analogously, the same difference exists between the *R*-D Unitary ESPRIT (*R*-D UE) to the *R*-D Unitary Tensor-ESPRIT (*R*-D UTE).

### 3.3 Proposed Approach: Closed-Form PARAFAC based Approach

For shift invariant arrays that obey (3.9),  $\mu_i^{(r)}$  can be estimated directly from  $\mathcal{X}$  via *R*-D Tensor-ESPRIT-type methods [44]. Alternatively, in this section, we can compute the PARAFAC decomposition of  $\mathcal{X}$ , which provides estimates for the array steering matrices  $\mathbf{A}^{(r)}$  in all the modes. It is more flexible with respect to the array geometry, since the shift invariance structure is not required. Therefore, the estimation of the spatial frequencies is performed by applying a peak search in Subsection 3.3.3. However, if in some dimensions a shift invariance structure is present, the spatial frequencies can be estimated via the shift invariance scheme in Subsection 3.3.4. As we demonstrate by means of simulations, this approach is more robust to positioning errors in the array, especially if the closed-form PARAFAC (CFP) algorithm is used to estimate the factors. Moreover, in Subsection 3.3.1, we present one important characteristic of the PARAFAC decomposition, which is to merge the dimensions. Due to the merging, we can increase considerably the maximum model order. In Subsection 3.3.2, we show how to separate the merged dimensions.

In the Section 3.4, we present a review of the closed-form PARAFAC (CFP) [84], and we discuss different implementations. Since the CFP produces multiple estimates, we compare different heuristics to choose the best estimate of the factors in Section 3.4.

#### 3.3.1 Merging Dimensions

The merging property is of great importance for the CFP in order to increase the maximum model order. As an example, let us consider a five-dimensional tensor represented by  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times M_3 \times M_4 \times N}$ . By using the representation of (3.8) in the absence of noise, we have

$$\mathcal{X} = \mathcal{I}_{5,d} \times_1 \mathbf{F}^{(1)} \times_2 \mathbf{F}^{(2)} \times_3 \mathbf{F}^{(3)} \times_4 \mathbf{F}^{(4)} \times_5 \mathbf{F}^{(5)}. \quad (3.20)$$

However, by merging dimensions, we can reduce (3.20) to a three-dimensional tensor  $\mathcal{X}^{(\text{mg})} \in \mathbb{C}^{M_1 \cdot M_2 \times M_3 \cdot M_4 \times N}$ , e.g.,

$$\mathcal{X}^{(\text{mg})} = \mathcal{I}_{3,d} \times_1 \mathbf{F}^{(1,2)} \times_2 \mathbf{F}^{(3,4)} \times_3 \mathbf{F}^{(5)}, \quad (3.21)$$

where

$$\mathbf{F}^{(p,q)} = \mathbf{F}^{(p)} \diamond \mathbf{F}^{(q)}. \quad (3.22)$$

If  $N$  is sufficiently large, i.e.,  $N \geq d_{\max}$ , where  $d_{\max}$  is the maximum model order for the CFP decomposition, then without merging  $d_{\max}$  is given by

$$d_{\max} = \max_{r=1,\dots,R} M_r, \quad (3.23)$$

while with merging

$$d_{\max} = \max_{p=1,2,\dots,R|p \neq q} \max_{q=1,2,\dots,R} M_p \cdot M_q. \quad (3.24)$$

Note that it is also possible to merge more than two dimensions.

Applying the CFP in (3.21) in the presence of noise, we obtain the following decomposition

$$\mathcal{X}^{(\text{mg})} \approx \mathcal{I}_{3,d} \times_1 \widehat{\mathbf{F}}^{(1,2)} \times_2 \widehat{\mathbf{F}}^{(3,4)} \times_3 \widehat{\mathbf{F}}^{(5)}. \quad (3.25)$$

Since we have merged some dimensions, we present the Least Squares Khatri-Rao Factorization (LSKRF) [86] as a more general and algebraic way to separate the merged dimensions in Subsection 3.3.2.

### 3.3.2 Least Squares Khatri-Rao Factorization (LSKRF)

In this subsection, we show how an estimate of the merged factor matrices from Subsection 3.3.1 can be refactorized into individual estimates for the separated dimensions.

First let us write our problem as how to estimate  $\widehat{\mathbf{f}}_i^{(p)}$  and  $\widehat{\mathbf{f}}_i^{(q)}$  such that

$$\widehat{\mathbf{f}}_i^{(p)} \diamond \widehat{\mathbf{f}}_i^{(q)} \approx \widehat{\mathbf{f}}_i^{(p,q)}, \quad (3.26)$$

where  $\widehat{\mathbf{f}}_i^{(p,q)} \in \mathbb{C}^{M_p \cdot M_q \times 1}$  is the  $i$ -th column of  $\widehat{\mathbf{F}}^{(p,q)}$  shown in (3.25), and where  $\widehat{\mathbf{f}}_i^{(p)} \in \mathbb{C}^{M_p \times 1}$  and  $\widehat{\mathbf{f}}_i^{(q)} \in \mathbb{C}^{M_q \times 1}$ .

To solve this problem, we reshape  $\widehat{\mathbf{f}}_i^{(p,q)}$  by using the unvec operator such that it becomes of size  $M_q \times M_p$ , and transform the problem into

$$\text{unvec}_{M_q \times M_p}(\widehat{\mathbf{f}}_i^{(p,q)}) \approx \widehat{\mathbf{f}}_i^{(q)} \cdot \left(\widehat{\mathbf{f}}_i^{(p)}\right)^T. \quad (3.27)$$

Since (3.27) should be a rank-one matrix, we can use a SVD-based rank-one approximation

$$\text{unvec}_{M_q \times M_p}(\widehat{\mathbf{f}}_i^{(p,q)}) = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^H, \quad (3.28)$$

The weighted right dominant singular vector  $\sqrt{\sigma_1} \cdot \mathbf{u}_1$  and the conjugate of the weighted left

dominant singular vector  $\sqrt{\sigma_1} \cdot \mathbf{v}_1^*$  are the estimates of  $\widehat{\mathbf{f}}_i^{(q)}$  and  $\widehat{\mathbf{f}}_i^{(p)}$ , respectively. Here,  $\sigma_1$  is the dominant singular value. Although there is one phase ambiguity inherent to the SVD, this does not influence the estimate of the spatial frequencies.

### 3.3.3 Estimation of Spatial Frequencies via Peak Search (PS)

Another way to estimate the spatial frequencies is performing a one-dimensional peak search (PS) by using the following function

$$\widehat{\mu}_i^{(r)} = \arg \max_{\mu_i^{(r)}} \left| \mathbf{a}^{(r)}(\mu_i^{(r)})^H \cdot \widehat{\mathbf{f}}_i^{(r)} \right|, \quad (3.29)$$

where  $\mathbf{a}^{(r)}(\mu_i^{(r)})$  denotes the array manifold of the  $r$ -th dimension. In case of one dimension being a two-dimensional arbitrary array, it is possible to modify (3.29) in order to estimate  $d$  separated two-dimensional peaks via a 2-D peak search, like in 2-D spectral MUSIC.

### 3.3.4 Estimation of Spatial Frequencies via Shift Invariance (SI)

If the array is shift invariant in the  $r$ -th mode, the estimation of the spatial frequency can be performed by exploiting this property as follows

$$\widehat{\mu}_i^{(r)} = \arg \left[ \left( \mathbf{J}_1^{(r)} \cdot \widehat{\mathbf{f}}_i^{(r)} \right)^H \cdot \mathbf{J}_2^{(r)} \cdot \widehat{\mathbf{f}}_i^{(r)} \right] \quad (3.30)$$

where the operator  $\arg(\cdot)$  returns the phase, and  $\mathbf{J}_1^{(r)} \in \mathbb{R}^{M_r^{(\text{sel})} \times M_r}$  and  $\mathbf{J}_2^{(r)} \in \mathbb{R}^{M_r^{(\text{sel})} \times M_r}$  are the selection matrices from the ESPRIT-type algorithms. The selection matrix  $\mathbf{J}_1^{(r)}$  selects the  $M_r^{(\text{sel})}$  elements of the first subarray and the selection matrix  $\mathbf{J}_2^{(r)}$  selects the  $M_r^{(\text{sel})}$  elements of the second subarray.  $M_r^{(\text{sel})}$  depends on the geometry of the array, and as an example for the ULA,  $M_r^{(\text{sel})} = M_r - 1$  corresponds to maximum overlap.

By exploiting the 2-D shift invariance as done by the 2-D ESPRIT-type algorithms, it is possible to estimate  $e^{j\widehat{\mu}_i^{(p)}}$  and  $e^{j\widehat{\mu}_i^{(q)}}$  from  $\widehat{\mathbf{f}}_i^{(p,q)}$  [44]. For instance, we can obtain the spatial frequencies by computing

$$\widehat{\mu}_i^{(p)} = \arg \left\{ \left[ \left( \mathbf{J}_1^{(p)} \otimes \mathbf{I}_{M_q} \right) \cdot \widehat{\mathbf{f}}_i^{(p,q)} \right]^H \cdot \left( \mathbf{J}_2^{(p)} \otimes \mathbf{I}_{M_q} \right) \cdot \widehat{\mathbf{f}}_i^{(p,q)} \right\}, \quad (3.31)$$

and

$$\widehat{\mu}_i^{(q)} = \arg \left\{ \left[ \left( \mathbf{I}_{M_p} \otimes \mathbf{J}_1^{(q)} \right) \cdot \widehat{\mathbf{f}}_i^{(p,q)} \right]^H \cdot \left( \mathbf{I}_{M_p} \otimes \mathbf{J}_2^{(q)} \right) \cdot \widehat{\mathbf{f}}_i^{(p,q)} \right\}, \quad (3.32)$$

where  $\otimes$  denotes the Kronecker product. Note that by using (3.31) and (3.32) the LSKRF is not needed.



### 3.4 Evaluation of Approaches

In this section we present simulation results demonstrating the performance of the proposed methods. The spatial frequencies  $\mu_i^{(r)}$  are drawn from a uniform distribution in  $[-\pi, \pi]$ . The source symbols are zero mean i.i.d. circularly symmetric complex Gaussian distributed with power equal to  $\sigma_s^2$  for all the sources. We define at the receiver

$$\text{SNR} = 10 \cdot \log_{10} \left( \frac{\sigma_s^2}{\sigma_n^2} \right), \quad (3.33)$$

where  $\sigma_n^2$  is the variance of the elements of the white noise tensor  $\mathcal{N}$ .

In order to compare the performance of the proposed technique to the ESPRIT based approaches, we compute the total RMSE of the estimated spatial frequencies  $\hat{\mu}_i$  as follows

$$\text{RMSE} = \sqrt{\text{E} \left\{ \sum_{r=1}^R \sum_{i=1}^d \left( \hat{\mu}_i^{(r)} - \mu_i^{(r)} \right)^2 \right\}}. \quad (3.34)$$

Since the CFP returns multiple solutions, one of the first problems is how to select the best solution. In [84], it was proposed to pick the combination of estimated factors which corresponds to the lowest reconstruction error, if compared to the noisy tensor. However, such approach does not necessarily minimize the estimation error for the factors. As we observe in the simulations, picking the factors according to the lowest residuals criterion may give more stable estimation results.

As a benchmark, we consider a genie-aided estimation scheme, where among the many estimates for one factor, the particular one with lowest estimation is selected. In Fig. 3.2, both lowest reconstruction error and lowest residuals present a performance very close to the genie case. Note also that since the simultaneous diagonalizations residuals error is an inherent output of the CFP, it is preferable to use it instead of computing the lowest reconstruction error.

In order to combine factors estimated in different diagonalizations processes, the permutation and scaling ambiguities should be solved. For this task, we apply the amplitude approach according to [100]. In Fig. 3.3, we compare the performance of CFP with three different possible integrations. The suffix LSKR SI stands for Least Squares Khatri-Rao Factorization (LSKRF) to separate merged dimensions as in (3.28) and spatial frequencies estimated by shift invariance (SI) according to (3.30). In case of LSKR PS, the LSKRF is also applied as in (3.28), but the spatial frequencies are estimated by using peak search (PS) according to (3.29). The suffix SI alone means that the dimensions are separated by using the shift invariance according to (3.30).

From Fig. 3.3, it is clear that for ULAs the estimation using the LSKRF with SI, the LSKRF with PS, or only the SI gives the same performance. In such a problem the LSKRF SI and SI are better, since they have a much lower computational complexity. Note that the increase of resolution done with PS does not give any gain in the estimation.

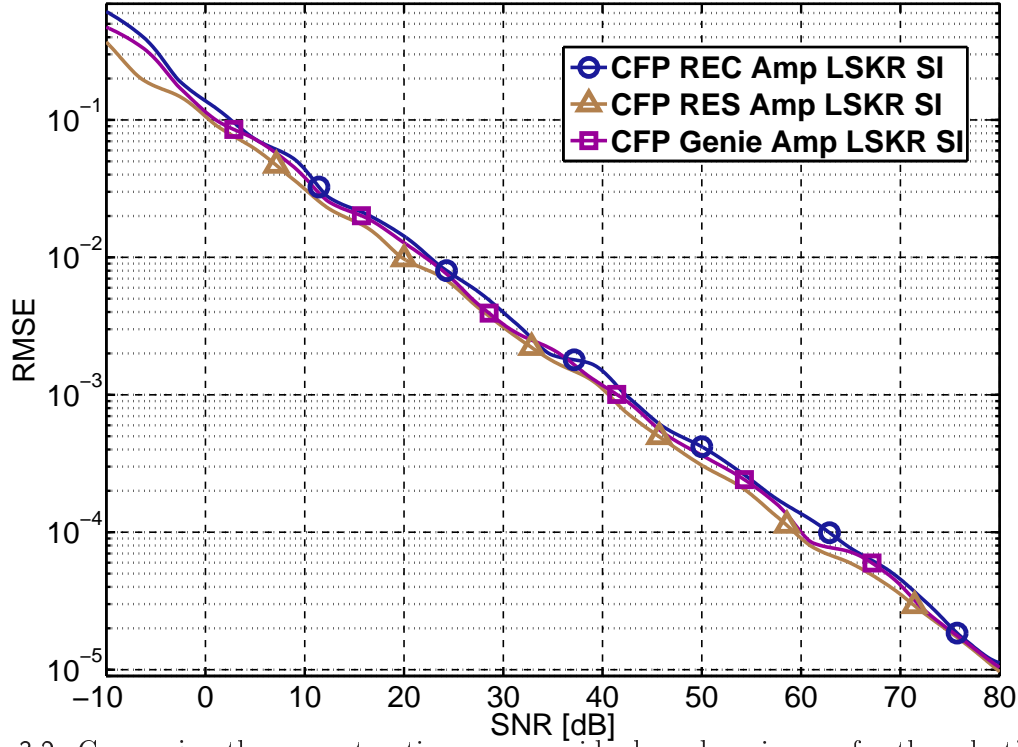


Fig. 3.2: Comparing the reconstruction error, residuals and genie case for the selection of the multiple solution generated by the Closed-Form PARAFAC. The array size, snapshots  $N$ , and  $d$  are the same as in Fig. 3.4

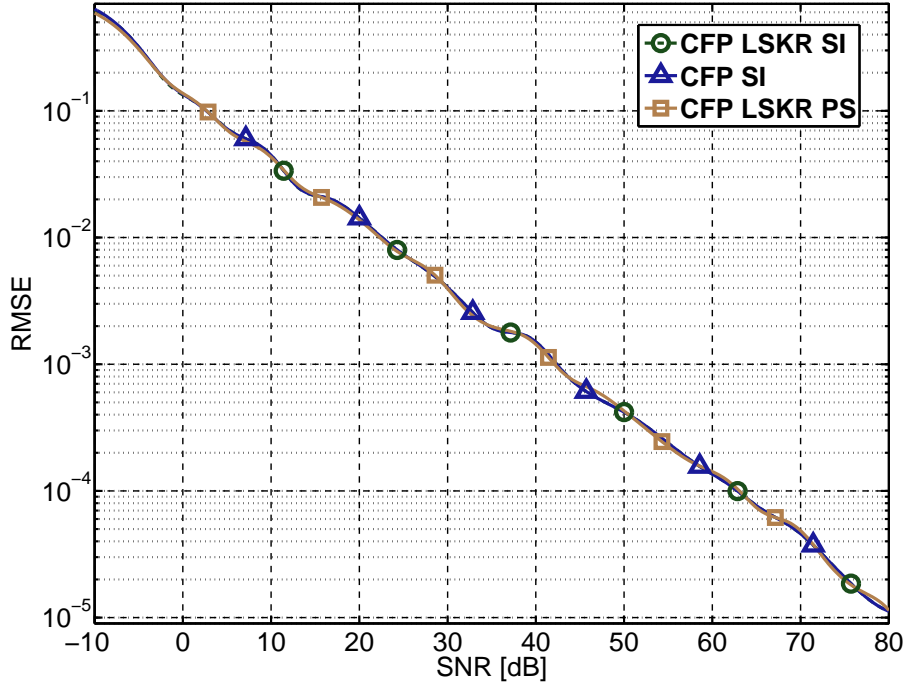


Fig. 3.3: RMSE versus SNR of the different integrations for the Closed-Form PARAFAC (CFP). In the scenario, the arrays are URA with perfectly known positions. The array size is  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$ .  $d$  is fixed to 4.

In Fig. 3.4,  $R$ -D SE,  $R$ -D UE,  $R$ -D UTE,  $R$ -D STE stands for the  $R$ -dimensional Standard ESPRIT,  $R$ -dimensional Unitary ESPRIT,  $R$ -dimensional Unitary Tensor-ESPRIT, and  $R$ -dimensional Standard Tensor-ESPRIT, respectively. Here we use the  $R$ -D ESPRIT-type algorithm with the LS solution [83, 44]. Moreover, for the  $R$ -D SE no FBA is applied, while for the other  $R$ -D ESPRIT-type algorithms FBA is applied according to Subsubsection 2.2.3.6.

We consider that we have a Uniform Rectangular Array (URA), i.e, the outer product of two Uniform Linear Array (ULAs) vectors, at the receiver (RX) represented by  $M_1$  and  $M_2$ , and another URA at the transmitter (TX) represented by  $M_3$  and  $M_4$ .  $M_5$  is related to the frequency bins for the estimation of the time delay of arrival (TDOA).

For all simulations in this chapter, we merge  $M_1$  and  $M_2$ , and also  $M_3$  and  $M_4$  for the CFP approach. By merging we have no degenerate case in the three new dimensions, while without merging four dimensions are degenerate. CFP computes multiple estimates for each factor. In particular, for  $T$  non-degenerate modes,  $T \cdot (T + 1)$  estimates for each factor are available. Consequently, after merging, we increase the number of estimates from 2 to 6, and the maximum possible model order from 7 to 9.

In Fig. 3.5 and 3.6, we consider the case that there is a URA with positioning errors at the RX. Therefore, for severe positioning errors, all the distances between all the consecutive antennas are different, which implies that the assumption of an outer product structure between the first and the second dimensions, which are merged in our CFP approach, is not valid anymore. These positioning errors are modeled by zero mean real-valued Gaussian random variables with variance  $\sigma_e^2$ . Note that the URA with positioning errors is only included at the RX side, whereas at the TX and frequency dimensions the arrays have a Vandermonde structure without any error. The error of the spatial frequencies at the third, fourth and fifth dimensions is depicted in Fig. 3.5. The estimation performed by the  $R$ -D ESPRIT-type based techniques is largely influenced by the errors at the first and second dimensions. On the other hand, since the CFP based approach is based on a very general decomposition, there is no degradation. For the parameters in the first and second dimensions, the estimation is equally degraded for all the schemes. Therefore, all the schemes have the same performance as shown in 3.6.

In Figs. 3.7 and 3.8, at the RX array, we have a Shift Invariant (SI) outer product based array (OPA), which varies until becoming a Non-Shift Invariant (NSI) outer product based array (OPA) according to a zero mean real Gaussian random variable  $n_{p,i}^{(r)}$  with variance  $\sigma_p^2$ . For instance, in Fig. 3.1,  $\Delta_i^{(r)} = \overline{\Delta}_i^{(r)} + n_{p,i}^{(r)}$ , where  $\overline{\Delta}_i^{(r)}$  is a constant. In contrast to Fig. 3.5, here the positions are considered exactly known, i.e., the deviation of the outer product structure is known. In Fig. 3.7 related to the RX, the CFP combined with the PS approach has the same performance as CFP combined with the SI approach, since in these dimensions we have the shift invariance property. On the other hand, in Fig. 3.8 related to the TX side and frequency bins, the CFP combined with the SI has the same error as the ESPRIT based approaches, and only the CFP combined with the PS has no degradation. The Multilinear Alternating Least Squares (MALs), which is the Alternating Least Squares for a

$M$  way tensor, presents the same performance than CFP when used in conjunction with the SI, LSKRF, and PS for the previous simulations.

In (2.7), we have introduced the model for correlated sources and in (2.37), we compute the signal correlation matrix as a function of  $\rho_s$ . We replace the  $\mathbf{S}$  in (3.6) by  $\mathbf{S}^{(c)}$  via (2.7) and (2.37). In Fig. 3.9,  $\sigma_s$  varies from 0 to 0.999, i.e., the correlation between the signals is increased. Although all the ESPRIT-type techniques, except the *R*-D SE, use FBA according to Subsubsection 2.2.3.6, they are outperformed by the CFP-MOS based and MALS based schemes, which do not use FBA. We compare the MALS to the CFP in a channel generated by the geometry-based channel modeling tool IlmProp [32]. The IlmProp generated scenario is shown in Fig. 3.10, and it is a NLOS (Non-line of Sight) scenario composed of 4 paths, i.e.,  $d = 4$ . At the RX and also at the TX, there is a URA  $3 \times 3$ , and 10 frequency bins are used. The carrier frequency  $f_0$  is  $2 \cdot 10^9$  Hz. The bandwidth is 80 MHz. 5 time snapshots are collected, and the time sampling is 1 ms. The space between the antennas is  $\frac{\lambda}{2}$ , where  $\lambda = \frac{c}{f_0}$ , and  $c$  is the speed of the light in the vacuum.

In Appendix J, we present more details about comparisons between the extracted spatial frequencies of IlmProp and the estimated frequencies between *R*-D Unitary Tensor-ESPRIT (UTE). Moreover, we provide a brief analysis of the errors caused in the estimation due to the Kronecker channel and narrowband signals assumption, when different parameters of the scenario vary.

In Fig. 3.11, the RMSE of the estimated spatial frequencies by applying MALS and CFP are compared. The MALS is more instable than CFP. Moreover, the MALS estimates have a limited accuracy for the RMSE oscillating around  $7 \cdot 10^{-1}$ , while the CFP reaches a limited accuracy for the RMSE lower than  $4 \cdot 10^{-3}$ . Note that the *R*-D STE has also a limited accuracy of  $3 \cdot 10^{-2}$  worse than the CFP.

The limited accuracy presented in all approaches in Fig. 3.11 is a bias, which is due to the violation of the narrowband signal assumption in the parameter estimation as shown in Appendix J.

### 3.5 Main Results of the Chapter

In this chapter, we have presented our proposed Closed-Form PARAFAC (CFP) based parameter estimation technique in combination with other techniques, such as the Least Squares Khatri-Rao Factorization (LSKRF), a Peak Search (PS), and the Shift Invariance (SI) [6].

For scenarios where the position of the antennas in a URA are exactly known, the performance of the CFP based approaches is similar to the ESPRIT based approaches. However, if positioning errors are assumed in some dimensions, the estimation with ESPRIT based techniques is degraded in all dimensions, while the CFP approaches have a degradation only in the dimensions with errors. Therefore, the CFP based approaches are more robust than the ESPRIT based techniques.

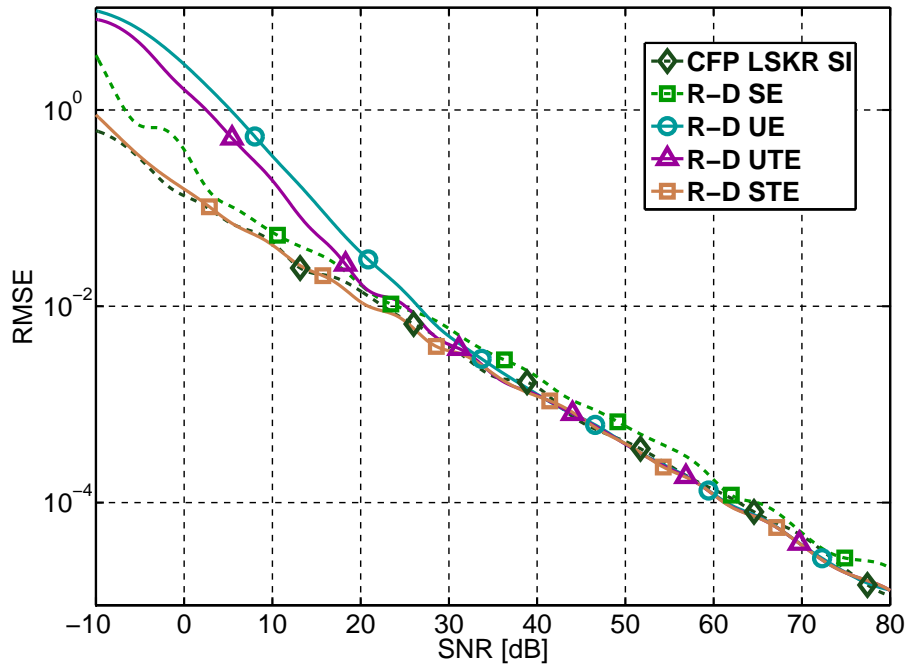


Fig. 3.4: RMSE versus SNR to compare the Closed-Form PARAFAC to the ESPRIT based techniques. In the scenario, URAs of sizes  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$  are applied.  $d$  is fixed to 4. The RMSE is averaged for the parameters in the five dimensions according to (3.34).

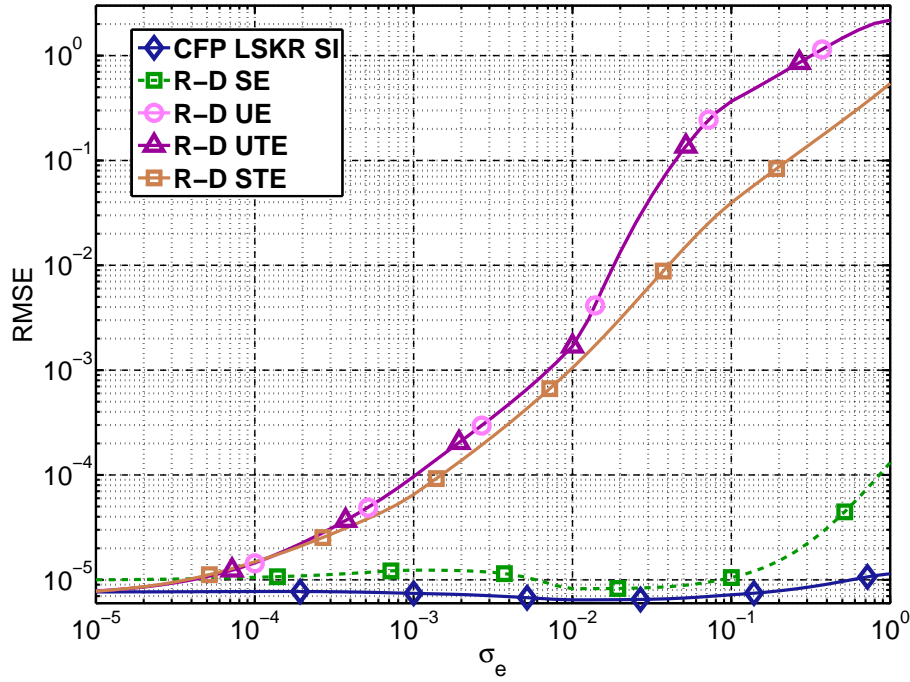


Fig. 3.5: RMSE versus  $\sigma_e$  to compare the Closed-Form PARAFAC to the ESPRIT based techniques in presence of positioning errors in some dimensions. For the RMSE only the parameters at the third, fourth and fifth dimensions without positioning errors are considered. In the scenario, URAs of sizes  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$  are applied.  $d$  is fixed to 4. The SNR is fixed to 80 dB.

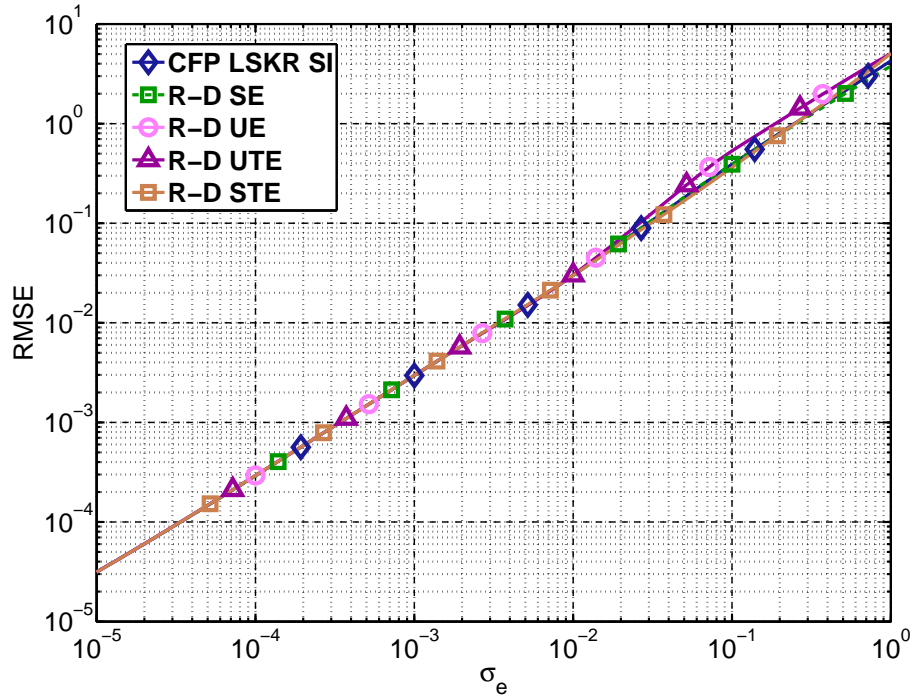


Fig. 3.6: RMSE versus  $\sigma_e$  to compare the Closed-Form PARAFAC to the ESPRIT based techniques in presence of positioning errors in some dimensions. For the RMSE only the parameters at the first and second dimensions with positioning errors are considered. In the scenario, URAs of sizes  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$  are applied.  $d$  is fixed to 4. The SNR is fixed to 80 dB.

Besides the robustness, the CFP is also applicable to a wide variety of arrays, e.g., the outer product based arrays (OPAs), and the mixed arrays composed of arbitrary arrays (AAs) combined to OPAs. We also propose to merge the dimensions in order to achieve a higher model order possible for the CFP approaches, such that all the rank-one tensors can be estimated. As an example, we have considered the case of merging two dimensions in Subsection 3.3.1.

Compared to the Multilinear Alternating Least Squares (MALS), which is an iterative PARAFAC decomposition, the CFP is also more robust for channel generated by the geometry-based channel modeling tool IlmProp.

Among the different techniques to estimate the spatial frequencies from the factor matrices, the combination of CFP and SI is only preferable for URAs without or with positioning errors, since it is computationally faster. In case that arrays are NSI (Non-Shift Invariant) OPAs, or mixed arrays, the CFP integrated to the Peak Search (PS) is the best approach, since the information about the different sensor positions can be included in the Peak Search algorithm.

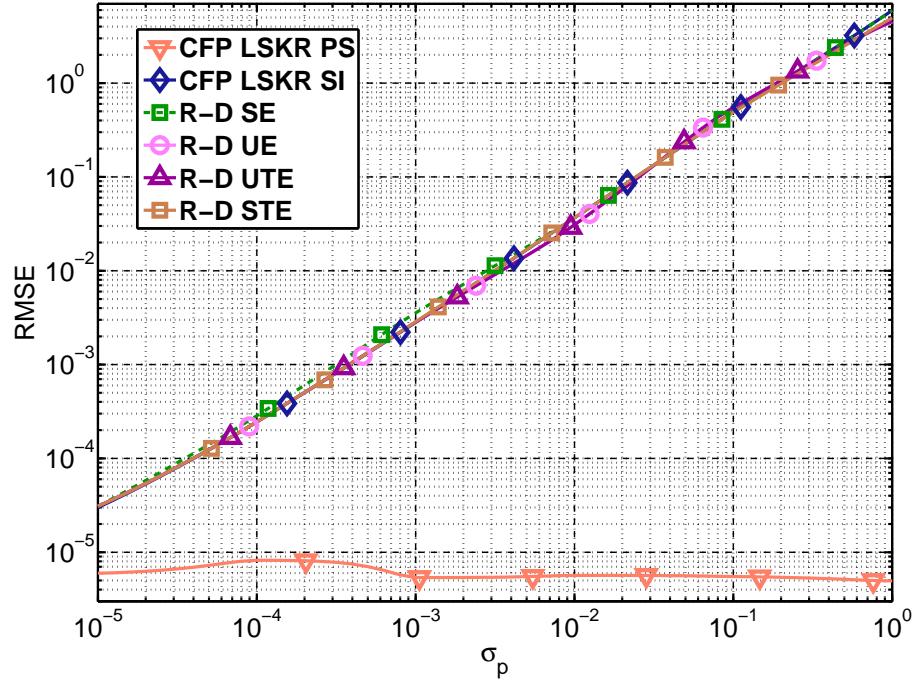


Fig. 3.7: RMSE versus  $\sigma_p$  to compare the Closed-Form PARAFAC to the ESPRIT based techniques in case that shift invariance property varies, but the outer product property is kept in the first and second dimensions. For the RMSE only the parameters at the first and second dimensions are considered. In the scenario, URAs of sizes  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$  are applied.  $d$  is fixed to 4. The SNR is fixed to 80 dB.

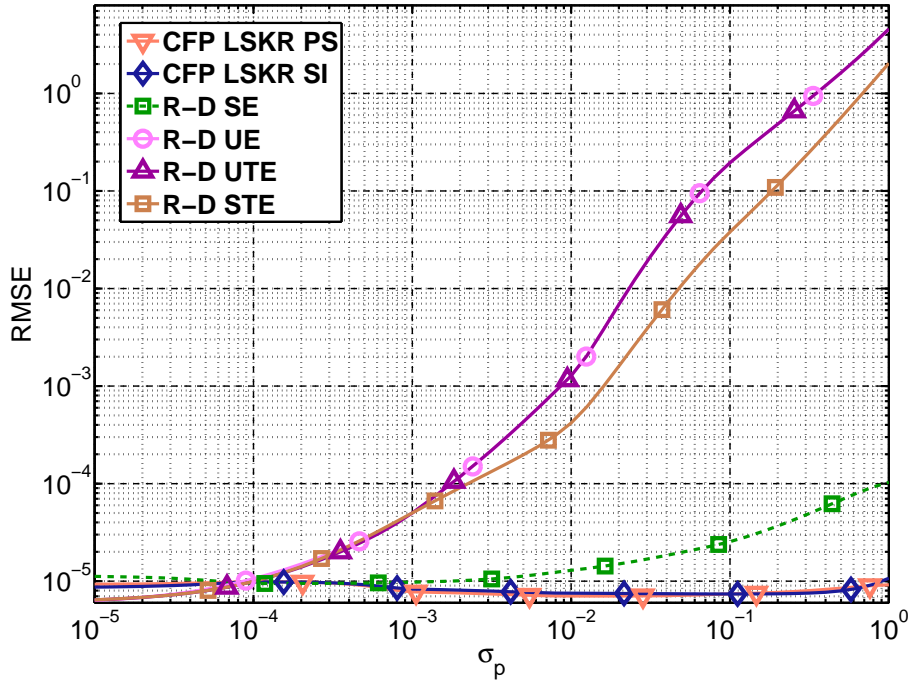


Fig. 3.8: RMSE versus  $\sigma_p$  to compare the Closed-Form PARAFAC to the ESPRIT based techniques in case that shift invariance property varies, but the outer product property is kept in the first and second dimensions. For the RMSE only the parameters at the third, fourth and fifth dimensions are considered. In the scenario, URAs of sizes  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$  are applied.  $d$  is fixed to 4. The SNR is fixed to 80 dB.

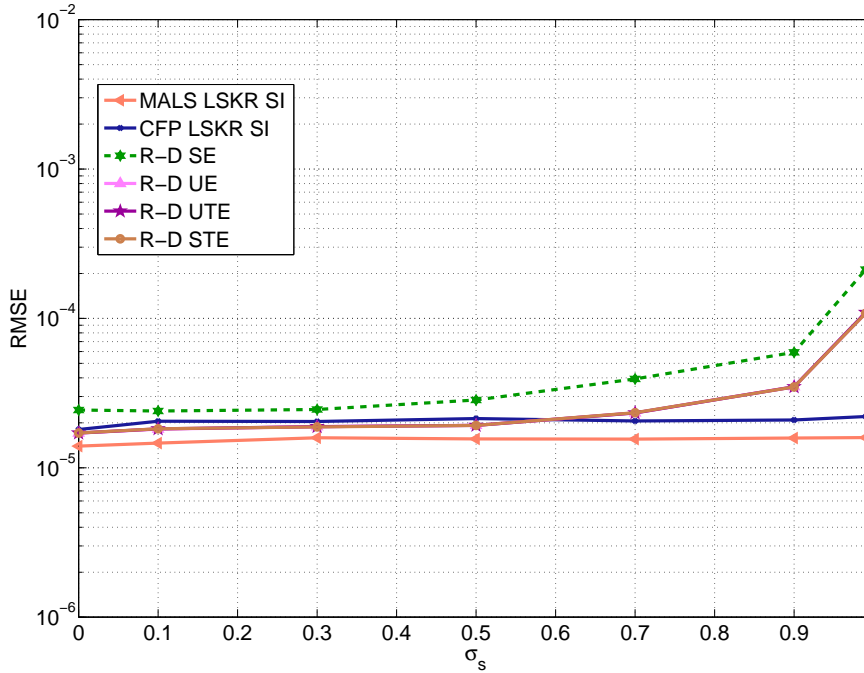


Fig. 3.9: RMSE versus  $\sigma_s$  to compare the Closed-Form PARAFAC to the MALS based and ESPRIT-type techniques. In the scenario, URAs of sizes  $M_1 = 3$ ,  $M_2 = 3$ ,  $M_3 = 3$ ,  $M_4 = 3$ ,  $M_5 = 7$ , and  $N = 5$  are applied.  $d$  is fixed to 4. The SNR is fixed to 80 dB.



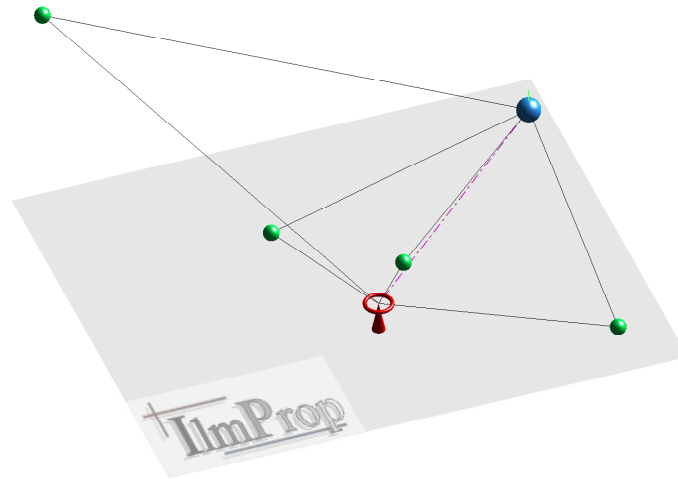


Fig. 3.10: The IlmProp [32] NLOS (non-line of sight) scenario composed of 4 paths. Both at the RX and at the TX URAs of size  $3 \times 3$  are placed. 10 frequency bins are used with a total bandwidth is 80 MHz. The carrier frequency is  $2 \cdot 10^9$  Hz, and 5 snapshots with 1 ms time sampling are collected.

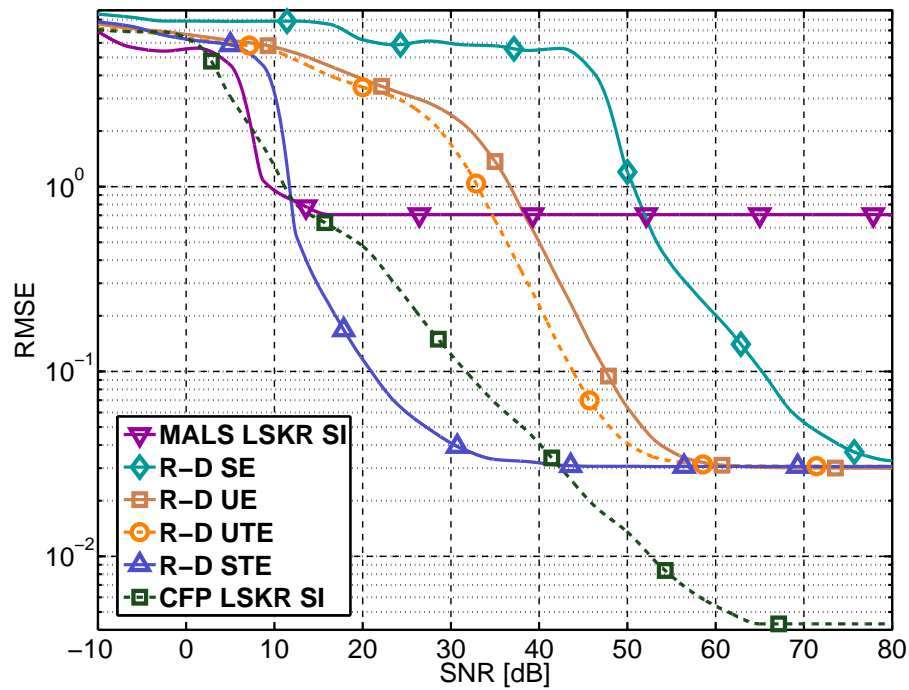


Fig. 3.11: RMSE versus SNR to compare the Closed-Form PARAFAC to the MALS based techniques. The IlmProp scenario is described in Fig. 3.10.



## 4. MULTI-DIMENSIONAL AND MATRIX-BASED SUBSPACE PREWHITENING

In Chapter 2, we have reviewed and proposed one-dimensional and multi-dimensional model order selection schemes for white and colored Gaussian noise, while in Chapter 3, we have assumed the model order as already being correctly estimated, and we have focused on the robust high-resolution parameter estimation of  $R$ -dimensional signals in the presence of white Gaussian noise.

Since colored noise is encountered in a variety of signal processing applications, such as audio processing, radar, sonar, EEG, MEG and medical imaging, and since the prewhitening step becomes essential for such problems, in this chapter we investigate matrix-based and tensor-based prewhitening schemes for high-resolution parameter estimation of  $R$ -dimensional signals in the presence of colored noise.

As shown in Chapter 2, for instance, and in several publications in the literature, if the noise is colored, the performance of parameter estimation techniques is substantially degraded. Therefore, in order to have an improved estimation of parameters in the presence of colored noise, we dedicate this chapter to propose one-dimensional and multi-dimensional subspace based prewhitening schemes.

Similarly to the other chapters, first we present in Section 4.1 a bibliographic review of the matrix-based prewhitening schemes found in the literature. Then, we show the proposed deterministic prewhitening scheme in [3], which is compared to the schemes in the literature.

Traditionally stochastic prewhitening techniques transform the colored noise into white noise keeping the SNR constant. In Section 4.1, we show our proposed deterministic approach for subspace prewhitening [3], where, by removing the noise correlation, an increase of the SNR is obtained. Consequently, in high noise correlation scenarios, where the subspace is prewhitened by our deterministic approach, there is a significant improvement in the parameter estimation accuracy. The proposed deterministic prewhitening assumes some structure of the noise correlation matrix as a function of the noise correlation coefficient. Therefore, we also propose solutions to estimate the correlation coefficients in Appendix 4.1.4.

As we mention in Section 2.2 of Chapter 2 for certain MIMO and EEG applications, the colored Kronecker noise or interference is present [52, 76]. For such a case, we propose in Section 4.2 the multi-dimensional prewhitening scheme called Sequential Generalized Singular Value Decomposition (S-GSVD), which takes into account the multi-dimensional structure of

the noise. Moreover, the S-GSVD is another generalization of the HOSVD as well as the Generalized HOSVD [97], however, in contrast to the Generalized HOSVD, which takes into account only one dimension between two tensors, the S-GSVD considers all the dimensions.

By exploiting the Kronecker structure of the noise correlation matrix for the estimation of the correlation factors, we achieve an improved accuracy compared to matrix-based prewhitening schemes. In addition, our S-GSVD approach is computationally more efficient than the classical matrix approach, since it has a lower complexity due to the  $n$ -mode GSVD operations [4].

Although in some applications, noise samples can be collected in order to estimate the noise statistics, e.g., in speech processing applications, the noise can be recorded in speechless frames [47], in some other applications the noise is only present, when the signal is present. Therefore, since the techniques in Section 4.1 and 4.2 are not suitable for this type of problem, we propose in Section 4.3 the *Iterative S-GSVD*.

Some open questions solved by us in this chapter include.

- 1) In case of severe noise correlation, can we prewhiten the noise by taking into account its structure such that its power gets reduced as a function of the level of the noise correlation?
  - 2) By applying such prewhitening to reduce of the noise power, is the noise obtained completely white? Moreover, is this scheme compatible to DOA estimation schemes, such as ESPRIT, MUSIC, or RARE?
  - 3) Once such a scheme is possible, how could we estimate in an efficient way the level of the noise correlation? How significant is the improvement caused by such a technique compared to the stochastic prewhitening schemes in the literature?
  - 4) For a data contaminated with colored noise with a multi-dimensional structure, can we propose multi-dimensional prewhitening schemes in order to provide a better estimation taking into account such structure? How is the computational performance compared to the traditional schemes?
  - 5) How can you apply prewhitening schemes when samples without signal components are not available? Moreover, can multi-dimensional prewhitening schemes also be applied in such a case? How is the difference of performance with and without the knowledge about the noise covariance structure when prewhitening schemes are applied?
-

## 4.1 Matrix-based Prewhitening

In practical applications using sensor arrays, the assumption that the noise of the sensors is uncorrelated may be not valid. For example, underwater noise components of a sonar system are in general spatially correlated [105]. Therefore, if no prewhitening step is applied, a severe degradation of the performance is observed.

Typically the prewhitening approaches require the estimation of the noise covariance matrix  $\mathbf{R}_{ww}$ , which is performed by collecting measurement samples in the absence of signal components. For example, in speech processing applications, the noise can be recorded in speechless frames [47]. The level of noise correlation ( $\rho$ ) depends on the specific application. For instance, for applications in physics [16] and in communications [65],  $\rho$  assumes values up to 0.99, while, for other applications, the correlation can assume smaller values.

In the stochastic prewhitening approaches of the literature [47, 89, 45], the data samples are multiplied by some prewhitening matrix,  $\mathbf{L}^{-1}$ , which transforms the correlated noise into white noise.

In Fig. 4.1, we present the profile of the noise singular values obtained by the SVD of the colored noise samples  $\mathbf{N}^{(c)}$ . In this case, the noise components are more concentrated in the main components region. Therefore, a degradation in the estimation of parameters extracted from the main components is expected.

By applying the stochastic prewhitening multiplying by  $\mathbf{L}^{-1}$ , the noise singular values become more equally distributed as shown in Fig. 4.2. Since the main components region is less affected by the noise, an improvement in the estimation of parameters extracted from the main components is expected.

In contrast to the stochastic prewhitening scheme, in our proposed deterministic approach, one sensor is used as the reference, and then, the correlated part of the noise is removed by taking into account its structure. In Fig. 4.3, we apply the deterministic prewhitening on the colored noise samples  $\mathbf{N}^{(c)}$ . We call our approach deterministic, since the prewhitening deterministic matrix  $\mathbf{D}$  has a certain structure and in this case depends only one parameter, which is the correlation coefficient. In Subsection 4.1.4, we propose also techniques to estimate these correlation parameters in terms of their amplitudes and phases. Note that the stochastic prewhitening schemes can be applied for any type of colored noise, while the deterministic prewhitening scheme can only be applied for colored noise with a certain structure.

In Subsection 4.1.5, we compare stochastic and deterministic prewhitening in computer simulations and demonstrate the improved performance of our deterministic approach. Here we restrict the application of the proposed deterministic prewhitening to ESPRIT-type algorithms. Nevertheless the deterministic prewhitening technique can be applied together with other subspace based parameter estimation techniques, like MUSIC, Root MUSIC, or RARE.

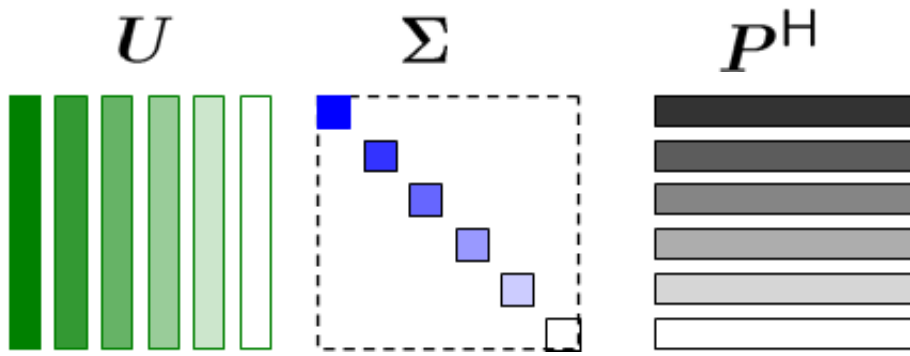


Fig. 4.1: Noise power profile in the presence of colored noise.  $N^{(c)} = U \cdot \Sigma \cdot P^H$

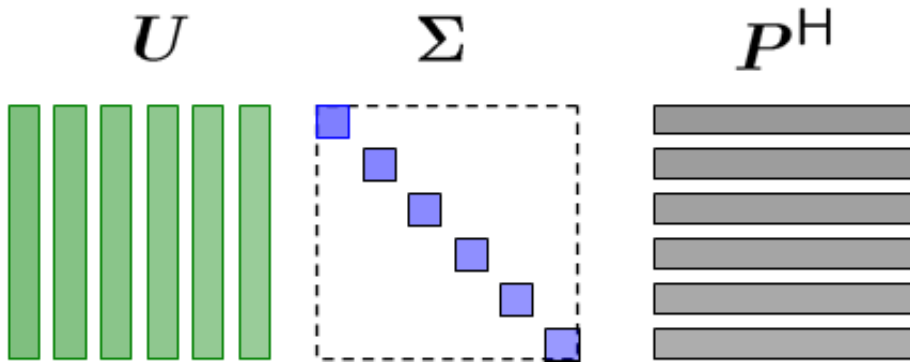


Fig. 4.2: Noise power profile in the presence of prewhitened noise via stochastic approaches.  
 $N = \hat{L}^{-1} \cdot N^{(c)}$

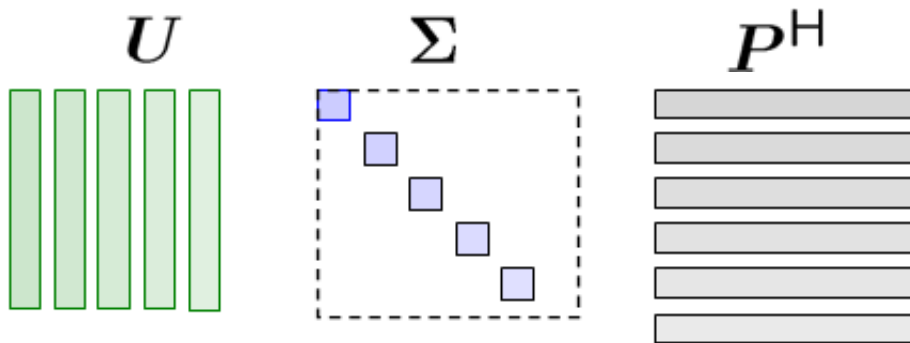


Fig. 4.3: Noise power profile in the presence of prewhitened noise via deterministic approach.  
 $V = D \cdot N^{(c)}$

### 4.1.1 Goal and Problem

Similarly to the previous chapter, our goal is to estimate the spatial frequencies. However, here we take into account the colored noise statistics to improve the estimation. In order to apply these statistics, we use prewhitening schemes.

First let us present our data model

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} + \mathbf{N}^{(c)}, \quad (4.1)$$

where  $\mathbf{N}^{(c)} = \mathbf{L} \cdot \mathbf{N}$  similarly as in Subsection 2.1.1. However, since we consider the same data model as in Subsubsection 2.1.2.1,  $\mathbf{A}$  has a Vandermonde structure as function of the spatial frequencies  $\mu_i$  similarly to (2.24).

We can state our prewhitening problem as to estimate all  $d$  spatial frequencies  $\mu_i$ , given  $\mathbf{X}$ , the statistics of  $\mathbf{N}^{(c)}$ , and  $d$ . Note that the statistics of  $\mathbf{N}^{(c)}$  are obtained with realizations in the absence of signals. Similarly as in Section 3.1, once all  $\mu_i$  are estimated,  $\mathbf{S}$  can be easily estimated.

### 4.1.2 Previous Approach: Stochastic Matrix-based Prewhitening

Here in this subsection, we show the stochastic matrix-based prewhitening using the matrix-based data model. The first step is to estimate the prewhitening matrix  $\mathbf{L}^{-1}$  by using samples with absence of signal components according to Subsubsection 4.1.2.1. With the prewhitening matrix, the data with signal samples has its noise prewhitened, and the signal subspace can be estimated. After that, in order to keep the original signal subspace, the estimated signal subspace should be dewhitened. The prewhitening and dewhitening process are explained in Subsubsection 4.1.2.2.

#### 4.1.2.1 Computation of the Stochastic Prewhitening Matrix

Here we consider the case that no signal components are present in the data. Therefore, we can estimate the prewhitening matrix  $\mathbf{L}$  according to [45, 47, 89] from the noise covariance matrix in the following way

$$\widehat{\mathbf{W}} = \widehat{\mathbf{L}} \cdot \widehat{\mathbf{L}}^H = \frac{1}{N_l} \cdot \mathbf{N}^{(c)} \cdot (\mathbf{N}^{(c)})^H \in \mathbb{C}^{M \times M}, \quad (4.2)$$

where  $N_l$  is the number of snapshots without signal components. Note that  $\widehat{\mathbf{L}}$  can, for example, be obtained by a Cholesky decomposition of  $\widehat{\mathbf{W}}$  or by an eigenvalue decomposition.

---

#### 4.1.2.2 Stochastic Prewhitening

In the classical approach [44], the prewhitening matrix  $\widehat{\mathbf{L}}$  from (4.2) is applied and the prewhitening step is performed by

$$\mathbf{X}' = \widehat{\mathbf{L}}^{-1} \cdot \mathbf{X}, \quad (4.3)$$

where  $\mathbf{X}'$  represents the prewhitened data matrix. Knowing the model order  $d$  and computing the SVD low-rank approximation of  $\mathbf{X}'$ , we can obtain the prewhitened subspace matrix  $\mathbf{U}^{[s]}$ . For the estimation of the spatial frequencies, the noise correlation should be taken into account (dewhitening), which is performed by

$$\mathbf{U}^{[s]'} = \widehat{\mathbf{L}} \cdot \mathbf{U}^{[s]} \quad (4.4)$$

where  $\mathbf{U}^{[s]}'$  can be used directly in the Standard ESPRIT. Alternatively, a similar procedure can be used for the  $R$ -D Unitary Tensor ESPRIT [44].

Note that instead of the matrix inversion in (4.37), the Generalized Singular Value Decomposition (GSVD) or the Generalized Eigenvalue Decomposition (GEVD) [45, 89] could have been applied. Except for critical cases that are not in the scope of this chapter, the performance of using inversion, the GSVD, and the GEVD are the same in terms of accuracy. However, in terms of computational complexity the GSVD and GEVD are always better than using the matrix inversion.

#### 4.1.3 Proposed Approach: Deterministic Matrix-based Prewhitening

Here we present the deterministic prewhitening, which was proposed by us in [3]. To demonstrate the estimation of the noise correlation coefficients, we consider the following specific correlation model [20]

$$n_{m+1}^{(c)}(n) = \rho \cdot n_m(n) + \sqrt{1 - |\rho|^2} \cdot n_{m+1}(n), \quad (4.5)$$

where  $n_{m+1}^{(c)}(n)$  is the colored noise sample in the  $m$ -th sensor for  $m = 1, 2, \dots, M - 1$  at the  $n$ -th instant for  $n = 1, \dots, N$ . Here,  $\rho \in \mathbb{C}$  represents the noise correlation coefficient between the sensors  $m$  and  $m + 1$ , such that  $0 \leq |\rho| \leq 1$ .

In the literature, we find three main stochastic approaches for the subspace prewhitening: the traditional one based on the Cholesky factorization of the noise covariance matrix [47], the GEVD approach [89, 45], and the GSVD approach [47, 89, 45].

In contrast to the stochastic approaches, a linear preprocessing with a deterministic matrix  $\mathbf{D}$  is applied to remove the noise correlation between two adjacent sensors  $m$  and  $m + 1$  in our proposed deterministic approach. First we consider the case where no signal component is present. Let us assume the noise covariance model in (2.36) to derive a preprocessing matrix  $\mathbf{D}$ . To this end, each element of the prewhitened noise matrix  $\mathbf{V} = \mathbf{D} \cdot \mathbf{N}^{(c)}$  is transformed as



$$v_m(n) = n_{m+1}(n) \cdot \sqrt{1 - |\rho|^2}. \quad (4.6)$$

Note that the greater  $|\rho|$ , the smaller is the variance of the deterministic prewhitened noise samples  $v_m(n)$ . Observe that the correlated noise  $n_{m+1}^{(c)}(n)$  in (4.5) is composed of a white noise component correlated with the noise at the previous sensor  $n_m(n)$  and another uncorrelated white component  $n_{m+1}(n)$ . Since the white noise component of  $n_m(n)$  is known from the previous sensor  $m$ , we can use this fact to obtain (4.6).

In order to decorrelate the noise as in (4.6), we propose a deterministic prewhitening matrix

$$\mathbf{D} = \begin{array}{cc} \underbrace{\mathbf{J}_2}_{(M-1) \times M} & -\widehat{\rho} \cdot \underbrace{\mathbf{J}_1}_{(M-1) \times M} \end{array}, \quad (4.7)$$

where  $\mathbf{J}_2 = [\mathbf{0}_{(M-1) \times 1} \quad \mathbf{I}_3] \in \mathbb{R}^{M-1 \times M}$  and  $\mathbf{J}_1 = [\mathbf{I}_3 \quad \mathbf{0}_{(M-1) \times 1}] \in \mathbb{R}^{M-1 \times M}$  are the selection matrices for the last  $M-1$  sensors and for the first  $M-1$  sensors, respectively, and  $\widehat{\rho}$  is an estimate of  $\rho$ . More information about the selection matrices  $\mathbf{J}_1$  and  $\mathbf{J}_2$  is shown in Subsubsection 2.1.2.1. In this section we assume that  $\widehat{\rho} = \rho$ , and in Subsection 4.1.4, we propose ways of calculating  $\widehat{\rho}$ .

Note that the same selection matrices  $\mathbf{J}_1$  and  $\mathbf{J}_2$  are used in the shift invariance equation with the following identity

$$\mathbf{J}_1 \cdot \mathbf{X} \cdot \Phi = \mathbf{J}_2 \cdot \mathbf{X}, \quad (4.8)$$

where  $\text{diag}(\Phi) = \{e^{j\mu_1}, \dots, e^{j\mu_d}\}$ .

Next, we consider the presence of signal components and the prewhitening matrix  $\mathbf{D}$  is applied in the following fashion

$$\mathbf{Y} = \mathbf{D} \cdot \mathbf{X} = \tilde{\mathbf{A}} \cdot (\Phi - \widehat{\rho} \cdot \mathbf{I}_d) \cdot \mathbf{S} + \mathbf{V}, \quad (4.9)$$

where  $\tilde{\mathbf{A}} = \mathbf{J}_1 \cdot \mathbf{A}$  has also a Vandermonde structure.<sup>1</sup>

**Lemma 1.** *Given (4.9), the first property that we prove is if applying the deterministic prewhitening to  $\mathbf{X}$  then obtained obtained matrix  $\mathbf{Y}$  satisfies the shift invariance equation, i.e.,*

$$\tilde{\mathbf{J}}_1 \cdot \mathbf{D} \cdot \mathbf{A} \cdot \Phi = \tilde{\mathbf{J}}_2 \cdot \mathbf{D} \cdot \mathbf{A}. \quad (4.10)$$

*Proof.*

$$\mathbf{D} \cdot \mathbf{A} = \mathbf{J}_2 \cdot \mathbf{A} - \widehat{\rho} \cdot \mathbf{J}_1 \cdot \mathbf{A}. \quad (4.11)$$

---

<sup>1</sup> Note that for the case that  $\rho$  is close to  $e^{j\mu_i}$  in magnitude and phase, then  $\sigma_s^2$ , the signal power for this particular  $\mu_i$  is reduced.

---

By replacing (4.8) in (4.11)

$$\begin{aligned}
\mathbf{D} \cdot \mathbf{A} &= \mathbf{J}_1 \cdot \mathbf{A} \cdot \Phi - \widehat{\rho} \cdot \mathbf{J}_1 \cdot \mathbf{A} \\
&= \mathbf{J}_1 \cdot \mathbf{A} \cdot (\Phi - \widehat{\rho} \cdot \mathbf{I}_d) \\
&= \tilde{\mathbf{A}} \cdot (\Phi - \widehat{\rho} \cdot \mathbf{I}_d)
\end{aligned} \tag{4.12}$$

Computing the shift invariance equation using (4.12),  $\tilde{\mathbf{J}}_1 \in \mathbb{R}^{(M-2) \times (M-1)}$ , and  $\tilde{\mathbf{J}}_2 \in \mathbb{R}^{(M-2) \times (M-1)}$ , gives us

$$\tilde{\mathbf{J}}_1 \cdot \tilde{\mathbf{A}} \cdot (\Phi - \widehat{\rho} \cdot \mathbf{I}_d) \cdot \Phi = \tilde{\mathbf{J}}_2 \cdot \tilde{\mathbf{A}} \cdot (\Phi - \widehat{\rho} \cdot \mathbf{I}_d), \tag{4.13}$$

which gives us that

$$\tilde{\mathbf{J}}_1 \cdot \tilde{\mathbf{A}} \cdot \Phi = \tilde{\mathbf{J}}_2 \cdot \tilde{\mathbf{A}} \tag{4.14}$$

□

**Lemma 2.** *Given (4.9), we desire to confirm if the resulting noise after applying the deterministic prewhitening is white, i.e.,*

$$\mathbb{E}\{\mathbf{V} \cdot \mathbf{V}^H\} = \beta \cdot \mathbf{I}_{M-1}, \tag{4.15}$$

where  $\beta = \sigma_n^2 \cdot (1 - |\rho|^2)$  as proven next.

*Proof.* First assume the following noise covariance matrix computed based on the model of [105] also used in (2.36)

$$\mathbf{R}_{nn} = \mathbf{L} \cdot \mathbf{L}^H = \begin{bmatrix} 1 & \rho^* & (\rho^2)^* & (\rho^3)^* & (\rho^4)^* \\ \rho & 1 & \rho^* & (\rho^2)^* & (\rho^3)^* \\ \rho^2 & \rho & 1 & \rho^* & (\rho^2)^* \\ \rho^3 & \rho^2 & \rho & 1 & \rho^* \\ \rho^4 & \rho^3 & \rho^2 & \rho^* & 1 \end{bmatrix}, \tag{4.16}$$

whose colored factor is given by

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \rho & \sqrt{1 - |\rho|^2} & 0 & 0 & 0 \\ \rho^2 & \rho \cdot \sqrt{1 - |\rho|^2} & \sqrt{1 - |\rho|^2} & 0 & 0 \\ \rho^3 & \rho^2 \cdot \sqrt{1 - |\rho|^2} & \rho \cdot \sqrt{1 - |\rho|^2} & \sqrt{1 - |\rho|^2} & 0 \\ \rho^4 & \rho^3 \cdot \sqrt{1 - |\rho|^2} & \rho^2 \cdot \sqrt{1 - |\rho|^2} & \rho \cdot \sqrt{1 - |\rho|^2} & \sqrt{1 - |\rho|^2} \end{bmatrix}. \tag{4.17}$$


---

Therefore, the prewhitened noise matrix  $\mathbf{V}$  is given by

$$\begin{aligned}
 \mathbf{V} &= \mathbf{D} \cdot \mathbf{N}^{(c)} = \mathbf{D} \cdot \mathbf{L} \cdot \mathbf{N} \\
 &= \mathbf{J}_2 \cdot \mathbf{L} \cdot \mathbf{N} - \widehat{\rho} \cdot \mathbf{J}_1 \cdot \mathbf{L} \cdot \mathbf{N} \\
 &= \begin{bmatrix} \rho & \sqrt{1-|\rho|^2} & 0 & 0 & 0 \\ \rho^2 & \rho \cdot \sqrt{1-|\rho|^2} & \sqrt{1-|\rho|^2} & 0 & 0 \\ \rho^3 & \rho^2 \cdot \sqrt{1-|\rho|^2} & \rho \cdot \sqrt{1-|\rho|^2} & \sqrt{1-|\rho|^2} & 0 \\ \rho^4 & \rho^3 \cdot \sqrt{1-|\rho|^2} & \rho^2 \cdot \sqrt{1-|\rho|^2} & \rho \cdot \sqrt{1-|\rho|^2} & \sqrt{1-|\rho|^2} \end{bmatrix} \cdot \mathbf{N} \\
 &\quad - \widehat{\rho} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \rho & \sqrt{1-|\rho|^2} & 0 & 0 & 0 \\ \rho^2 & \rho \cdot \sqrt{1-|\rho|^2} & \sqrt{1-|\rho|^2} & 0 & 0 \\ \rho^3 & \rho^2 \cdot \sqrt{1-|\rho|^2} & \rho \cdot \sqrt{1-|\rho|^2} & \sqrt{1-|\rho|^2} & 0 \end{bmatrix} \cdot \mathbf{N}. \tag{4.18}
 \end{aligned}$$

By using the assumption that  $\widehat{\rho} = \rho$  in (4.18) it gives us

$$\mathbf{V} = \left\{ \begin{bmatrix} 0 & \sqrt{1-|\rho|^2} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{1-|\rho|^2} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{1-|\rho|^2} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{1-|\rho|^2} \end{bmatrix} \right\} \cdot \mathbf{N} = \sqrt{1-|\rho|^2} \cdot \mathbf{J}_2 \cdot \mathbf{N} \tag{4.19}$$

By computing the covariance matrix of  $\mathbf{V}$

$$\begin{aligned}
 \mathbb{E}\{\mathbf{V} \cdot \mathbf{V}^H\} &= (1-|\rho|^2) \cdot \mathbf{J}_2 \cdot \mathbb{E}\{\mathbf{N} \cdot \mathbf{N}^H\} \cdot \mathbf{J}_2^T \\
 &= \sigma_n^2 \cdot (1-|\rho|^2) \cdot \mathbf{J}_2 \cdot \mathbf{J}_2^T \\
 &= \sigma_n^2 \cdot (1-|\rho|^2) \cdot \mathbf{I}_{M-1}. \tag{4.20}
 \end{aligned}$$

□

On the other hand, the covariance matrix of the prewhitened noise via stochastic approaches is given by

$$\mathbb{E}\{\mathbf{N} \cdot \mathbf{N}^H\} = \sigma_n^2 \mathbf{I}_M \tag{4.21}$$

The prewhitened noise power  $P_{\text{pwt}}$  is given by

$$P_{\text{pwt}} = \mathbb{E}\{v_m(n) \cdot v_m^*(n)\} = \sigma_n^2 \cdot (1-|\rho|^2), \tag{4.22}$$

where  $P_n = \mathbb{E}\{n_m(n) \cdot n_m^*(n)\} = \sigma_n^2$ .

Since  $\frac{P_{\text{pwt}}}{P_n} = 1-|\rho|^2$ , we have  $P_{\text{pwt}} < P_n$  for  $|\rho|^2 > 0$ , which means that this approach always gives a better SNR, for  $|\rho| > 0$  than in the white noise scenario. It is possible to observe this behavior by considering the SNR

$$\text{SNR}_{\text{pwt}} = 10 \cdot \log_{10} \left[ \frac{\sigma_s^2 \cdot |e^{j\mu_i} - \rho|^2}{\sigma_n^2 \cdot (1 - |\rho|^2)} \right]. \quad (4.23)$$

If the correlation coefficient  $|\rho|$  is close to 1, it implies that  $\text{SNR}_{\text{pwt}}$  approaches infinity. Therefore, the higher the correlation, the better the parameter estimation for the deterministic approach, and this is shown in the simulations in Subsection 4.1.5. Note that for such a gain, it is necessary that the correlation  $\rho$  should not be in phase to the signal.

The drawback of the deterministic approach is that the array aperture is reduced from  $M$  to  $M - 1$  sensors. This leads to a minor performance degradation, which becomes visible for small  $M$  and low correlations.

In summary, the objective of the *stochastic prewhitening* approaches is given a certain colored noise matrix  $\mathbf{N}^{(c)}$  and an estimate of the stochastic prewhitening matrix  $\mathbf{L}^{-1}$ , to prewhiten the noise in (4.1), such that the elements of the prewhitened noise  $\mathbf{N} = \mathbf{L}^{-1} \cdot \mathbf{N}^{(c)}$  have zero mean and variance  $\sigma_w^2$ , where  $\sigma_w^2$  denotes the noise power. In our *deterministic approach*, we propose to use  $\mathbf{D}$ , instead of the prewhitening matrix  $\mathbf{L}^{-1}$ , such that  $\mathbf{V} = \mathbf{D} \cdot \mathbf{W}^{(c)}$ , and the elements of  $\mathbf{V}$  have zero mean and variance  $(1 - |\rho|^2) \cdot \sigma_n^2$ , where  $(1 - |\rho|^2) \cdot \sigma_n^2$  denotes the noise power after applying the proposed deterministic prewhitening.

#### 4.1.4 Estimation of the Correlation Coefficient

Since for the deterministic approach in Subsection 4.1.3, the estimation of  $\rho$  is necessary, we propose different ways of performing this estimation from measurements taken in the absence of signal components. First we represent  $\rho$  as a function of its phase and magnitude, such that  $\rho = |\rho| \cdot e^{j\phi}$ .

Let us first take the sample estimate to obtain the phase and magnitude of  $\rho$

$$\hat{\rho} = \frac{\sum_{n=1}^N n_{m+1}^{(c)}(n) \cdot [n_m^{(c)}(n)]^*}{\sum_{n=1}^N n_m^{(c)}(n) \cdot [n_m^{(c)}(n)]^*}, \quad (4.24)$$

where  $m = 1, \dots, M - 1$ . Note that the numerator in (4.24) is  $N$  times the element of  $\hat{\mathbf{R}}_{nn}$  in row  $m + 1$  and column  $m$ . The sample estimate is applicable to arbitrary colored noise models. Next we show that for the considered specific correlated noise model, it is possible to improve the estimation of the correlation considerably. To this end, we propose two other techniques: the ESPRIT based phase estimation and the magnitude estimation of  $\rho$ .

For  $M = 3$ , the colored noise samples can be written as

$$\mathbf{N}^{(c)} = \begin{bmatrix} 1 & 0 & 0 \\ \rho & \sqrt{1 - |\rho|^2} & 0 \\ \rho^2 & \rho \cdot \sqrt{1 - |\rho|^2} & \sqrt{1 - |\rho|^2} \end{bmatrix} \cdot \mathbf{N}. \quad (4.25)$$

Note that this linear transformation has a specific structure that can effectively be exploited to estimate  $\rho$ .

For instance, the first column of the transformation matrix, which has the strongest power, has a Vandermonde structure with rate equal to  $\rho$ . Therefore, an ESPRIT based approach can be applied for the estimation of the phase of  $\rho$ , as shown in the following shift invariance equation

$$\mathbf{J}_2 \cdot \mathbf{u}_1^{(c)} = \mathbf{J}_1 \cdot \mathbf{u}_1^{(c)} \cdot \widehat{\rho}, \quad (4.26)$$

where the SVD of  $\mathbf{N}^{(c)}$  is given by  $\mathbf{U}^{(c)} \cdot \boldsymbol{\Sigma}^{(c)} \cdot (\mathbf{P}^{(c)})^H$  and  $\mathbf{u}_1^{(c)}$  is the first column of the matrix  $\mathbf{U}^{(c)}$ . The estimated phase shift  $e^{j\widehat{\phi}}$  is given by  $\frac{\widehat{\rho}}{|\widehat{\rho}|}$ .

In Figure 4.4, we compare the sample estimate in (4.24) with the ESPRIT based approach in (4.26) for the estimation of  $e^{j\phi}$ . We note that the performance of the ESPRIT based approach in (4.26) is far better than the performance of the sample estimate in (4.24).

Since we have already estimated the phase shift between two consecutive sensors, this information can be applied in order to phase align the correlated noise samples. Therefore, we consider the case that  $\rho \in \mathbb{R}$ , and the outputs of two consecutive sensors are given by

$$\begin{aligned} x_m(n) &= n_m^{(c)}(n) \\ x_{m+1}(n) &= n_{m+1}(n) \cdot \sqrt{1 - \rho^2} + n_m^{(c)}(n) \cdot \rho. \end{aligned} \quad (4.27)$$

The noise power can be estimated by

$$\widehat{\sigma}_w^2 = \frac{1}{N} \cdot \sum_{n=1}^N x_m^2(n) = \frac{1}{N} \cdot \sum_{n=1}^N [n_m^{(c)}(n)]^2. \quad (4.28)$$

Given the noise model, the following expression can be derived

$$\mathbb{E}\{[x_{m+1}(n) - x_m(n)]^2\} = 2 \cdot \sigma_n^2 \cdot (1 - \rho). \quad (4.29)$$

Since  $\sigma_w^2$  was estimated in (4.28) and using the expression in (4.29), the magnitude estimation of  $\rho$  is calculated according to

$$\widehat{\rho} = 1 - \frac{1}{2 \cdot N \cdot \widehat{\sigma}_n^2} \cdot \sum_{n=1}^N [x_{m+1}(n) - x_m(n)]^2 \quad (4.30)$$

In Figure 4.5, we compare the performance of the magnitude estimation in (4.30) with the sample estimate in (4.24), and the magnitude estimation outperforms significantly the estimation by the sample estimate. Note that both approaches have a better estimation accuracy when the number of samples is increased.

#### 4.1.5 Evaluation of Approaches

We consider the data symbols  $s_i(n)$  as being ZMCSCG distributed. The performance comparison is based on the spatial frequency estimation with standard ESPRIT (SE) [89] and for each realization, the spatial frequency for each source is chosen randomly from a uniform distribution in the interval from  $-\frac{\pi}{2}$  to  $\frac{\pi}{2}$ . In addition, we assume that  $\rho$  is known, since we are

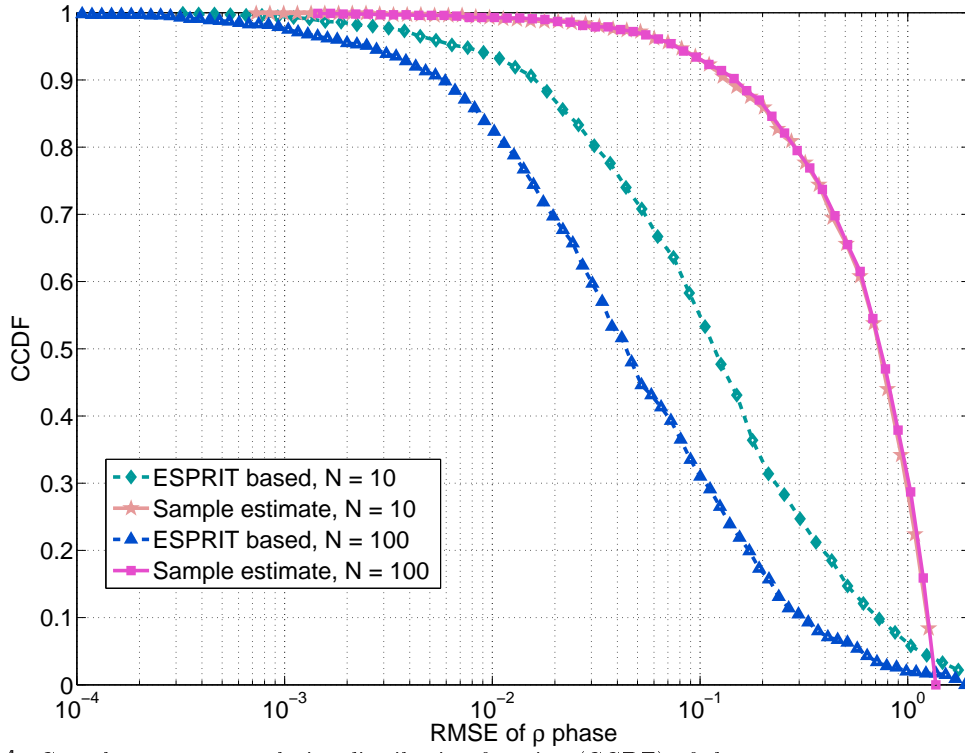


Fig. 4.4: Complementary cumulative distribution function (CCDF) of the root mean square error (RMSE) of  $\frac{\hat{\rho}}{|\hat{\rho}|}$  considering a only noise system with  $M = 11$  sensors, and  $N = 10$  and  $N = 100$  snapshots.

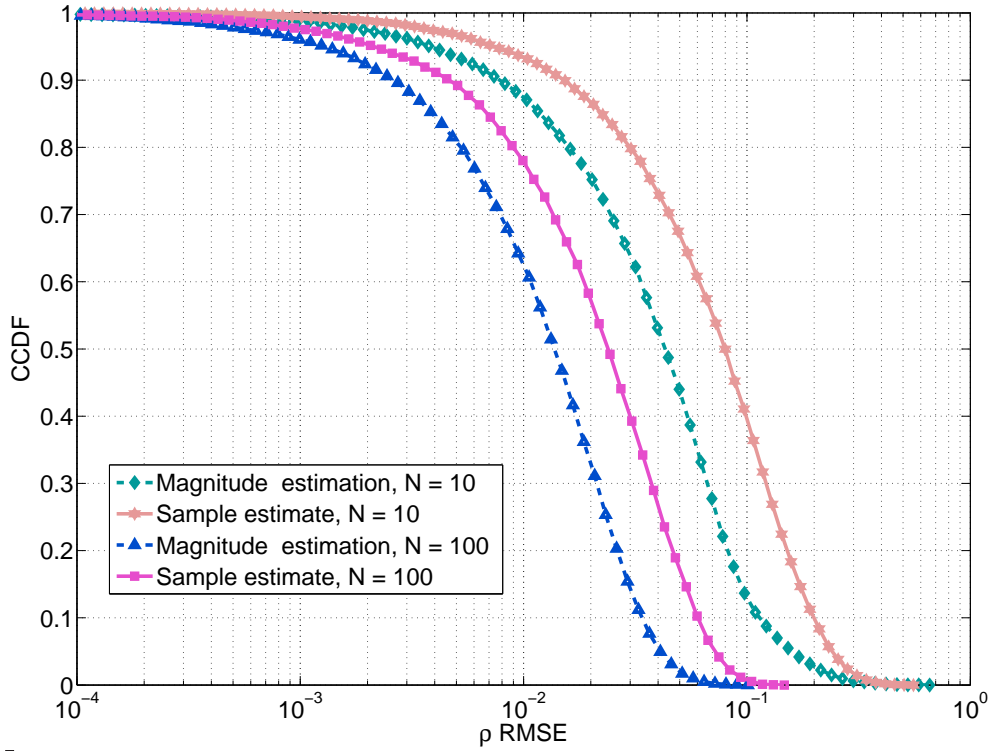


Fig. 4.5: Complementary cumulative distribution function (CCDF) of the root mean square error (RMSE) of  $|\hat{\rho}|$  considering a only noise system with  $M = 2$  sensors, and  $N = 10$  and  $N = 100$  snapshots.

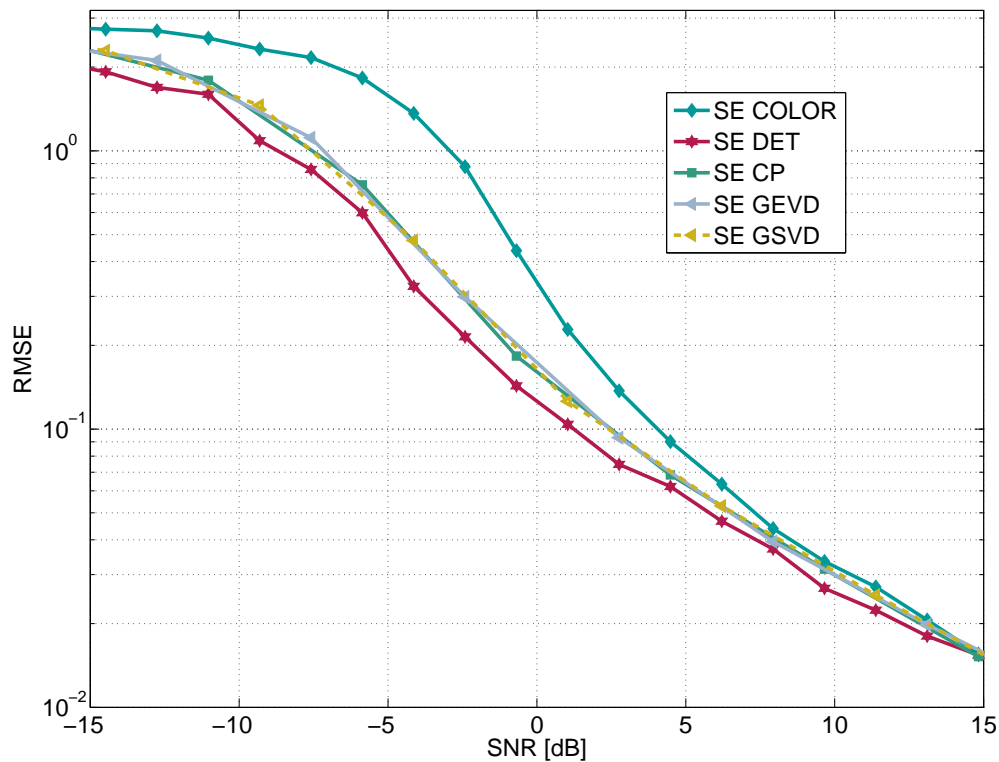


Fig. 4.6: RMSE of the spatial frequencies vs. SNR considering a system with  $N = 10$  snapshots, with  $M = 10$  sensors and with  $\rho = 0.7$ .  $d = 3$  sources are present.

interested in evaluating the different prewhitening schemes and not the estimation of  $\rho$ , which may vary, for example, for a different number of snapshots according to Section 4.1.4. The notation in the legends is the following: SE Color stands for the estimation without using any prewhitening, SE DET for the deterministic approach proposed here, SE CP for the classical prewhitening in [47], SE GEVD for the prewhitening scheme in [89, 45], and SE GSVD for the prewhitening in [47, 89, 45].

In Figure 4.6, the RMSE of the spatial frequencies is plotted versus the SNR. For this scenario,  $\rho = 0.7$  and all the prewhitening techniques outperform SE Color. Note that the SE DET outperforms all the other prewhitening techniques.

In order to observe the performance as a function of  $\rho$ , we fix the SNR to 1 dB in Figure 4.7. Note that a considerable improvement by using all types of prewhitening is only observed for  $\rho > 0.3$ . Also note that the stochastic prewhitening schemes tend to keep the noise power  $\sigma_w^2$  constant for all values of  $\rho$ . On the other hand, for the deterministic approach, the greater the correlation, the greater the gain obtained, which is expected according to (4.23). Therefore, in Figure 4.7, the deterministic prewhitening outperforms significantly the stochastic approaches for  $\rho > 0.7$ , and only slightly for  $0.4 \leq \rho \leq 0.7$ . As a drawback, we note that for  $\rho \leq 0.4$ , the stochastic prewhitening techniques slightly outperform SE DET, and for  $\rho < 0.3$ , the estimation without prewhitening also slightly outperforms SE DET. This phenomenon is due to the aperture reduction already mentioned. In Fig. 4.8, we consider only one source with

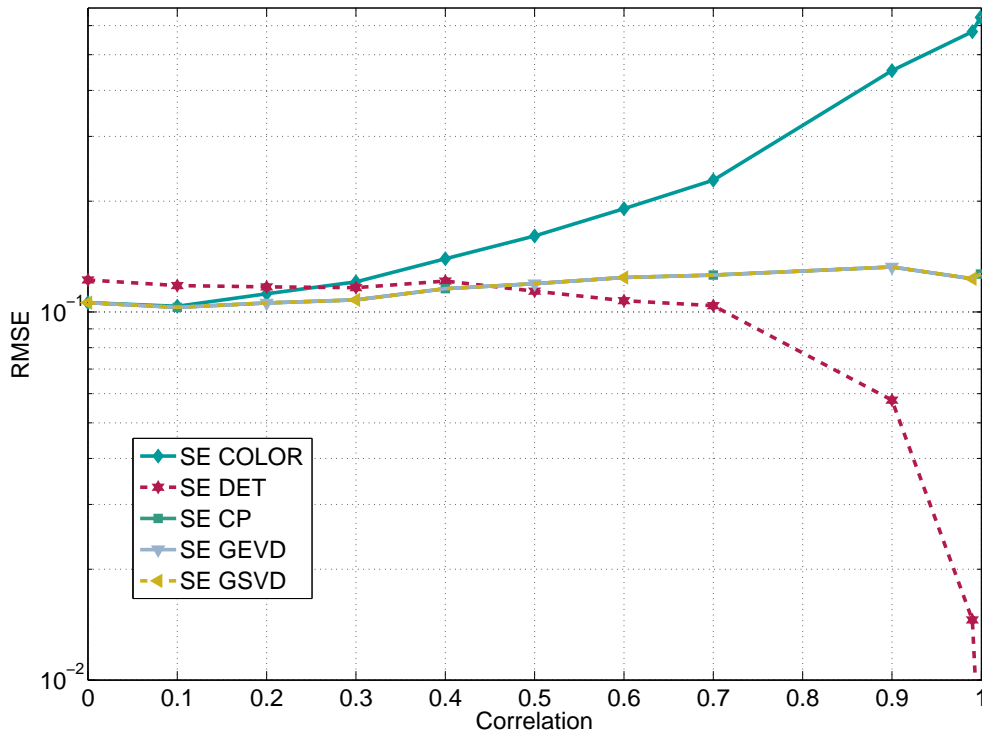


Fig. 4.7: RMSE of the spatial frequencies vs. correlation factor  $\rho$  considering a system with  $N = 10$  snapshots and with  $M = 10$  sensors.  $d = 3$  sources are present. The SNR is fixed to 1 dB.

spatial frequency  $\mu_1$ , and we consider that the phase between the noise correlation coefficient  $\frac{\rho}{|\rho|}$  and the spatial frequency of the source  $\frac{\mu_1}{|\mu_1|}$  is getting closer. The angular distance is computed by  $e_{\rho, \mu_1} = \left| \frac{\mu_1}{|\mu_1|} - \frac{\rho}{|\rho|} \right|$ . Such problem is mentioned in (4.9). In general, all schemes are affected when the noise correlation and the spatial frequency are in phase. Therefore, not only the deterministic prewhitening scheme has the performance degraded for such a scenario, but also and mainly the stochastic prewhitening schemes are much more degraded. In addition, for very close distances, the performance using the deterministic prewhitening is close to the performance of using no prewhitening.

Also in Fig. 4.8, the performance of all schemes becomes better, when the phase of the noise becomes closer to the phase of the signal. Since the noise phase is the same of the signal, one of the noise components improves the estimation instead of degrading the estimation. This noise component is the first column of  $\mathbf{L}$  in (4.16).

#### 4.1.6 Main Results of the Section

We propose a deterministic prewhitening technique, which outperforms the prewhitening techniques presented in the literature in case of high noise correlation. Observe that in general we have three cases: First there is the case of a small noise correlation, when the prewhitening step for the simulated scenario does not give a significant improvement. For an intermediate



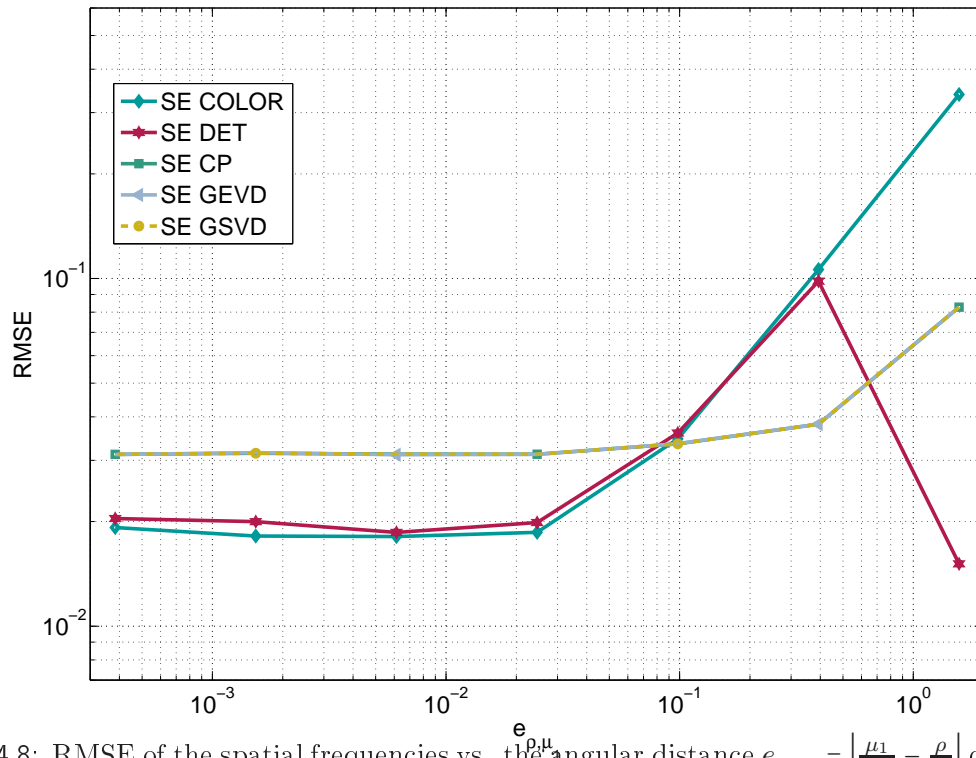


Fig. 4.8: RMSE of the spatial frequencies vs. the angular distance  $e_{\rho, \mu_1} = \left| \frac{\mu_1}{|\mu_1|} - \frac{\rho}{|\rho|} \right|$  considering a system with  $N = 10$  snapshots and with  $M = 10$  sensors.  $d = 1$  source is present. The SNR is fixed to 1 dB. The correlation coefficient is set to  $\rho = 0.9$ .

level of noise correlation, the stochastic prewhitening slightly outperforms the deterministic. Finally, for a high noise correlation, the proposed deterministic approach outperforms significantly the stochastic approaches.

Moreover, in Subsection 4.1.4, we propose an ESPRIT based phase estimation together with the proposed magnitude estimation to obtain the correlation coefficients. Therefore, depending on the estimated level of correlation, it is possible to switch between no prewhitening, the deterministic and stochastic prewhitening.

## 4.2 Multi-dimensional Prewhitening

In this section, we present our proposed multi-dimensional prewhitening scheme called Sequential GSVD (S-GSVD) based prewhitening. Similarly to the previous sequential, we also assume that noise samples are available in the absence of signal components.

One important application of the S-GSVD (Sequential Generalized Singular Value Decomposition) based prewhitening scheme is the parameter estimation in environments with multi-dimensional colored noise or interference with Kronecker correlation structure, i.e., the overall noise covariance matrix can be expressed as a Kronecker product of the noise covariance matrices in the separate dimensions as in (2.43) in Chapter 2. Since the S-GSVD based prewhitening approach requires knowledge of the noise correlation, we propose an efficient scheme to estimate the correlation factors which exploits this Kronecker structure.

Comparing the generalized HOSVD presented in [97] to our proposed S-GSVD based prewhitening approach, we observe that in the former the GSVD was only applied to one unfolding, while in our approach the sequential GSVDs are applied to all unfoldings.

### 4.2.1 Goal and Problem

In this section, the following data model is assumed

$$\mathcal{X} = \mathcal{I}_{R+1,d} \times_{r=1}^R \mathbf{A}^{(r)} \times_{R+1} \mathbf{S}^T + \mathcal{N}^{(c)}, \quad (4.31)$$

where  $\mathcal{N}^{(c)}$  is defined in (2.43). Note that (4.31) is similar to the data model in Chapter 3 in (3.9), except that in (3.9) the noise is white, while here the noise is colored.

Therefore, the parameter estimation problem in the presence of colored noise can be stated as: given the measurement tensor  $\mathcal{X}$ , the model order  $d$ , and noise samples  $\mathcal{N}^{(c)}$ , we desire to estimate all the spatial frequencies  $\mu_i^{(r)}$ , which in total are  $R \cdot d$  variables. Obviously that  $\mathcal{X}$  and  $\mathcal{N}^{(c)}$  are obtained from different realizations. Similarly as in Section 3.1, once all  $\mu_i$  are estimated,  $\mathbf{S}$  can be easily estimated.

### 4.2.2 Previous Approaches: Stochastic Prewhitening and Generalized HOSVD

#### 4.2.2.1 Generalized HOSVD of [97]

The generalized HOSVD proposed in [97] could also be applied to the prewhitening task. The HOSVD is defined via the SVDs of all unfoldings of a tensor, while in the generalized HOSVD one of the SVDs is replaced by the GSVD of the  $t$ -th unfoldings of two tensors. In our case, we consider the tensors  $\mathcal{X}$  and  $\mathcal{N}^{(c)}$ . The GSVD is explained in more detail in (G.1) and in (G.2) in Appendix G

---

Since the generalized HOSVD considers only one dimension for the GSVD [97], we can apply it only in the case that the  $t$ -th dimension of the noise is correlated. Therefore, applying the generalized HOSVD of  $\mathcal{X}$  and  $\mathcal{N}^{(c)}$  in the  $t$ -th dimension, we obtain

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_t \mathbf{G} \dots \times_R \mathbf{U}_R \times_{R+1} \mathbf{U}_{R+1} \quad (4.32)$$

$$\mathcal{N}^{(c)} = \mathcal{T} \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \dots \times_t \mathbf{G} \dots \times_R \mathbf{V}_R \times_{R+1} \mathbf{V}_{R+1}, \quad (4.33)$$

Note that since  $\mathcal{N}^{(c)}$  and  $\mathcal{X}$  cannot be observed simultaneously, they represent different realizations of the noise process and the signal plus noise process, respectively.

Similarly to the HOSVD-based low-rank approximation of  $\mathcal{X}$  [44], we might apply the generalized HOSVD-based low-rank approximation to (4.32) in order to obtain

$$\mathcal{U}^{[s]} = \mathcal{S}^{[s]} \times_1 \mathbf{U}_1^{[s]} \times_2 \mathbf{U}_2^{[s]} \dots \times_t \mathbf{G}^{[s]} \dots \times_R \mathbf{U}_R^{[s]}, \quad (4.34)$$

where  $\mathcal{U}^{[s]} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$ ,  $\mathcal{S}^{[s]} \in \mathbb{C}^{p_1 \times p_2 \times \dots \times p_R \times d}$ ,  $\mathbf{U}_i^{[s]} \in \mathbb{C}^{M_i \times p_i}$ , and  $\mathbf{G}^{[s]} \in \mathbb{C}^{M_t \times p_t}$ , such that  $p_i = \min(M_i, d)$ . The subspace tensor  $\mathcal{U}^{[s]}$  can be used in the  $R$ -D Standard Tensor-ESPRIT algorithm. In this section, we show briefly how the matrix-based prewhitening approach can be used to prewhiten multi-dimensional data for the tensor-based signal subspace estimation schemes developed in [44].

#### 4.2.2.2 Prewhitening matrix estimation

In this subsection, we consider the case that no signal components are present in the data. Therefore, we can estimate the prewhitening matrix according to [45, 47, 89] from the noise covariance matrix in the following way

$$\widehat{\mathbf{W}}_{\text{mtx}} = \widehat{\mathbf{L}}_{\text{mtx}} \cdot \widehat{\mathbf{L}}_{\text{mtx}}^H = \frac{1}{\alpha \cdot N_l} \cdot [\mathcal{N}^c]_{(R+1)}^T \cdot [\mathcal{N}^c]_{(R+1)}^* \in \mathbb{C}^{M \times M}, \quad (4.35)$$

where  $\alpha$  is chosen such that  $\text{tr}(\widehat{\mathbf{W}}_{\text{mtx}}) = M$  and  $N_l$  is the number of snapshots without signal components as in (4.2). Note that  $\widehat{\mathbf{L}}_{\text{mtx}}$  can, for example, be obtained by a Cholesky decomposition of  $\widehat{\mathbf{W}}_{\text{mtx}}$  or by an eigenvalue decomposition. Since we assume an ergodic Kronecker noise model, if  $N_l$  approaches infinity, then  $\widehat{\mathbf{L}}_{\text{mtx}}$  satisfies the following expression

$$\mathbf{L}_{\text{mtx}} = \mathbf{L}_1 \otimes \mathbf{L}_2 \dots \otimes \mathbf{L}_R. \quad (4.36)$$

The equation (4.35) is applicable to an arbitrary noise correlation model. Note that if  $R = 1$ , then  $\mathbf{L}_{\text{mtx}} = \mathbf{L}$  as in (4.2).

---

### 4.2.2.3 Matrix prewhitening

In the classical prewhitening approach to prewhiten the subspace tensor defined in [44], the prewhitening matrix  $\widehat{\mathbf{L}}_{\text{mtx}}$  from (4.35) is applied and the prewhitening step is performed by

$$[\mathbf{x}'']_{(R+1)}^T = \widehat{\mathbf{L}}_{\text{mtx}}^{-1} \cdot [\mathbf{x}]_{(R+1)}^T, \quad (4.37)$$

where  $\mathbf{x}''$  represents the prewhitened data tensor using the matrix-based approach. Computing the HOSVD-based low-rank approximation of  $\mathbf{x}''$  according to [44], we obtain the prewhitened subspace tensor  $\mathbf{u}^{[s]}$ . Finally, for the estimation of the spatial frequencies, the noise correlation should be taken into account (dewhitening), which is performed by

$$[\mathbf{u}^{[s]''}]_{(R+1)}^T = \widehat{\mathbf{L}}_{\text{mtx}} \cdot [\mathbf{u}^{[s]}]_{(R+1)}^T, \quad (4.38)$$

where  $\mathbf{u}^{[s]''}$  can be used directly in the  $R$ -D Standard Tensor-ESPRIT algorithm [44] to obtain the harmonic retrieval parameters. Alternatively, a similar procedure can be used for the  $R$ -D Unitary Tensor ESPRIT [44].

## 4.2.3 Proposed Approaches

### 4.2.3.1 Prewhitening Correlation Factors Estimation

In this subsubsection, we consider the case that no signal components are present in the data. Nevertheless, if the noise covariance has a Kronecker structure, we can obtain a better estimate of the noise model by considering each  $[\mathcal{N}^{(c)}]_{(i)}$  for  $i = 1, 2, \dots, R$ . Based on (2.45) it is possible to estimate the correlation factor  $\mathbf{L}_i$  by applying an eigenvalue decomposition (EVD) or a Cholesky decomposition. For example, in case of using the EVD of  $\widehat{\mathbf{W}}_i$ , we can estimate the factors by performing

$$\widehat{\mathbf{W}}_i = \frac{M_i}{\alpha \cdot N_l \cdot M} \cdot [\mathcal{N}^{(c)}]_{(i)} \cdot [\mathcal{N}^{(c)}]_{(i)}^H, \quad (4.39)$$

where  $\alpha$  is chosen such that  $\text{tr}(\widehat{\mathbf{W}}_i) = M_i$  and  $N_l$  is the number of snapshots without signal components as in (4.35). Since the EVD of  $\widehat{\mathbf{W}}_i$  is given by

$$\widehat{\mathbf{W}}_i = \mathbf{Q}_i \cdot \mathbf{\Lambda} \cdot \mathbf{Q}_i^H, \quad (4.40)$$

we can choose

$$\widehat{\mathbf{L}}_i = \mathbf{Q}_i \cdot \mathbf{\Lambda}^{\frac{1}{2}}. \quad (4.41)$$

### 4.2.3.2 Tensor Prewhitening Scheme: $n$ -mode Products using Matrix Inversions

In this subsubsection, we show briefly how to apply the prewhitening using the correlation factors. In the simulations, the prewhitening techniques shown here are used in conjunction

with  $R$ -D Standard Tensor-ESPRIT [44]. In the same way, it is also possible to use the multi-dimensional prewhitening schemes in conjunction with  $R$ -D Unitary Tensor-ESPRIT [44].

First let us propose the prewhitening scheme based on  $n$ -mode products using matrix inversions with the correlation factor estimation proposed in (4.41). The expression below shows the prewhitened tensor

$$\mathcal{X}' = \mathcal{X} \times_1 \widehat{\mathbf{L}}_1^{-1} \times_2 \widehat{\mathbf{L}}_2^{-1} \dots \times_R \widehat{\mathbf{L}}_R^{-1}. \quad (4.42)$$

Replacing (3.6) with the noise model from (2.44) in (4.42), we obtain that:

$$\mathcal{X}' = \mathcal{A} \times_1 \widehat{\mathbf{L}}_1^{-1} \times_2 \widehat{\mathbf{L}}_2^{-1} \dots \times_R \widehat{\mathbf{L}}_R^{-1} \times_{R+1} \mathbf{S}^T + \mathcal{N}, \quad (4.43)$$

where after prewhitening the multi-dimensional noise becomes white. Moreover,  $\mathcal{X}'$  can be represented by its HOSVD as follows

$$\mathcal{X}' = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_R \mathbf{U}_R \times_{R+1} \mathbf{U}_{R+1}, \quad (4.44)$$

where  $\mathcal{S} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times N}$  and  $\mathbf{U}_i \in \mathbb{C}^{M_i \times M_i}$ , such that  $i = 1, 2, \dots, R$ . The HOSVD-based low-rank approximation can be applied to (4.44). Therefore, the corresponding subspace tensor is equal to

$$\mathcal{U}^{[s]'} = \mathcal{S}^{[s]} \times_1 \mathbf{U}_1^{[s]} \times_2 \mathbf{U}_2^{[s]} \dots \times_R \mathbf{U}_R^{[s]}, \quad (4.45)$$

where  $\mathcal{S}^{[s]} \in \mathbb{C}^{p_1 \times p_2 \times \dots \times p_R \times d}$ ,  $\mathbf{U}_i^{[s]} \in \mathbb{C}^{M_i \times p_i}$ , and  $\mathcal{U}^{[s]'} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$ , such that  $p_i = \min(M_i, d)$  for  $i = 1, 2, \dots, R$ . We assume that  $d \leq N$ .

Similarly to the derivation of  $R$ -D Standard Tensor-ESPRIT in [44], there is an invertible matrix  $\mathbf{T}$ , such that

$$\mathcal{U}^{[s]'} \approx \mathcal{A} \times_1 \widehat{\mathbf{L}}_1^{-1} \times_2 \widehat{\mathbf{L}}_2^{-1} \dots \times_R \widehat{\mathbf{L}}_R^{-1} \times_{R+1} \mathbf{T}. \quad (4.46)$$

Therefore, to obtain the correct subspace is necessary to dewhiten the estimated subspace by performing

$$\mathcal{U}^{[s]} = \mathcal{U}^{[s]'} \times_1 \widehat{\mathbf{L}}_1 \times_2 \widehat{\mathbf{L}}_2 \dots \times_R \widehat{\mathbf{L}}_R. \quad (4.47)$$

#### 4.2.3.3 Tensor Prewhitening Scheme: S-GSVD

In the S-GSVD based prewhitening approach, the explicit prewhitening by matrix inversion can be avoided, and additionally it allows the sequential implementation of (4.42). Let  $\tilde{\mathcal{X}}^{(i)}$  be the tensor which has been prewhitened up to the  $i$ -th dimension, where  $i = 1, 2, \dots, R$ , and  $\tilde{\mathcal{X}}^{(0)} = \mathcal{X}$ . To perform the prewhitening in the  $i$ -th dimension, we calculate the GSVD of  $\left[ \tilde{\mathcal{X}}^{(i-1)} \right]_{(i)}^H$  and  $\widehat{\mathbf{L}}_i^H$ , which gives us

$$\begin{aligned}\widehat{\mathbf{L}}_i &= \mathbf{V}_i \cdot \boldsymbol{\Xi}_i^{(L)\top} \cdot \mathbf{U}_i^{(L)\text{H}} \\ \left[ \tilde{\boldsymbol{\chi}}^{(i-1)} \right]_{(i)} &= \mathbf{V}_i \cdot \boldsymbol{\Xi}_i^{(X)\top} \cdot \mathbf{U}_i^{(X)\text{H}},\end{aligned}\quad (4.48)$$

where  $\mathbf{U}_i^{(X)} \in \mathbb{C}^{\frac{M \cdot N}{M_i} \times \frac{M \cdot N}{M_i}}$  and  $\mathbf{U}_i^{(L)} \in \mathbb{C}^{M_i \times M_i}$  are unitary matrices, and the matrices  $\boldsymbol{\Xi}_i^{(X)} \in \mathbb{R}^{\frac{M \cdot N}{M_i} \times M_i}$  and  $\boldsymbol{\Xi}_i^{(L)} \in \mathbb{R}^{M_i \times M_i}$  have all elements equal to zero, except for the main diagonals. Note that the S-GSVD is related to the technique of Subsection 4.2.3.2, since each  $n$ -mode GSVD would be equivalent to one  $n$ -mode multiplication by  $\widehat{\mathbf{L}}_i^{-1}$  of the technique of Subsection 4.2.3.2. Indeed, we have that

$$\begin{aligned}\left[ \tilde{\boldsymbol{\chi}}^{(i)} \right]_{(i)} &= \widehat{\mathbf{L}}_i^{-1} \cdot \left[ \tilde{\boldsymbol{\chi}}^{(i-1)} \right]_{(i)} \\ &= \mathbf{U}_i^{(L)} \cdot \boldsymbol{\Xi}_i^{(L)-1} \cdot \boldsymbol{\Xi}_i^{(X)\top} \cdot \mathbf{U}_i^{(X)\text{H}},\end{aligned}\quad (4.49)$$

which is equivalent to

$$\tilde{\boldsymbol{\chi}}^{(i)} = \tilde{\boldsymbol{\chi}}^{(i-1)} \times_i \widehat{\mathbf{L}}_i^{-1}. \quad (4.50)$$

Since (4.48) works for cases that  $\widehat{\mathbf{L}}_i$  is rank deficient, then it should be used instead of (4.50).

After applying  $R$  GSVDs in (4.48) and rebuilding the prewhitened tensor using (4.49), we obtain  $\tilde{\boldsymbol{\chi}}^{(R)}$  which is equal to  $\boldsymbol{\chi}'$  in (4.42).

Since  $\tilde{\boldsymbol{\chi}}^{(0)}$  is prewhitened sequentially in the S-GSVD approach,  $\mathbf{U}_i^{(L)}$  in (4.49) is an approximation of  $\mathbf{U}_i$  in (4.44). With every step in the sequential prewhitening, the approximation becomes better until for the last GSVD, e.g., when  $i = R$ ,  $\mathbf{U}_R^{(L)} = \mathbf{U}_R$ . At the end of this section, it is explained how to improve the accuracy of the S-GSVD in order to obtain the same performance as in Subsection 4.2.3.2.

Since  $\mathbf{U}_i^{(L)}$  is a unitary matrix, we can rewrite  $\tilde{\boldsymbol{\chi}}^{(R)}$  in the following fashion

$$\tilde{\boldsymbol{\chi}}^{(R)} = \boldsymbol{\mathcal{S}}' \times_1 \mathbf{U}_1^{(L)} \times_2 \mathbf{U}_2^{(L)} \dots \times_R \mathbf{U}_R^{(L)} \times_{R+1} \mathbf{U}_{R+1}. \quad (4.51)$$

where  $\boldsymbol{\mathcal{S}}' \in \mathbb{C}^{M_1 \times M_2 \dots \times M_R \times N}$ .  $\mathbf{U}_{R+1}$  is calculated by applying the SVD in the  $R+1$ -mode unfolding.

If we multiply (4.48) by  $\mathbf{U}_i^{(L)}$  from the right side, we can conclude that

$$\widehat{\mathbf{L}}_i \cdot \mathbf{U}_i^{(L)} = \mathbf{V}_i \cdot \boldsymbol{\Xi}_i^{(L)\top} = \mathbf{V}_i \cdot \boldsymbol{\Xi}_i^{(L)}. \quad (4.52)$$

Therefore, the low-rank approximation can be applied to (4.51) and using (4.52) an approximated dewhitened subspace tensor can be directly obtained

$$\tilde{\boldsymbol{\mathcal{U}}}^{[s]} = \boldsymbol{\mathcal{S}}^{[s]'} \times_1 \left( \mathbf{V}_1^{[s]} \cdot \boldsymbol{\Xi}_1^{(L)[s]} \right) \times_2 \left( \mathbf{V}_2^{[s]} \cdot \boldsymbol{\Xi}_2^{(L)[s]} \right) \dots \times_R \left( \mathbf{V}_R^{[s]} \cdot \boldsymbol{\Xi}_R^{(L)[s]} \right), \quad (4.53)$$

where  $\boldsymbol{\mathcal{S}}^{[s]'} \in \mathbb{C}^{p_1 \times p_2 \dots \times p_R \times d}$ ,  $\mathbf{V}_i^{[s]} \in \mathbb{C}^{M_i \times p_i}$ , and  $\boldsymbol{\Xi}_i^{(L)[s]} \in \mathbb{C}^{p_i \times p_i}$ . This  $\tilde{\boldsymbol{\mathcal{U}}}^{[s]}$  is the subspace

tensor which can be used in the  $R$ -D Standard Tensor-ESPRIT algorithm. Note that the order of the dimensions in the sequential prewhitening scheme can be chosen arbitrarily. The best accuracy is achieved if we begin in the dimension with the highest noise correlation and proceed to the less correlated dimensions sequentially. The core tensor  $\mathcal{S}'$ , which is truncated to  $\mathcal{S}^{[s]}'$  in (4.53), can be obtained via the  $i$ -th mode product between the tensor  $\tilde{\mathcal{X}}^{(R)}$  in (4.51) and the matrix  $\left(\mathbf{U}_i^{(L)}\right)^H$  for  $i = 1, \dots, R + 1$ .

As mentioned previously, since the S-GSVD is applied sequentially, the terms  $\mathbf{U}_i^{(L)}$  are an approximation of  $\mathbf{U}_i$ . To obtain an equality,  $\mathbf{U}_i$  can be obtained via an SVD as the left singular vectors of  $\left[\tilde{\mathcal{X}}^{(R)}\right]_{(i)}$ . Therefore, (4.53) has to be modified by

$$\mathbf{u}^{[s]} = \mathcal{S}^{[s]} \times_1 \left(\hat{\mathbf{L}}_1 \cdot \mathbf{U}_1^{[s]}\right) \times_2 \left(\hat{\mathbf{L}}_2 \cdot \mathbf{U}_2^{[s]}\right) \dots \times_{R-1} \left(\hat{\mathbf{L}}_{R-1} \cdot \mathbf{U}_{R-1}^{[s]}\right) \times_R \left(\mathbf{V}_R^{[s]} \cdot \Xi_R^{(L)[s]}\right), \quad (4.54)$$

where  $\mathcal{S}^{[s]} \in \mathbb{C}^{p_1 \times p_2 \dots \times p_R \times d}$  and  $\mathbf{U}_i^{[s]} \in \mathbb{C}^{M_i \times p_i}$ . By performing the modification, the accuracy is exactly the same as in Subsection 4.2.3.2, although additionally  $R - 1$  SVDs are computed, which leads to a mild increase in complexity as compared to the low complexity version of the S-GSVD.

#### 4.2.4 Evaluation of Approaches

In the proposed multi-dimensional approach in (4.42), we need to invert  $R$  matrices of a much smaller size than the matrix  $\hat{\mathbf{L}}_{\text{mtx}}$  in (4.37) to prewhiten the data. Consequently, (4.42) is also computationally much more efficient than (4.37), since the complexity of an inversion of an  $M \times M$  matrix is approximately  $O(M^3)$ . Moreover, in the S-GSVD based prewhitening approach in (4.48) no explicit matrix inversion is required.

Note that the S-GSVD in its low complexity form in (4.53) only requires the computation of GSVDs for the unfoldings with noise correlation and SVDs for the unfoldings without noise correlation. However, in its high accuracy form in (4.54) additional SVDs should be computed. In this subsection we present simulation results demonstrating the performance of the proposed methods. The spatial frequencies  $\mu_i^{(r)}$  are drawn from a uniform distribution in  $[-\pi, \pi]$ . The source symbols are zero mean i.i.d. circularly symmetric complex Gaussian distributed with power equal to  $\sigma_s^2$  for all the sources. We define at the receiver

$$\text{SNR} = 10 \cdot \log_{10} \left( \frac{\sigma_s^2}{\sigma_n^2} \right), \quad (4.55)$$

where  $\sigma_n^2$  is the variance of the elements of the white noise tensor  $\mathcal{N}$  and  $\text{tr}(\mathbf{W}_i) = M_i$ . In addition,  $\mathbf{W}_i$  varies as a function of the correlation coefficient  $\rho_i$ .

In order to evaluate the prewhitening schemes integrated into the  $R$ -D Standard Tensor-

Abbreviation	Algorithm
<i>R</i> -D SE Color	The Standard ESPRIT algorithm is applied without prewhitening.
<i>R</i> -D SE MP E	The Standard ESPRIT algorithm is applied together with the matrix-based prewhitening scheme based on (4.37) and (4.38) with the matrix-based estimate of the correlation matrix from (4.35).
<i>R</i> -D SE TP E	The Standard ESPRIT algorithm is applied together with the matrix-based prewhitening scheme based on (4.37) and (4.38), where the correlation factors matrix is computed via (4.36) based on the factor estimation from (4.41).
<i>R</i> -D STE Color	The <i>R</i> -D Standard Tensor-ESPRIT algorithm is applied without prewhitening.
<i>R</i> -D STE MP E	The matrix prewhitening is calculated according to (4.37) and (4.38) with the matrix-based estimate of the correlation matrix from (4.35).
<i>R</i> -D STE S-GSVD I	The proposed multi-dimensional schemes with correlation factors estimation (4.41), and we use the S-GSVD prewhitening based on (4.48), (4.49) and (4.53).
<i>R</i> -D STE S-GSVD II	The proposed multi-dimensional schemes with correlation factors estimation (4.41), and we use the S-GSVD prewhitening based on (4.48) and (4.49). However, we include $R - 1$ SVDs at the end of the S-GSVD in order to improve the estimation, and (4.54) should be computed.
<i>R</i> -D STE TP CI	The correlation factors are known, and we apply (4.42) and (4.46) to perform the multi-dimensional prewhitening.
G-HOSVD	The generalized HOSVD-based low-rank approximation is applied in (4.34).

Tab. 4.1: Notation of the legends used in all figures, where  $R = 3$  for Fig. 4.11 and  $R = 2$  for the other figures.

ESPRIT [44], we compute the total RMSE of the estimated spatial frequencies  $\widehat{\mu}_i$  as follows

$$\text{RMSE} = \sqrt{\text{E} \left\{ \sum_{r=1}^R \sum_{i=1}^d \left( \widehat{\mu}_i^{(r)} - \mu_i^{(r)} \right)^2 \right\}}. \quad (4.56)$$

Initially let us describe the notation of the legends used in all figures. The prefix *R*-D STE stands for *R*-D Standard Tensor-ESPRIT [44], and SE stands for Standard ESPRIT. In Table 4.1, the acronym of each applied technique is shown. Moreover, for  $N_l \geq 10^2$ , the correlation factors estimation based on (4.41) gives an estimate, which is very close to the true correlation factors used in all figures. The suffix in Table 4.1 MP stands for matrix prewhitening, since the prewhitening matrix is computed without taking into account the Kronecker structure as shown in (4.35), while TP stands for the tensor prewhitening, where the correlation factor matrices  $\mathbf{L}_i$  are computed. Finally, the suffices E and CI mean estimated and correlation information, respectively. In the first one the prewhitening factor matrices is computed by using noise samples, while in the second the prewhitening factor matrices are considered known. For the G-HOSVD, S-GSVD I and the S-GSVD II, samples with the absence of signal components are used in the schemes.



In Figure 4.9, 2-D SE Color and 2-D STE Color present almost the same performance, since for the colored noise the major part of the noise is concentrated in the signal eigenvectors, and consequently the HOSVD-based low-rank approximation does not lead to a significant denoising. For this reason, when a prewhitening scheme is applied, it is expected that the noise gets almost equally distributed over all eigenvectors. Therefore, since the prewhitening takes the multi-dimensional structure into account, the estimation is improved much more. The 2-D SE MP E and 2-D SE TP E give a similar performance, since the 2-D SE does not take into account the multi-dimensional structure. The 2-D STE MP E has a limited accuracy of approximately  $10^{-2}$ , which is caused by the estimation of the prewhitening matrix which does not take into account the Kronecker structure. 2-D STE S-GSVD I has a performance close to the optimal, however it is slightly worse, since the prewhitening is applied sequentially. 2-D STE S-GSVD II achieves the optimal performance, since it is on top of 2-D STE TP CI, which considers the correlation factors as being known.

In Figure 4.10, the number of sources is increased from  $d = 3$  to  $d = 4$ , and in general all the curves get closer to each other, which means that the gain obtained by using the prewhitening is reduced. Moreover, for the matrix-based approach 2-D STE MP E, the limited accuracy is increased to  $2 \cdot 10^{-2}$ . Therefore, increasing the number of sources, the performance of the matrix-based approach gets drastically worse.

In Figure 4.11, the number of dimensions has been increased, and the array size is changed to  $M_1 = 5$ ,  $M_2 = 5$ , and  $M_3 = 5$ . This requires an additional correlation factor  $\rho_3$ , which is set to 0.95. In general, we can observe that the gain caused by the multi-dimensional prewhitening is significantly increased, if we compare it to Fig. 4.10, for example. In addition, the matrix-based approach 3-D STE MP E has an even worse performance, since its limited accuracy is approximately  $9 \cdot 10^{-2}$ . The limited accuracy comes from the fact that the multi-dimensional structure of the colored noise is not taken into account in the matrix-based prewhitening scheme. Moreover, for such scenarios, the multi-dimensional prewhitening enhances the accuracy significantly even for high SNR levels, while for scenarios in Fig. 4.10 for high SNR levels the improvement caused by the multi-dimensional prewhitening vs. no prewhitening is very small.

In Figure 4.12, the SNR is fixed to 30 dB. For the case that no prewhitening is applied, i.e., 2-D STE Color, the RMSE should be constant. Since we consider a high SNR regime, the matrix-based prewhitening 2-D STE MP E has a limited accuracy for each  $N_l$ . Moreover, even for  $N_l$  very large, the accuracy is much worse than applying no prewhitening. Note also that the matrix-based prewhitening for  $N_l$  lower than  $M_1 \cdot M_2 = 25$  leads to a rank-deficient estimate of  $\mathbf{L}_{\text{mtx}}$ . Therefore, for the simulations  $\mathbf{L}_{\text{mtx}}^{-1}$  in (4.37) is replaced by  $\mathbf{L}_{\text{mtx}}^+$ . On the other hand, since the SNR is very high, the multi-dimensional prewhitening schemes show no accuracy limitation even for a very small  $N_l$ , since they remain almost constant for  $N_l$  varying from  $10^2$  to  $2 \cdot 10^4$ .

In Figure 4.13, the same scenario as in Fig. 4.12 is considered, but the SNR is fixed to 0 dB. Since we consider a low SNR regime, the matrix-based prewhitening 2-D STE MP E

has an improvement in its estimate until  $N_l = 10^4$ , after this value no improvement is observed. Similarly to Fig. 4.12, the multi-dimensional prewhitening schemes show no accuracy limitation even for a very small  $N_l$ , since they remain almost constant for  $N_l$  varying from  $10^2$  to  $2 \cdot 10^4$ .

In Figure 4.14, a scenario with an intermediate SNR = 10 dB is selected, and the correlation levels  $\rho_1$  and  $\rho_2$  are equal to  $\rho$ , which varies from 0 to 0.999. Since the matrix-based prewhitening technique 2-D STE MP E has an accuracy limitation, then for high correlation levels, the performance is worse than the one using no prewhitening. The approaches using the correlation factors give a better estimate compared to the matrix-based prewhitening or no prewhitening cases, when the correlation levels are higher. Therefore, for intermediate SNRs the gains are substantial only for high correlation levels, i.e.,  $\rho > 0.6$ .

In Figure 4.15, we have the same scenario as in Fig. 4.14, except that instead of an intermediate SNR level, we have a low SNR level of -5 dB. For such scenarios, even for low correlations levels  $\rho > 0.3$ , there is a significant improvement.

In Figure 4.16, the array size varies, and as expected the RMSE becomes smaller. Note that there is always a significant improvement of the techniques based on correlation factors compared to the matrix-based prewhitening and no prewhitening.

In Figure 4.17, the number of snapshots  $N$  increases, and as expected the RMSE becomes smaller for all approaches. Note that increasing the samples the gain obtained by the proposed multi-dimensional schemes is greater than the one obtained in 2-D STE Color and 2-D STE MP E.

In Figure 4.18, the G-HOSVD is included, and due to its definition we consider only one correlation factor in the first dimension. The correlation level is 0.9. The proposed multi-dimensional techniques outperform the G-HOSVD as shown in Fig. 4.18.

In Figure 4.19, we increase the number of sources from 2 to 4, and the G-HOSVD becomes even closer to the 2-D STE Color.

In Figure 4.20, the number of sources is increased from 4 to 5 and the G-HOSVD superposes the 2-D STE Color, as expected. Note that for  $d \geq M_i$ , there is no accuracy limitation for the 2-D STE MP E. The performance of the techniques with factors estimation, e.g., 2-D STE TP S-GSVD II, is the same as the one in which the factors are assumed to be known, i.e., 2-D STE TP CI. Note that the proposed multi-dimensional techniques outperforms all the other techniques.

In Figures 4.18, 4.19, and 4.20,  $N_l$  for the generalized HOSVD is set to  $10^2$ , since for larger  $N_l$ , the computational complexity of the generalized HOSVD can be very high. Note that the generalized HOSVD is not included in Figs. 4.9- 4.17, since it is not possible to prewhiten more than one dimension using the generalized HOSVD. Therefore, the inclusion of the generalized HOSVD for such scenarios would not be a fair comparison.

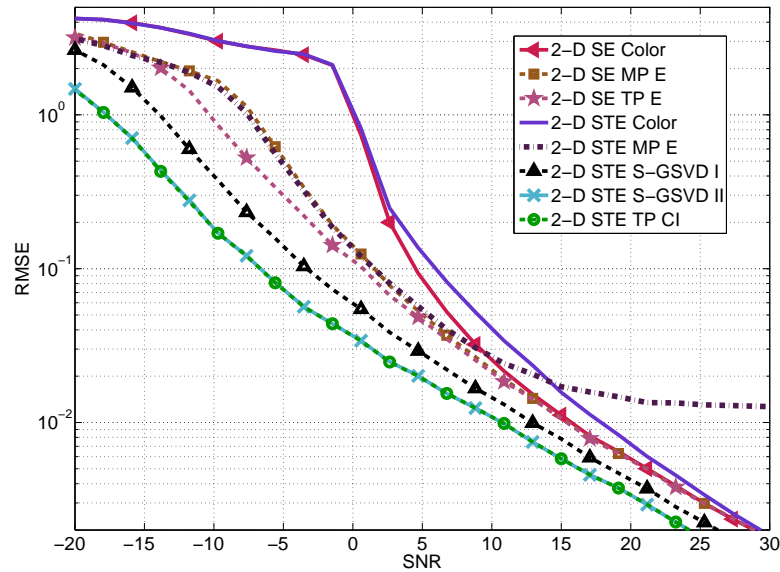


Fig. 4.9: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard ESPRIT and 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N$  is set to 20 and the number of sources  $d = 3$ . The correlation levels are  $\rho_1 = 0.9$  for  $\mathbf{L}_1$  and  $\rho_2 = 0.95$  for  $\mathbf{L}_2$ . The number of snapshots  $N_l$  is set to  $10^4$ .

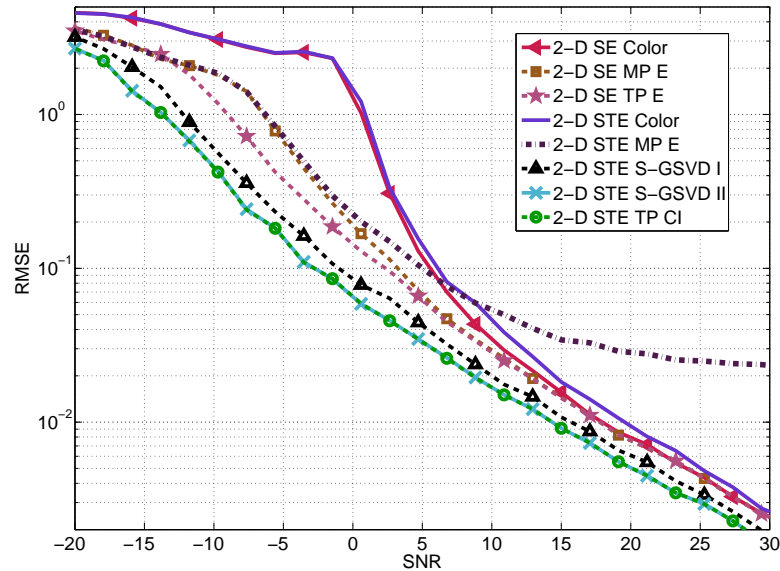


Fig. 4.10: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard ESPRIT and 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N$  is set to 20. The correlation levels are  $\rho_1 = 0.9$  for  $\mathbf{L}_1$  and  $\rho_2 = 0.95$  for  $\mathbf{L}_2$ . The number of snapshots  $N_l$  is set to  $10^4$ . In contrast to Fig. 4.9, the number of sources is set to  $d = 4$ .

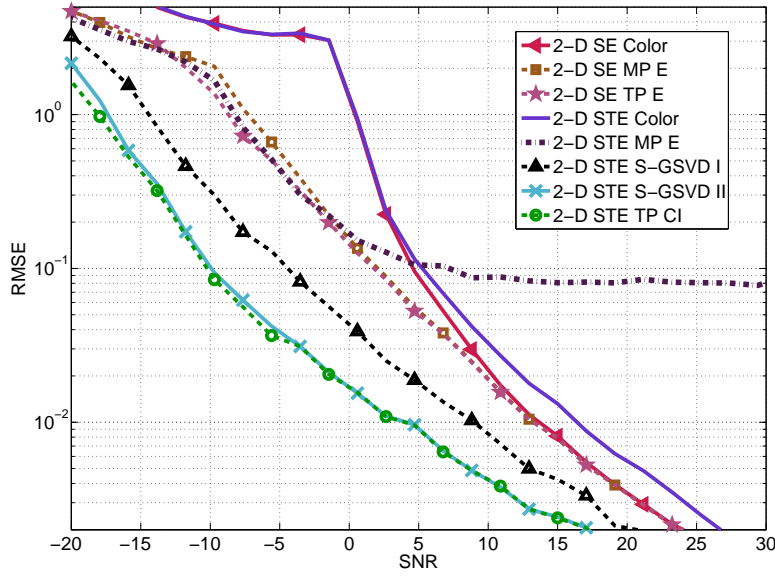


Fig. 4.11: Comparing the different types of prewhitening schemes in conjunction with 3-D Standard ESPRIT and 3-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here the number of snapshots  $N$  is set to 20 and  $d = 4$ . The number of snapshots  $N_l$  is set to  $10^4$ . The array size is set to  $M_1 = 5$ ,  $M_2 = 5$  and  $M_3 = 5$ , and the correlation levels are  $\rho_1 = 0.9$  for  $\mathbf{L}_1$ ,  $\rho_2 = 0.95$  for  $\mathbf{L}_2$ , and  $\rho_3 = 0.95$  for  $\mathbf{L}_3$ .

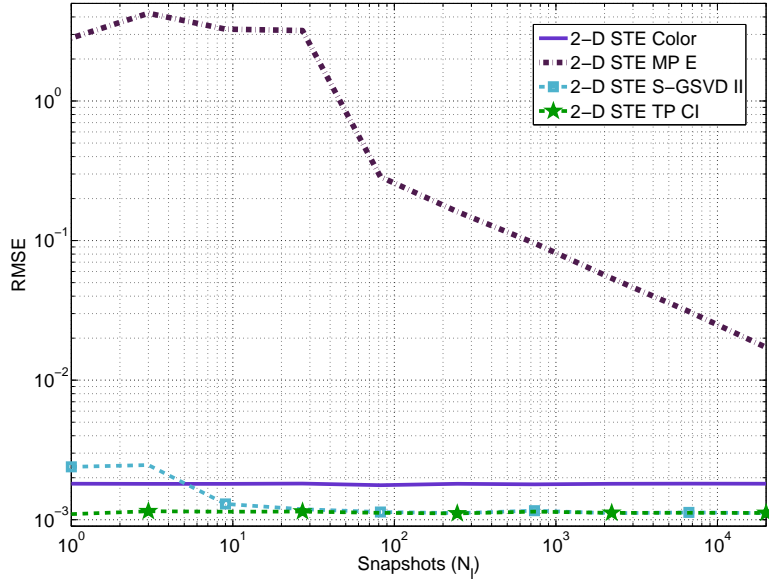


Fig. 4.12: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N$  is set to 20. The correlation levels are  $\rho_1 = 0.9$  for  $\mathbf{L}_1$  and  $\rho_2 = 0.95$  for  $\mathbf{L}_2$ . The number of sources is set to  $d = 4$ . The SNR is fixed to 30 dB, and the snapshots  $N_l$  vary from 1 to  $2 \cdot 10^4$  samples.

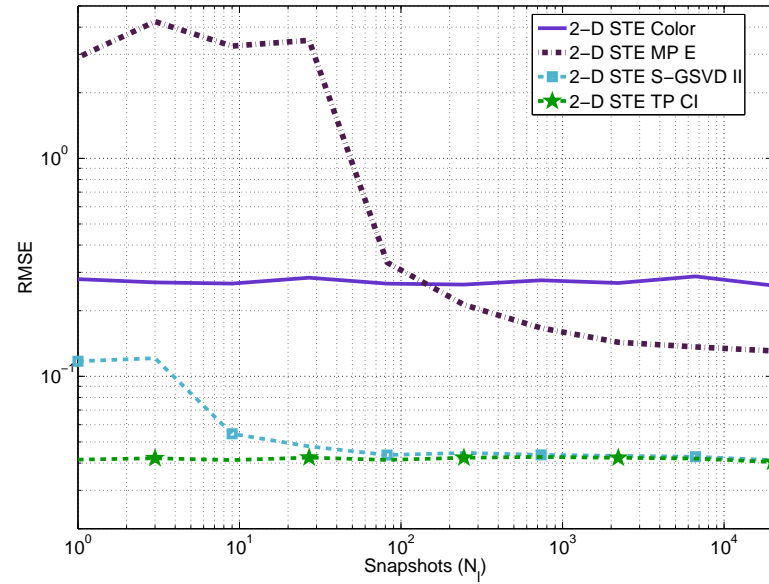


Fig. 4.13: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of sources is set to  $d = 4$ . The SNR is fixed to 0 dB, and the snapshots  $N_l$  vary from 1 to  $2 \cdot 10^4$  samples.

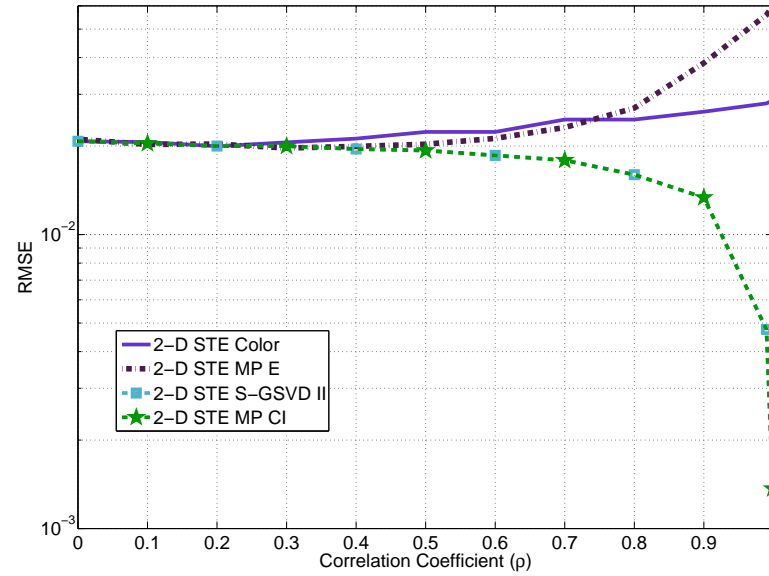


Fig. 4.14: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N$  is set to 20. The number of snapshots  $N_l$  is set to  $10^4$ . The number of sources is equal to 4. The SNR is fixed to 10 dB, and the correlation levels  $\rho_1$  and  $\rho_2$  are equal to  $\rho$ , where  $0 \leq \rho < 1$ .

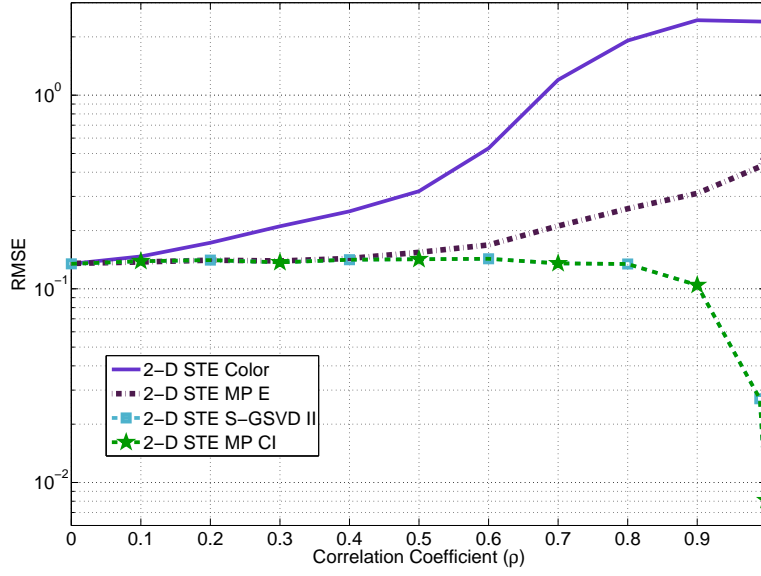


Fig. 4.15: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N$  is set to 20. The number of snapshots  $N_l$  is set to  $10^4$ . The number of sources is equal to 4. The SNR is fixed to  $-5$  dB, and the correlation levels  $\rho_1$  and  $\rho_2$  are equal to  $\rho$ , where  $0 \leq \rho < 1$ .

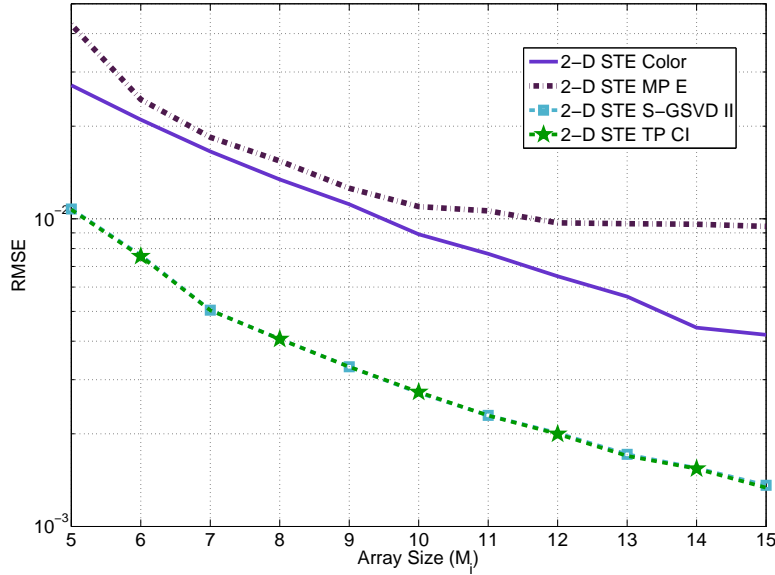


Fig. 4.16: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. The number of snapshots  $N$  is set to 20. The correlation levels are  $\rho_1 = 0.9$  for  $\mathbf{L}_1$  and  $\rho_2 = 0.95$  for  $\mathbf{L}_2$ . The number of snapshots  $N_l$  is set to  $10^4$ . The number of sources is equal to 4. However, the SNR is fixed to 10 dB, and the array size  $M_1$  and  $M_2$  equal to  $M_i$ , which varies from 5 to 15.

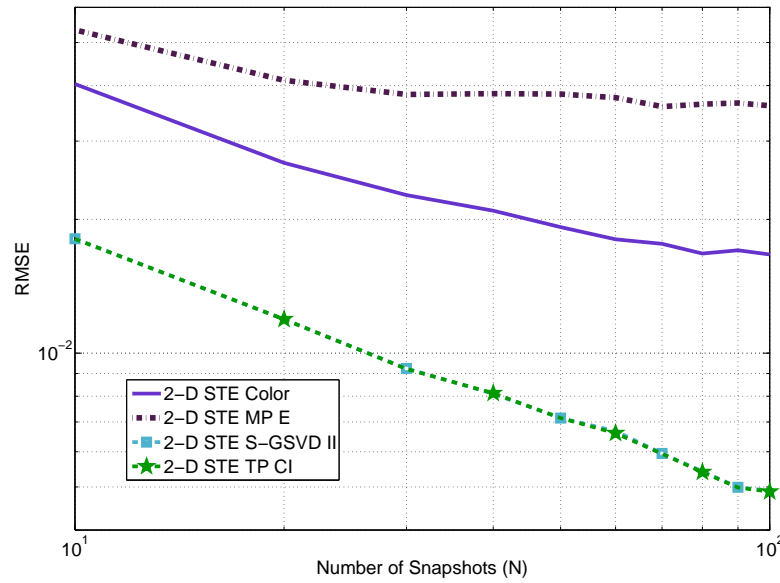


Fig. 4.17: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The correlation levels are  $\rho_1 = 0.9$  for  $\mathbf{L}_1$  and  $\rho_2 = 0.95$  for  $\mathbf{L}_2$ . The number of snapshots  $N_l$  is set to  $10^4$ . The number of sources is equal to 4. The SNR is fixed to 10 dB, and the number of snapshots  $N$  varies from 10 to 100.

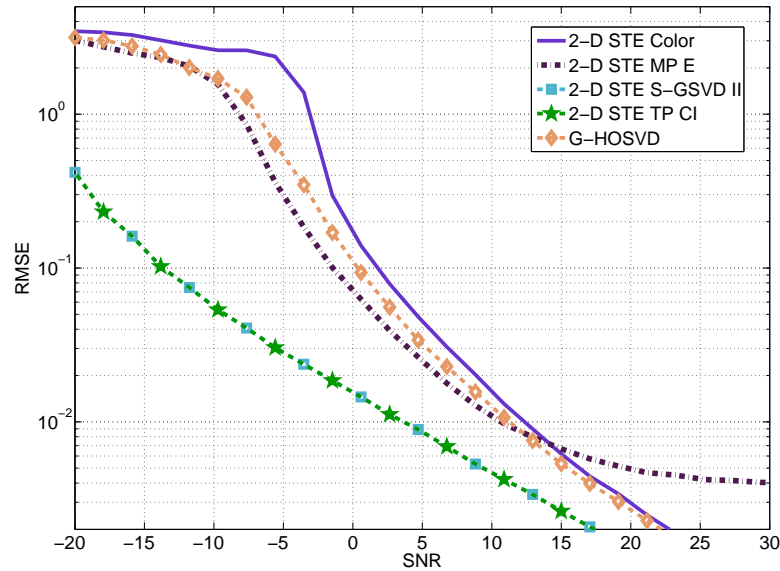


Fig. 4.18: Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N_l$  is set to  $10^4$  and the number of snapshots  $N$  is set to 20. The number of sources  $d$  is equal to 2, and we set  $\rho_1 = 0.9$  and  $\rho_2 = 0$ .

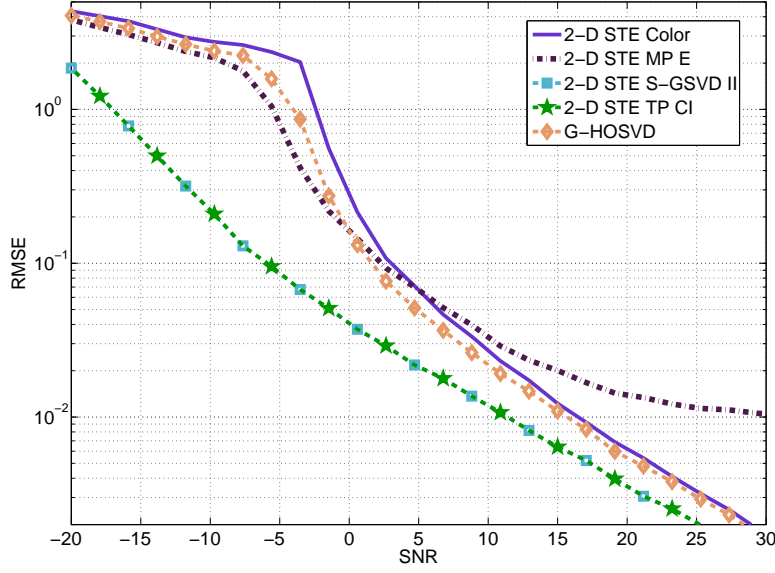


Fig. 4.19: Comparing the different types of whitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N_l$  is set to  $10^4$  and the number of snapshots  $N$  is set to 20. We set  $\rho_1 = 0.9$  and  $\rho_2 = 0$ . The number of sources  $d$  is set to 4.

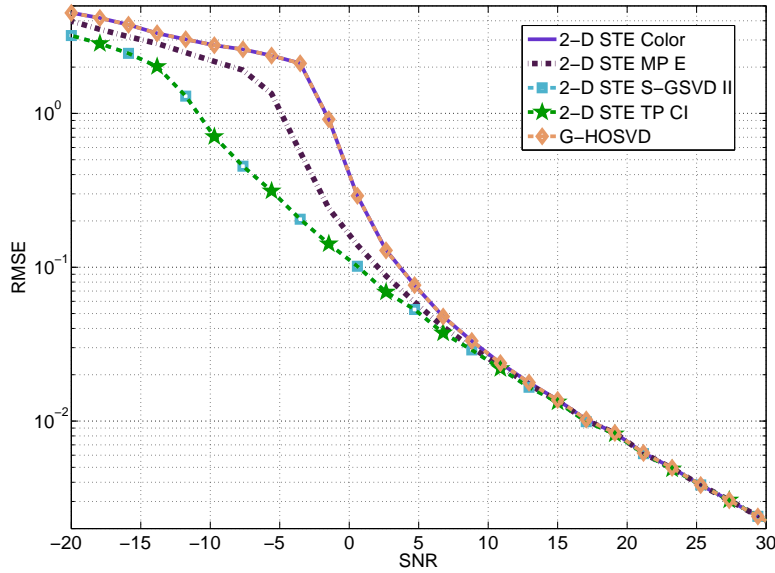


Fig. 4.20: Comparing the different types of whitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size  $M_1 = 5$  and  $M_2 = 5$ . The number of snapshots  $N_l$  is set to  $10^4$  and the number of snapshots  $N$  is set to 20. We set  $\rho_1 = 0.9$  and  $\rho_2 = 0$ . The number of sources  $d$  is set to 5.



### 4.2.5 Main Results of the Section

In this section, we have proposed the S-GSVD based prewhitening, where we take into account the Kronecker tensor structure of the colored noise. The first benefit of the S-GSVD is the superior estimation accuracy, since the S-GSVD is based on the estimation of the prewhitening correlation factors proposed in Subsection 4.2.3.1. The proposed prewhitening correlation factors can be integrated separately to other prewhitening approaches, e.g., in Subsection 4.2.2.3 via (4.36), in Subsection 4.2.3.2, and in Subsection 4.2.3.3, and for all of them the same accuracy on the estimation is obtained. In addition, besides the improved estimation accuracy of the S-GSVD based prewhitening approach, our proposed scheme also has a lower computational complexity version, which requires only the computation of  $n$ -mode GSVDs.

### 4.3 Iterative Multi-dimensional Prewhitening

As shown in the previous section, the subspace-based parameter estimation schemes can be significantly improved via the S-GSVD based prewhitening technique in environments with multi-dimensional colored noise with Kronecker structure, which is found in EEG applications [52] as well as in MIMO applications [76].

In general, for EEG and MIMO applications, noise samples can only be collected in the presence of the desired signal components. Therefore, in order to improve the subspace-based parameter estimation in such problems, we propose a multi-dimensional prewhitening scheme called Iterative S-GSVD (I-S-GSVD) [5]. The I-S-GSVD has the low complexity of the S-GSVD and at the same time, it provides a performance close to the one obtained in the case that the second order statistics of the noise are estimated from samples where the desired signal components are absent.

#### 4.3.1 Goal and Problem

Here we have the same data model as in Subsection 4.2.1. However, we state our problem in a different way: given the measurement tensor  $\mathcal{X}$ , the model order  $d$ , we desire to estimate all the spatial frequencies  $\mu_i^{(r)}$  taking into account the noise structure, which in total are  $R \cdot d$  variables. Moreover, together with the the spatial frequencies  $\mu_i^{(r)}$ , we also estimate  $\mathbf{S}$ , and the noise factors  $\mathbf{L}_i$  for  $i = 1, \dots, R$ . The information about the noise variance can also be used to compute the Cramér-Rao Lower Bound.

#### 4.3.2 Previous Approach

As shown in Section 4.2, the matrix-based prewhitening scheme [89, 47] degrades significantly the parameter estimation in the case where only a small number of samples is available in the absence of the desired signal components. Also as shown in [4], the S-GSVD based prewhitening scheme can handle such a case successfully. Therefore, we do not include an iterative version of the matrix-based prewhitening scheme here.

#### 4.3.3 Proposed Approach: Iterative S-GSVD (I-S-GSVD)

Inspired by [69], where a deterministic version of the Expectation Maximization (EM) algorithm is proposed, we propose the *iterative S-GSVD (I-S-GSVD)*. Similarly to the deterministic EM [69], a certain structure of the data is assumed in our scheme. Our proposed multi-dimensional prewhitening scheme is applied to improve the estimation of the spatial frequencies. Before starting the iterative algorithm, first let us define  $K$  and  $k$  as the maximum number of iterations in the iterative approach and the iteration index, respectively.

0) Set  $k = 1$ .

---

- 1) Given  $\mathcal{X}$  compute the subspace tensor  $\mathcal{U}^{[s]} \in \mathbb{C}^{M_1 \times \dots \times M_R \times d}$  via a truncated HOSVD [29]-based low-rank approximation according to [44].
- 2) Given  $\mathcal{U}^{[s]}$  estimate the spatial frequencies  $\widehat{\mu}_{k,i}^{(r)}$  for  $r = 1, \dots, R$  and for  $i = 1, \dots, d$ , for instance, via  $R$ -D Standard-Tensor ESPRIT ( $R$ -D STE) [44]. Note that  $\widehat{\mu}_{k,i}^{(r)}$  is the estimate of  $\mu_i^{(r)}$  at the  $k$ -th iteration.
- 3) From  $\widehat{\mu}_{k,i}^{(r)}$ , compute  $\widehat{\mathcal{A}}$  according to the structure of our data model in (4.31). Using  $\mathcal{X}$  and  $\widehat{\mathcal{A}}$ , calculate  $\widehat{\mathcal{S}} = \left( [\mathcal{X}]_{(R+1)} \cdot [\widehat{\mathcal{A}}]_{(R+1)}^+ \right)^T$ .
- 4) Given  $\widehat{\mathcal{A}}$  and  $\widehat{\mathcal{S}}$ , the estimate of the noise tensor  $\mathcal{N}^{(c)}$  can be compute via  $\mathcal{N}^{(c)} = \mathcal{X} - \widehat{\mathcal{A}} \times_{R+1} \widehat{\mathcal{S}}^T$ .
- 5) From  $\mathcal{N}^{(c)}$ , we can estimate via Prewhitening Correlation Factors Estimation (PCFE) in Subsubsection 4.2.3.1 the correlation factors  $\widehat{\mathcal{L}}_r$  according to [4].
- 6) Increment  $k$ .
- 7) According to Subsubsection 4.2.3.3, compute the subspace tensor  $\mathcal{U}^{[s]} \in \mathbb{C}^{M_1 \times \dots \times M_R \times d}$  via the S-GSVD II of  $\mathcal{X}$  and  $\widehat{\mathcal{L}}_r$ , for  $r = 1, \dots, R$ .
- 8) Given  $\mathcal{U}^{[s]}$ , estimate the spatial frequencies  $\widehat{\mu}_{k,i}^{(r)}$  for  $r = 1, \dots, R$  and for  $i = 1, \dots, d$  for instance via  $R$ -D Standard-Tensor ESPRIT ( $R$ -D STE) [44].
- 9) Compute the variation of the spatial frequencies estimates  $E_\mu^{(k)} = \sqrt{\sum_{r=1}^R \sum_{i=1}^d \left( \widehat{\mu}_{k,i}^{(r)} \right)^2 - \left( \widehat{\mu}_{k-1,i}^{(r)} \right)^2}$ . If the variation  $E_\mu^{(k)}$  is zero or smaller than a certain defined threshold, then the algorithm stops. Otherwise, increment  $k$  and go to item 3).

#### 4.3.4 Evaluation of Approaches

Here we generate our samples based on the data model of (4.31), where the spatial frequencies  $\mu_i^{(r)}$  are drawn from a uniform distribution in  $[-\pi, \pi]$ . The source symbols are zero mean i.i.d. circularly symmetric complex Gaussian distributed with power equal to  $\sigma_s^2$  for all the sources. The SNR at the receiver is defined as  $\text{SNR} = 10 \cdot \log_{10} \left( \frac{\sigma_s^2}{\sigma_n^2} \right)$ , where  $\sigma_n^2$  is the variance of the elements of the white noise tensor  $\mathcal{N}$  in (2.43).

In order to compare the performance of the proposed prewhitening scheme combined with the  $R$ -D Standard-Tensor ESPRIT ( $R$ -D STE), we compute the total RMSE of the estimated spatial frequencies  $\widehat{\mu}_i^{(r)}$ .

In Fig. 4.21,  $R$ -D STE w/o PWT means  $R$ -D STE without any prewhitening.  $R$ -D STE I-S-GSVD stands for the  $R$ -D STE combined with S-GSVD II [4], while the unknown noise is estimated deterministically with the iterative approach proposed in Subsection 4.3.3.  $R$ -D

STE S-GSVD stands for the  $R$ -D STE applied together with the S-GSVD II [4], and assuming that  $N$  samples in the absence of signal components are available.

Also in Fig. 4.21, there is almost no difference of performance between the I-S-GSVD without any information about the noise statistics and the S-GSVD which requires second order statistics of the noise. In general, this is valid for high and intermediate SNR regimes. Similarly to the S-GSVD, the proposed I-S-GSVD outperforms significantly the estimation without prewhitening for intermediate and high noise correlation levels.

From Fig. 4.21 to Fig. 4.22, we reduce the SNR from 20 dB to 5 dB. In this case, the improvement caused by the S-I-GSVD is greater, since the noise degradation is higher. For instance, at  $\rho_i = 0.9$ , the RMSE of the S-I-GSVD is 10 times smaller than the STE without prewhitening in Fig. 4.21, while in Fig. 4.22 it is 20 times.

In Fig. 4.23, we consider the same scenario as in Fig. 4.21, but we evaluate the SNR versus number of iterations. For  $k = 2$  iterations, a performance very close to the converged RMSE is obtained. Therefore, in Fig. 4.21, we have considered  $k = 2$ . By reducing the SNR from 20 dB to 5 dB as from Fig. 4.22 to Fig. 4.24, the number of iterations should be increased to  $k = 3$  to have a performance close to the converged RMSE. Moreover, while in Fig. 4.23, the curve of the I-S-GSVD superposes the S-GSVD, in Fig. 4.24, there is a gap. Therefore, by reducing the SNR, the performance of the I-S-GSVD gets closer to the performance of using no prewhitening scheme. In addition, the proposed I-S-GSVD outperforms significantly the estimation without prewhitening for intermediate and high noise correlation levels as exemplified in Fig. 4.25, where the total RMSE is computed by varying the SNR from -35 dB to 40 dB.

In Figs. 4.26 and 4.27, we reduce the number of sources from 3 to 2 compared to Fig. 4.24. In this case, the gain of the I-S-GSVD with respect to the case without prewhitening is more significant. Such a behavior is also seen in Section 4.2 for the S-GSVD.

In Figs. 4.28 and 4.29, we increase the array size  $M_i$ , where  $i = 1, \dots, 5$ , and the number of snapshots compared to the previous figures from 5 to 7. In particular, comparing Figs. 4.24 and 4.29, where the only difference is the increase of the tensor size, the gap between the I-S-GSVD and the S-GSVD remains.

### 4.3.5 Main Results of the Section

We have proposed the iterative S-GSVD (I-S-GSVD), which can be applied in cases where the knowledge of the noise covariance information is not available. The I-S-GSVD even without knowledge of the noise covariance information is able to improve the subspace based parameter estimation very close to estimation obtained using the S-GSVD with knowledge of the noise covariance information.

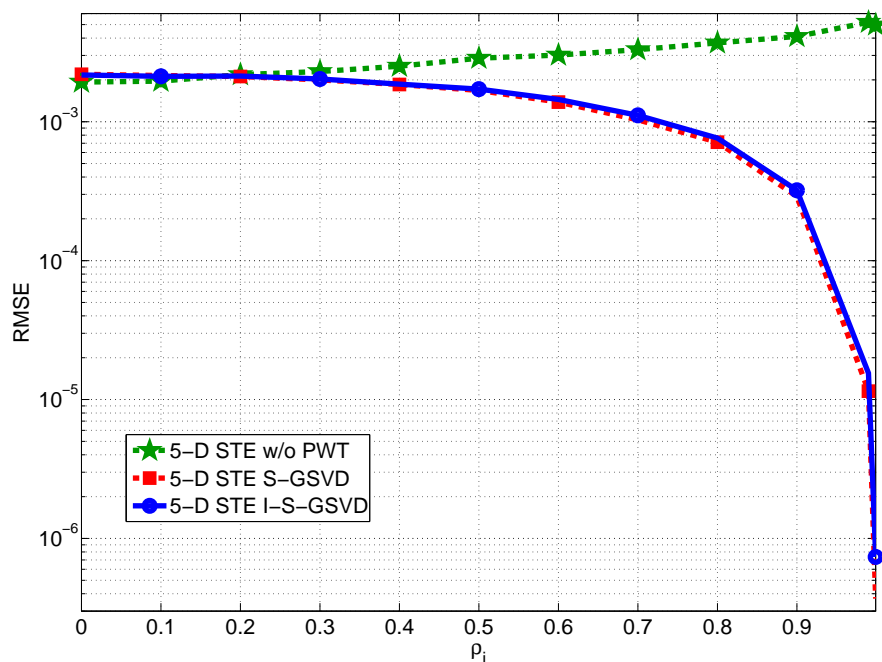


Fig. 4.21: Total RMSE of the 5 estimated spatial frequencies versus  $\rho_i$ , for  $i = 1, \dots, 5$ . The SNR and the number of sources  $d$  are set to 20 dB and 3, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ . The I-S-GSVD is stopped at the second iteration, i.e.,  $k = 2$ .

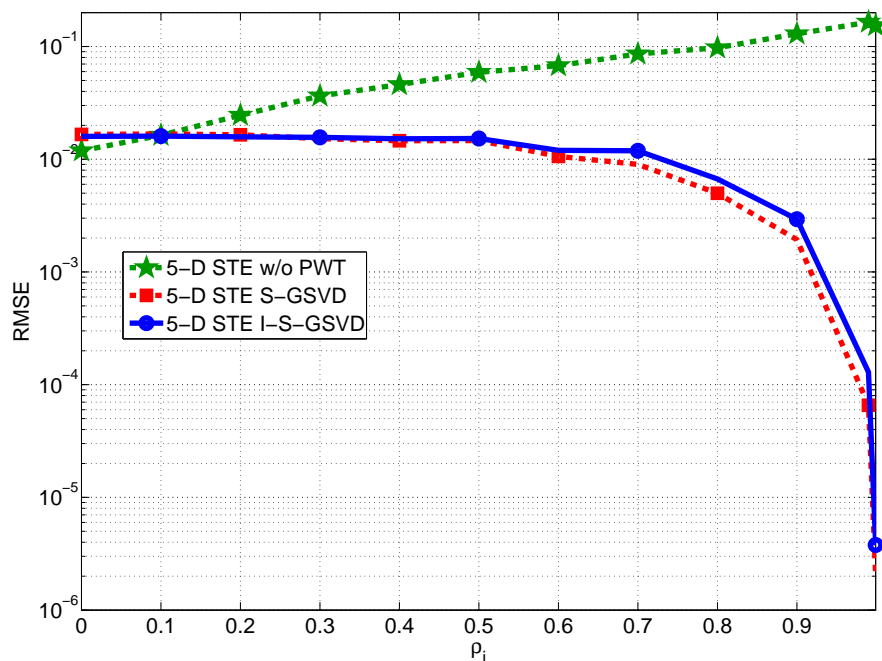


Fig. 4.22: Total RMSE of the 5 estimated spatial frequencies versus  $\rho_i$ , for  $i = 1, \dots, 5$ . The SNR and the number of sources  $d$  are set to 5 dB and 3, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ . The I-S-GSVD is stopped at the third iteration, i.e.,  $k = 3$ .

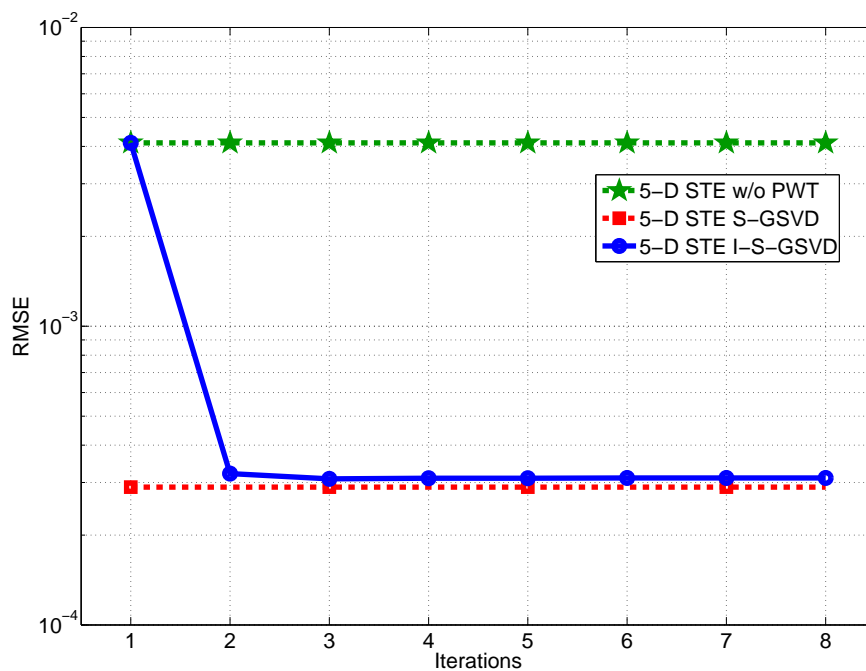


Fig. 4.23: Total RMSE of the 5 estimated spatial frequencies versus the number of iterations  $k$ . The SNR and the number of sources  $d$  are set to 20 dB and 3, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ .

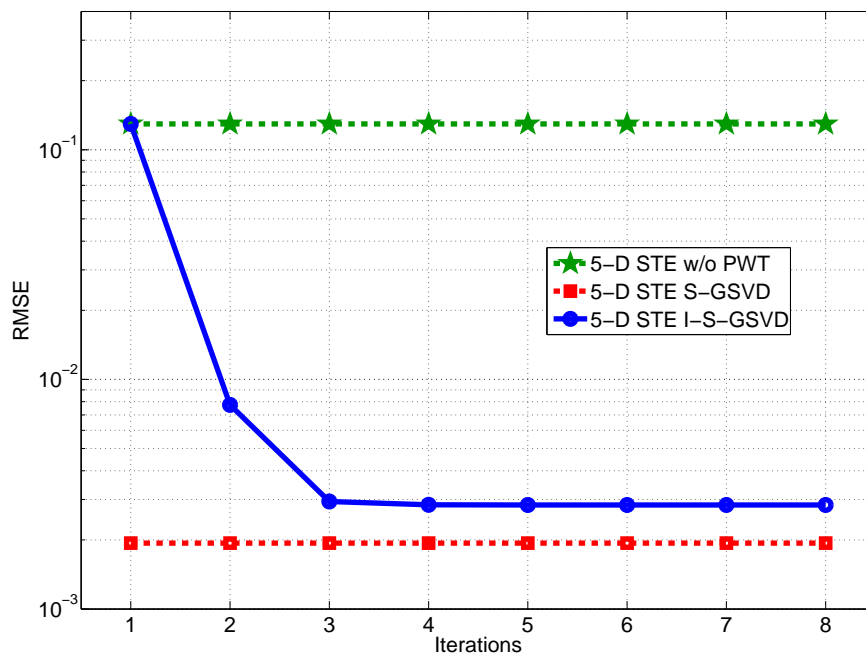


Fig. 4.24: Total RMSE of the 5 estimated spatial frequencies versus the number of iterations  $k$ . The SNR and the number of sources  $d$  are set to 5 dB and 3, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ .

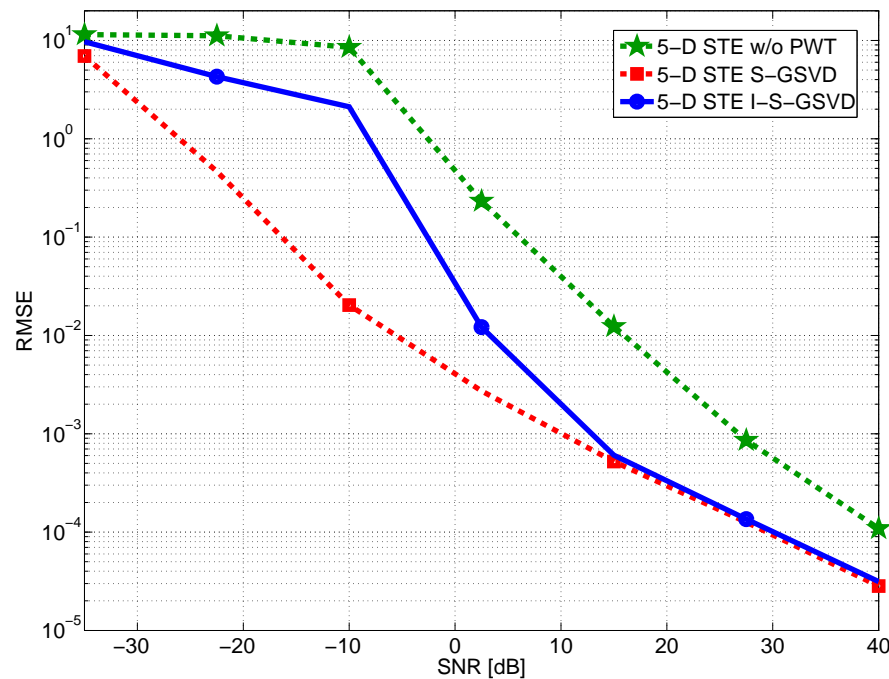


Fig. 4.25: Total RMSE of the 5 estimated spatial frequencies versus the SNR is depicted. The number of sources  $d$  and the correlation coefficient  $\rho_i$  are set to 3 and to 0.9, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ . The I-S-GSVD is stopped after the third iteration, i.e.,  $K = 3$ .

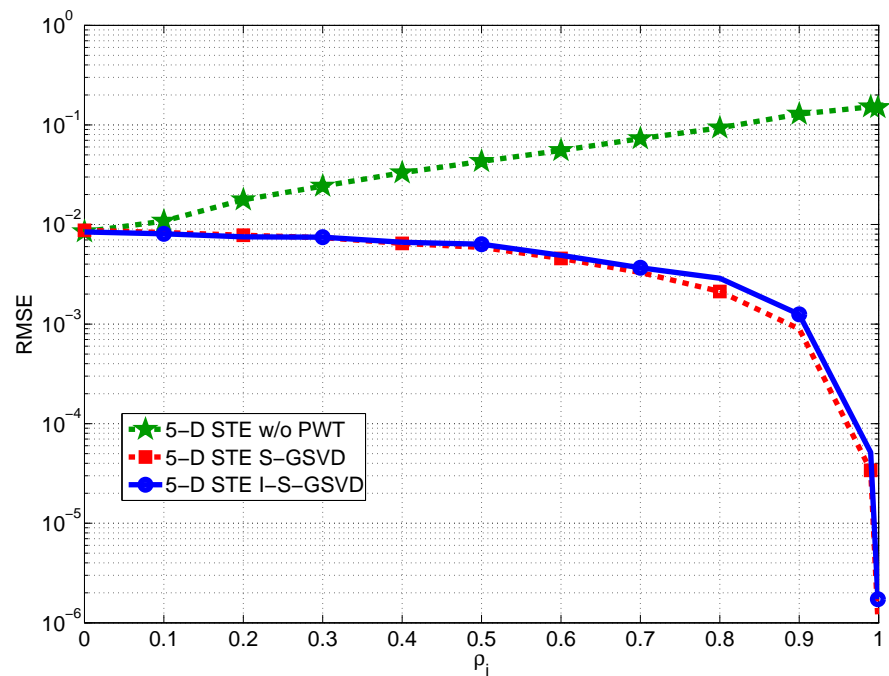


Fig. 4.26: Total RMSE of the 5 estimated spatial frequencies versus  $\rho_i$  for  $i = 1, \dots, 5$  is depicted. The SNR and the number of sources  $d$  are set to 5 dB and 2, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ . The I-S-GSVD is stopped after the third iteration, i.e.,  $K = 3$ .

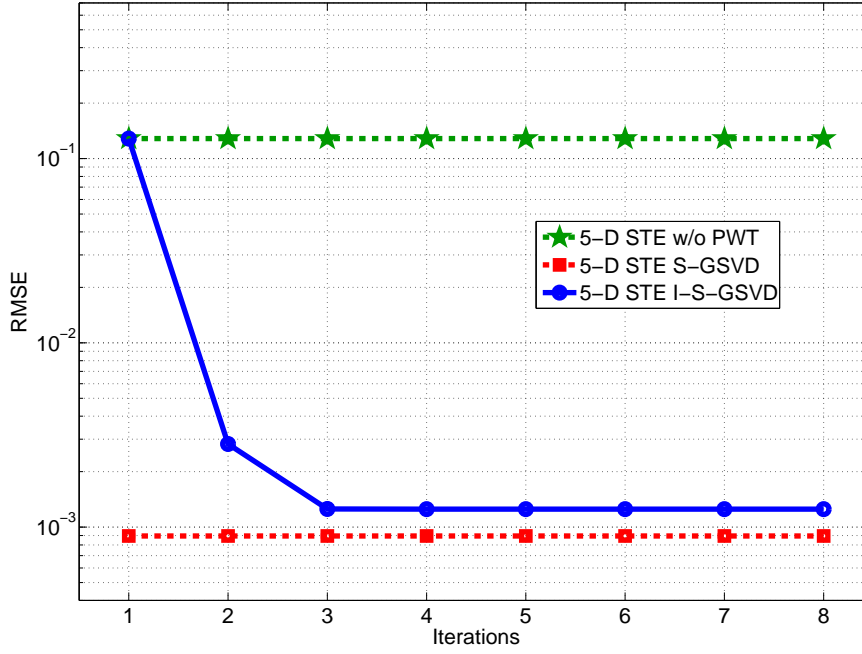


Fig. 4.27: Total RMSE of the 5 estimated spatial frequencies versus the number of iterations  $k$  is depicted. The SNR and the number of sources  $d$  are set to 5 dB and 2, respectively. The array size is  $M_i = 5$ , for  $i = 1, \dots, 5$ , and  $N = 5$ . The noise correlation  $\rho_i$  is equal to 0.9. The same scenario as in Fig. 4.26 is considered.

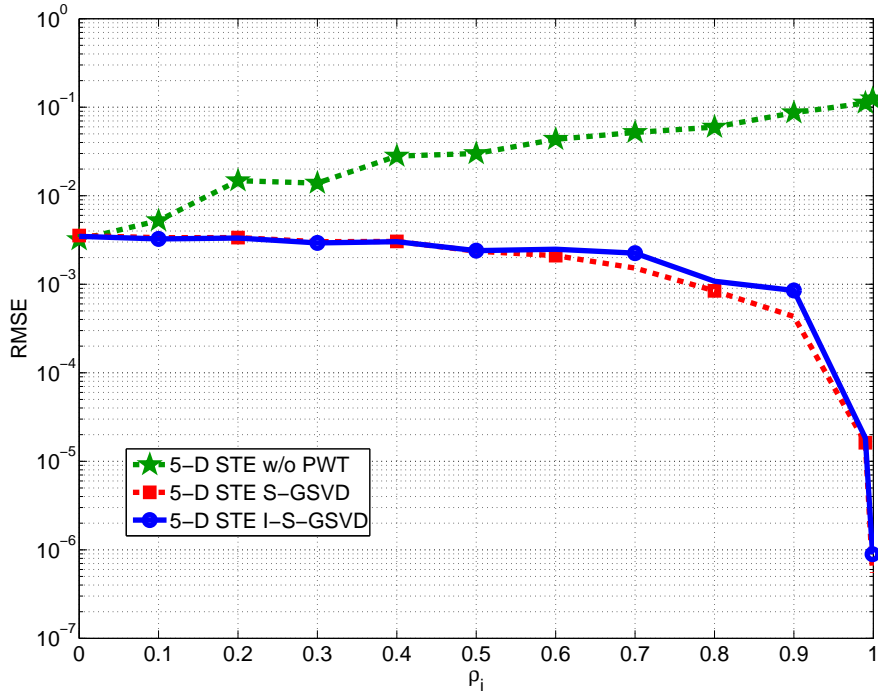


Fig. 4.28: Total RMSE of the 5 estimated spatial frequencies versus  $\rho_i$  for  $i = 1, \dots, 5$  is depicted. The SNR and the number of sources  $d$  are set to 5 dB and 3, respectively. The array size is  $M_i = 7$ , for  $i = 1, \dots, 5$ , and  $N = 7$ . The I-S-GSVD is stopped after the third iteration, i.e.,  $K = 3$ .



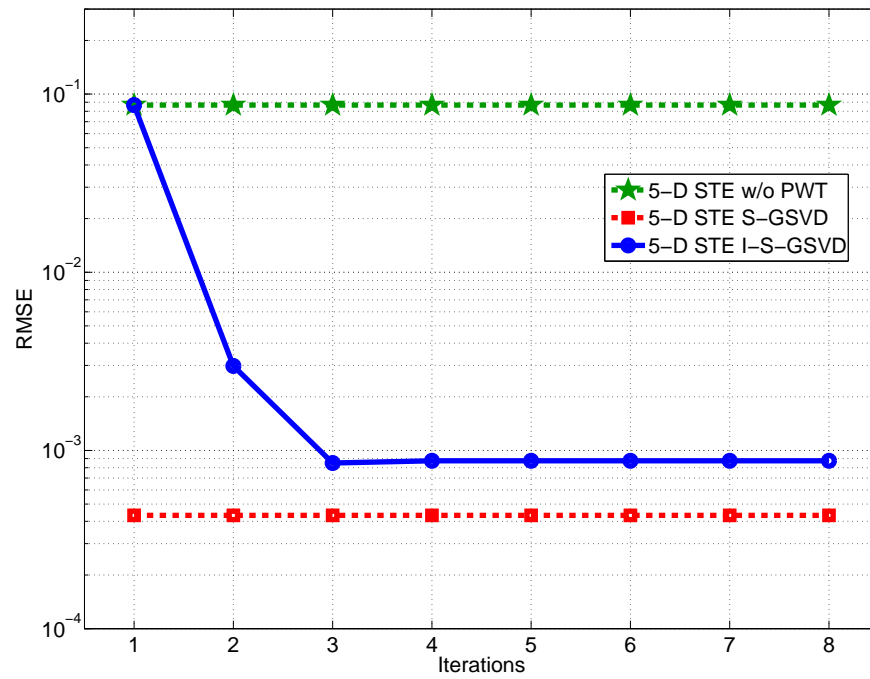


Fig. 4.29: Total RMSE of the 5 estimated spatial frequencies versus the number of iterations  $k$  is depicted. The SNR and the number of sources  $d$  are set to 5 dB and 3, respectively. The array size is  $M_i = 7$ , for  $i = 1, \dots, 5$ , and  $N = 7$ . The noise correlation  $\rho_i$  is equal to 0.9. The same scenario as in Fig. 4.28 is considered.

#### 4.4 Main Results of the Chapter

In this chapter, we have proposed matrix-based and tensor-based subspace prewhitening schemes, where we take into account some structure of the colored noise in order to improve the subspace prewhitening. Next we can summarize our important results according to the type of colored noise.

First, for matrix-based colored noise with a certain structure, the deterministic prewhitening scheme proposed by us in Subsection 4.1.3 should be applied. Particularly, for severe colored noise scenarios, the deterministic prewhitening is very appealing, since it gains is inversely proportional to level of the noise correlation.

For multi-dimensional scenarios, where the multi-dimensional colored noise has a Kronecker structure, we have proposed the Sequential GSVD (S-GSVD) in Subsubsection 4.2.3.3. The S-GSVD has both a very a much lower computational complexity and a significant improvement in the estimation of the spatial frequency if compared to the schemes proposed in the literature.

For prewhitening schemes, noise samples without signal components are required. However, we propose to incorporate the S-GSVD to an iterative algorithm, which gives us the iterative S-GSVD (I-S-GSVD). By using the I-S-GSVD, it is possible to obtain jointly a very good estimation of the signal parameters and of the noise covariance. The performance of the

I-S-GSVD without the knowledge of the noise covariance information is comparable to the S-GSVD with the noise covariance information.

## 5. CONCLUDING REMARKS

One of the most challenging tasks involving model order selection (MOS) is certainly to find in the literature the technique which gives the highest Probability of correct Detection (PoD). The first reason is that in the literature most of the techniques are compared to AIC and MDL only, which means that all the state-of-the-art MOS techniques should be implemented and compared. Second is that the model order selection problem is encountered in for several scientific fields as mentioned in Chapter 2. Third that when a MOS technique is proposed in the literature, usually its performance is better only for a certain types of scenario, i.e., with certain number of sensors and snapshots. Therefore, varying the scenario, as shown in Chapter 2, one MOS technique is outperformed by another one.

Instead of only investigating a best technique for a certain array configuration, we focused on a more general objective, which is proposing a model order technique with highest PoD when the data model is contaminated by complex-valued white or colored Gaussian noise. In this sense, our first main contribution is our proposed modified Exponential Fitting Test (M-EFT) [2, 1, 7], which outperforms all the MOS techniques found in the literature for scenarios with white Gaussian noise. Particularly, when the number of sensors is much greater than the number of snapshots, which is usual for scenarios of multi-dimensional applications, there is a huge gap between our M-EFT and second best MOS approach in the literature. In addition, for such multi-dimensional scenarios, the AIC and MDL should be modified to our proposed 1-D AIC and 1-D MDL [2, 1, 7]. Moreover, for scenarios, where severe colored noise is present, we have confirmed that the MOS scheme with best performance is RADOI [80].

Since multi-dimensional data are found in several scientific fields, we have extended our proposed M-EFT to its multi-dimensional form called *R-D EFT* [2, 1]. The *R-D EFT* presents a performance even better than the M-EFT. Therefore, for multi-dimensional scenarios with white Gaussian noise, the *R-D EFT* is the best approach to estimate the model order. For the case of colored noise, we have proposed the Closed-Form PARAFAC based Model Order Selection (CFP-MOS) scheme, which similarly to RADOI estimates model order even when the noise is severely correlated. Moreover, CFP-MOS has the same multi-dimensional gain as in our proposed multi-dimensional schemes *R-D EFT*, *R-D AIC*, and *R-D MDL* for white Gaussian noise. Therefore, for colored noise scenarios in multi-dimensional data, CFP-MOS achieves the highest PoD.

In Chapters 3 and 4, we concentrate on applications involving multi-dimensional estimation of spatial frequencies, which is a major topic in several signal processing applications such as seismology, radar, sonar and MIMO systems. Doubtless the contributions in Chapter 2 are

fundamental for these two chapters, since the estimation of the model order is a required step in the multi-dimensional parameter estimation techniques presented.

Our proposed closed-form PARAFAC based parameter estimation (CFP-PE) scheme are general, since the array is only required to be an Outer Product based Array (OPA) or a mixed array, which is partially an OPA and partially a Non-Outer Product based Array NOPA. Moreover, for realistic array, positioning errors are expected, and in contrast to  $R$ -D Standard Tensor-based ESPRIT, our CFP-PE is very robust to such scenarios, since the dimensions are decomposed only assuming the Kronecker structure. Finally, for scenarios generated with the IlmProp, where the narrowband assumption is violated, the CFP-PE, which is based on the HOSVD and is non-iterative, outperforms significantly the iterative MALS approach proposed in the literature.

In order to improve the performance of multi-dimensional estimation of spatial frequencies in scenarios with colored noise, we have investigated in Chapter 4 the one-dimensional and multi-dimensional prewhitening schemes. For scenarios, where the noise is severely correlated, then our deterministic matrix-based prewhitening scheme can exploit the structure of the noise in order to give a significant improvement in the estimation of parameters.

Since in multi-dimensional parameter estimation applications, the colored noise or interference has also a multi-dimensional structure, we have proposed a multi-dimensional prewhitening scheme called Sequential GSVD (S-GSVD). The S-GSVD has a very low computational complexity compared to matrix-based approaches, and the performance is also much superior, even when a small number of noise samples is available.

Due to the good performance of the S-GSVD, we have proposed also the iterative S-GSVD (I-S-GSVD) for scenarios where samples without signal components are not available, and, therefore, a joint estimation of the parameters and the noise statistics should be performed. While in such scenarios the matrix-based approach degrades the estimation of parameters compared to the case without prewhitening, the I-S-GSVD provides a significant improvement.

In summary, multi-dimensional model order selection schemes, such as  $R$ -D EFT and CFP-MOS, are essential for the estimation of parameters of MIMO channels as well as for several other scientific fields in the presence of white and colored Gaussian noise, respectively. For the estimation of spatial frequencies, once the model order is estimated, by applying robust multi-dimensional parameter estimation schemes such CFP-PE, it is possible to obtain a better estimation even when modeling errors are present. Finally, for multi-dimensional applications, in order to have an improved estimation of parameter in the presence of colored noise, the S-GSVD and the I-S-GSVD should be applied.

## 5.1 Outlook and Future Works

Due to the improved accuracy, reliability, and robustness, the application of multidimensional techniques has become much more common in the literature. In addition, in some problems only by using the multidimensional structure of the data, the desired solution can be obtained.

---

Although in this thesis several contributions have been proposed, we list below some open tasks that should be investigated

- 1) In Chapter 3, the CFP-PE has been shown to be robust to data with modeling errors, whereas in Section 4.2, the S-GSVD have been presented as a generalization of the GSVD for two tensors. Therefore, one important task is the development of a new decomposition that combines the CFP and the S-GSVD of two tensors.
  - 2) In audio applications, it is possible to collect samples in the absence of signal components. For these applications, the model order selection techniques in Chapter 2 can be integrated to the prewhitening schemes in Chapter 4. Therefore, another important task is to improve the performance of model order selection schemes by taking into account the information of the noise statistics.
  - 3) Although in the thesis we simulate with PARAFAC data model, it is important to validate the proposed techniques by using measurements.
  - 4) In this thesis, we have considered white and colored Gaussian noise, which are assumed in several applications, however, the extension of the proposed model order techniques for other types of noise should be also investigated.
-



## Appendices





## Appendix A

# PROBABILITIES TO EVALUATE THE MODEL ORDER SELECTION SCHEMES

For the model order estimation, three types of probabilities are considered, which are the probability of correct detection (PoD), the probability of false alarm (Pfa) and the probability of non-detection (PnD).

The PoD is defined as the probability that the correct model order  $d$  is estimated, which means that  $P_{\text{oD}} = \Pr[\hat{d} = d]$ , where  $\hat{d}$  is the estimated value of the model order for a certain MOS scheme. The probability of false alarm (Pfa) is the probability that  $\hat{d} > d$ . Therefore, we can write  $P_{\text{fa}} = \Pr[\hat{d} > d]$ . Finally, the probability of non-detection (PnD) is the probability that  $\hat{d} < d$ . Consequently, we can also write  $P_{\text{nd}} = \Pr[\hat{d} < d]$ . It goes without saying that one of the three cases should be satisfied, which means that

$$P_{\text{oD}} + P_{\text{fa}} + P_{\text{nd}} = 1. \quad (\text{A.1})$$

In the literature, the model order selection techniques are also classified according to their type of error, i.e., by overestimation or by underestimation. According to [80], for radar applications, the error by overestimation is preferred to the error by underestimation. On the other hand, for the estimation of the spatial frequencies via  $R$ -D ESPRIT-type algorithms, the error of the model order degrades considerably the estimation. Therefore, it is preferred to have model order selection techniques that have the highest PoD for different scenarios in conjunction with  $R$ -D ESPRIT-type algorithms.

As an example of the different estimation errors of model order selection techniques, we can compute numerically via Monte Carlo simulations the three probabilities using their respective previous definitions. In Fig. A.1, we show the PoD versus the SNR of several model order selection techniques, which are described in Section 2.1. In Fig. A.2, the Pfa versus the SNR is plotted, while in Fig. A.3, the PnD versus the SNR is depicted.

According to Figs. A.1, A.2, and A.3, the schemes RADOI and ESTER fail the estimation in low SNR regimes by overestimation, while the other schemes fail by underestimation. Note that we set the Pfa of EFT and M-EFT to  $10^{-2}$  in order to obtain the respective threshold coefficients.

In Fig. A.4, we present the PoD versus the Pfa. This is also used to evaluate the model

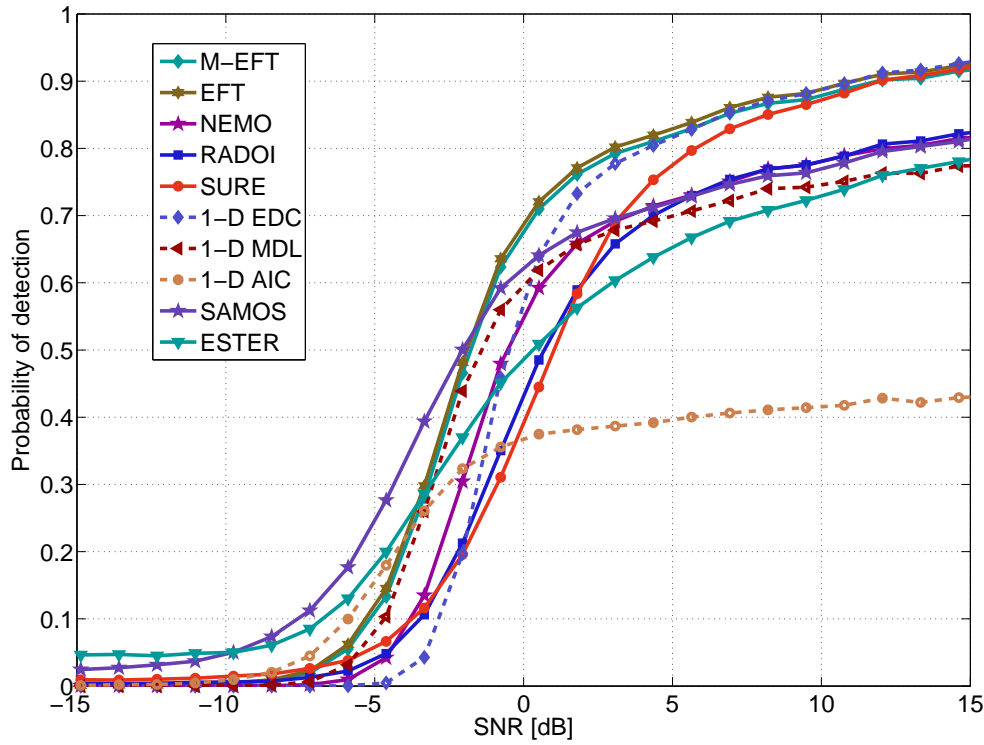


Fig. A.1: PoD versus SNR for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We set the number of sources  $d = 3$ .

order scheme according to the application, since it is possible to distinguish if the model order schemes fails by overestimation or by underestimating of the model order.

In Fig. A.5, we consider the case that  $d = 0$ , i.e., that there is no source in the data, and the model order selection schemes should be able to detect that. In practice, to detect  $d = 0$ , the following schemes have worked: M-EFT, EFT, NEMO, 1-D MDL, and 1-D EDC. Although 1-D AIC also works, but it has a perceivable Pfa. ESTER, SAMOS, RADOI, and SURE fail for this type of scenario, although theoretically ESTER and SAMOS are the only schemes that should not be applicable to this scenario.

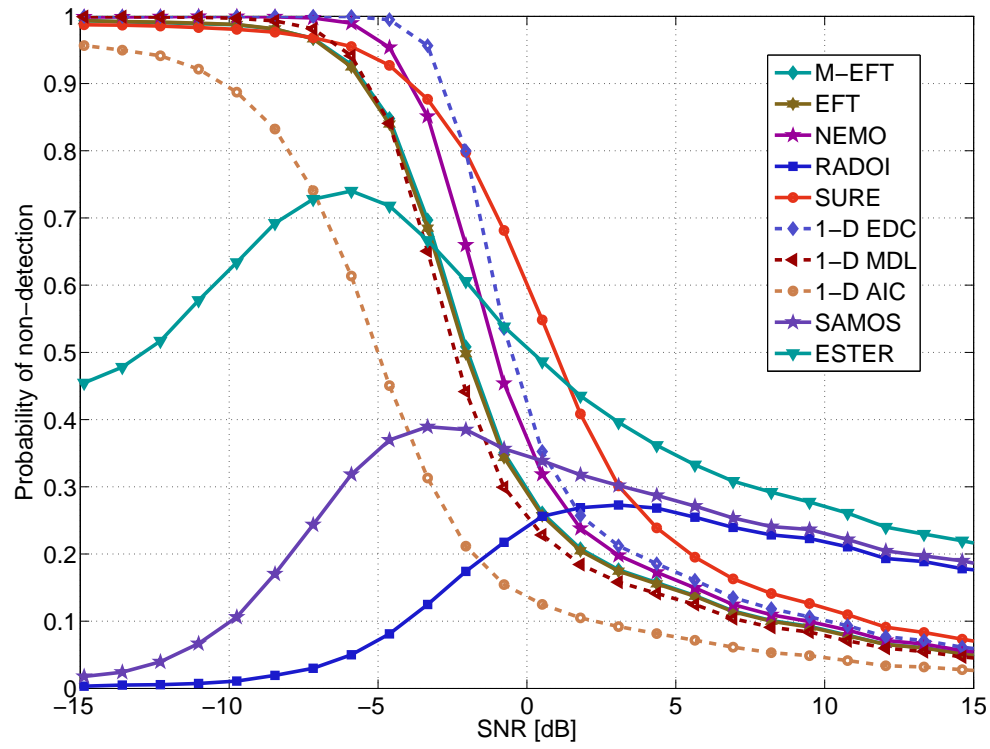


Fig. A.2: Pfa versus SNR for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We set the number of sources  $d = 3$ .

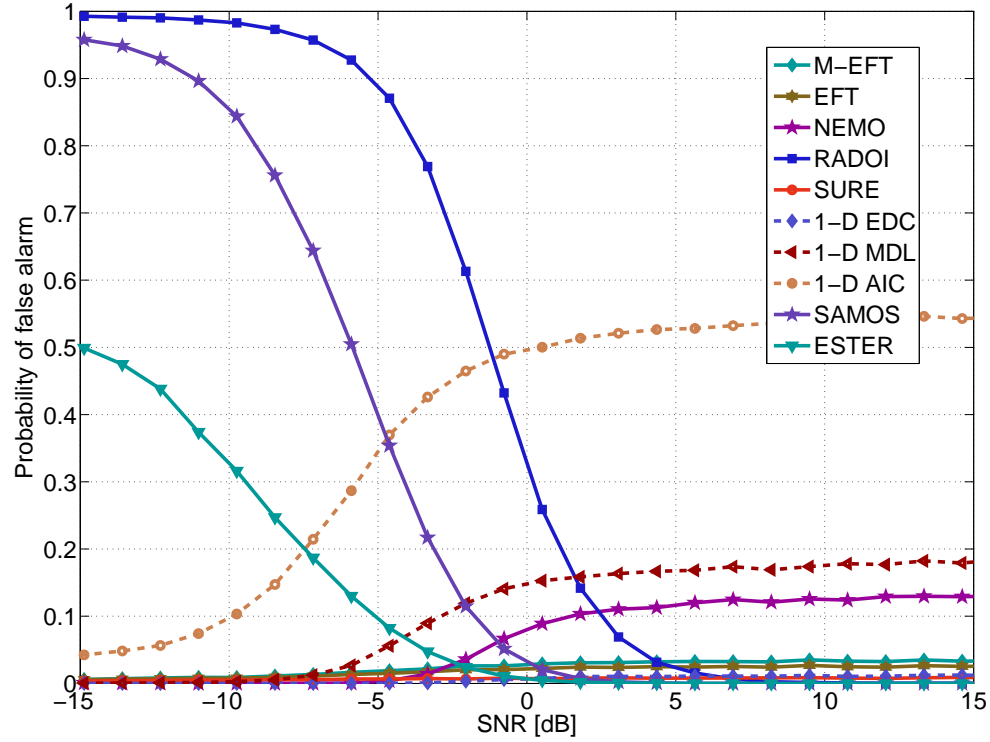


Fig. A.3: Pnd versus SNR for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We set the number of sources  $d = 3$ .

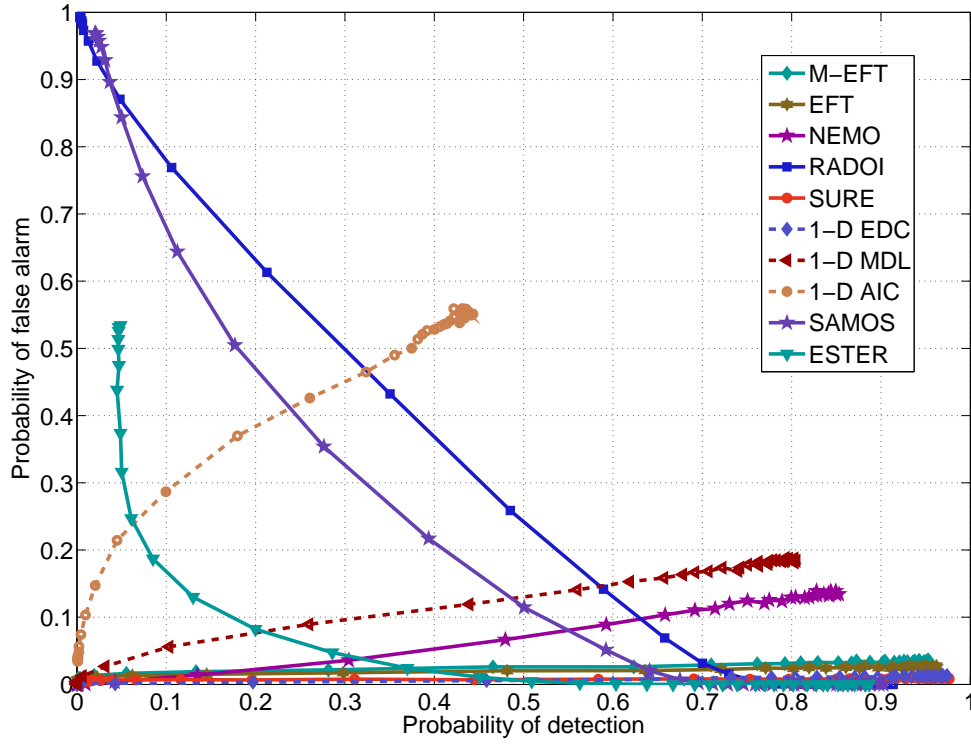


Fig. A.4: PoD versus Pfa for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We set the number of sources  $d = 3$ .

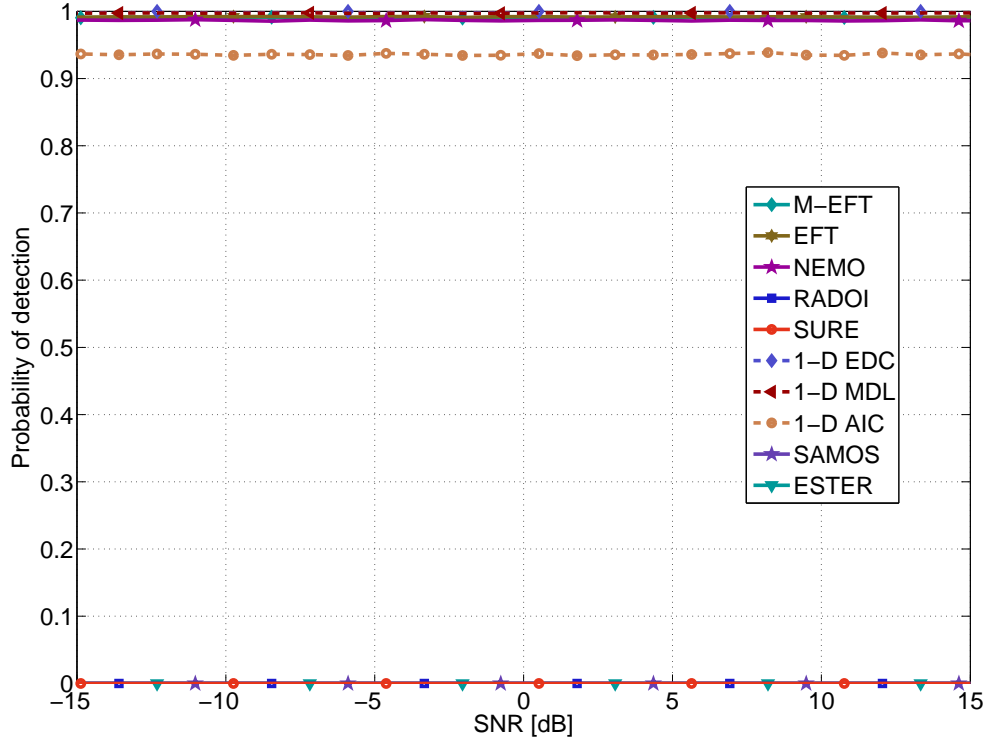


Fig. A.5: PoD versus SNR for different model order selection schemes. Here we consider an array of size  $M = 10$  with  $N = 12$  snapshots. We set the number of sources  $d = 3$ .

## Appendix B

# 1-D INFORMATION THEORETIC CRITERIA - EXPLANATION

In this Appendix, we present the derivation of the 1-D Information Theoretic Criteria proposed in [7] based on [98]. According to [10]

$$\text{AIC}(P) = -2 \cdot \log(f(\mathbf{X}|\Theta)) + 2 \cdot f_p, \quad (\text{B.1})$$

while according to [82]

$$\text{MDL}(P) = -2 \cdot \log(f(\mathbf{X}|\Theta)) + \frac{1}{2} \cdot f_p \cdot \log(N), \quad (\text{B.2})$$

where  $\Theta$  are the estimates for the eigenvectors, eigenvalues and noise according to [98],  $f_p$  denotes the amount of free parameters, and  $\log(f(\mathbf{X}|\Theta))$  is the log-likelihood function. Also according to [98]  $\Theta = \{\lambda_1, \dots, \lambda_{\min(M,N)}, \sigma_n^2, \mathbf{U}_1, \dots, \mathbf{U}_{\min(M,N)}\}$ , where  $\sigma_n^2$  is the noise variance, and  $\lambda_i$  and  $\mathbf{U}_i$  are the eigenvalues and the eigenvectors, respectively, computed as in (2.11).

According to [11], by the asymptotic theory for principal component analysis, the log-likelihood function is given by

$$\log(f(\mathbf{X}|\Theta)) = -N \cdot \log\left(\prod_{i=P+1}^{\alpha} \lambda_i\right) + N \cdot (\alpha - P) \cdot \log\left(\frac{1}{k} \cdot \sum_{i=P+1}^{\alpha} \lambda_i\right) \quad (\text{B.3})$$

where  $\alpha = \min(M, N)$ . Note here we consider both cases where  $M \leq N$  or  $M \geq N$  in the log-likelihood function proposed by [11], while in [98] only the case  $N > M$  is considered.

From (B.3) it is straightforward to find

$$\log(f(\mathbf{X}|\Theta)) = -N \cdot (\alpha - P) \cdot \log\left(\frac{\mathbf{a}(P)}{\mathbf{g}(P)}\right), \quad (\text{B.4})$$

where  $\mathbf{a}(P)$  and  $\mathbf{g}(P)$  are the arithmetic and geometric mean of the  $P$  smallest eigenvalues, respectively. In [11], all the noise or error eigenvalues are assumed as being equal. Therefore, (B.4) should be replaced in (B.1) and in (B.2).

In order to compute  $f_p$ ,  $P$  real-valued eigenvalues,  $P$  complex-valued eigenvectors  $\in \mathbb{C}^{M \times 1}$

and one real-valued noise variance are considered as free parameters of  $\Theta$  in [98]. Since complex-valued counts twice, then we have  $f_p = P + 1 + 2 \cdot M \cdot P = (2 \cdot M + 1) \cdot P + 1$ . Still according to [98], the eigenvectors are not independently adjusted. Therefore, we have to subtract by  $2 \cdot P$  and by  $2 \cdot \frac{1}{2} \cdot P \cdot (P - 1)$ , due to their unit norm, and their mutual orthogonality, respectively. Consequently,  $f_p = P \cdot (2 \cdot M - P) + 1$ . Since this constant is included for all  $P$  in the AIC(k) cost function, we can remove it, and we have only  $f_p = P \cdot (2 \cdot M - P)$ , which should be replaced in (B.1) and in (B.2).

In the derivation of expressions for AIC and MDL in [98], the assumption  $N > M$  is done. However, in the case that  $M > N$ , the rank of the subspace of  $\mathbf{U}$  is equal to  $\alpha$ , where  $\alpha = \min(M, N)$ . Therefore, since from  $M$  dimensions, only  $N$  are considered for this case, we have  $f_p = P + 1 + 2 \cdot \alpha \cdot P = (2 \cdot M + 1) \cdot P + 1$ . Similarly to [98], considering the unit norm and the mutual orthogonality of the eigenvectors, and ignoring the constant part, we obtain  $f_p = P \cdot (2 \cdot \alpha - P)$ , which should be replaced in (B.1) and in (B.2). Note that (B.4) and  $f_p = P \cdot (2 \cdot \alpha - P)$  are used together, or the version proposed in [98] is used. Any attempt to use (B.4) together with [98] or only using  $f_p = P \cdot (2 \cdot \alpha - P)$  together with [98] gives a worse Probability of Detection (PoD) than our 1-D ITC and [98].

## Appendix C

# PROPERTIES OF EIGENVALUES WITH A WISHART DISTRIBUTION

In this Appendix, we present some properties for the eigenvalues with a Wishart distribution. These properties are necessary in Appendix D to derive the Exponential Fitting Test (EFT) and Modified Exponential Fitting Test (M-EFT).

Since we desire to generalize the M-EFT for multi-dimensional applications, we consider the  $r$ -th unfolding of the noise tensor  $\mathcal{N} \in \mathbb{C}^{M_1 \times \dots \times M_R \times M_{R+1}}$ , where each element  $n_{m_1, \dots, m_{R+1}}$  is a zero mean circularly symmetric (ZMCS) complex Gaussian random variable with variance  $\sigma_n^2$ . For the sake of notation, we define  $M_{R+1} = N$  and  $M = \prod_{k=1}^{R+1} M_k$ . Therefore, eigenvalues with a Wishart distribution are obtained from the covariance matrix as follows

$$\frac{M_r}{M} \cdot \mathbb{E} \left\{ [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \right\} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H \quad (\text{C.1})$$

where  $\mathbb{E} \{ \cdot \}$  is the expected value operator, and the diagonal matrix  $\mathbf{\Lambda}$  contains the eigenvalues  $\{\lambda_1, \dots, \lambda_\alpha\}$ . We define  $\alpha = \min \left( M_r, \frac{M}{M_r} \right)$ .

**Lemma 3.** *Let us consider a complex Gaussian random matrix  $[\mathcal{N}]_{(r)} \in \mathbb{C}^{M_r \times \frac{M}{M_r}}$ , where  $n_{(r)}(i, j)$  indicates the element of  $[\mathcal{N}]_{(r)}$  in row  $i$  and column  $j$ . The following equality holds*

$$\sum_{k=1}^{\alpha} \lambda_k = M_r \cdot \sigma_n^2, \quad (\text{C.2})$$

where  $\lambda_k$  are calculated using the expected value operator according to (C.1).

*Proof.* First let us represent (C.1) as a sum of rank-one matrices as follows

$$\mathbb{E} \left\{ [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \right\} = \sum_{i=1}^{\frac{M}{M_r}} \mathbb{E} \left\{ \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i) \right\}, \quad (\text{C.3})$$

where  $\mathbf{n}_{(r)}(i)$  is the  $i$ -th column of the matrix  $\mathcal{N}_{(r)}$ .

It is straightforward that

$$\mathbb{E} \left\{ \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i) \right\} = \sigma_n^2 \cdot \mathbf{I}_{M_r}, \quad (\text{C.4})$$

where  $\mathbf{I}_{M_r}$  is an identity matrix of size  $M_r \times M_r$ .

Therefore, replacing (C.4) in (C.3), we have that

$$\begin{aligned} \mathbb{E} \left\{ [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \right\} &= \sum_{i=1}^{\frac{M}{M_r}} \sigma_n^2 \cdot \mathbf{I}_{M_r} \\ &= \frac{M}{M_r} \cdot \sigma_n^2 \cdot \mathbf{I}_{M_r} \end{aligned} \quad (\text{C.5})$$

Applying the trace operator in (C.5)

$$\text{tr} \left( \mathbb{E} \left\{ [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \right\} \right) = M \cdot \sigma_n^2, \quad (\text{C.6})$$

while also applying the trace operator to the right hand side of (C.1)

$$\begin{aligned} \text{tr}(\mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^H) &= \text{tr}(\mathbf{\Lambda} \cdot \mathbf{U} \cdot \mathbf{U}^H) \\ &= \text{tr}(\mathbf{\Lambda}) \\ &= \sum_{k=1}^{\alpha} \lambda_k. \end{aligned} \quad (\text{C.7})$$

Therefore, replacing (C.6) and (C.7) in (C.1), we obtain (C.2).  $\square$

By means of simulations, we can compute the sum of the square of the eigenvalues via the following expression

$$\left( \frac{M_r}{M} \right)^2 \cdot \text{tr} \left( \mathbb{E} \left\{ [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \cdot [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \right\} \right) = \sum_{k=1}^{\alpha} \lambda_k^2. \quad (\text{C.8})$$

**Lemma 4.** *Let us consider a complex Gaussian random matrix  $[\mathcal{N}]_{(r)} \in \mathbb{C}^{M_r \times \frac{M}{M_r}}$ , where  $n_{(r)}(i, j)$  indicates the element of  $[\mathcal{N}]_{(r)}$  in row  $i$  and column  $j$ . The following equality holds*

$$\sum_{k=1}^{\alpha} \lambda_k^2 = \frac{M_r^2}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \cdot \sigma_n^4, \quad (\text{C.9})$$

where  $\lambda_k$  are calculated using the expected value operator according to (C.8), and  $\gamma = 1$  if  $n_{m_1, \dots, m_{R+1}} \in \mathbb{R}$ , or  $\gamma = 0$ , if  $n_{m_1, \dots, m_{R+1}} \in \mathbb{C}$ .

*Proof.* Similarly to (C.3), we represent the first member as product of rank-one matrices as follows

$$[\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \cdot [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H = \sum_{i=1}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i) \cdot \sum_{t=1}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(t) \cdot \mathbf{n}_{(r)}^H(t), \quad (\text{C.10})$$

where  $\mathbf{n}_{(r)}(t)$  is the  $t$ -th column of the matrix  $[\mathcal{N}]_{(r)}$ . Therefore, we can separate (C.10) into



two cases. The first one, when  $i = t$

$$\mathbf{A}_1 = \sum_{i=1}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i) \cdot \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i), \quad (\text{C.11})$$

and the second one, when  $i \neq t$

$$\mathbf{A}_2 = \sum_{i=1}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i) \cdot \sum_{t=1|t \neq i}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(t) \cdot \mathbf{n}_{(r)}^H(t). \quad (\text{C.12})$$

Since  $i \neq t$ , then the product of the two sums are uncorrelated, therefore,

$$\begin{aligned} \mathbb{E}\{\mathbf{A}_2\} &= \mathbb{E}\left\{\sum_{i=1}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(i) \cdot \mathbf{n}_{(r)}^H(i)\right\} \cdot \mathbb{E}\left\{\sum_{t=1|t \neq i}^{\frac{M}{M_r}} \mathbf{n}_{(r)}(t) \cdot \mathbf{n}_{(r)}^H(t)\right\} \\ &= \frac{M}{M_r} \cdot \left(\frac{M}{M_r} - 1\right) \cdot \sigma_n^4 \cdot \mathbf{I}_{M_r}. \end{aligned} \quad (\text{C.13})$$

Now we just have to compute  $\mathbb{E}\{\mathbf{A}_1\}$ . Before that, we compute

$$\mathbb{E}\{n_{m_1, \dots, m_{R+1}} \cdot n_{m_1, \dots, m_{R+1}}^* \cdot n_{m_1, \dots, m_{R+1}} \cdot n_{m_1, \dots, m_{R+1}}^*\} = (2 + \gamma) \cdot \sigma_n^4, \quad (\text{C.14})$$

where  $n_{m_1, \dots, m_{R+1}} \in \mathbb{R}$ , then  $\gamma = 1$ , otherwise, if  $n_{m_1, \dots, m_{R+1}} \in \mathbb{C}$ , then  $\gamma = 0$ . (C.14) is the fourth order moment of  $n_{m_1, \dots, m_{R+1}}$ .

Applying the expected value operator in (C.11), it is straightforward that

$$\mathbb{E}\{\mathbf{A}_1\} = \frac{M}{M_r} \cdot (M_r + 1 + \gamma) \cdot \sigma_n^4 \cdot \mathbf{I}_{M_r}, \quad (\text{C.15})$$

where one element of the diagonal of  $\mathbb{E}\{\mathbf{A}_1\}$  is  $\mathbb{E}\{a(i, i)\} = (2 + \gamma) \cdot \sigma_n^4 + (M_r - 1) \cdot \sigma_n^4$ .

By adding (C.13) and (C.15), we obtain

$$\mathbb{E}\{\mathbf{A}_1\} + \mathbb{E}\{\mathbf{A}_2\} = \frac{M}{M_r} \cdot \left(\frac{M}{M_r} + M_r + \gamma\right) \cdot \sigma_n^4 \cdot \mathbf{I}_{M_r}. \quad (\text{C.16})$$

Due to (C.11) and (C.12), we can find the expected value of (C.10)

$$\mathbb{E}\left\{[\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \cdot [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H\right\} = \frac{M}{M_r} \cdot \left(\frac{M}{M_r} + M_r + \gamma\right) \cdot \sigma_n^4 \cdot \mathbf{I}_{M_r}. \quad (\text{C.17})$$

Applying the trace operator in (C.18)

$$\text{tr}\left(\mathbb{E}\left\{[\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H \cdot [\mathcal{N}]_{(r)} \cdot [\mathcal{N}]_{(r)}^H\right\}\right) = M \cdot \left(\frac{M}{M_r} + M_r + \gamma\right) \cdot \sigma_n^4. \quad (\text{C.18})$$

Therefore, replacing (C.18) in (C.8), we obtain (C.9). The result of (C.9) has been checked by means of simulations.

□



## Appendix D

# MODIFIED EXPONENTIAL FITTING TEST II - DERIVATION

In this Appendix, we present the derivation of the Modified Exponential Fitting Test II [7], which includes both the real-valued and the complex-valued case. In order to obtain the expression for the M-EFT proposed in [2], it is only necessary to restrict the general expression to the complex-valued case.

Similarly to Appendix C, we assume that there is only noise. Since we desire to generalize the M-EFT for multi-dimensional applications, we consider the  $r$ -th unfolding of the noise tensor  $\mathcal{N} \in \mathbb{C}^{M_1 \times \dots \times M_R \times M_{R+1}}$ , where  $M_r$  stands for the number of sensors in the  $r$ -th dimension and where each element  $n_{m_1, \dots, m_{R+1}}$  is a zero mean circularly symmetric (ZMCS) complex Gaussian random variable with variance  $\sigma_n^2$ . For the sake of notation, we define  $M_{R+1} = N$  and  $M = \prod_{k=1}^{R+1} M_k$ .

In Fig. D.1, we validate the assumption that the noise eigenvalues are approximated by an exponential curve by considering an example.

Before showing the derivation of the M-EFT in Section D.2, we present a short review of the EFT in Section D.1 similarly as in [42, 79].

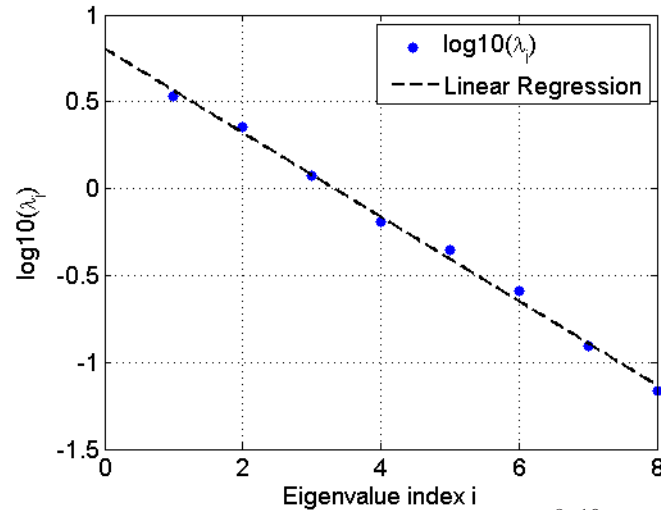


Fig. D.1: In this scenario,  $\mathcal{N} \in \mathbb{C}^{8 \times 10}$ .

## D.1 Review of the EFT

Here the matrix-based data model is assumed, i.e.,  $M_{R+1} = N$  and  $M = \prod_{k=1}^{R+1} M_k$ , and only the  $R+1$ -mode unfolding of  $\mathcal{N}$  is taken into account. First we present the three equations, from which the rate  $q$  of the exponential profile can be derived. The first one is the assumption of the exponential profile, which is given by

$$\lambda_k = \lambda_1 \cdot q^{k-1}, \quad (\text{D.1})$$

where  $q$  is the decreasing rate of the exponential curve used for the approximation,  $\lambda_1$  is the largest noise eigenvalue, and  $k$  is the index to indicate the selected eigenvalue.

The second equation is the sum of the noise eigenvalues, which is given by

$$\sum_{k=1}^M \lambda_k = M \cdot \sigma_n^2, \quad (\text{D.2})$$

The third equation is related to the variance of the noise eigenvalues, which is given by

$$\frac{1}{N} \cdot \sum_{k=1}^M (\lambda_k - \sigma_n^2)^2 = \frac{M \cdot \sigma_n^4}{N}. \quad (\text{D.3})$$

By using the three equations, the following equality is obtained

$$\frac{(1-q) \cdot (1-q^M)}{(1-q^M) \cdot (1+q)} = \frac{M+N}{M \cdot N}, \quad (\text{D.4})$$

where, according to [42, 79], using  $q = e^{-2 \cdot a}$ , we can rewrite

$$\frac{M \cdot \text{th}(a) - \text{th}(a)}{M \cdot \text{th}(a)} = \frac{1}{N}, \quad (\text{D.5})$$

where  $\text{th}(\cdot)$  is the tangent hyperbolic operator. By truncating the operator tangent hyperbolic  $\text{th}(\cdot)$  in a forth order expansion, the following expression is obtained

$$a = \sqrt{\frac{15}{2 \cdot (M^2 + 2)} \left[ 1 - \frac{4 \cdot M \cdot (M^2 + 2)}{5 \cdot N \cdot (M^2 - 1)} \right]}. \quad (\text{D.6})$$

The derivation to obtain (D.4) from (D.1), (D.2), and (D.3) is more specific than the one that we show in the next section, since it assumes white complex-valued noise and also that  $M > N$ , and is already presented by the authors of [42, 79]. Therefore, we consider here only their final results.

## D.2 Derivation of the M-EFT II

Similarly to the EFT, the Exponential Fitting Test [42, 79], on which the Modified Exponential Fitting Test II is based, is derived according to three equations. In order to simplify and also to generalize the derivation, we use our expressions of Lemmas 3 and 4 derived in Appendix C.

In this section we propose the derivation of the M-EFT II, which considers the cases that the noise can be real-valued or complex-valued,  $M$  and  $N$  can assume any values, and the multi-dimensional notation, i.e., for a certain  $r$ -mode unfolding of  $\mathbf{N}$ , we have that  $M = M_r$  and  $N = \frac{M}{M_r}$  in comparison to Section D.1.

Our first equation similarly to [42, 79] for the exponential profile approximation is given by

$$\lambda_k^{(r)} = \lambda_1^{(r)} \cdot q^{(r)} \left( M_r, \frac{M}{M_r} \right)^{k-1}, \quad (\text{D.7})$$

where  $q^{(r)}(M_r, \frac{M}{M_r})$  is the rate of the exponential profile that should be estimated. Note that  $q^{(r)}(M_r, \frac{M}{M_r})$  is a function of  $r$ ,  $M_r$ , and  $M$ . However, for the sake of notation, we write  $q^{(r)}(M_r, \frac{M}{M_r})$  just as  $q$ .

Our second equation is slightly modified for the sum of the eigenvalues according to Lemma 3 in Appendix C

$$\sum_{k=1}^{\alpha} \lambda_k^{(r)} = M_r \cdot \sigma_n^2, \quad (\text{D.8})$$

where  $\alpha = \min\left(M_r, \frac{M}{M_r}\right)$ .

Our third equation is also modified instead of variance, we estimate the sum of the square eigenvalues according to Lemma 4 in Appendix C

$$\sum_{k=1}^{\alpha} \lambda_k^2 = \frac{M_r^2}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \cdot \sigma_n^4. \quad (\text{D.9})$$

Note that using the variance expression in [42, 79], two expressions are found, one for  $\frac{M}{M_r} < M_r$ , and another one for  $\frac{M}{M_r} > M_r$ . Since we desire a more general expression, we prefer to use (D.9) instead of [42, 79]. By replacing (D.7) in (D.8), we obtain that

$$\lambda_1^{(r)} + \lambda_1^{(r)} \cdot q + \dots + \lambda_1^{(r)} \cdot q^{\alpha-1} = M_r \cdot \sigma_n^2 \quad (\text{D.10})$$

$$\lambda_1^{(r)} \cdot \left( \frac{1 - q^{\alpha}}{1 - q} \right) = M_r \cdot \sigma_n^2. \quad (\text{D.11})$$

From (D.11), we find that

$$\lambda_1^{(r)} = \left( \frac{1 - q}{1 - q^{\alpha}} \right) \cdot M_r \cdot \sigma_n^2. \quad (\text{D.12})$$

By replacing (D.7) in (D.9), we obtain that

$$\sum_{k=1}^{\alpha} \left( \lambda_1^{(r)} \right)^2 \cdot q^{2 \cdot (k-1)} = \frac{M_r^2}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \cdot \sigma_n^4, \quad (\text{D.13})$$

and also replacing (D.11) in (D.9), we obtain that

$$\begin{aligned} \sum_{k=1}^{\alpha} \left[ \left( \frac{1-q}{1-q^{\alpha}} \right) \cdot M_r \cdot \sigma_n^2 \right]^2 \cdot q^{2 \cdot (k-1)} &= \frac{M_r^2}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \cdot \sigma_n^4 \\ \left( \frac{1-q}{1-q^{\alpha}} \right)^2 \cdot M_r^2 \cdot \sigma_n^4 \cdot \sum_{k=1}^{\alpha} q^{2 \cdot (k-1)} &= \frac{M_r^2}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \cdot \sigma_n^4 \\ \left( \frac{1-q}{1-q^{\alpha}} \right)^2 \cdot \sum_{k=1}^{\alpha} q^{2 \cdot (k-1)} &= \frac{1}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \end{aligned} \quad (\text{D.14})$$

It is straightforward that

$$\sum_{k=1}^{\alpha} q^{2 \cdot (k-1)} = \frac{1 - q^{2 \cdot \alpha}}{1 - q^2}. \quad (\text{D.15})$$

Replacing (D.15) in (D.14)

$$\left( \frac{1-q}{1-q^{\alpha}} \right)^2 \cdot \frac{1 - q^{2 \cdot \alpha}}{1 - q^2} = \frac{1}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right) \quad (\text{D.16})$$

Since

$$\frac{1 - q^{2 \cdot \alpha}}{1 - q^2} = \left( \frac{1 - q^{\alpha}}{1 - q} \right) \cdot \left( \frac{1 + q^{\alpha}}{1 + q} \right), \quad (\text{D.17})$$

we can rewrite (D.16) as

$$\left( \frac{1-q}{1-q^{\alpha}} \right) \cdot \left( \frac{1+q^{\alpha}}{1+q} \right) = \frac{1}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right), \quad (\text{D.18})$$

we define the function  $C\left(M_r, \frac{M}{M_r}\right) = \frac{1}{M} \cdot \left( \frac{M}{M_r} + M_r + \gamma \right)$ , and for the sake of notation, we represent  $C(M_r, \frac{M}{M_r})$  just as  $C$ .

(D.18) can be written in a polynomial form as a function of  $C$

$$(C - 1) \cdot q^{\alpha+1} + (C + 1) \cdot q^{\alpha} - (C + 1) \cdot q + 1 - C = 0. \quad (\text{D.19})$$

As we have proposed in [7], a real-valued solution of  $q$  in the interval  $(0, 1)$  should be chosen.

A very good numerical approximation for (D.19) is proposed in [2] based on [42, 79]

$$q(\alpha_r, \beta_r) = \exp \left\{ - \sqrt{ \frac{30}{\alpha_r^2 + 2} - \sqrt{ \frac{900}{(\alpha_r^2 + 2)^2} - \frac{720\alpha_r}{\beta_r(\alpha_r^4 + \alpha_r^2 - 2)} } } \right\}, \quad (\text{D.20})$$

where  $\alpha_r = \min\left(M_r, \frac{M}{M_r}\right)$  and  $\beta_r = \max\left(M_r, \frac{M}{M_r}\right)$ . In Fig. D.2, we compare M-EFT based

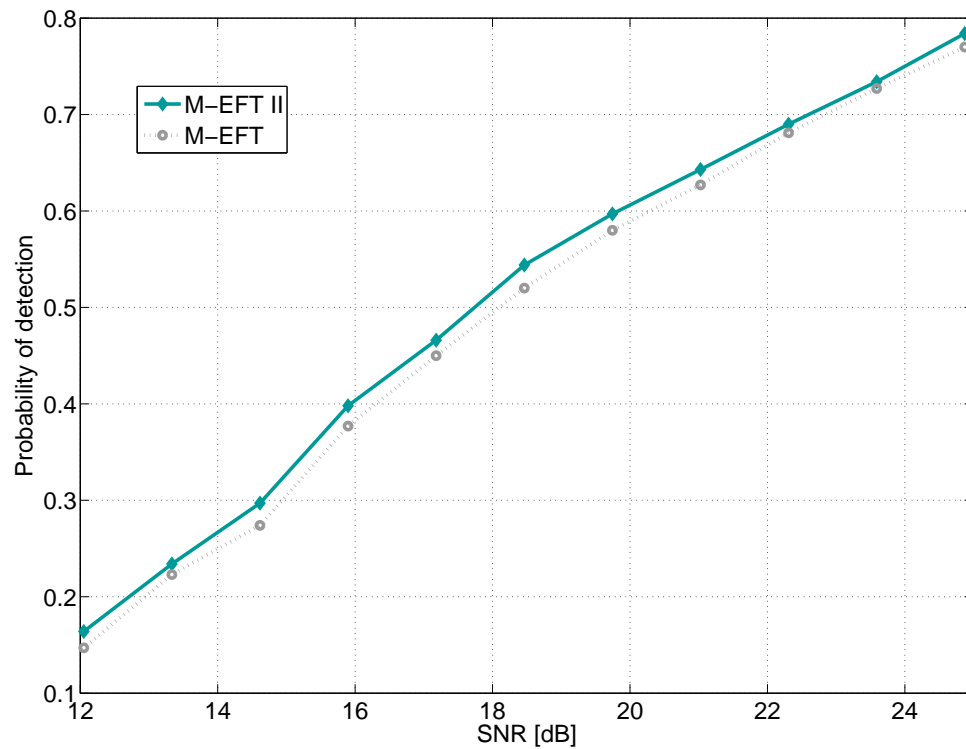


Fig. D.2: Comparing the M-EFT II and M-EFT for the case of real noise. Probability of detection vs. SNR for an array of size  $M_1 = 5$ . The number of snapshots  $N$  is set to 6 and the number of sources  $d = 3$ .

on (D.19), which is referred as M-EFT II, to M-EFT based on (D.20), which is referred as M-EFT. In this simulation we consider ZMCS real-valued Gaussian noise. A small improvement is achieved in terms of PoD by taking into account  $\gamma$  via the expression in (D.19) instead of (D.20). Since  $M < N$  in Fig. D.2, the performance of the M-EFT and the EFT is the same. Therefore, we just include the M-EFT.





## Appendix E

# TENSOR CALCULUS

In this Appendix, we present the notation applied, as well as notions of tensor calculus.

In order to facilitate the distinction between scalars, matrices, and tensors, the following notation is used: Scalars are denoted as italic letters ( $a, b, \dots, A, B, \dots, \alpha, \beta, \dots$ ), column vectors as lower-case bold-face letters ( $\mathbf{a}, \mathbf{b}, \dots$ ), matrices as bold-face capitals ( $\mathbf{A}, \mathbf{B}, \dots$ ), and tensors are written as bold-face calligraphic letters ( $\mathcal{A}, \mathcal{B}, \dots$ ). Lower-order parts are consistently named: the  $(i, j)$ -element of the matrix  $\mathbf{A}$ , is denoted as  $a_{i,j}$  and the  $(i, j, k)$ -element of a third order tensor  $\mathcal{X}$  as  $x_{i,j,k}$ . The  $n$ -mode vectors of a tensor are obtained by varying the  $r$ -th index within its range  $(1, 2, \dots, M_r)$  and keeping all the other indices fixed. The applied notation can be easily visualized in Figs. E.1, E.2, E.3, and E.4.

We use the superscripts  $T, H, ^{-1}, +$ , and  $*$  for transposition, Hermitian transposition, matrix inversion, the Moore-Penrose pseudo inverse of matrices, and complex conjugation, respectively. Moreover the Khatri-Rao product (columnwise Kronecker product) is denoted by  $\mathbf{A} \diamond \mathbf{B}$  and the outer product is indicated by  $\mathbf{a} \circ \mathbf{b}$ .

For the sake of simplicity, we consider in our definitions and citations a 3-way or 3-dimensional tensor of the form of Fig. E.4.

$$x \in \mathbb{C}^{1 \times 1}$$

Scalar

Fig. E.1: Notation for scalars

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \in \mathbb{C}^{M \times 1}$$

Vector

Fig. E.2: Notation for vectors

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,M_2} \\ \vdots & \ddots & \vdots \\ x_{M_1,1} & \cdots & x_{M_1,M_2} \end{bmatrix} \in \mathbb{C}^{M_1 \times M_2}$$

Matrix

Fig. E.3: Notation for matrices

$$\mathcal{X} = \begin{array}{|c|c|c|} \hline & x_{1,1,M_3} & \cdots & x_{1,M_2,M_3} \\ \hline & \vdots & \ddots & \vdots \\ \hline x_{1,1,1} & \cdots & x_{1,M_2,1} & x_{M_1,M_2,M_3} \\ \hline \vdots & \ddots & \vdots & \\ \hline x_{M_1,1,1} & \cdots & x_{M_1,M_2,1} & \cdots \\ \hline \end{array} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$$

Tensor

Fig. E.4: Visualization of a tensor representation

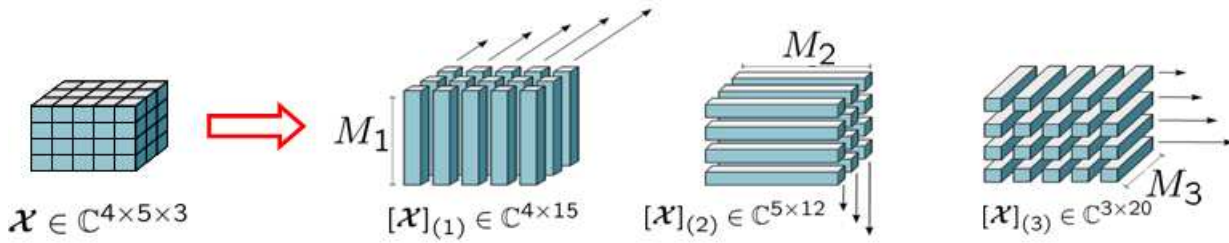


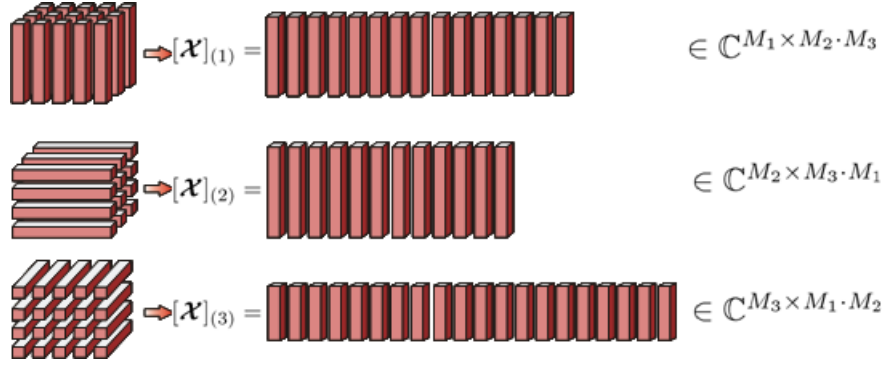
Fig. E.5:  $n$ -mode vectors of  $\mathbf{X} \in \mathbb{C}^{4 \times 5 \times 3}$

## E.1 $r$ -mode vectors and $r$ -mode unfoldings

The  $r$ -mode vector definition is important to represent  $\mathcal{X}$  in the matrix form, which is also referred as  $n$ -mode unfolding of  $\mathcal{X}$ . Therefore, the set of all  $r$ -mode vectors for a certain  $r$  forms a  $r$ -unfolding matrix. Later the definition of unfoldings is very useful in order to define tensor operations.

First we define an  $r$ -mode vector of an  $(M_1 \times M_2 \times \dots \times M_R)$ -dimensional tensor  $\mathcal{X}$  as an  $M_r$ -dimensional vector obtained from  $\mathcal{X}$  by varying the index  $m_r$  and keeping the other indices fixed.

For instance, by computing according to Figs. E.5 and E.6 all 1-mode vectors of  $\mathcal{X}$  shown in Fig. E.4, we obtain the 1-mode unfolding of  $\mathcal{X}$  given by

Fig. E.6:  $n$ -mode vectors of  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$ 

$$[\mathcal{X}]_{(1)} = \begin{bmatrix} x_{1,1,1} & x_{1,1,2} & \dots & x_{1,1,M_3} & x_{1,2,1} & x_{1,2,2} & \dots & x_{1,M_2,M_3} \\ x_{2,1,1} & x_{2,1,2} & \dots & x_{2,1,M_3} & x_{2,2,1} & x_{2,2,2} & \dots & x_{2,M_2,M_3} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \dots \\ x_{M_1,1,1} & x_{M_1,1,2} & \dots & x_{M_1,1,M_3} & x_{M_1,2,1} & x_{M_1,2,2} & \dots & x_{M_1,M_2,M_3} \end{bmatrix}, \quad (\text{E.1})$$

where  $[\mathcal{X}]_{(1)} \in \mathbb{C}^{M_1 \times M_2 \cdot M_3}$ . Similarly, the 2-mode unfolding of  $\mathcal{X}$ ,  $[\mathcal{X}]_{(2)}$ , is given by

$$[\mathcal{X}]_{(2)} = \begin{bmatrix} x_{1,1,1} & x_{2,1,1} & \dots & x_{M_1,1,1} & x_{1,1,2} & x_{2,1,2} & \dots & x_{M_1,1,M_3} \\ x_{1,2,1} & x_{2,2,1} & \dots & x_{M_1,2,1} & x_{1,2,2} & x_{2,2,2} & \dots & x_{M_1,2,M_3} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \dots \\ x_{1,M_2,1} & x_{2,M_2,1} & \dots & x_{M_1,M_2,1} & x_{1,M_2,2} & x_{2,M_2,2} & \dots & x_{M_1,M_2,M_3} \end{bmatrix}, \quad (\text{E.2})$$

where  $[\mathcal{X}]_{(2)} \in \mathbb{C}^{M_2 \times M_1 \cdot M_3}$ . In the same way, the 3-mode unfolding of  $\mathcal{X}$ ,  $[\mathcal{X}]_{(3)}$ , is given by

$$[\mathcal{X}]_{(3)} = \begin{bmatrix} x_{1,1,1} & x_{1,2,1} & \dots & x_{1,M_2,1} & x_{2,1,1} & x_{2,2,1} & \dots & x_{M_1,M_2,1} \\ x_{1,1,2} & x_{1,2,2} & \dots & x_{1,M_2,2} & x_{2,1,2} & x_{2,2,2} & \dots & x_{M_1,M_2,2} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \dots \\ x_{1,1,M_3} & x_{1,2,M_3} & \dots & x_{1,M_2,M_3} & x_{2,1,M_3} & x_{2,2,M_3} & \dots & x_{M_1,M_2,M_3} \end{bmatrix}, \quad (\text{E.3})$$

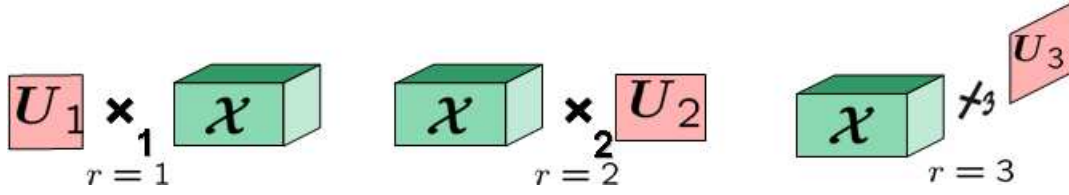
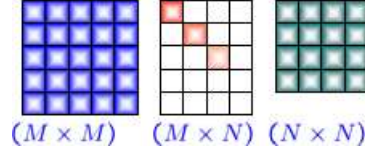
where  $[\mathcal{X}]_{(3)} \in \mathbb{C}^{M_3 \times M_1 \cdot M_2}$ .

## E.2 Tensor operators

The tensor operations we use are consistent with [29]:

**The inner product** of a tensor  $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  and a tensor  $\mathcal{B} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  is given by the element-wise product, which is computed as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \dots \sum_{m_R=1}^{M_R} a_{m_1, m_2, \dots, m_R} \cdot b_{m_1, m_2, \dots, m_R}^* \quad (\text{E.4})$$

Fig. E.7:  $n$ -mode productFig. E.8: Full SVD:  $\mathbf{X} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^H$ 

Note that by computing the inner product of a tensor with itself gives us the square of the higher norm of a tensor, which is represented by  $\|\cdot\|_H^2$ .

**The  $r$ -mode product** of a tensor  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  and a matrix  $\mathbf{U} \in \mathbb{C}^{J_r \times M_r}$  is denoted as  $\mathcal{X} \times_r \mathbf{U} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times J_r \times \dots \times M_R}$ . It is obtained by multiplying all  $r$ -mode vectors of  $\mathcal{X}$  from the left-hand side by the matrix  $\mathbf{U}$ , which give us that

$$[\mathcal{X} \times_r \mathbf{U}]_{(r)} = \mathbf{U} \cdot [\mathcal{X}]_{(r)}. \quad (\text{E.5})$$

In Fig. E.7, graphical representations of possible  $r$ -mode products by  $\mathbf{U}_1$ ,  $\mathbf{U}_2$ , and  $\mathbf{U}_3$  are shown.

**The higher order SVD (HOSVD)** of a tensor  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  is given by

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_R \mathbf{U}_R, \quad (\text{E.6})$$

where  $\mathcal{S} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  is the core-tensor which satisfies the all-orthogonality conditions [29] and  $\mathbf{U}_r \in \mathbb{C}^{M_r \times M_r}$ ,  $r = 1, 2, \dots, R$  are the unitary matrices of  $r$ -mode singular vectors.

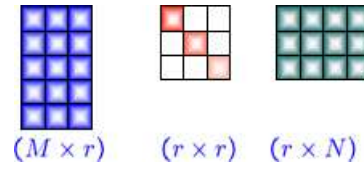
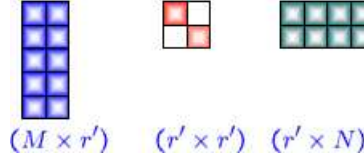
Since the HOSVD is defined in (E.6), next we present the concept of low-rank HOSVD approximation, which is done by analogy with the low-rank SVD approximation.

In Fig. E.8, we present the full size SVD of  $\mathbf{X}$ , where since  $\mathbf{X}$  is rank deficient, some of its eigenvalues are equal to zero. The full SVD of  $\mathbf{X}$  is given by following expression

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^H, \quad (\text{E.7})$$

where  $\mathbf{X} \in \mathbb{C}^{M \times N}$ ,  $\mathbf{U} \in \mathbb{C}^{M \times M}$ ,  $\mathbf{\Sigma} \in \mathbb{C}^{M \times N}$ , and  $\mathbf{V} \in \mathbb{C}^{N \times N}$ .

Since  $\mathbf{X}$  in (E.7) is rank deficient with rank equal to  $r$ , then an economy size SVD can be defined. In this case, by reducing the number of columns of  $\mathbf{U}$  and  $\mathbf{V}$ , we still have the same data, i.e., without losses, but occupying less memory as shown in Fig. E.9. The economy size

Fig. E.9: Economy size SVD:  $\mathbf{X} = \mathbf{U}_s \cdot \mathbf{\Sigma}_s \cdot \mathbf{V}_s^H$ Fig. E.10: Low-rank approximation SVD:  $\mathbf{X} \approx \mathbf{U}'_s \cdot \mathbf{\Sigma}'_s \cdot \mathbf{V}'_s^H$ 

SVD of  $\mathbf{X}$  is computed as

$$\mathbf{X} = \mathbf{U}_s \cdot \mathbf{\Sigma}_s \cdot \mathbf{V}_s^H, \quad (\text{E.8})$$

where  $\mathbf{X} \in \mathbb{C}^{M \times N}$ ,  $\mathbf{U}_s \in \mathbb{C}^{M \times r}$ ,  $\mathbf{\Sigma}_s \in \mathbb{C}^{r \times r}$ , and  $\mathbf{V}_s \in \mathbb{C}^{N \times r}$ . Note that in (E.8)  $r$  is the number of eigenvalues that are different from zero.

Finally, since  $\mathbf{X}$  in (E.8) only  $r'$  components may be necessary to characterize the data, the low-rank approximation is applied in Fig. E.10, and represented by

$$\mathbf{X} \approx \mathbf{U}'_s \cdot \mathbf{\Sigma}'_s \cdot \mathbf{V}'_s^H, \quad (\text{E.9})$$

where  $\mathbf{X} \in \mathbb{C}^{M \times N}$ ,  $\mathbf{U}'_s \in \mathbb{C}^{M \times r'}$ ,  $\mathbf{\Sigma}'_s \in \mathbb{C}^{r' \times r'}$ , and  $\mathbf{V}'_s \in \mathbb{C}^{N \times r'}$ .

Similarly to the SVD, we can have the economy size and low-rank approximation HOSVD (or truncated HOSVD). First let us consider the full size HOSVD given by

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \quad (\text{E.10})$$

where  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$ ,  $\mathcal{S} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$ ,  $\mathbf{U}_1 \in \mathbb{C}^{M_1 \times M_1}$ ,  $\mathbf{U}_2 \in \mathbb{C}^{M_2 \times M_2}$ , and  $\mathbf{U}_3 \in \mathbb{C}^{M_3 \times M_3}$ .

Since in Fig. E.11 the core tensor  $\mathcal{S} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$  contains zeros when  $m_r > d_r$  for  $r = 1, \dots, R$ , it is possible to replace it by a smaller tensor without zeros  $\mathcal{S}^{[s]} \in \mathbb{C}^{t_1 \times t_2 \times \dots \times t_R}$ , where we define  $t_r = \min(d_r, M_r)$ .  $d_r$  is the number of nonzero elements in the main diagonal of the matrix obtained via the following product  $\mathcal{S}^{[s]} \cdot (\mathcal{S}^{[s]})^H$ . Therefore, we define the economy size HOSVD by reducing the dimensions of the core tensor as in Fig. E.11, and also of the *columns* of the matrices  $\mathbf{U}_r$ .

$$\mathcal{X} = \mathcal{S}^{[s]} \times_1 \mathbf{U}_1^{[s]} \times_2 \mathbf{U}_2^{[s]} \times_3 \mathbf{U}_3^{[s]}, \quad (\text{E.11})$$

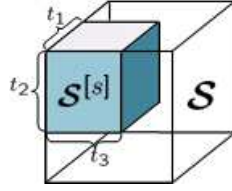


Fig. E.11: Economy size and low-rank approximation HOSVD of the core tensor  $\mathcal{S}$

where  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$ ,  $\mathcal{S}^{[s]} \in \mathbb{C}^{t_1 \times t_2 \times \dots \times t_R}$ ,  $\mathbf{U}_1^{[s]} \in \mathbb{C}^{M_1 \times t_1}$ ,  $\mathbf{U}_2^{[s]} \in \mathbb{C}^{M_2 \times t_2}$ , and  $\mathbf{U}_3^{[s]} \in \mathbb{C}^{M_3 \times t_3}$ ,

Finally, the low-rank HOSVD approximation of (E.11) is given by

$$\mathcal{X} = \mathcal{S}^{[s]'} \times_1 \mathbf{U}_1^{[s]'} \times_2 \mathbf{U}_2^{[s]'} \times_3 \mathbf{U}_3^{[s]'}, \quad (\text{E.12})$$

where  $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}$ ,  $\mathcal{S}^{[s]'} \in \mathbb{C}^{t'_r \times t'_r \times \dots \times t'_r}$ ,  $\mathbf{U}_1^{[s]'} \in \mathbb{C}^{M_1 \times t'_r}$ ,  $\mathbf{U}_2^{[s]'} \in \mathbb{C}^{M_2 \times t'_r}$ , and  $\mathbf{U}_3^{[s]'} \in \mathbb{C}^{M_3 \times t'_r}$ . We define  $t'_r = \min(d'_r, M_r)$ .

## Appendix F

# IMPORTANT DEFINITIONS AND RELATIONS FOR TENSORS

In this appendix, we present important relations that are used in tensors. The Kronecker product and the Khatri-Rao product are commonly used in other sections. Therefore, we start defining both.

Given two matrices  $\mathbf{A} \in \mathbb{C}^{M_1 \times N_1}$  and  $\mathbf{B} \in \mathbb{C}^{M_2 \times N_2}$ , the Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$  denoted by  $\mathbf{A} \otimes \mathbf{B}$  is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1} \cdot \mathbf{B} & a_{1,2} \cdot \mathbf{B} & \dots & a_{1,N_1} \cdot \mathbf{B} \\ a_{2,1} \cdot \mathbf{B} & a_{2,2} \cdot \mathbf{B} & \dots & a_{2,N_1} \cdot \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M_1,1} \cdot \mathbf{B} & a_{M_1,2} \cdot \mathbf{B} & \dots & a_{M_1,N_1} \cdot \mathbf{B} \end{bmatrix} \in \mathbb{C}^{(M_1 \cdot M_2) \times (N_1 \cdot N_2)}, \quad (\text{F.1})$$

where the resulting matrix has  $M_1 \times N_1$  blocks each with size  $M_2 \times N_2$ .

Given two matrices  $\mathbf{A} \in \mathbb{C}^{M_1 \times N}$  and  $\mathbf{B} \in \mathbb{C}^{M_2 \times N}$ , the Khatri-Rao product also known as column-wise Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$  denoted by  $\mathbf{A} \diamond \mathbf{B}$  is defined as

$$\mathbf{A} \diamond \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \dots & \mathbf{a}_N \otimes \mathbf{b}_N \end{bmatrix} \in \mathbb{C}^{(M_1 \cdot M_2) \times N}, \quad (\text{F.2})$$

where  $\mathbf{a}_i$  and  $\mathbf{b}_i$  denotes the  $i$ -th columns of the matrices  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, and

$$\mathbf{a}_i \otimes \mathbf{b}_i = \begin{bmatrix} a_{1,i} \cdot \mathbf{b}_i \\ a_{2,i} \cdot \mathbf{b}_i \\ \dots \\ a_{M_1,i} \cdot \mathbf{b}_i \end{bmatrix} \in \mathbb{C}^{(M_1 \cdot M_2) \times 1}. \quad (\text{F.3})$$

The following important identity relates the  $\text{vec}(\cdot)$  operator to the Kronecker products [40]. Given the matrices  $\mathbf{A}_1 \in \mathbb{C}^{M_1 \times M_2}$ ,  $\mathbf{A}_2 \in \mathbb{C}^{M_2 \times M_3}$ , and  $\mathbf{A}_3 \in \mathbb{C}^{M_3 \times M_4}$ ,

$$\text{vec}(\mathbf{A}_1 \cdot \mathbf{A}_2 \cdot \mathbf{A}_3) = (\mathbf{A}_3^T \otimes \mathbf{A}_1) \cdot \text{vec}(\mathbf{A}_2). \quad (\text{F.4})$$

**Lemma 5.** *An important relation between the Kronecker product and the Khatri-Rao product can be found by using a selection matrix  $\mathbf{J} \in \mathbb{R}^{N^2 \times N}$ . Given two matrices  $\mathbf{A} \in \mathbb{C}^{M_1 \times N}$  and*

---

$\mathbf{B} \in \mathbb{C}^{M_2 \times N}$ , the relation between the Khatri-Rao product and the Kronecker product is given by

$$\mathbf{A} \diamond \mathbf{B} = (\mathbf{A} \otimes \mathbf{B}) \cdot \mathbf{J} \in \mathbb{C}^{(M_1 \cdot M_2) \times N}, \quad (\text{F.5})$$

where  $\mathbf{J}$  is the  $r$ -mode unfolding of the identity tensor  $\mathcal{I}_{3,N} \in \mathbb{R}^{N \times N \times N}$  for  $r = 1, 2$ , or  $3$ .

*Proof.*

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{B} & \mathbf{a}_2 \otimes \mathbf{B} & \dots & \mathbf{a}_N \otimes \mathbf{B} \end{bmatrix} \in \mathbb{C}^{(M_1 \cdot M_2) \times N^2}, \quad (\text{F.6})$$

By comparing (F.2) to (F.6), we can take the  $i$ -th blocks of both equations, which are given by  $\mathbf{a}_i \otimes \mathbf{b}_i$  and by  $\mathbf{a}_i \otimes \mathbf{B}$ , respectively. Therefore, the selection vector  $\mathbf{j}_i$  can be used to select only the  $i$ -th column of  $\mathbf{a}_i \otimes \mathbf{B}$  as follows

$$\mathbf{a}_i \otimes \mathbf{b}_i = (\mathbf{a}_i \otimes \mathbf{B}) \cdot \mathbf{j}_i \in \mathbb{C}^{(M_1 \cdot M_2) \times 1}, \quad (\text{F.7})$$

where  $\mathbf{j}_i \in \mathbb{R}^{N \times 1}$  has all elements equal to zero, except at the  $i$ -th position which is equal to 1. Therefore, by concatenating all the  $N$  blocks are in (F.2), we have the following identity

$$\begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \dots & \mathbf{a}_N \otimes \mathbf{b}_N \end{bmatrix} = \begin{bmatrix} (\mathbf{a}_1 \otimes \mathbf{B}) \cdot \mathbf{j}_1 & (\mathbf{a}_2 \otimes \mathbf{B}) \cdot \mathbf{j}_2 & \dots & (\mathbf{a}_N \otimes \mathbf{B}) \cdot \mathbf{j}_N \end{bmatrix}, \quad (\text{F.8})$$

which can be written as

$$\mathbf{A} \diamond \mathbf{B} = (\mathbf{A} \otimes \mathbf{B}) \cdot \mathbf{J} \in \mathbb{C}^{(M_1 \cdot M_2) \times N}, \quad (\text{F.9})$$

where

$$\mathbf{J} = \begin{bmatrix} \mathbf{j}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{j}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{j}_N \end{bmatrix} \in \mathbb{R}^{N^2 \times N}, \quad (\text{F.10})$$

where  $\mathbf{0} \in \mathbb{R}^{N \times 1}$  is a zero vector. Note that  $\mathbf{J} = [\mathcal{I}_{3,N}]_{(r)}$  for  $r = 1, 2, 3$ , i.e.,  $\mathbf{J}$  can be also obtained as the 1-mode, 2-mode or 3-mode unfolding of the identity tensor  $\mathcal{I}_{3,N} \in \mathbb{R}^{N \times N \times N}$ , since  $[\mathcal{I}_{3,N}]_{(1)} = [\mathcal{I}_{3,N}]_{(2)} = [\mathcal{I}_{3,N}]_{(3)}$ . The identity tensor has elements that are equal to one when the indices  $i_1 = i_2 = i_3$  and zero otherwise. More information about tensor operations is provided in Appendix E.

□

In Fig. F.1, we show a representation of a 3-dimensional identity tensor. The main diagonal, where all the indices  $i_1 = i_2 = i_3$ , contains ones, and the other elements with zeros. In Appendix E, the  $r$ -mode product is defined. Using this definition here we depict the rep-



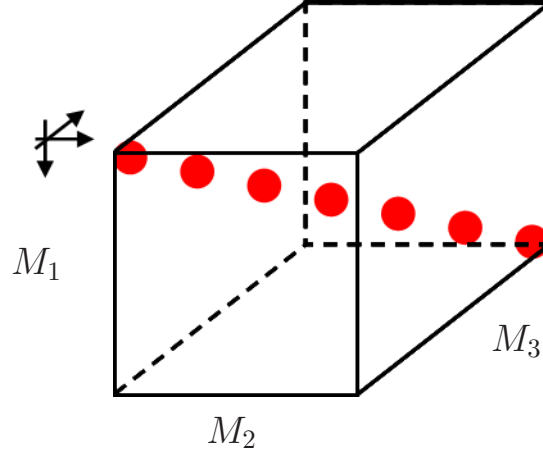


Fig. F.1: Representation of the identity tensor  $\mathcal{I}_{R,N}$ , where  $R = 3$ .

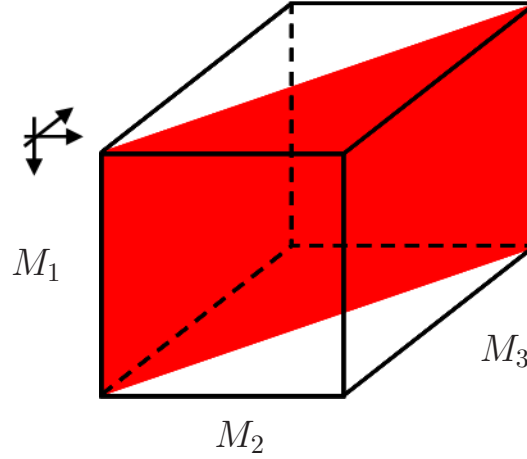


Fig. F.2: Representation of  $\mathcal{I}_{3,N} \times_1 \mathbf{F}^{(1)}$ .

representation of each  $n$ -mode product between the identity tensor and a certain factor matrix  $\mathbf{F}^{(r)}$  for  $r = 1, 2, 3$ .

In Fig. F.2, we define  $\mathcal{X}' = \mathcal{I}_{3,N} \times_1 \mathbf{F}^{(1)}$ , whose 1-mode unfolding can be represented as

$$\begin{aligned} [\mathcal{X}']_{(1)} &= [\mathcal{I}_{3,N} \times_1 \mathbf{F}^{(1)}]_{(1)}, \\ &= \begin{bmatrix} \mathbf{F}_1^{(1)} \cdot \mathbf{J}_1^{(U)} & \mathbf{F}_1^{(1)} \cdot \mathbf{J}_2^{(U)} & \dots & \mathbf{F}_1^{(1)} \cdot \mathbf{J}_N^{(U)} \end{bmatrix}, \end{aligned} \quad (\text{F.11})$$

where the selection matrix  $\mathbf{J}_i^{(U)} \in \mathbb{R}^{N \times N}$  has the element  $j_{i,i} = 1$ , and the other elements equal to zero. Therefore, the product  $\mathbf{F}_i^{(1)} \cdot \mathbf{J}_i^{(U)}$  gives us a matrix with the  $i$ -th column equal to the  $i$ -th column of  $\mathbf{F}^{(1)}$ , and the other columns have only zero elements. After applying the product and computing the inverse unfolding of  $[\mathcal{X}']_{(1)}$  according to Fig. E.6 in Appendix E, we obtain the tensor in Fig. F.2.

In Figs. F.3 and F.4, we have  $\mathcal{X}'' = \mathcal{I}_{3,N} \times_2 \mathbf{F}^{(2)}$  and  $\mathcal{X}''' = \mathcal{I}_{3,N} \times_3 \mathbf{F}^{(3)}$ , i.e., the same

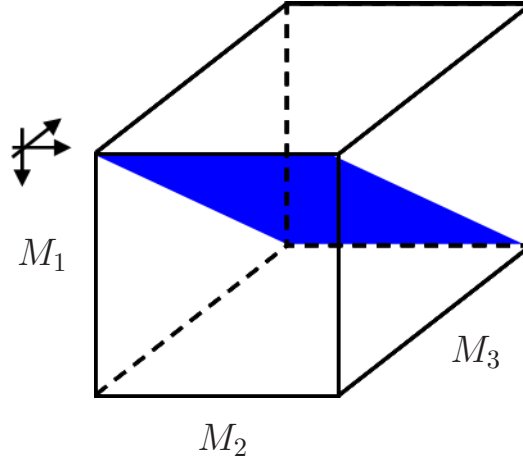


Fig. F.3: Representation of  $\mathcal{I}_{3,N} \times_2 \mathbf{F}^{(2)}$ .

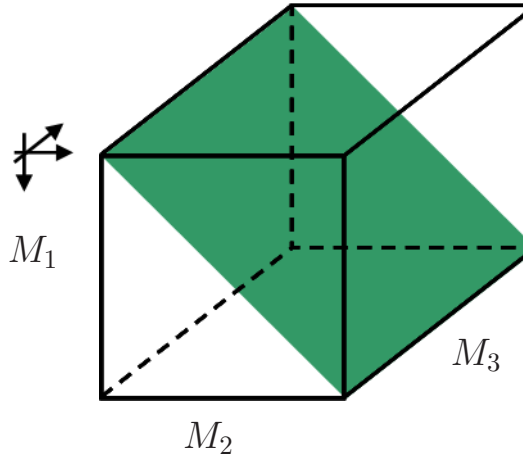


Fig. F.4: Representation of  $\mathcal{I}_{3,N} \times_3 \mathbf{F}^{(3)}$ .

procedure done for  $\mathcal{X}'$  in Fig. F.2 is repeated for  $n = 2$ , and for  $n = 3$ , and different structures are obtained for the tensors.

Based on Fig F.3 and by computing  $[\mathcal{I}_{3,N} \times_2 \mathbf{F}^{(2)}]_{(3)}$ , we can find the following structure

$$\begin{aligned} [\mathcal{X}'']_{(3)} &= [\mathcal{I}_{3,N} \times_2 \mathbf{F}^{(2)}]_{(3)}, \\ &= \begin{bmatrix} \mathbf{J}_1^{(U)} \cdot (\mathbf{F}^{(2)})^T & \mathbf{J}_2^{(U)} \cdot (\mathbf{F}^{(2)})^T & \dots & \mathbf{J}_N^{(U)} \cdot (\mathbf{F}^{(2)})^T \end{bmatrix}, \end{aligned} \quad (\text{F.12})$$

since  $\mathbf{J}_1^{(U)} = (\mathbf{J}_1^{(U)})^T$ , we obtain

$$\begin{aligned} [\mathcal{X}'']_{(3)} &= [\mathcal{I}_{3,N} \times_2 \mathbf{F}^{(2)}]_{(3)} \\ &= \begin{bmatrix} (\mathbf{F}^{(2)} \cdot \mathbf{J}_1^{(U)})^T & (\mathbf{F}^{(2)} \cdot \mathbf{J}_2^{(U)})^T & \dots & (\mathbf{F}^{(2)} \cdot \mathbf{J}_N^{(U)})^T \end{bmatrix} \\ &= (\mathbf{I}_N \diamond \mathbf{F}^{(2)})^T, \end{aligned} \quad (\text{F.13})$$

which can be also written according to (F.9) as

$$[\mathcal{X}'']_{(3)} = [\mathcal{I}_{3,N}]_{(3)}^T \cdot \left( \mathbf{I}_N \otimes \mathbf{F}^{(2)} \right)^T. \quad (\text{F.14})$$

Similarly by computing  $[\mathcal{X}'']_{(1)}$ , we have that

$$[\mathcal{X}'']_{(1)} = [\mathcal{I}_{3,N}]_{(1)}^T \cdot \left( \mathbf{F}^{(2)} \otimes \mathbf{I}_N \right)^T. \quad (\text{F.15})$$

Now we define

$$\mathcal{X} = \mathcal{I}_{3,N} \times_1 \mathbf{F}^{(1)} \times_2 \mathbf{F}^{(2)} \times_3 \mathbf{F}^{(3)}. \quad (\text{F.16})$$

The case in (F.16) can be seen as the superposition of the cases  $\mathcal{X}'$ ,  $\mathcal{X}''$ , and  $\mathcal{X}'''$ . Therefore, by using the previous results, the following identities hold for each unfolding of  $\mathcal{X}$ .

$$\begin{aligned} [\mathcal{X}]_{(3)} &= \mathbf{F}^{(3)} \cdot \left[ \left( \mathbf{F}^{(1)} \diamond \mathbf{F}^{(2)} \right) \right]^T \\ [\mathcal{X}]_{(2)} &= \mathbf{F}^{(2)} \cdot \left[ \left( \mathbf{F}^{(3)} \diamond \mathbf{F}^{(1)} \right) \right]^T \\ [\mathcal{X}]_{(1)} &= \mathbf{F}^{(1)} \cdot \left[ \left( \mathbf{F}^{(2)} \diamond \mathbf{F}^{(3)} \right) \right]^T, \end{aligned} \quad (\text{F.17})$$

which can be applied to estimate iteratively all the factor matrices as done in ALS in Subsection 2.2.2 and in (3.10) in Subsection 3.2.

By applying the  $\text{vec}(\cdot)$  on the  $[\mathcal{X}]_{(3)}$  according to (F.17), the following expression can be found

$$\text{vec}([\mathcal{X}]_{(3)}) = \text{vec} \left( \mathbf{F}^{(3)} \cdot [\mathcal{I}_{3,N}]_{(3)} \cdot \left[ \left( \mathbf{F}^{(1)} \otimes \mathbf{F}^{(2)} \right) \right]^T \right), \quad (\text{F.18})$$

where applying (F.4), we can rewrite as

$$\text{vec}([\mathcal{X}]_{(3)}) = \left( \mathbf{F}^{(1)} \otimes \mathbf{F}^{(2)} \otimes \mathbf{F}^{(3)} \right) \cdot \text{vec}([\mathcal{I}_{3,N}]_{(3)}). \quad (\text{F.19})$$

The expression in (F.19) is the basis for the CORCONDIA as shown in Subsection 2.2.2.



## Appendix G

### GENERALIZED SVD (GSVD)

The Generalized SVD (GSVD) [39] of two matrices  $\mathbf{A}^H \in \mathbb{C}^{M \times N}$  with  $M > N$  and  $\mathbf{B}^H \in \mathbb{C}^{P \times N}$  is a very important technique for optimization problems and for prewhitening, particularly, when these two matrices  $\mathbf{A}$  and  $\mathbf{B}$  are rank deficient. In Figs. G.1 and G.2, we have depicted the GSVD of  $\mathbf{A}^H$  and  $\mathbf{B}^H$ , where for  $\mathbf{A}^H$  we have the following expression

$$\mathbf{A}^H = \mathbf{U}^{(A)H} \cdot \mathbf{\Xi}^{(A)T} \cdot \mathbf{V}^H, \quad (\text{G.1})$$

and for  $\mathbf{B}^H$  we have that

$$\mathbf{B}^H = \mathbf{U}^{(B)H} \cdot \mathbf{\Xi}^{(B)T} \cdot \mathbf{V}^H, \quad (\text{G.2})$$

where  $\mathbf{\Xi}^{(A)} \in \mathbb{C}^{N \times M}$  and  $\mathbf{\Xi}^{(B)} \in \mathbb{C}^{P \times M}$  have all elements equal to zero, except the main diagonals.  $\mathbf{U}^{(A)} \in \mathbb{C}^{M \times M}$  and  $\mathbf{U}^{(B)} \in \mathbb{C}^{P \times P}$  are orthogonal matrices and  $\mathbf{V} \in \mathbb{C}^{N \times N}$  is invertible. The vectors of  $\mathbf{V}$  are referred to the generalized singular vectors of the pair  $(\mathbf{A}^H, \mathbf{B}^H)$ .

The diagonal elements of  $\mathbf{\Xi}^{(A)}$  are represented by  $\xi_i^{(A)}$  and in the same way the elements of  $\mathbf{\Xi}^{(B)}$  are represented by  $\xi_i^{(B)}$  for  $i = 1, \dots, q$ , where  $q = \min(P, N)$ . The generalized singular

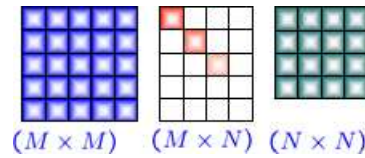


Fig. G.1: The decomposition  $\mathbf{A}^H = \mathbf{U}^{(A)H} \cdot \mathbf{\Xi}^{(A)T} \cdot \mathbf{V}^H$  is obtained from the GSVD of  $\mathbf{A}^H$  and  $\mathbf{B}^H$

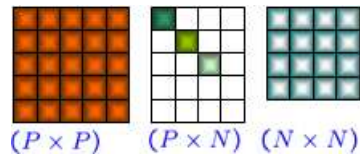


Fig. G.2: The decomposition  $\mathbf{B}^H = \mathbf{U}^{(B)H} \cdot \mathbf{\Xi}^{(B)T} \cdot \mathbf{V}^H$  is obtained from GSVD of  $\mathbf{A}^H$  and  $\mathbf{B}^H$

values are defined as  $\sigma(\mathbf{A}, \mathbf{B}) = \left\{ \frac{\xi_1^{(A)}}{\xi_1^{(B)}}, \frac{\xi_2^{(A)}}{\xi_2^{(B)}}, \dots, \frac{\xi_q^{(A)}}{\xi_q^{(B)}} \right\}$ .

Note that in the case that  $\mathbf{A}$  and  $\mathbf{B}$  are not rank deficient, we have that

$$\mathbf{B}^{-1} \cdot \mathbf{A} = \mathbf{U}^{(B)\text{H}} \cdot \mathbf{\Xi}^{(B)-1} \cdot \mathbf{\Xi}^{(A)} \cdot \mathbf{U}^{(A)}, \quad (\text{G.3})$$

if  $\mathbf{B}^{-1}$  exists.

## Appendix H

# APPLICATIONS CONSIDERING KRONECKER COLORED NOISE

In this Appendix, we show how the data model assumed is related to the data model used in MIMO and EEG applications.

In military applications using MIMO systems, besides the transmitter and the receiver, an interferer may exist, which is referred as the jammer. By considering a scenario with receiver, transmitter and jammer, in [76], the following data model is assumed

$$\mathbf{Y} = \mathbf{H} \cdot \mathbf{S} + \mathbf{H}^{(J)} \cdot \mathbf{S}^{(J)}, \quad (\text{H.1})$$

where  $\mathbf{Y} \in \mathbb{C}^{M_{\text{RX}} \times N}$  are the received symbols by the receiver array with  $M_{\text{RX}}$  antennas during  $N$  time instants,  $\mathbf{H} \in \mathbb{C}^{M_{\text{RX}} \times M_{\text{TX}}}$  is the channel between the transmitter and the receiver, and  $\mathbf{S} \in \mathbb{C}^{M_{\text{TX}} \times N}$  are the transmitted symbols from the transmitter array with  $M_{\text{TX}}$  antennas during  $N$  time instants. The term  $\mathbf{H}^{(J)} \cdot \mathbf{S}^{(J)}$  is related to the jammer, and are the matrix  $\mathbf{H}^{(J)} \in \mathbb{C}^{M_{\text{J}} \times M_{\text{TX}}}$  is the channel between the jammer and the receiver and  $\mathbf{S}^{(J)} \in \mathbb{C}^{M_{\text{J}} \times N}$  are the transmitted symbols by the jammer. Note that compared to the data model presented in Chapters 2 and 4, we have that

$$\mathbf{N}^{(c)} = \mathbf{H}^{(J)} \cdot \mathbf{S}^{(J)}. \quad (\text{H.2})$$

The symbols  $\mathbf{S}^{(J)}$  transmitted by the jammer are assumed to be correlated only in the time dimension, and uncorrelated in the space dimension [76]. Therefore, as in [76], by computing

---

the covariance matrix of two certain  $i$ -th and  $j$ -th columns of  $\mathbf{S}^{(J)}$  gives us

$$\begin{aligned}
\mathbb{E} \left\{ \mathbf{s}_i^{(J)} \cdot \left( \mathbf{s}_j^{(J)} \right)^H \right\} &= \mathbb{E} \left\{ \begin{bmatrix} s_{i,1}^{(J)} \\ s_{i,2}^{(J)} \\ \vdots \\ s_{i,M_J}^{(J)} \end{bmatrix} \cdot \begin{bmatrix} \left( s_{j,1}^{(J)} \right)^* & \left( s_{j,2}^{(J)} \right)^* & \cdots & \left( s_{j,M_J}^{(J)} \right)^* \end{bmatrix} \right\} \\
&= \begin{bmatrix} \mathbb{E} \left\{ \mathbf{s}_{i,1}^{(J)} \cdot \left( s_{j,1}^{(J)} \right)^* \right\} & 0 & 0 & 0 \\ 0 & \mathbb{E} \left\{ \mathbf{s}_{i,2}^{(J)} \cdot \left( s_{j,2}^{(J)} \right)^* \right\} & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \mathbb{E} \left\{ \mathbf{s}_{i,M_J}^{(J)} \cdot \left( s_{j,M_J}^{(J)} \right)^* \right\} \end{bmatrix} \\
&= \rho_{(i,j)}^{(J)} \cdot \mathbf{I}_{M_J}. \tag{H.3}
\end{aligned}$$

Therefore, the matrix  $\mathbb{E} \left\{ \mathbf{s}_i^{(J)} \cdot \left( \mathbf{s}_j^{(J)} \right)^H \right\}$  in (H.3) is the correlation coefficient  $\rho_{(i,j)}^{(J)}$  of the  $i$  and  $j$  temporal snapshots times the identity matrix  $\mathbf{I}_{M_J}$ . As in [76], here we define the correlation matrix  $\mathbf{Q}_N$ , whose the element at the position  $(i, j)$  is equal to  $\rho_{(i,j)}^{(J)}$ .

By applying the  $\text{vec}(\cdot)$  operator in  $\mathbf{S}^{(J)}$ , we can compute the correlation matrix

$$\begin{aligned}
\mathbb{E} \left\{ \mathbf{S}^{(J)} \cdot \left( \mathbf{S}^{(J)} \right)^H \right\} &= \begin{bmatrix} \mathbf{s}_1^{(J)} \\ \mathbf{s}_2^{(J)} \\ \vdots \\ \mathbf{s}_N^{(J)} \end{bmatrix} \cdot \begin{bmatrix} \left( \mathbf{s}_1^{(J)} \right)^H & \left( \mathbf{s}_2^{(J)} \right)^H & \cdots & \left( \mathbf{s}_N^{(J)} \right)^H \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E} \left\{ \mathbf{s}_1^{(J)} \cdot \left( \mathbf{s}_1^{(J)} \right)^H \right\} & \mathbb{E} \left\{ \mathbf{s}_1^{(J)} \cdot \left( \mathbf{s}_2^{(J)} \right)^H \right\} & \cdots & \mathbb{E} \left\{ \mathbf{s}_1^{(J)} \cdot \left( \mathbf{s}_N^{(J)} \right)^H \right\} \\ \mathbb{E} \left\{ \mathbf{s}_2^{(J)} \cdot \left( \mathbf{s}_1^{(J)} \right)^H \right\} & \mathbb{E} \left\{ \mathbf{s}_2^{(J)} \cdot \left( \mathbf{s}_2^{(J)} \right)^H \right\} & \cdots & \mathbb{E} \left\{ \mathbf{s}_2^{(J)} \cdot \left( \mathbf{s}_N^{(J)} \right)^H \right\} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbb{E} \left\{ \mathbf{s}_N^{(J)} \cdot \left( \mathbf{s}_1^{(J)} \right)^H \right\} & \mathbb{E} \left\{ \mathbf{s}_N^{(J)} \cdot \left( \mathbf{s}_2^{(J)} \right)^H \right\} & \cdots & \mathbb{E} \left\{ \mathbf{s}_N^{(J)} \cdot \left( \mathbf{s}_N^{(J)} \right)^H \right\} \end{bmatrix} \tag{H.4}
\end{aligned}$$

Applying (H.3) in (H.4), we have that

$$\mathbb{E} \left\{ \mathbf{S}^{(J)} \cdot \left( \mathbf{S}^{(J)} \right)^H \right\} = \begin{bmatrix} \rho_{(1,1)}^{(J)} \cdot \mathbf{I}_{M_J} & \rho_{(1,2)}^{(J)} \cdot \mathbf{I}_{M_J} & \cdots & \rho_{(1,N)}^{(J)} \cdot \mathbf{I}_{M_J} \\ \rho_{(2,1)}^{(J)} \cdot \mathbf{I}_{M_J} & \rho_{(2,2)}^{(J)} \cdot \mathbf{I}_{M_J} & \cdots & \rho_{(2,N)}^{(J)} \cdot \mathbf{I}_{M_J} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{(N,1)}^{(J)} \cdot \mathbf{I}_{M_J} & \rho_{(N,2)}^{(J)} \cdot \mathbf{I}_{M_J} & \cdots & \rho_{(N,N)}^{(J)} \cdot \mathbf{I}_{M_J} \end{bmatrix} \tag{H.5}$$

The equation (H.5) can be represented in a compact form by using the Kronecker product as follows

$$\mathbb{E} \left\{ \mathbf{S}^{(J)} \cdot \left( \mathbf{S}^{(J)} \right)^H \right\} = \mathbf{Q}_N \otimes \mathbf{I}_{M_J}. \tag{H.6}$$



The total covariance matrix  $\mathbf{Q}$  of  $\mathbf{N}^{(c)}$  is defined as

$$\mathbf{Q} = \mathbb{E} \left\{ \text{vec}(\mathbf{N}^{(c)}) \cdot [\text{vec}(\mathbf{N}^{(c)})]^H \right\} \quad (\text{H.7})$$

$$= \mathbb{E} \left\{ \text{vec}(\mathbf{H}^{(J)} \cdot \mathbf{S}^{(J)}) \cdot [\text{vec}(\mathbf{H}^{(J)} \cdot \mathbf{S}^{(J)})]^H \right\}, \quad (\text{H.8})$$

using  $\mathbf{N}^{(c)}$  according to (H.2). According to [50], we have that

$$\text{vec}(\mathbf{H}^{(J)} \cdot \mathbf{S}^{(J)}) = \mathbf{I}_N \otimes \mathbf{H}^{(J)} \cdot \text{vec}(\mathbf{S}^{(J)}), \quad (\text{H.9})$$

where (H.9) can be obtained via (F.4), and in this case  $\mathbf{A}_1 = \mathbf{H}_J$ ,  $\mathbf{A}_2 = \mathbf{S}_J$ , and  $\mathbf{A}_3 = \mathbf{I}_N$ .

By replacing (H.9) in (H.8), it give us

$$\begin{aligned} \mathbf{Q} &= \mathbb{E} \left\{ \mathbf{I}_N \otimes \mathbf{H}^{(J)} \cdot \text{vec}(\mathbf{S}^{(J)}) \cdot \text{vec}(\mathbf{S}^{(J)})^H \cdot (\mathbf{I}_N \otimes \mathbf{H}^{(J)})^H \right\} \\ &= \mathbf{I}_N \otimes \mathbf{H}^{(J)} \cdot \mathbb{E} \left\{ \text{vec}(\mathbf{S}^{(J)}) \cdot \text{vec}(\mathbf{S}^{(J)})^H \right\} \cdot (\mathbf{I}_N \otimes \mathbf{H}^{(J)})^H \\ &= \mathbf{Q}_N \otimes \left[ \mathbf{H}^{(J)} \cdot (\mathbf{H}^{(J)})^H \right] \\ &= \mathbf{Q}_N \otimes \mathbf{Q}^{(J)}, \end{aligned} \quad (\text{H.10})$$

where  $\mathbf{Q}^{(J)} = \mathbf{H}^{(J)} \cdot (\mathbf{H}^{(J)})^H$  as done in [76].

By comparing to (2.43), we have that  $\mathbf{Q}_N = \mathbf{L}_1 \cdot \mathbf{L}_1^H$ , and that  $\mathbf{Q}^{(J)} = \mathbf{L}_2 \cdot \mathbf{L}_2^H$ . In order to have a tensor structure, a third dimension related to the trials or realizations can be included according to [52].

The similar problem is found in EEG/MEG applications [52], where the following data model is given

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{S} + \mathbf{N}^{(c)}, \quad (\text{H.11})$$

where  $\mathbf{A} \in \mathbb{C}^{M \times d}$  is the matrix with unit activity of  $d$  sources, such that each source is composed of three Euclidean location parameters and three Euclidean orientation parameters,  $\mathbf{S} \in \mathbb{C}^{d \times N}$  gives the time-varying amplitude of each source, and  $\mathbf{N}^{(c)}$  is the colored noise matrix.

The correlation matrix  $\mathbf{Q} = \mathbb{E} \left\{ \text{vec}(\mathbf{N}^{(c)}) \cdot [\text{vec}(\mathbf{N}^{(c)})]^H \right\}$  is of size  $M \cdot N \times M \cdot N$ , which means that many trials are necessary to estimate  $\mathbf{Q}$ . Therefore, by assuming the Kronecker product of the covariances in time and in space, a lower computational complexity and a better performance can be obtained as proposed in [52].



## Appendix I

# MULTI-DIMENSIONAL COLORED NOISE

In this appendix, we show the derivation for the equivalence between (2.43), (2.44), and (2.45), which is also the basis for (4.41). It is possible to derive (2.45) by first calculating the covariance matrix of each unfolding of the noise tensor as follows

$$\begin{aligned} \mathbb{E}\left\{[\mathcal{N}^c]_{(i)} \cdot [\mathcal{N}^c]_{(i)}^H\right\} = \\ \mathbf{L}_i \cdot \mathbb{E}\{[\mathcal{N}]_{(i)} \cdot (\mathbf{L}_{i+1} \otimes \dots \otimes \mathbf{L}_R \otimes \mathbf{I}_N \otimes \mathbf{L}_1 \dots \otimes \mathbf{L}_{i-1}) \cdot \\ (\mathbf{L}_{i+1} \otimes \dots \otimes \mathbf{L}_R \otimes \mathbf{I}_N \otimes \mathbf{L}_1 \dots \otimes \mathbf{L}_{i-1})^H \cdot [\mathcal{N}]_{(i)}^H\} \cdot \mathbf{L}_i^H. \end{aligned} \quad (\text{I.1})$$

Since the elements of  $[\mathcal{N}]_{(i)}$  are independent and identically distributed (i.i.d.), we can show that

$$\begin{aligned} \mathbb{E}\{[\mathcal{N}]_{(i)} \cdot (\mathbf{L}_{i+1} \otimes \dots \otimes \mathbf{L}_R \otimes \mathbf{I}_N \otimes \mathbf{L}_1 \dots \otimes \mathbf{L}_{i-1}) \cdot \\ (\mathbf{L}_{i+1} \otimes \dots \otimes \mathbf{L}_R \otimes \mathbf{I}_N \otimes \mathbf{L}_1 \dots \otimes \mathbf{L}_{i-1})^H \cdot [\mathcal{N}]_{(i)}^H\} = \alpha \cdot \mathbf{I}_{M_i}, \end{aligned} \quad (\text{I.2})$$

Therefore, replacing (I.2) in (I.1), we realize that  $\mathbb{E}\{[\mathcal{N}^c]_{(i)} \cdot [\mathcal{N}^c]_{(i)}^H\} = \alpha \cdot \mathbf{L}_i \cdot \mathbf{L}_i^H$ . The proof for (I.2) is given in Lemma 6.

**Lemma 6.** *Let us consider a complex Gaussian random matrix  $\mathbf{N} \in \mathbb{C}^{M \times N}$  and also a constant matrix  $\mathbf{F} \in \mathbb{C}^{N \times T}$ , where  $n_{i,j}$  indicates the element of  $\mathbf{N}$  in row  $i$  and column  $j$ , and  $f_{i,j}$  indicates the element of  $\mathbf{F}$  in row  $i$  and column  $j$ , respectively. The following equality holds*

$$\mathbb{E}\{\mathbf{N} \cdot \mathbf{F} \cdot \mathbf{F}^H \cdot \mathbf{N}^H\} = \alpha \cdot \mathbf{I}_M. \quad (\text{I.3})$$

where  $\alpha = \text{tr}(\mathbf{F} \cdot \mathbf{F}^H) \cdot \sigma_n^2$

*Proof.* First let us represent the product of two matrices as a sum of rank-one matrices as follows:

$$\mathbf{N} \cdot \mathbf{F} = \sum_{k=1}^N \begin{bmatrix} n_{1,k} \\ n_{2,k} \\ \vdots \\ n_{M,k} \end{bmatrix} \cdot [f_{k,1} \ f_{k,2} \ \dots \ f_{k,T}]. \quad (\text{I.4})$$

Also the term  $(\mathbf{N} \cdot \mathbf{F})^H$  can be written as

$$(\mathbf{N} \cdot \mathbf{F})^H = \mathbf{F}^H \cdot \mathbf{N}^H = \sum_{p=1}^N \begin{bmatrix} f_{p,1}^* \\ f_{p,2}^* \\ \vdots \\ f_{p,T}^* \end{bmatrix} \cdot [n_{1,p}^* \quad n_{2,p}^* \quad \dots \quad n_{M,p}^*]. \quad (\text{I.5})$$

Therefore, the product of (I.4) and (I.5) can be represented as the product of two sums of matrices:

$$\mathbf{N} \cdot \mathbf{F} \cdot \mathbf{F}^H \cdot \mathbf{N}^H = \sum_{k=1}^N \begin{bmatrix} n_{1,k} \\ n_{2,k} \\ \vdots \\ n_{M,k} \end{bmatrix} \cdot \begin{bmatrix} f_{k,1} \\ f_{k,2} \\ \vdots \\ f_{k,T} \end{bmatrix}^T \cdot \sum_{p=1}^N \begin{bmatrix} f_{p,1}^* \\ f_{p,2}^* \\ \vdots \\ f_{p,T}^* \end{bmatrix} \cdot \begin{bmatrix} n_{1,p}^* \\ n_{2,p}^* \\ \vdots \\ n_{M,p}^* \end{bmatrix}^T. \quad (\text{I.6})$$

In order to facilitate the derivation, we separate (I.6) into two terms. The first term is represented by  $\mathbf{B}_1$  for the cases when  $p = k$ , and the second term is represented by  $\mathbf{B}_2$  for the cases when  $p \neq k$ .

$$\mathbf{N} \cdot \mathbf{F} \cdot \mathbf{F}^H \cdot \mathbf{N}^H = \mathbf{B}_1 + \mathbf{B}_2, \quad (\text{I.7})$$

where the first term is given by

$$\mathbf{B}_1 = \sum_{k=1}^N \begin{bmatrix} n_{1,k} \\ n_{2,k} \\ \vdots \\ n_{M,k} \end{bmatrix} \cdot \begin{bmatrix} f_{k,1} \\ f_{k,2} \\ \vdots \\ f_{k,T} \end{bmatrix}^T \cdot \begin{bmatrix} f_{k,1}^* \\ f_{k,2}^* \\ \vdots \\ f_{k,T}^* \end{bmatrix} \cdot \begin{bmatrix} n_{1,k}^* \\ n_{2,k}^* \\ \vdots \\ n_{M,k}^* \end{bmatrix}^T. \quad (\text{I.8})$$

Since the product between the two vectors in the middle gives a scalar, we can rewrite (I.8) in the following way

$$\mathbf{B}_1 = \sum_{k=1}^N \sum_{t=1}^T |f_{k,t}|^2 \cdot \begin{bmatrix} n_{1,k} \\ n_{2,k} \\ \vdots \\ n_{M,k} \end{bmatrix} \cdot \begin{bmatrix} n_{1,k}^* \\ n_{2,k}^* \\ \vdots \\ n_{M,k}^* \end{bmatrix}^T. \quad (\text{I.9})$$

Applying the expected value operator, we obtain

$$\mathbb{E}\{\mathbf{B}_1\} = \sum_{k=1}^N \sum_{t=1}^T |f_{k,t}|^2 \cdot \sigma_n^2 \cdot \mathbf{I}_M. \quad (\text{I.10})$$

While applying the expected value operator to  $\mathbf{B}_2$

$$\mathbb{E}\{\mathbf{B}_2\} = \mathbf{0}_{M \times M}, \quad (\text{I.11})$$

since the noise samples  $n_{i,j}$  are uncorrelated. In addition, the expected value of the sum is equal to the sum of the expected values. Therefore, we can apply the expected value to each matrix  $\mathbf{B}_1$  and  $\mathbf{B}_2$  in (I.7) separately. Moreover, using that  $\sum_{k=1}^N \sum_{t=1}^T |f_{k,t}|^2 = \text{tr}(\mathbf{F} \cdot \mathbf{F}^H)$ , we find that

$$\mathbb{E}\{\mathbf{N} \cdot \mathbf{F} \cdot \mathbf{F}^H \cdot \mathbf{N}^H\} = \text{tr}(\mathbf{F} \cdot \mathbf{F}^H) \cdot \sigma_n^2 \cdot \mathbf{I}_M. \quad (\text{I.12})$$

Since  $\text{tr}(\mathbf{F} \cdot \mathbf{F}^H)$  and  $\sigma_n^2$  are constants, just one constant  $\alpha$  can represent them.  $\square$

The expression in (I.2) is shown by inserting  $\mathbf{F} = (\mathbf{L}_{i+1} \otimes \dots \otimes \mathbf{L}_R \otimes \mathbf{I}_N \otimes \mathbf{L}_1 \dots \otimes \mathbf{L}_{i-1}) \cdot (\mathbf{L}_{i+1} \otimes \dots \otimes \mathbf{L}_R \otimes \mathbf{I}_N \otimes \mathbf{L}_1 \dots \otimes \mathbf{L}_{i-1})^H$  and  $\mathbf{N} = [\mathcal{N}]_{(i)}$  in Lemma 6.



## Appendix J

# ERROR ANALYSIS FOR THE ESTIMATION OF PARAMETERS OF MIMO CHANNELS

In this appendix, we present an *IlmProp* interface to extract the spatial frequencies from MIMO channels generated by *IlmProp* [32, 30], and we compare the extracted spatial frequencies to the estimated ones using different parameter estimation approaches.

In general, narrow bandwidth signals are assumed to derive parameter estimation schemes. However, in practice, such assumption is only an approximation, which results in errors. To evaluate the influence of these errors by using measurements is complicated, since other sources of errors are present, such as arrays calibration, synchronization between transmitter (TX) and receiver (RX), and presence of noise and interferes. Therefore, in order to isolate only the desired sources of errors and to estimate their impact, we apply our MIMO channel applying a flexible geometry-based Multi-User MIMO Channel Modeling tool called *IlmProp* [32, 30].

In order to provide our error analysis of the high-resolution estimation of parameters, let us consider the MIMO channel generated by *Ilmprop* in Fig. J.1. In this scenario, we have a  $7 \times 7$  Uniform Rectangular Array (URA) at the transmitter (TX) and another  $7 \times 7$  URA at the receiver (RX). Our model order  $d$  is equal to 5, since we have 5 paths. Moreover, we consider 10 frequency bins, five snapshots, bandwidth and carrier frequency equal to 600 kHz and 2 GHz, respectively. The sample rate is equal to 1 ms. First let us assume that the transmitter, the receiver and the scatters are static.

For the given scenario, we represent our collected samples by the tensor  $\mathcal{X} \in \mathbb{C}^{7 \times 7 \times 7 \times 7 \times 5 \times 10}$ , where the first two dimensions are related to the  $7 \times 7$  receiver (RX) array, third and fourth dimensions are related to the  $7 \times 7$  transmitter (TX) array, the fifth dimension is related to the time snapshots, and the last dimension is related to the frequency bins. Of course that the tensor can be reshaped such that the position of the time and frequency dimensions are interchanged, which gives us  $\mathcal{X} \in \mathbb{C}^{7 \times 7 \times 7 \times 7 \times 10 \times 5}$ . By applying Forward-Backward Averaging (FBA) [63, 87], briefly reviewed in Subsection 2.2.3.6, we can virtually double the number of snapshots to  $\mathcal{X} \in \mathbb{C}^{7 \times 7 \times 7 \times 7 \times 10 \times 10}$ .

To estimate the spatial frequencies, we apply the  $R$ -dimensional Unitary Tensor-ESPRIT (UTE). However, any other multi-dimensional harmonic retrieval technique can be applied with similar results. In order to evaluate only the errors of Kronecker model and narrow band

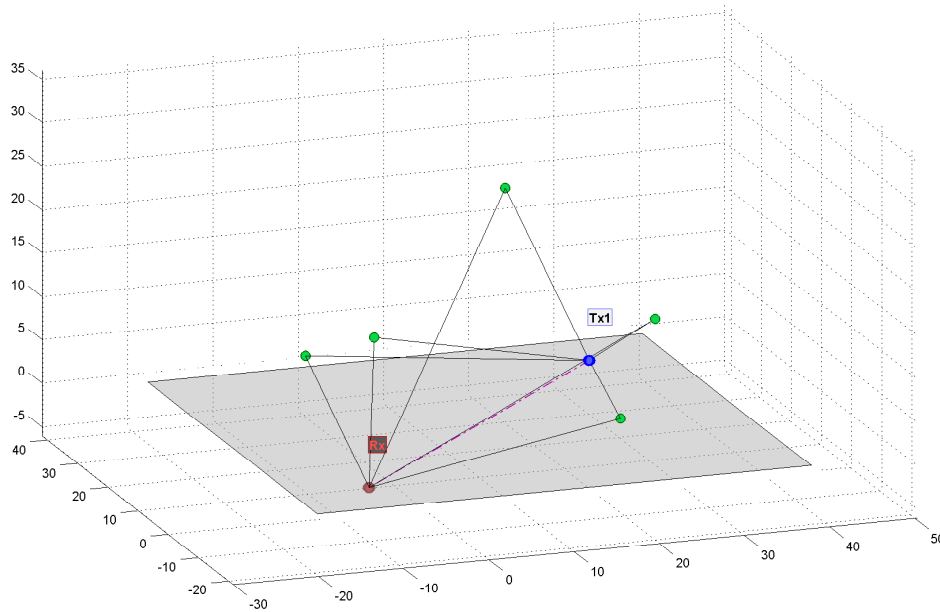


Fig. J.1: Scenario generated by *IlmProp* composed of a  $7 \times 7$  URA at RX,  $7 \times 7$  URA at TX, and 5 scatters.

assumptions, we consider that no noise is present. In Fig. J.2, we compute the error between the extracted spatial frequencies and the estimated ones by *R-D UTE* related to the azimuth and elevation at the RX, i.e., Direction-of-Arrival (DOA), while in Fig. J.3, we compute the errors for the azimuth and elevation at the TX, i.e., Direction-of-Departure (DOD).

In Fig. J.4, we consider a complex-valued random fading coefficient over each path. Therefore, due to the fading, the paths with smaller TDOA do not have a higher power, as it is expected when only path loss is taken into account. [38]

The estimation in Figs. J.2, J.3 and J.4 seems to be perfect, but actually errors are obtained as shown in scenario 1 of Table J.1. These errors are consequence of the assumptions of Kronecker channel and narrow band signals not being satisfied. The parameters of the scenario, e.g., i.e., type of paths, array size, bandwidth, play an important role in the errors computed in Table J.1.

Other scenarios are also considered in Table J.1 in order to evaluate the impact of each parameter over the assumption errors. For instance, comparing scenario 1 to scenario 2, the bandwidth is increased from 0.6 MHz to 7.2 MHz. Consequently, the errors of estimation degrades. Such degradation is expected, since the fact of increasing the bandwidth goes against the narrow band assumption.

Comparing scenario 3 to scenario 1 in Table J.1, we have changed the array size from  $7 \times 7$  to  $2 \times 2$ . Since, we observe only a slight improvement, the influence of this change is almost irrelevant.

The carrier frequency is varied from 2 GHz to 0.2 GHz, when we compare scenario 1 to scenario 4. A drastic degradation on the estimation is obtained, since by reducing the



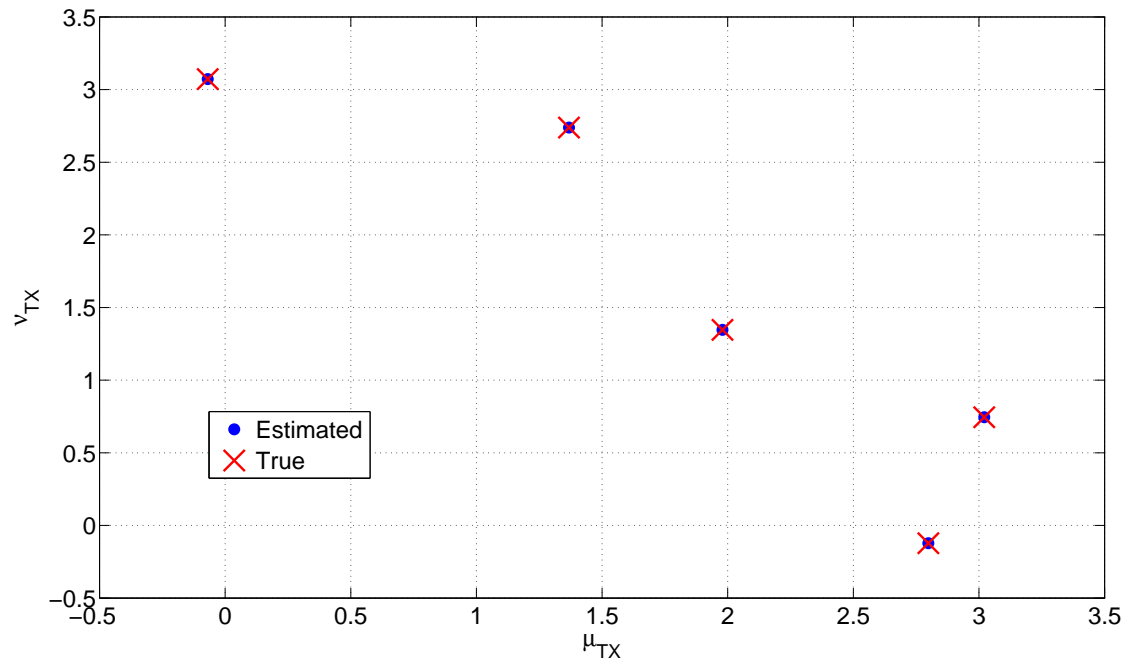


Fig. J.2: Spatial frequencies related to the azimuth and elevation at the TX, i.e., Direction-of-Departure (DOD).

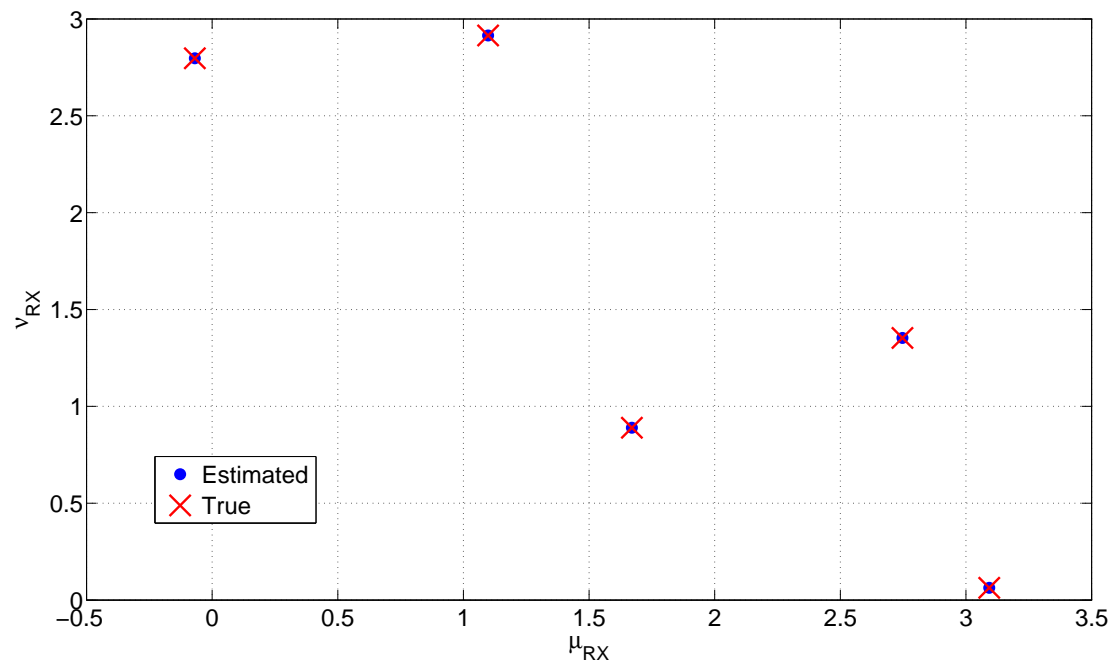


Fig. J.3: Spatial frequencies related to the azimuth and elevation at the RX, i.e., Direction-of-Arrival (DOA).

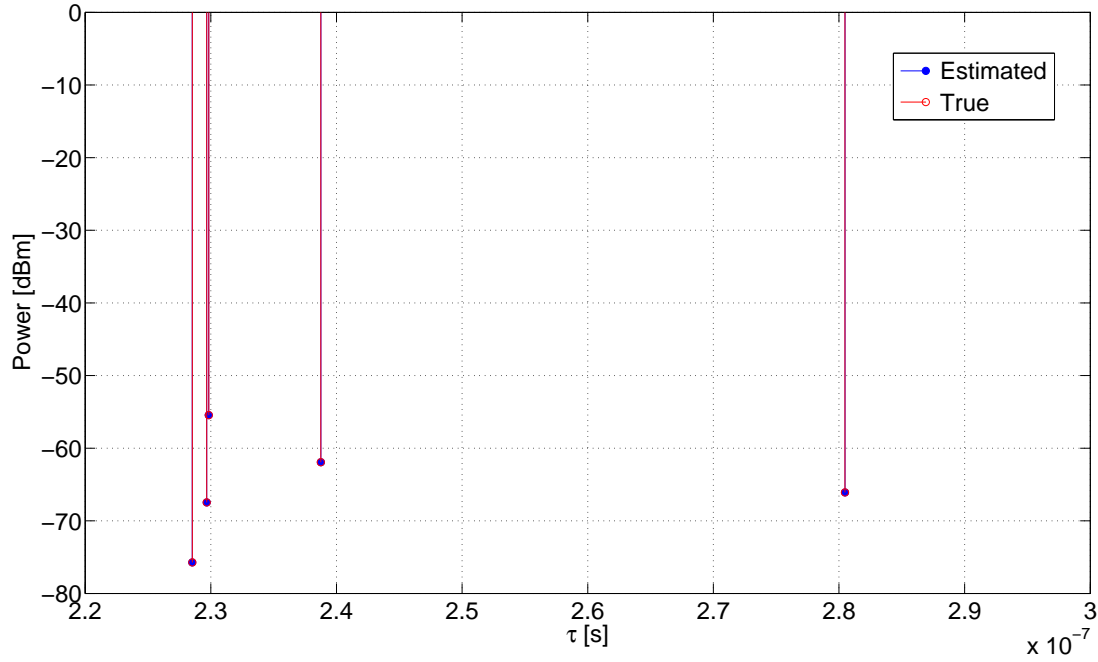


Fig. J.4: Spatial frequencies related to the frequency bins dimension, i.e., Time Delay-of-Arrival (TDOA).

	<b>TX RMSE</b>	<b>RX RMSE</b>	<b><math>\tau</math> RMSE</b>
<b>Scenario 1</b>	$3.2 \cdot 10^{-5}$	$1.1 \cdot 10^{-5}$	$2.1 \cdot 10^{-5}$
<b>Scenario 2</b>	$5.5 \cdot 10^{-3}$	$2.3 \cdot 10^{-3}$	$3.8 \cdot 10^{-4}$
<b>Scenario 3</b>	$2.2 \cdot 10^{-5}$	$3.9 \cdot 10^{-5}$	$6.7 \cdot 10^{-5}$
<b>Scenario 4</b>	$1.2 \cdot 10^{-2}$	$3.4 \cdot 10^{-3}$	$1.6 \cdot 10^{-4}$
<b>Scenario 5</b>	$7.6 \cdot 10^{-4}$	$7.9 \cdot 10^{-4}$	$2.8 \cdot 10^{-6}$
<b>Scenario 6</b>	$2.8 \cdot 10^{-6}$	$4.0 \cdot 10^{-6}$	$2.5 \cdot 10^{-6}$

Tab. J.1: Comparison between the extracted spatial frequencies of the *IlmProp* generated channel to the spatial frequencies estimated via *R-D UTE*.

carrier frequency, the relation between bandwidth and carrier frequency increase. Again this modification in the scenario goes against the assumption of narrow band signals.

Comparing scenario 1 to scenario 5 in Table J.1, the number of frequency bins is increased from 10 to 50. As expected the error in the estimation of  $\tau$  becomes smaller, since the number of units in the frequency dimension increase. However, interestingly, the estimation of DOA and DOD is slightly degraded.

The number of snapshots is changed from 10 to 50 comparing scenario 1 in Table J.1 to scenario 6, and as expected, the error of estimation becomes smaller.

Note that by applying spatial smoothing [44], only one snapshot could have been applied to compute the errors in Table J.1. However, since finding the best spatial smoothing parameter  $L$  is an open problem, we preferred not to digress from our main topic, which is the analysis of the impact of the violation of the narrow band signals and Kronecker assumptions.

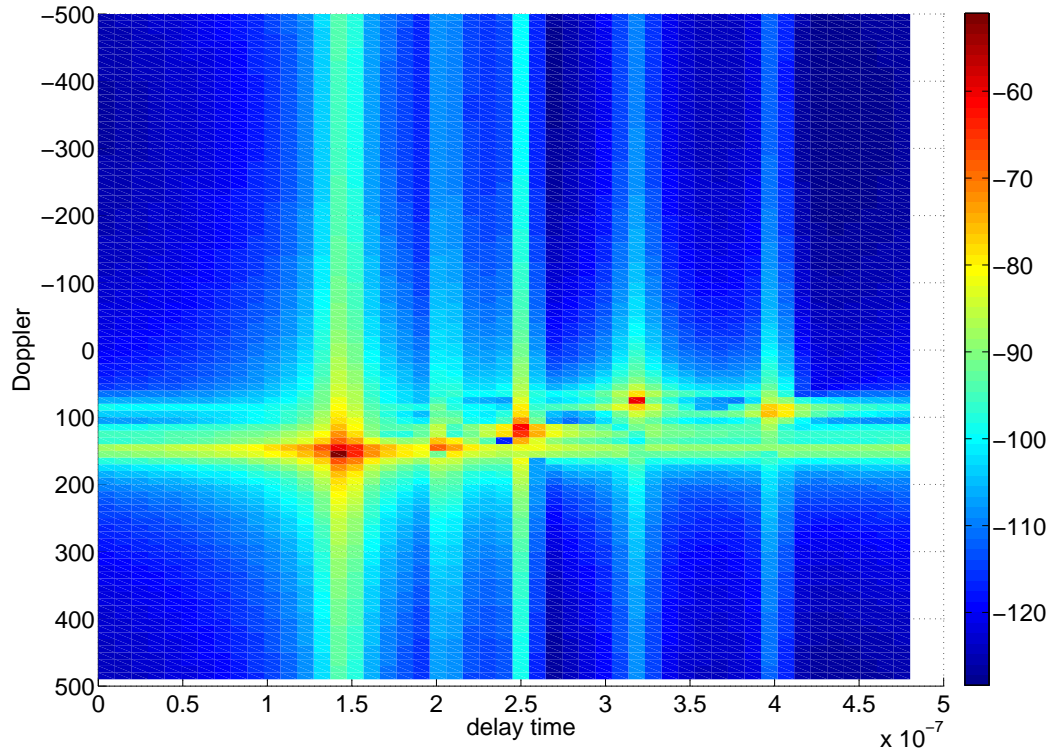


Fig. J.5: Bello function: Time Delay of Arrival (TDOA) vs. Doppler Shift (DS) using the Bello Explorer, which is a tool integrated with the *IlmProp*. The color bar is scaled in dBm.

Number of snapshots (N)	TX RMSE	RX RMSE	$\tau$ RMSE
2	$9.3 \cdot 10^{-2}$	$2.4 \cdot 10^{-1}$	$1.1 \cdot 10^{-4}$
5	$8.1 \cdot 10^{-3}$	$8.1 \cdot 10^{-2}$	$2.5 \cdot 10^{-5}$
20	$5.6 \cdot 10^{-3}$	$3.0 \cdot 10^{-2}$	$9.1 \cdot 10^{-5}$
40	$2.5 \cdot 10^{-2}$	$2.8 \cdot 10^{-1}$	$1.6 \cdot 10^{-4}$

Tab. J.2: Comparison between the extracted spatial frequencies of the *IlmProp* generated channel to the spatial frequencies estimated via *R-D UTE* in case that the MIMO channel is time-varying.

In our next example, we consider that the TX is moving with 8 km/h, and that the scatters and the RX are static. Since TX is moving, the Doppler shift caused by the TX speed should be taken into account.

In Fig. J.5, the model order can be visually estimated by the number of peaks of the Bello function of Time Delay of Arrival (TDOA) vs. Doppler Shift (DS). Unfortunately, in practice and also with *IlmProp*, a scenario like in Fig. J.5, where the model order is visually estimated, is unrealistic. Therefore, using Bello function to estimate the model order and the spatial frequencies is unlikely. The scenario in Fig. J.5 has 100 snapshots, 50 frequency bins, a carrier frequency  $f_0 = 2 \cdot 10^9$  Hz, wavelength equal to 0.15, a  $3 \times 3$  Uniform Rectangular Array (URA) at the RX and a  $3 \times 3$  URA at the TX, and time sampling equal to 1 ms.

In Table J.2 we consider again the scenario 1, but with the TX at a speed of 8 km/h. From Table J.2 and similarly to the result in Table J.1, by increasing the number of snapshots, we have that the estimation of the spatial frequencies becomes better. However, since the transmitter is moving, the angles and the size of the paths are changing. Therefore, by increasing the number of snapshots more than a certain limit, the variations caused in the paths degrade the estimation of the spatial frequencies as shown in Table J.2 for the case that  $N = 40$ . Hence, an optimal number of snapshots should be considered when Doppler shift is estimated.

In all chapters, we consider always that the TX, RX and scatters are static. Otherwise, a greater error in the estimation of other parameters model order, DOA, DOD, and TDOA, if compared to the static case. Moreover, the problem of estimating an optimal number of snapshots should also be solved. The reader is referred to [53] as an option of technique to estimate the Doppler shift.

## LIST OF FIGURES

2.1	Eigenvalues profile in case that $\text{SNR} \rightarrow \infty$ and $N \rightarrow \infty$ , and considering $M = 8$ and $d = 2$ . . . . .	16
2.2	Eigenvalues profile in case that $N \rightarrow \infty$ , and considering $\text{SNR} = 0$ dB, $M = 8$ and $d = 2$ . . . . .	16
2.3	Eigenvalues profile considering $\text{SNR} = 0$ dB, $M = 8$ , $N = 10$ and $d = 2$ . . . . .	17
2.4	Typical profile of the eigenvalues. The last $P-1$ eigenvalues are used to estimate the $(M - P)$ -th eigenvalue, denoted by a yellow square. The EFT method considers the gap between the measured eigenvalue and the predicted one. . .	19
2.5	Functional dependency between the probability of false alarm $P_{\text{fa}}(P)$ and the threshold $\eta_P$ . The plot corresponds to $M = 5$ , $N = 6$ , and $P = 1$ . . . . .	21
2.6	Selection of the first $M - 1$ sensors multiplying by the matrix $\mathbf{J}_1$ . . . . .	23
2.7	Selection of the last $M - 1$ sensors multiplying by the matrix $\mathbf{J}_2$ . . . . .	23
2.8	Probability Density Function (PDF) of the relative value $\frac{\lambda_2}{E\{\lambda_2\}}$ for an array of size $M_1 = 350$ in an only white noise scenario. The number of snapshots $N$ is set to 10. In the EFT and M-EFT, the correction version is used. . . . .	27
2.9	Probability Density Function (PDF) of the relative value $\frac{\lambda_{10}}{E\{\lambda_{10}\}}$ for an array of size $M_1 = 350$ in an only white noise scenario. The number of snapshots $N$ is set to 10. In the EFT and M-EFT, the correction version is used. . . . .	28
2.10	Probability Density Function (PDF) of the relative value $\frac{\lambda_2}{E\{\lambda_2\}}$ for an array of size $M_1 = 350$ in an only white noise scenario. The number of snapshots $N$ is set to 10. In the EFT and M-EFT, the prediction version is used. . . . .	29
2.11	Probability Density Function (PDF) of the relative value $\frac{\lambda_{10}}{E\{\lambda_{10}\}}$ for an array of size $M_1 = 350$ in an only white noise scenario. The number of snapshots $N$ is set to 10. In the EFT and M-EFT, the prediction version is used. . . . .	30
2.12	Profile of the expected-valued eigenvalues for an array of size $M = 8$ , $N = 10$ , $d = 2$ and $\text{SNR} = 25$ dB. . . . .	30
2.13	Probability of correct Detection (PoD) vs. SNR considering $M = 100$ , $N = 10$ and $d = 3$ . Comparison between the proposed 1-D ITC and ITC. . . . .	32

---

2.14	Probability of correct Detection (PoD) vs. SNR considering $M = 8$ , $N = 10$ and $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques found in the literature. . . . .	33
2.15	Probability of correct Detection (PoD) vs. SNR considering $M = 100$ , $N = 10$ and $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques found in the literature. . . . .	34
2.16	Probability of correct Detection (PoD) vs. SNR considering $M = 100$ , $N = 10$ and $d = 3$ . Comparison between the proposed M-EFT and the state-of-the-art eigenvalue based matrix-based model order selection techniques in the literature. . . . .	35
2.17	Probability of correct Detection (PoD) vs. SNR considering $M = 8$ , $N = 10$ and $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques and the subspace based ones. . . . .	36
2.18	Probability of correct Detection (PoD) vs. SNR considering $M = 100$ , $N = 10$ and $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques and the subspace based ones. . . . .	37
2.19	Probability of correct Detection vs. SNR for an array of size $M = 15$ . The number of snapshots $N$ is set to 10 and the number of sources $d = 3$ . Comparison between the state-of-the-art eigenvalue based matrix-based model order selection techniques and the subspace based ones. . . . .	38
2.20	Probability of correct Detection (PoD) vs. SNR, considering $\rho = 0.9$ , $M = 8$ , $N = 10$ and $d = 2$ . . . . .	39
2.21	Probability of correct Detection (PoD) vs. SNR, considering $\rho = 0.9$ , $M = 8$ , $N = 10$ and $d = 4$ . . . . .	39
2.22	Probability of correct Detection vs. SNR for an array of size $M_1 = 10$ . The number of snapshots $N$ is set to 20 and the number of sources $d = 3$ . . . . .	40
2.23	Probability of correct Detection vs. SNR for an array of size $M_1 = 10$ . The number of snapshots $N$ is set to 20 and the number of sources $d = 3$ . . . . .	40
2.24	PoD versus $\rho_s$ for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ . . . . .	41
2.25	PoD versus $\rho_s$ for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ . . . . .	41
2.26	Pfa versus $\rho_s$ for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ . . . . .	42
2.27	Pfa versus $\rho_s$ for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ . . . . .	42
2.28	Pnd versus $\rho_s$ for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ . . . . .	43

---

2.29	Pnd versus $\rho_s$ for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ .	43
2.30	PoD versus Pfa for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ .	44
2.31	PoD versus Pfa for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We fix the number of sources $d = 3$ .	44
2.32	CORE CONSistency in % vs. the number of components considering a scenario with a data model of $M_1 = 7, M_2 = 7, M_3 = 7, M_4 = 7$ and $d = 3$ components and with SNR = 50 dB. The break point distance between the components 3 and 4 is approximately 99.6 %.	51
2.33	Comparison between the global eigenvalues profile and the $R$ -mode eigenvalues profile for a scenario with array size $M_1 = 4, M_2 = 4, M_3 = 4, M_4 = 4, M_5, d = 1$ and SNR = 0 dB.	53
2.34	Adaptive definition of the global eigenvalues - 1 <sup>st</sup> eigenvalue set	53
2.35	Adaptive definition of the global eigenvalues - 1 <sup>st</sup> and 2 <sup>nd</sup> eigenvalue sets	54
2.36	Adaptive definition of the global eigenvalues - 1 <sup>st</sup> , 2 <sup>nd</sup> and 3 <sup>rd</sup> eigenvalue sets	54
2.37	Probability of correct Detection v.s. $\Delta$ (PoD $_{\Delta}$ ) for the threshold computations of the T-CORCONDIA Fix. Varying $\Delta$ from 0 to 50 with steps of 0.1. The curve is estimated using the four first $\Delta$ thresholds, i.e., $d$ is varied from 1 to 5. In this example scenario, the data has a size of $M_1 = 7, M_2 = 7, M_3 = 7$ and $M_4 = 7$ . The SNR varies from -20 dB to 50 dB.	58
2.38	Probability of correct Detection (PoD) vs. SNR for an array of size $M_1 = 5, M_2 = 7, M_3 = 9$ and $M_4 = 10$ . The number of sources $d$ is set to 3. $R$ -D MOS versus matrix-based MOS.	65
2.39	Probability of correct Detection (PoD) vs. SNR considering a system with a data model of $M_1 = 7, M_2 = 7, M_3 = 7, M_4 = 7$ and $d = 3$ sources.	66
2.40	Probability of correct Detection (PoD) vs. SNR considering a system with a data model of $M_1 = 7, M_2 = 7, M_3 = 7, M_4 = 7$ and $d = 4$ sources.	66
2.41	Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5, M_1 = 5, M_2 = 5, M_3 = 5, M_4 = 5, M_5 = 5$ , and $N = 5$ presence of white noise. We fixed $d = 2$ .	67
2.42	Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5, M_1 = 5, M_2 = 5, M_3 = 5, M_4 = 5, M_5 = 5$ , and $N = 5$ presence of white noise. We fixed $d = 3$ .	68

---

2.43	Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5$ , $M_1 = 5$ , $M_2 = 5$ , $M_3 = 5$ , $M_4 = 5$ , $M_5 = 5$ , and $N = 5$ presence of colored noise, where $\rho_1 = 0.9$ , $\rho_2 = 0.95$ , $\rho_3 = 0.85$ , and $\rho_4 = 0.8$ . We fixed $d = 2$ . . . . .	69
2.44	Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5$ , $M_1 = 5$ , $M_2 = 5$ , $M_3 = 5$ , $M_4 = 5$ , $M_5 = 5$ , and $N = 5$ presence of colored noise, where $\rho_1 = 0.9$ , $\rho_2 = 0.95$ , $\rho_3 = 0.85$ , and $\rho_4 = 0.8$ . We fixed $d = 3$ . . . . .	70
2.45	Probability of correct Detection (PoD) vs. SNR for an array of size $M_1 = 5$ , $M_2 = 7$ , and $M_3 = 9$ . The number of snapshots $N$ is set to 10 and the number of sources $d = 3$ . No FBA is applied. . . . .	71
2.46	Probability of correct Detection (PoD) vs. SNR for an array of size $M_1 = 5$ , $M_2 = 7$ , and $M_3 = 9$ . The number of snapshots $N$ is set to 10 and the number of sources $d = 3$ . FBA is applied. . . . .	72
2.47	Probability of correct Detection (PoD) vs. SNR for an array of size $M_1 = 5$ , $M_2 = 7$ , and $M_3 = 9$ . The number of snapshots $N$ is set to 10 and the number of sources $d = 5$ . FBA is applied. . . . .	73
2.48	Probability of correct Detection (PoD) vs. SNR for an array of size $M_1 = 5$ , $M_2 = 7$ , and $M_3 = 9$ . The number of snapshots $N$ is set to 10 and the number of sources $d = 9$ . FBA is applied. . . . .	74
2.49	The case of having the tensor with one dimension much higher than the other dimensions. The one-dimensional solution is the best one. Probability of correct Detection (PoD) vs. SNR for an array of size of $M_1 = 20$ and $M_2 = 5$ . The number of snapshots $N$ is set to 10 and the number of sources $d = 4$ . . . . .	75
3.1	An example of a 2-D outer product based array (OPA) of size $4 \times 4$ is illustrated. All the distances $\Delta_i^{(r)}$ for $i = 1, 2, 3$ and for $r = 1, 2$ may assume different values. . . . .	80
3.2	Comparing the reconstruction error, residuals and genie case for the selection of the multiple solution generated by the Closed-Form PARAFAC. The array size, snapshots $N$ , and $d$ are the same as in Fig. 3.4 . . . . .	88
3.3	RMSE versus SNR of the different integrations for the Closed-Form PARAFAC (CFP). In the scenario, the arrays are URA with perfectly known positions. The array size is $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ . $d$ is fixed to 4. . . . .	88
3.4	RMSE versus SNR to compare the Closed-Form PARAFAC to the ESPRIT based techniques. In the scenario, URAs of sizes $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ are applied. $d$ is fixed to 4. The RMSE is averaged for the parameters in the five dimensions according to (3.34). . . . .	91



3.5	RMSE versus $\sigma_e$ to compare the Closed-Form PARAFAC to the ESPRIT based techniques in presence of positioning errors in some dimensions. For the RMSE only the parameters at the third, fourth and fifth dimensions without positioning errors are considered. In the scenario, URAs of sizes $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ are applied. $d$ is fixed to 4. The SNR is fixed to 80 dB. . . . .	91
3.6	RMSE versus $\sigma_e$ to compare the Closed-Form PARAFAC to the ESPRIT based techniques in presence of positioning errors in some dimensions. For the RMSE only the parameters at the first and second dimensions with positioning errors are considered. In the scenario, URAs of sizes $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ are applied. $d$ is fixed to 4. The SNR is fixed to 80 dB. . . . .	92
3.7	RMSE versus $\sigma_p$ to compare the Closed-Form PARAFAC to the ESPRIT based techniques in case that shift invariance property varies, but the outer product property is kept in the first and second dimensions. For the RMSE only the parameters at the first and second dimensions are considered. In the scenario, URAs of sizes $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ are applied. $d$ is fixed to 4. The SNR is fixed to 80 dB. . . . .	93
3.8	RMSE versus $\sigma_p$ to compare the Closed-Form PARAFAC to the ESPRIT based techniques in case that shift invariance property varies, but the outer product property is kept in the first and second dimensions. For the RMSE only the parameters at the third, fourth and fifth dimensions are considered. In the scenario, URAs of sizes $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ are applied. $d$ is fixed to 4. The SNR is fixed to 80 dB. . . . .	94
3.9	RMSE versus $\sigma_s$ to compare the Closed-Form PARAFAC to the MALS based and ESPRIT-type techniques. In the scenario, URAs of sizes $M_1 = 3$ , $M_2 = 3$ , $M_3 = 3$ , $M_4 = 3$ , $M_5 = 7$ , and $N = 5$ are applied. $d$ is fixed to 4. The SNR is fixed to 80 dB. . . . .	94
3.10	The IlmProp [32] NLOS (non-line of sight) scenario composed of 4 paths. Both at the RX and at the TX URAs of size $3 \times 3$ are placed. 10 frequency bins are used with a total bandwidth is 80 MHz. The carrier frequency is $2 \cdot 10^9$ Hz, and 5 snapshots with 1 ms time sampling are collected. . . . .	95
3.11	RMSE versus SNR to compare the Closed-Form PARAFAC to the MALS based techniques. The IlmProp scenario is described in Fig. 3.10. . . . .	95
4.1	Noise power profile in the presence of colored noise. $\mathbf{N}^{(c)} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{P}^H$ . . . . .	100
4.2	Noise power profile in the presence of prewhitened noise via stochastic approaches. $\mathbf{N} = \widehat{\mathbf{L}}^{-1} \cdot \mathbf{N}^{(c)}$ . . . . .	100
4.3	Noise power profile in the presence of prewhitened noise via deterministic approach. $\mathbf{V} = \mathbf{D} \cdot \mathbf{N}^{(c)}$ . . . . .	100

---

4.4	Complementary cumulative distribution function (CCDF) of the root mean square error (RMSE) of $\frac{\hat{\rho}}{ \hat{\rho} }$ considering a only noise system with $M = 11$ sensors, and $N = 10$ and $N = 100$ snapshots. . . . .	108
4.5	Complementary cumulative distribution function (CCDF) of the root mean square error (RMSE) of $ \hat{\rho} $ considering a only noise system with $M = 2$ sensors, and $N = 10$ and $N = 100$ snapshots. . . . .	108
4.6	RMSE of the spatial frequencies vs. SNR considering a system with $N = 10$ snapshots, with $M = 10$ sensors and with $\rho = 0.7$ . $d = 3$ sources are present. . .	109
4.7	RMSE of the spatial frequencies vs. correlation factor $\rho$ considering a system with $N = 10$ snapshots and with $M = 10$ sensors. $d = 3$ sources are present. The SNR is fixed to 1 dB. . . . .	110
4.8	RMSE of the spatial frequencies vs. the angular distance $e_{\rho, \mu_1} = \left  \frac{\mu_1}{ \mu_1 } - \frac{\rho}{ \rho } \right $ considering a system with $N = 10$ snapshots and with $M = 10$ sensors. $d = 1$ source is present. The SNR is fixed to 1 dB. The correlation coefficient is set to $\rho = 0.9$ . . . . .	111
4.9	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard ESPRIT and 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N$ is set to 20 and the number of sources $d = 3$ . The correlation levels are $\rho_1 = 0.9$ for $\mathbf{L}_1$ and $\rho_2 = 0.95$ for $\mathbf{L}_2$ . The number of snapshots $N_l$ is set to $10^4$ . . . . .	121
4.10	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard ESPRIT and 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N$ is set to 20. The correlation levels are $\rho_1 = 0.9$ for $\mathbf{L}_1$ and $\rho_2 = 0.95$ for $\mathbf{L}_2$ . The number of snapshots $N_l$ is set to $10^4$ . In contrast to Fig. 4.9, the number of sources is set to $d = 4$ . . . . .	121
4.11	Comparing the different types of prewhitening schemes in conjunction with 3-D Standard ESPRIT and 3-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here the number of snapshots $N$ is set to 20 and $d = 4$ . The number of snapshots $N_l$ is set to $10^4$ . The array size is set to $M_1 = 5$ , $M_2 = 5$ and $M_3 = 5$ , and the correlation levels are $\rho_1 = 0.9$ for $\mathbf{L}_1$ , $\rho_2 = 0.95$ for $\mathbf{L}_2$ , and $\rho_3 = 0.95$ for $\mathbf{L}_3$ . . . . .	122
4.12	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N$ is set to 20. The correlation levels are $\rho_1 = 0.9$ for $\mathbf{L}_1$ and $\rho_2 = 0.95$ for $\mathbf{L}_2$ . The number of sources is set to $d = 4$ . The SNR is fixed to 30 dB, and the snapshots $N_l$ vary from 1 to $2 \cdot 10^4$ samples. . . . .	122

4.13	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of sources is set to $d = 4$ . The SNR is fixed to 0 dB, and the snapshots $N_l$ vary from 1 to $2 \cdot 10^4$ samples. . . . .	123
4.14	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N$ is set to 20. The number of snapshots $N_l$ is set to $10^4$ . The number of sources is equal to 4. The SNR is fixed to 10 dB, and the correlation levels $\rho_1$ and $\rho_2$ are equal to $\rho$ , where $0 \leq \rho < 1$ . . . . .	123
4.15	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N$ is set to 20. The number of snapshots $N_l$ is set to $10^4$ . The number of sources is equal to 4. The SNR is fixed to -5 dB, and the correlation levels $\rho_1$ and $\rho_2$ are equal to $\rho$ , where $0 \leq \rho < 1$ . . . . .	124
4.16	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. The number of snapshots $N$ is set to 20. The correlation levels are $\rho_1 = 0.9$ for $\mathbf{L}_1$ and $\rho_2 = 0.95$ for $\mathbf{L}_2$ . The number of snapshots $N_l$ is set to $10^4$ . The number of sources is equal to 4. However, the SNR is fixed to 10 dB, and the array size $M_1$ and $M_2$ equal to $M_i$ , which varies from 5 to 15. . . . .	124
4.17	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The correlation levels are $\rho_1 = 0.9$ for $\mathbf{L}_1$ and $\rho_2 = 0.95$ for $\mathbf{L}_2$ . The number of snapshots $N_l$ is set to $10^4$ . The number of sources is equal to 4. The SNR is fixed to 10 dB, and the number of snapshots $N$ varies from 10 to 100. . . . .	125
4.18	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N_l$ is set to $10^4$ and the number of snapshots $N$ is set to 20. The number of sources $d$ is equal to 2, and we set $\rho_1 = 0.9$ and $\rho_2 = 0$ . . . . .	125
4.19	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N_l$ is set to $10^4$ and the number of snapshots $N$ is set to 20. We set $\rho_1 = 0.9$ and $\rho_2 = 0$ . The number of sources $d$ is set to 4. . . . .	126

---

4.20	Comparing the different types of prewhitening schemes in conjunction with 2-D Standard Tensor-ESPRIT for the estimation of the spatial frequencies. Here we consider an array of size $M_1 = 5$ and $M_2 = 5$ . The number of snapshots $N_l$ is set to $10^4$ and the number of snapshots $N$ is set to 20. We set $\rho_1 = 0.9$ and $\rho_2 = 0$ . The number of sources $d$ is set to 5. . . . .	126
4.21	Total RMSE of the 5 estimated spatial frequencies versus $\rho_i$ , for $i = 1, \dots, 5$ . The SNR and the number of sources $d$ are set to 20 dB and 3, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . The I-S-GSVD is stopped at the second iteration, i.e., $k = 2$ . . . . .	131
4.22	Total RMSE of the 5 estimated spatial frequencies versus $\rho_i$ , for $i = 1, \dots, 5$ . The SNR and the number of sources $d$ are set to 5 dB and 3, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . The I-S-GSVD is stopped at the third iteration, i.e., $k = 3$ . . . . .	131
4.23	Total RMSE of the 5 estimated spatial frequencies versus the number of iterations $k$ . The SNR and the number of sources $d$ are set to 20 dB and 3, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . . . . .	132
4.24	Total RMSE of the 5 estimated spatial frequencies versus the number of iterations $k$ . The SNR and the number of sources $d$ are set to 5 dB and 3, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . . . . .	132
4.25	Total RMSE of the 5 estimated spatial frequencies versus the SNR is depicted. The number of sources $d$ and the correlation coefficient $\rho_i$ are set to 3 and to 0.9, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . The I-S-GSVD is stopped after the third iteration, i.e., $K = 3$ . . . . .	133
4.26	Total RMSE of the 5 estimated spatial frequencies versus $\rho_i$ for $i = 1, \dots, 5$ is depicted. The SNR and the number of sources $d$ are set to 5 dB and 2, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . The I-S-GSVD is stopped after the third iteration, i.e., $K = 3$ . . . . .	133
4.27	Total RMSE of the 5 estimated spatial frequencies versus the number of iterations $k$ is depicted. The SNR and the number of sources $d$ are set to 5 dB and 2, respectively. The array size is $M_i = 5$ , for $i = 1, \dots, 5$ , and $N = 5$ . The noise correlation $\rho_i$ is equal to 0.9. The same scenario as in Fig. 4.26 is considered. .	134
4.28	Total RMSE of the 5 estimated spatial frequencies versus $\rho_i$ for $i = 1, \dots, 5$ is depicted. The SNR and the number of sources $d$ are set to 5 dB and 3, respectively. The array size is $M_i = 7$ , for $i = 1, \dots, 5$ , and $N = 7$ . The I-S-GSVD is stopped after the third iteration, i.e., $K = 3$ . . . . .	134
4.29	Total RMSE of the 5 estimated spatial frequencies versus the number of iterations $k$ is depicted. The SNR and the number of sources $d$ are set to 5 dB and 3, respectively. The array size is $M_i = 7$ , for $i = 1, \dots, 5$ , and $N = 7$ . The noise correlation $\rho_i$ is equal to 0.9. The same scenario as in Fig. 4.28 is considered. .	135

A.1	PoD versus SNR for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We set the number of sources $d = 3$ .	144
A.2	Pfa versus SNR for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We set the number of sources $d = 3$ .	145
A.3	Pnd versus SNR for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We set the number of sources $d = 3$ .	145
A.4	PoD versus Pfa for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We set the number of sources $d = 3$ .	146
A.5	PoD versus SNR for different model order selection schemes. Here we consider an array of size $M = 10$ with $N = 12$ snapshots. We set the number of sources $d = 3$ .	146
D.1	In this scenario, $\mathbf{N} \in \mathbb{C}^{8 \times 10}$ .	153
D.2	Comparing the M-EFT II and M-EFT for the case of real noise. Probability of detection vs. SNR for an array of size $M_1 = 5$ . The number of snapshots $N$ is set to 6 and the number of sources $d = 3$ .	157
E.1	Notation for scalars	159
E.2	Notation for vectors	159
E.3	Notation for matrices	160
E.4	Visualization of a tensor representation	160
E.5	$n$ -mode vectors of $\mathbf{X} \in \mathbb{C}^{4 \times 5 \times 3}$	160
E.6	$n$ -mode vectors of $\mathbf{X} \in \mathbb{C}^{M_1 \times M_2 \times M_3}$	161
E.7	$n$ -mode product	162
E.8	Full SVD: $\mathbf{X} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^H$	162
E.9	Economy size SVD: $\mathbf{X} = \mathbf{U}_s \cdot \mathbf{\Sigma}_s \cdot \mathbf{V}_s^H$	163
E.10	Low-rank approximation SVD: $\mathbf{X} \approx \mathbf{U}'_s \cdot \mathbf{\Sigma}'_s \cdot \mathbf{V}'_s{}^H$	163
E.11	Economy size and low-rank approximation HOSVD of the core tensor $\mathcal{S}$	164
F.1	Representation of the identity tensor $\mathcal{I}_{R,N}$ , where $R = 3$ .	167
F.2	Representation of $\mathcal{I}_{3,N} \times_1 \mathbf{F}^{(1)}$ .	167
F.3	Representation of $\mathcal{I}_{3,N} \times_2 \mathbf{F}^{(2)}$ .	168
F.4	Representation of $\mathcal{I}_{3,N} \times_3 \mathbf{F}^{(3)}$ .	168

---

G.1	The decomposition $\mathbf{A}^H = \mathbf{U}^{(A)H} \cdot \mathbf{\Xi}^{(A)T} \cdot \mathbf{V}^H$ is obtained from the GSVD of $\mathbf{A}^H$ and $\mathbf{B}^H$ . . . . .	171
G.2	The decomposition $\mathbf{B}^H = \mathbf{U}^{(B)H} \cdot \mathbf{\Xi}^{(B)T} \cdot \mathbf{V}^H$ is obtained from GSVD of $\mathbf{A}^H$ and $\mathbf{B}^H$ . . . . .	171
J.1	Scenario generated by <i>IlmProp</i> composed of a $7 \times 7$ URA at RX, $7 \times 7$ URA at TX, and 5 scatters. . . . .	182
J.2	Spatial frequencies related to the azimuth and elevation at the TX, i.e., Direction-of-Departure (DOD). . . . .	183
J.3	Spatial frequencies related to the azimuth and elevation at the RX, i.e., Direction-of-Arrival (DOA). . . . .	183
J.4	Spatial frequencies related to the frequency bins dimension, i.e., Time Delay-of-Arrival (TDOA). . . . .	184
J.5	Bello function: Time Delay of Arrival (TDOA) vs. Doppler Shift (DS) using the Bello Explorer, which is a tool integrated with the <i>IlmProp</i> . The color bar is scaled in dBm. . . . .	185

## LIST OF TABLES

2.1	Classification of matrix-based model order selection techniques . . . . .	12
2.2	Penalty functions for Information Theoretic Criteria . . . . .	18
2.3	Penalty functions for 1-D Information Theoretic Criteria . . . . .	26
2.4	Summarized table comparing characteristics of the matrix-based model order selection schemes. . . . .	37
2.5	Classification of multi-dimensional model order selection techniques . . . . .	46
2.6	Penalty functions for $R$ -D Information Theoretic Criteria . . . . .	56
2.7	Summarized table comparing characteristics of the multi-dimensional model order selection schemes. . . . .	71
4.1	Notation of the legends used in all figures, where $R = 3$ for Fig. 4.11 and $R = 2$ for the other figures. . . . .	118
J.1	Comparison between the extracted spatial frequencies of the <i>IlmProp</i> generated channel to the spatial frequencies estimated via $R$ -D UTE. . . . .	184
J.2	Comparison between the extracted spatial frequencies of the <i>IlmProp</i> generated channel to the spatial frequencies estimated via $R$ -D UTE in case that the MIMO channel is time-varying. . . . .	185





# GLOSSARY OF ACRONYMS, SYMBOLS AND NOTATION

## Acronyms

<b>1-D</b>	one-dimensional
<b>2-D</b>	two-dimensional
<b>3-D</b>	three-dimensional
<b>AIC</b>	Akaike's Information Criterion
<b>BIC</b>	Bayesian Information Criterion
<b>CAIC</b>	Consistent AIC
<b>CRLB</b>	Cramér-Rao lower bound
<b>CORCONDIA</b>	Core Consistency Diagnostics
<b>CFP</b>	Closed-Form PARAFAC
<b>CFP-MOS</b>	Closed-Form PARAFAC based Model Order Selection
<b>CFP-PE</b>	Closed-Form PARAFAC based Parameter Estimation
<b>CP</b>	Criterion for Prediction
<b>DOA</b>	Direction Of Arrival
<b>DOD</b>	Direction Of Departure
<b>EDC</b>	Efficient Detection Criterion
<b>EEG</b>	Electroencephalography
<b>EFT</b>	Exponential Fitting Test
<b>EM</b>	Expectation Maximization
<b>ESPRIT</b>	Estimation of Signal Parameters by Rotational Invariance Techniques
<b>ESTER</b>	ESTimation ERror
<b>EVD</b>	Eigenvalue Decomposition
<b>FBA</b>	Forward Backward Averaging
<b>GAIC</b>	Geometric AIC
<b>GBIC</b>	Geometric BIC
<b>GMDL</b>	Geometric MDL
<b>GESE</b>	GEneralized Eigenvalues using signal Subspace Eigenvectors
<b>GEVD</b>	Generalized Eigenvalue decomposition
<b>GSVD</b>	Generalized Singular Value Decomposition
<b>HOSVD</b>	Higher Order Singular Value Decomposition
<b>I-S-GSVD</b>	Iterative Sequential Generalized Singular Value Decomposition
<b>ICA</b>	Independent Component Analysis
<b>IMT Advanced</b>	International Mobile Telecommunications-Advanced
<b>ITC</b>	Information Theoretic Criteria
<b>ITU</b>	International Telecommunication Union
<b>LS</b>	Least Squares
<b>LSKRF</b>	Least Squares Khatri-Rao Factorization
<b>M-EFT</b>	Modified Exponential Fitting Test
<b>MALS</b>	Multilinear Alternating Least Squares
<b>MEG</b>	Magnetoencephalography

---

---

<b>MDL</b>	Minimum Description Length
<b>MIMO</b>	Multiple Input Multiple Output
<b>ML</b>	maximum likelihood
<b>MOS</b>	Model Order Selection
<b>MPC</b>	Multi-Path Component
<b>MUSIC</b>	MUltiple SIgnal Classification
<b>NEMO</b>	Nadakuditi Edelman Model Order selection
<b>OPA</b>	Outer Product based Array
<b>PARAFAC</b>	Parallel Factor Analysis
<b>PDF</b>	Probability Density Function
<b>PDL</b>	predictive description length
<b>PoD</b>	Probability of correct Detection
<b>PS</b>	Peak Search
<b>QALS</b>	Quadrilinear Alternating Least Squares
<b><i>R</i>-D AIC</b>	<i>R</i> -dimensional Akaike's Information Criterion
<b><i>R</i>-D MDL</b>	<i>R</i> -dimensional Minimum Description Length
<b><i>R</i>-D EFT</b>	<i>R</i> -dimensional Exponential Fitting Test
<b><i>R</i>-D SE</b>	<i>R</i> -dimensional Standard ESPRIT
<b><i>R</i>-D STE</b>	<i>R</i> -dimensional Standard-Tensor ESPRIT
<b><i>R</i>-D UTE</b>	<i>R</i> -dimensional Unitary-Tensor ESPRIT
<b><i>R</i>-D UE</b>	<i>R</i> -dimensional Unitary ESPRIT
<b>RADAR</b>	Radio detection and ranging
<b>RARE</b>	Rank reduction estimator
<b>RELFIT</b>	Relative Fitness
<b>RESID</b>	residuals
<b>RMSE</b>	Root Mean Squared Error
<b>RX</b>	Receiver
<b>S-GSVD</b>	Sequential Generalized Singular Value Decomposition
<b>SAGE</b>	Space-Alternating Generalized EM
<b>SAMOS</b>	Subspace-based Automatic Model Order Selection
<b>SI</b>	Shift Invariance
<b>SLS</b>	Structured Least Squares
<b>SMD</b>	simultaneous matrix diagonalization
<b>SNR</b>	Signal to Noise Ratio
<b>SONAR</b>	Sound navigation and ranging
<b>SSC</b>	Surface Selection Criterion
<b>SSD</b>	Shortest Data Description
<b>SURE</b>	Stein's Unbiased Risk Estimator
<b>SVD</b>	singular value decomposition
<b>T-CORCONDIA</b>	Threshold CORCONDIA
<b>TALS</b>	Trilinear Alternating Least Squares
<b>TDOA</b>	Time Delay Of Arrival
<b>TS-SLS</b>	Tensor-Structured Structured Least Squares
<b>TUIL</b>	Techinsche Universität Ilmenau
<b>TX</b>	Transmitter
<b>UMCS</b>	users of MIMO channel sounders
<b>ULA</b>	Uniform Linear Array
<b>URA</b>	Uniform Rectangular Array
<b>VTRS</b>	Variance of Transformed Rotational Submatrix
<b>ZMCSG</b>	Zero Mean Circular Symmetric Complex Gaussian
<b>WIM</b>	WINNER channel Model
<b>WINNER</b>	Wireless World Initiative New Radio
<b>WWRF</b>	Wireless World Research Forum

---

## BIBLIOGRAPHY

### Own Publications

- [1] J. P. C. L. da Costa, M. Haardt, and F. Roemer, "Robust methods based on HOSVD for estimating the model order in PARAFAC models," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM'08)*, Darmstadt, Germany, July 2008.
- [2] J. P. C. L. da Costa, M. Haardt, F. Roemer, and G. Del Galdo, "Enhanced model order estimation using higher-order arrays," in *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2007.
- [3] J. P. C. L. da Costa, F. Roemer, and M. Haardt, "Deterministic prewhitening to improve subspace parameter estimation techniques in severely colored noise environments," in *Proc. 54th International Scientific Colloquium (IWK'09)*, Ilmenau, Germany, Sept. 2009.
- [4] —, "Sequential GSVD based prewhitening for multidimensional HOSVD based subspace estimation," in *Proc. ITG Workshop on Smart Antennas (WSA'09)*, Berlin, Germany, Feb. 2009.
- [5] —, "Iterative sequential GSVD (I-S-GSVD) based prewhitening for multidimensional HOSVD based subspace estimation without knowledge of the noise covariance information," in *Proc. ITG Workshop on Smart Antennas (WSA'10)*, Bremen, Germany, Feb. 2010.
- [6] J. P. C. L. da Costa, F. Roemer, M. Weis, and M. Haardt, "Robust  $R$ -D parameter estimation via closed-form PARAFAC," in *Proc. ITG Workshop on Smart Antennas (WSA'10)*, Bremen, Germany, Feb. 2010.
- [7] J. P. C. L. da Costa, A. Thakre, F. Roemer, and M. Haardt, "Comparison of model order selection techniques for high-resolution parameter estimation algorithms," in *Proc. 54th International Scientific Colloquium (IWK'09)*, Ilmenau, Germany, Oct. 2009.

### References by Other Authors

- [8] "Guidelines for evaluation of radio interface technologies for IMT-advanced," *Report ITU-R M.2135*, 2008.
- [9] R. J. Adock, "The analyst," *Philosophical Magazine*, vol. 5, pp. 53–54, 1878.
- [10] H. Akaike, "Information theory and extension of the maximum likelihood principle," *2nd Int. Symp. Inform. Theory suppl. Problems of Control and Inform. Theory*, pp. 267–281, 1973.

- [11] T. W. Anderson, "Asymptotic theory for principal component analysis," *Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, 1963.
- [12] A. Bab-Hadiashar and N. Gheissari, "Model selection for range segmentation of curved objects," in *8th European Conference on Computer Vision*, 2004, pp. 83–94.
- [13] R. Badeau, B. David, and G. Richard, "Selecting the modeling order for the ESPRIT high resolution method: an alternative approach," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004)*, Montreal, Canada, May 2004.
- [14] H. Bozdogan, "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions," in *Psychometrika*, vol. 52, 1987, pp. 345–370.
- [15] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics*, vol. 17, pp. 274–286, 2003.
- [16] L. Cao and D. J. Wu, "Noise-induced transport in a periodic system driven by Gaussian white noises with intensive cross-correlation," *Physics Letters A*, vol. 291, pp. 371–375, Dec. 2001.
- [17] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decompositions," *Psychometrika*, vol. 35, pp. 283–319, Sept. 1970.
- [18] A. L. Cauchy, *Oeuvres*, 1829, vol. 9, no. 2.
- [19] E. Ceulemans and H. A. L. Kiers, "Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method," *British Journal of Mathematical and Statistical Psychology*, vol. 59, no. 1, pp. 133–150, 2006.
- [20] Y. H. Chen and Y. S. Lin, "DOA estimation by fourth order cumulants in unknown noise environments," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1993)*, vol. 4, 1993, pp. 296–299.
- [21] D. Chickering and D. Heckerman, "Efficient approximation for the marginal likelihood of bayesian networks with hidden variables," in *Machine Learning*, vol. 29, no. 2, 1997.
- [22] A. Cichicki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, 1st ed. Wiley, 2009.
- [23] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [24] P. Comon and J. M. F. ten Berge, "Generic and typical ranks of three-way arrays," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Las Vegas, USA, Apr. 2008, pp. 3313–3316.
- [25] N. Czink, *The Random-Cluster model - A stochastic MIMO channel model for broadband wireless communications systems of the 3rd generation and beyond*. Vienna, Austria: Ph.D. Dissertation, Vienna University of Technology, Dec 2007.

- [26] J. P. C. L. da Costa, G. A. Borges, and R. Zelenovsky, "Study of techniques for estimation of signal frequency and direction-of-arrival applied to smart-antennas," in *Proc. VII Brazilian Symposium on Intelligent Automation*, São Luis, Brazil, Sept. 2005.
  - [27] J. P. C. L. da Costa, S. B. Puttini, and R. Zelenovsky, "Simulation of the intelligent hearing aid," in *Proc. IV Congresso Ibero-americano de Tecnologias Assistivas para Portadores de Deficiência (IBERDISCAP)*, Vitória, Brazil, Feb. 2006.
  - [28] —, "Study of viability for the intelligent hearing aid," in *Proc. IV Congresso Ibero-americano de Tecnologias Assistivas para Portadores de Deficiência (IBERDISCAP)*, Vitória, Brazil, Feb. 2006.
  - [29] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21(4), 2000.
  - [30] G. Del Galdo, "IlmProp, a flexible geometry-based Multi-User MIMO channel modeling tool," <http://www.tu-ilmenau.de/ilmprop/>, 2004.
  - [31] G. Del Galdo, *Geometry-based Channel Modeling for Multi-User MIMO systems and applications*. Ilmenau, Germany: Ph.D. Dissertation, Ilmenau University of Technology, May 2007.
  - [32] G. Del Galdo, M. Haardt, and C. Schneider, "Geometry-based channel modeling in MIMO scenarios in comparison with channel sounder measurements," *Advances in Radio Science - Kleinheubacher Berichte*, vol. 2, pp. 117–126, 2004.
  - [33] M. Drozd, P. Husar, A. Nowakowski, and G. Henning, "Detecting evoked potentials with SVD- and ICA- based statistical models," *IEEE Engineering in Medicine and Biology Magazine*, pp. 51–58, Jan. 2005.
  - [34] J. A. Fessler and A. O. Hero, "Space-Alternating Generalized Expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 1–14, 1994.
  - [35] T. Fu and X. Gao, "Simultaneous diagonalization with similarity transformation for non-defective matrices," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, vol. IV, Toulouse, France, May 2006, pp. 1137–1140.
  - [36] J. Ghasemi and A. Niazi, "Two- and three-way chemometrics methods applied for spectrophotometric determination of lorazepam in pharmaceutical formulations and biological fluids," *Analytica Chimica Acta*, vol. 533, pp. 169–177, 2005.
  - [37] N. Gheissari and A. Bab-Hadiashar, "Detecting cylinders in 3D range data using model selection criteria," in *Proc. 5th International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, Ottawa, Canada, June 2005, pp. 51–58.
  - [38] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.
  - [39] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1991.
  - [40] A. Graham, *Kronecker products and matrix calculus*. Chichester, U.K.: Ellis Horwood Ltd., 1981.
  - [41] P. J. Green and D. P. Taylor, "Dynamic signal enumeration algorithm for smart antennas," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1307–1314, June 2002.
-

- [42] J. Grouffaud, P. Larzabal, and H. Clergeot, "Some properties of ordered eigenvalues of a wishart matrix: application in detection test and model order selection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, vol. 5, May 1996, pp. 2463 – 2466.
  - [43] M. Haardt and J. A. Nossék, "Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Processing*, vol. 46, pp. 161–169, Jan. 1998.
  - [44] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198 – 3213, July 2008.
  - [45] M. Haardt, R. S. Thomä, and A. Richter, "Multidimensional high-resolution parameter estimation with applications to channel sounding," in *High-Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Chen, Eds. Marcel Dekker, New York, NY, 2004, pp. 255–338, chapter 5.
  - [46] M. Haardt, rapporteur, "Future mobile and wireless radio systems: Challenges in European research," in *Report on the FP 7 Consultation Meeting, European Commission, Information Society and Media*, Brussels, Belgium, Feb. 2008.
  - [47] P. C. Hansen and S. H. Jensen, "Prewhitening for rank-deficient noise in subspace methods for noise reduction," *IEEE Trans. Signal Processing*, vol. 53, pp. 3718–3726, Oct. 2005.
  - [48] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers in phonetics*, vol. 16, pp. 1–84, Dec. 1970.
  - [49] Z. He, A. Cichocki, and K. Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, in press 2009.
  - [50] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
  - [51] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal Educational Psychology*, vol. 24, pp. 417–441, 498–520, 1933.
  - [52] H. M. Huizenga, J. C. de Munck, L. J. Waldorp, and R. P. P. P. Grasman, "Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 6, pp. 533 – 539, June 2002.
  - [53] S. S. Jeng, C. W. Tsung, and Y. F. Lu, "Estimation and cancellation of doppler shift for an OFDM smart antenna system," in *Vehicular Technology Conference (VTC'07)*, 2007.
  - [54] J. S. Jiang and M. A. Ingram, "Robust detection of number of sources using the transformed rotational matrix," in *Wireless Communications and Networking Conference WCNC IEEE*, vol. 1, 2004, pp. 501 – 506.
  - [55] K. Kanatani, "Geometric information criterion for model selection," in *Int. J. Comput. Vision*, Jan. 1998, pp. 171–189.
-

- [56] ———, “Model selection for geometric inference,” in *Digital Image Computing-Techniques and Applications (DICTA2002)*, Melbourne, Australia, Jan. 2002.
  - [57] R. Kass and A. Raftery, “Bayes factors,” *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.
  - [58] H. A. L. Kiers and E. Ceulemans, “Choosing the number of components in three-way component model,” in *Three-way methods in Chemistry and Psychology (TRICAP’06)*, 2006.
  - [59] H. A. L. Kiers and A. der Kinderen, “A fast method for choosing the numbers of components in Tucker3 analysis,” *British journal of mathematical and statistical psychology*, vol. 56, no. 1, pp. 119–125, May 2003.
  - [60] P. M. Kroonenberg and J. de Leeuw, “Principal component analysis of three-mode data by means of alternating least squares algorithms,” *Psychometrika*, vol. 45, no. 1, pp. 69–97, Mar. 1980.
  - [61] P. M. Kroonenberg and T. H. A. van der Voort, “Multiplicatieve decompositie van interacties bij oordelen over de werkelijkheidswaarde van televisiefilms [multiplicative decomposition of interactions for judgements of realism of television films],” *Kwantitatieve Methoden*, vol. 8, no. 1, pp. 117–144, 1987.
  - [62] D. Kuhling, A. Ibing, and V. Jungnickel, “ $12 \times 12$  MIMO-OFDM realtime implementation for 3GPP LTE+ on a cell processor,” in *Proc. Wireless Conference*, June 2008, pp. 1–5.
  - [63] A. Lee, “Centrohermitian and skew-centrohermitian matrices,” *Linear algebra and its applications*, vol. 29, pp. 205–210, 1980.
  - [64] Y.-O. Li, T. Adali, and V. D. Calhoun, “Estimating the number of independent components for functional magnetic resonance imaging data,” *Human Brain Mapping*, vol. 28, pp. 1251–1266, 2007.
  - [65] T. Liu and S. Gazor, “Adaptive MLSD receiver employing noise correlation,” *IEEE Proc.-Commun.*, vol. 53, pp. 719–724, Oct. 2006.
  - [66] X. Liu and N. Sidiropoulos, “PARAFAC techniques for high-resolution array processing,” in *High-Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Chen, Eds. Marcel Dekker, New York, NY, 2004, pp. 111–150, chapter 3.
  - [67] X. Luciani, S. Mounier, H. H. M. Paraquetti, R. Redon, Y. Lucas, A. Bois, L. D. Lacerda, M. Raynaud, and M. Ripert, “Tracing of dissolved organic matter from the SEPETIBA bay (Brazil) by PARAFAC analysis of total luminescence matrices,” *Marine environmental research*, vol. 65, pp. 148–157, 2008.
  - [68] C. L. Mallows, “Some comments on CP,” in *Technometrics*, vol. 15, 1973, pp. 661–675.
  - [69] M. I. Miller and D. R. Fuhrmann, “Maximun likelihood narrow-band direction finding and the EM algorithm,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 1560–1577, 1990.
  - [70] T. Minka, “Automatic choice of dimensionality for PCA,” *Advances in Neural Information Processing Systems (NIPS’00)*, pp. 598–604, 2000.
-

- [71] G. E. Moore, "Cramming more components onto integrated circuits," in *Electronics Magazine*, no. 4, 1965.
  - [72] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Transactions of Signal Processing*, vol. 56, pp. 2625–2638, July 2008.
  - [73] M. Narandzic, C. Schneider, and R. S. Thomä, "WINNER wideband MIMO system-level channel model - comparison with other reference models," in *Proc. 54th International Scientific Colloquium (IWK'09)*, Ilmenau, Germany, Sept. 2009.
  - [74] A. Niazi and A. Yazdanipour, "PLS and PARAFAC applied to determination of nescapine in biological fluids by excitation - emission matrix fluorescence," *Pharmaceutical Chemistry Journal*, vol. 41, 2007.
  - [75] J.-M. Papy, L. De Lathauwer, and S. Van Huffel, "A shift invariance-based order-selection technique for exponential data modeling," *IEEE Signal Processing Letters*, vol. 14, pp. 473–476, July 2007.
  - [76] B. Park and T. F. Wong, "Training sequence optimization in MIMO systems with colored noise," in *Military Communications Conference (MILCOM 2003)*, Gainesville, USA, Oct. 2003.
  - [77] C. Pierdzioch, J. Doepke, and D. Hartmann, "Forecasting stock market volatility with macroeconomic variable in real time," *Journal of Economics & Business*, vol. 60, pp. 256–276, 2008.
  - [78] S. U. Pillai and B. H. Kwon, "GEESE (generalized eigenvalues utilizing signal subspace eigenvectors) - a new technique for direction finding," in *Proc. of the 22nd annual Asilomar Conference on Signals, Systems, and Computers (ASILOMAR 1988)*, Pacific Grove, CA, USA, Nov. 1988.
  - [79] A. Quinlan, J. Barbot, P. Larzabal, and M. Haardt, "Model order selection for short data: An exponential fitting test (EFT)," *EURASIP Journal on Applied Signal Processing*, 2007, special Issue on Advances in Subspace-based Techniques for Signal Processing and Communications.
  - [80] E. Radoi and A. Quinquis, "A new method for estimating the number of harmonic components in noise with application in high resolution radar," *EURASIP Journal on Applied Signal Processing*, pp. 1177–1188, 2004.
  - [81] C. B. Ribeiro, *Propagation parameter estimation in MIMO systems*. Helsinki, Finland: Ph.D. Dissertation, Helsinki University of Technology, Apr 2008.
  - [82] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
  - [83] F. Roemer, *Efficient Multi-Dimensional DOA-Estimation with spherical antenna arrays*. Ilmenau, Germany: Diploma Thesis, Ilmenau University of Technology, Oct 2006.
  - [84] F. Roemer and M. Haardt, "A closed-form solution for multilinear PARAFAC decompositions," in *Proc. 5-th IEEE Sensor Array and Multich. Sig. Proc. Workshop (SAM 2008)*, Darmstadt, Germany, July 2008, pp. 487 – 491.
-



- [85] —, “A closed-form solution for parallel factor (PARAFAC) analysis,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Las Vegas, USA, Apr. 2008, pp. 2365–2368.
  - [86] —, “Tensor-Based channel estimation (TENCE) for Two-Way relaying with multiple antennas and spatial reuse,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Sig. Proc. (ICASSP 2009)*, Taipei, Taiwan, Apr. 2009.
  - [87] F. Roemer, M. Haardt, and G. Del Galdo, “Higher order SVD based subspace estimation to improve multi-dimensional parameter estimation algorithms,” in *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2006, pp. 961–965.
  - [88] R. Roy and T. Kailath, “ESPRIT - Estimation of signal parameter via rotational invariance techniques,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 984–995, July 1989.
  - [89] —, “ESPRIT - Estimation of signal parameters via rotational invariance techniques,” in *Signal Processing Part II: Control Theory and Applications*, L. Auslander, F. A. Grünbaum, J. W. Helton, T. Kailath, P. Khargonekar, and S. Mitter, Eds. Springer-Verlag, 1990, pp. 369–411.
  - [90] E. Sanchez and B. R. Kowalski, “Tensorial resolution: a direct trilinear decomposition,” *J. Chemom.*, vol. 4, pp. 29–45, 1990.
  - [91] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, pp. 276–280, Mar. 1986.
  - [92] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.
  - [93] S. Silverstein, “A new use of ESPRIT for detection of signal parameters via rotational invariance techniques,” *IEEE Signals Processing Letters*, vol. 1, pp. 147–149, Oct. 1994.
  - [94] M. Steinbauer, A. Molisch, and E. Bonek, “The double-directional radio channel,” *IEEE Antennas and Propagation Magazine*, vol. 43, no. 4, pp. 51–63, 2001.
  - [95] M. O. Ulfarsson and V. Solo, “Rank selection in noisy PCA with SURE and random matrix theory,” in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Las Vegas, USA, Apr. 2008.
  - [96] S. Valaee and P. Kabal, “An information theoretic approach to source enumeration in array signal processing,” *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1171–1178, 2004.
  - [97] J. Vandewalle, L. De Lathauwer, and P. Comon, “The generalized higher order singular value decomposition and the oriented signal-to-signal ratios of pairs of signal tensors and their use in signal processing,” in *European Conference on Circuit Theory and Design*, Cracow, Poland, Sept. 2003.
  - [98] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 387–392, 1985.
  - [99] M. Wax and I. Ziskind, “Detection of the number of coherent signals by the mdl principle,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1190–1196, 1989.
-

- [100] M. Weis, F. Roemer, M. Haardt, D. Jannek, and P. Husar, "Multi-dimensional Space-Time-Frequency component analysis of event-related EEG data using closed-form PARAFAC," in *Proc. IEEE Int. Conf. Acoustics, Speech and Sig. Proc. (ICASSP 2009)*, Taipei, Taiwan, Apr. 2009.
  - [101] J. Winters, Motia, "Understanding MIMO: Part I," in *CommsDesing An EE Times Community*, July 2009.
  - [102] A. Wittneben and B. Rankov, "MIMO signaling for low rank channels," in *Proc. International Symposium on Electromagnetic Theory*, May 2004.
  - [103] H.-T. Wu, J.-F. Yang, and F.-K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1325–1333, 1995.
  - [104] G. Xu, R. H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Processing*, vol. 42, pp. 102–112, Jan. 1994.
  - [105] Q. T. Zhang and K. M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise," *IEEE Transactions on Signal Processing*, vol. 41, pp. 1652–1662, Apr. 1993.
  - [106] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, pp. 1–25, 1986.
-

## THESES

- T1: Multi-dimensional parameter estimation techniques are based on the manipulation of a summation of rank one tensors contaminated by additive measurement noise. Each rank one tensor characterizes a source or a principal component and from the multi-dimensional structure of each rank one tensor, a set of parameters that are of practical interest can be estimated.
- T2: The number of sources or principal components is a required information to extract these parameters. Therefore, the estimation of this number from noisy observations is an important problem, which is also known as model order selection.
- T3: In the case where the data does not possess a multi-dimensional structure and is contaminated only by additive white Gaussian noise, the model order selection problem should be solved by the Modified Exponential Fitting Test (M-EFT), where the white Gaussian noise eigenvalues are predicted by approximating it by an exponential profile.
- T4: For applications, where the data has a multi-dimensional structure and is contaminated by white Gaussian noise, the model order selection problem should be solved via  $R$ -dimensional Exponential Fitting Test ( $R$ -D EFT), which combines the eigenvalues of different matrix representations of the tensor, called *global eigenvalues*, in order to improve the estimation. The same concept of the global eigenvalues can be also used for multi-dimensional extensions of Akaike's Information Criterion (AIC) and Minimum Description Length (MDL).
- T5: For the case that the data has no multi-dimensional structure and the noise is severely correlated, the RADOI technique should be applied to solve the model order selection problem. However, if the data has a multi-dimensional structure, the closed-form PARAFAC based model order selection (CFP-MOS) scheme, which is based on the multiple estimates of the rank one tensors obtained from the closed-form PARAFAC, should be used.
- T6: By modeling the data as a superposition of undamped complex exponentials sampled on a multi-dimensional lattice, it is possible to estimate the corresponding spatial frequencies in each dimension. The total number of spatial frequencies is the product of the model order and the dimensions with the specific complex exponential structure.
-

- T7: From a practical standpoint it is attractive to find closed-form algorithms to extract these spatial frequencies since they feature a low computational complexity while often still performing close to the Cramér-Rao lower bound. The closed-form PARAFAC based parameter estimation (CFP-PE) scheme is applicable for scenarios, where the shift invariance structure is not present. Moreover, since the CFP-PE decouples the dimensions, errors from one dimension do not affect the others. The CFP-PE does not require forward-backward averaging (FBA) in case of correlated sources. Finally, the CFP-PE is more robust for scenarios, where the narrow-band assumption is violated, and, therefore, it provides a more accurate estimation than the other schemes.
- T8: Most parameter estimation schemes implicitly assume the noise to be white. Therefore, their performance degrades in the case where the noise samples are mutually correlated. To avoid such degradation, prewhitening schemes should be applied.
- T9: In some applications, the colored noise may exhibit a specific structure. In this case, a deterministic prewhitening scheme, which takes advantage of this structure, outperforms previous stochastic prewhitening approaches.
- T10: The colored noise also can present a multi-dimensional structure for MIMO and EEG applications. Therefore, the multi-dimensional structure of the noise should be taken into account in the prewhitening process. The Sequential Generalized Singular Value Decomposition (S-GSVD), besides being a technique with a low computational complexity, provides an estimation of the signal subspace taking into account the multi-dimensional structure of the colored noise.
- T11: Since in some applications it is impossible to obtain samples of the noise process in the absence of desired signal components, the second order statistics of the noise should be estimated jointly with the parameters of the principal components in an iterative fashion. For this task, the Iterative S-GSVD (I-S-GSVD), which does not require and prior knowledge about the noise statistics, is proposed. The I-S-GSVD provides a parameter estimation accuracy similar to the S-GSVD where the second order statistics of the noise are estimated separately. The I-S-GSVD inherits the low computational complexity of the S-GSVD.
-

## ERKLÄRUNG

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich<sup>1</sup> geholfen:

1: X

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungsbzw. Beratungsdiensten ( Promotionsberater oder anderer Personen ) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalte der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch angesehen wird und den erfolglosen Abbruch des Promotionsverfahrens zu Folge hat.

Ilmenau, den 16. Dezember 2009

João Paulo Carvalho Lustosa da Costa

---

<sup>1</sup> Unzutreffendes bitte streichen.

---