



KTH Electrical Engineering

Exploiting Prior Information in Parametric Estimation Problems for Multi-Channel Signal Processing Applications

PETTER WIRFÄLT

Doctoral Thesis in Signal Processing
Stockholm, Sweden 2013

TRITA-EE 2013:040
ISSN 1653-5146
ISBN 978-91-7501-916-1

Signal Processing
School of Electrical Engineering
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden
E-mail: wirfalt@kth.se

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av Teknologie doktorsexamen i signalbehandling fredagen den 6 december 2013 klockan 13.15 i Q2, Osqldas väg 10, Stockholm.

Exploiting Prior Information in Parametric Estimation Problems for Multi-Channel Signal Processing Applications

Copyright © 2013 by Petter Wirfält except where otherwise stated. All rights reserved.

Printed by Universitetsservice US-AB.

Abstract

This thesis addresses a number of problems all related to parameter estimation in sensor array processing. The unifying theme is that some of these parameters are known before the measurements are acquired. We thus study how to improve the estimation of the unknown parameters by incorporating the knowledge of the known parameters; exploiting this knowledge successfully has the potential to dramatically improve the accuracy of the estimates.

For covariance matrix estimation, we exploit that the true covariance matrix is Kronecker and Toeplitz structured. We then devise a method to ascertain that the estimates possess this structure. Additionally, we can show that our proposed estimator has better performance than the state-of-art when the number of samples is low, and that it is also efficient in the sense that the estimates have Cramér-Rao lower Bound (CRB) equivalent variance.

In the direction of arrival (DOA) scenario, there are different types of prior information; first, we study the case when the location of some of the emitters in the scene is known. We then turn to cases with additional prior information, i.e. when it is known that some (or all) of the source signals are uncorrelated. As it turns out, knowledge of some DOA combined with this latter form of prior knowledge is especially beneficial, giving estimators that are dramatically more accurate than the state-of-art. We also derive the corresponding CRBs, and show that under quite mild assumptions, the estimators are efficient.

Finally, we also investigate the frequency estimation scenario, where the data is a one-dimensional temporal sequence which we model as a spatial multi-sensor response. The line-frequency estimation problem is studied when some of the frequencies are known; through experimental data we show that our approach can be beneficial. The second frequency estimation paper explores the analysis of pulse spin-locking data sequences, which are encountered in nuclear resonance experiments. By introducing a novel modeling technique for such data, we develop a method for estimating the interesting parameters of the model. The technique is significantly faster than previously available methods, and provides accurate estimation results.

Keywords Array signal processing, covariance matrix, damped sinusoids, direction of arrival estimation, frequency estimation, Kronecker, NQR, NMR, pa-

parameter estimation, persymmetric, signal processing algorithms, structured covariance estimation, Toeplitz

Sammanfattning

Denna doktorsavhandling behandlar parameterestimeringsproblem inom flerkanalssignalbehandling. Den gemensamma förutsättningen för dessa problem är att det finns information om de sökta parametrarna redan innan data analyseras; tanken är att på ett så finurligt sätt som möjligt använda denna kunskap för att förbättra skattningarna av de okända parametrarna.

I en uppsats studeras kovariansmatrisskattning när det är känt att den sanna kovariansmatrisen har Kronecker- och Toeplitz-struktur. Baserat på denna kunskap utvecklar vi en metod som säkerställer att även skattningarna har denna struktur, och vi kan visa att den föreslagna skattaren har bättre prestanda än existerande metoder. Vi kan också visa att skattarens varians når Cramér-Rao-gränsen (CRB).

Vi studerar vidare olika sorters förhandskunskap i riktningbestämningsscenario: först i det fall då riktningarna till ett antal av sändarna är kända. Sedan undersöker vi fallet då vi även vet något om kovariansen mellan de mottagna signalerna, nämligen att vissa (eller alla) signaler är okorrelerade. Det visar sig att just kombinationen av förkunskap om både korrelation och riktning är speciellt betydelsefull, och genom att utnyttja denna kunskap på rätt sätt kan vi skapa skattare som är mycket noggrannare än tidigare möjligt. Vi härleder även CRB för fall med denna förhandskunskap, och vi kan visa att de föreslagna skattarna är effektiva.

Slutligen behandlar vi även frekvensskattning. I detta problem är data en en-dimensionell temporal sekvens som vi modellerar som en spatiell fler-kanalssignal. Fördelen med denna modelleringsstrategi är att vi kan använda liknande metoder i estimatorerna som vid sensor-signalbehandlingsproblemen. Vi utnyttjar återigen förhandskunskap om källsignalerna: i ett av bidragen är antagandet att vissa frekvenser är kända, och vi modifierar en existerande metod för att ta hänsyn till denna kunskap. Genom att tillämpa den föreslagna metoden på experimentell data visar vi metodens användbarhet. Det andra bidraget inom detta område studerar data som erhålls från exempelvis experiment inom kärnmagnetisk resonans. Vi introducerar en ny modelleringsmetod för sådan data och utvecklar en algoritm för att skatta de önskade parametrarna i denna modell. Vår algoritm är betydligt snabbare än existerande metoder, och skattningarna är tillräckligt noggranna för typiska tillämpningar.

Nyckelord Array, signalbehandling, kovariansmatris, dämpad sinus, riktning-

bestämning, frekvensskattning, Kronecker, NQR, NMR, parameterestimering, per-
symmetrisk, algoritm, strukturerad kovariansmatris, Toeplitz

*Per ardua
ad gloriam*

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Prof. Magnus Jansson, who gave me this opportunity as well as supported, refined, and helped me develop my ideas through countless discussions. Your enthusiasm with regards to, and willingness to confront, the challenging theoretical problems I have faced has been a constant source of motivation. I also hope your attention to detail has permeated at least some parts of this thesis.

I also want to thank my other co-authors: Prof. Petre Stoica, I was most impressed by your efficient handling of the manuscripts. Dr Erik Gudmundsson, I have immensely enjoyed our interactions both inside and outside academia. Dr Guillaume Bouleux, thank you for the rewarding collaboration.

Prof. Björn Ottersten, thank you for inviting me to the department and for opening my eyes to a new world through the course you taught me. Prof. Peter Händel, I hope our entrepreneurial ambitions come to fruit. Thank you also to Prof. Joakim Jaldén and Prof. Mats Bengtsson — I appreciate the fact that both of you have always been available for discussions of topics, big and small.

Thank you Prof. Abdelhak M. Zoubir for taking the time to act as the faculty opponent during the defense of my dissertation; I would also like to thank Prof. Tomas McKelvey, Prof. Håkan Hjalmarsson, and Prof. Sven Nordebo for constituting the grading committee.

My time at the department would not have been the very enjoyable time it was but for the ever evolving cadre of fellow Ph. D students. A heartfelt thank you to all of you, and I hope to be able to see you again. A special thank you to Dennis for, amongst other things, proofreading this thesis.

I would also like to thank the financial sponsors that has made my stay at the Signal Processing department possible: the Swedish Research Council (VR) and the European Research Council (ERC).

Last but certainly not least, my family: Vincent, Olle, and my lovely wife Ingrid — thank you for all the encouragement and for enduring the sacrifices. My parents, for always being there. My brother, for complementing my abilities.

Contents

Abstract	i
Sammanfattning	iii
Acknowledgments	vii
Contents	ix
Nomenclature	xiii
I Introduction	1
Introduction	1
1 Sensor array processing	3
1.1 Performance bounds in SAP	5
2 Direction of arrival estimation	7
2.1 Uniform linear array at receiver	8
2.2 Prior information	9
2.3 Conclusions and future work	16
3 Frequency estimation	17
3.1 Sinusoidal estimation with prior information	17
3.2 Pulse spin-locking data sequence estimation	19
3.3 Conclusions and future work	22
4 Kronecker structured covariance matrix estimation	25
4.1 Accounting for Kronecker structure with persymmetric factor matrices	26
4.2 Accounting for Toeplitz structured factor matrices	27
5 Contributions	31
References	33

II Included papers 39

A Optimal Prior Knowledge-Based Direction of Arrival Estimation A1

- 1 Introduction A1
- 2 Data model A3
- 3 DOA estimation A4
 - 3.1 General signal correlation A5
 - 3.2 Uncorrelated sources A6
- 4 Using prior knowledge A7
 - 4.1 PLEDGE A7
 - 4.2 PLEDGE: general case A8
 - 4.3 PLEDGE: uncorrelated sources A8
 - 4.4 Other methods A9
- 5 PLEDGE CRB A10
 - 5.1 Correlated sources A11
 - 5.2 Uncorrelated sources A12
 - 5.3 Comparison to no-prior bounds A12
- 6 Simulated data examples A13
 - 6.1 Uncorrelated sources A14
 - 6.2 Unknown source correlation A16
 - 6.3 Error in prior knowledge A17
 - 6.4 Varying known-source properties A18
 - 6.5 A note on the computational complexity of the algorithms A19
- 7 Experimental data application A21
- 8 Conclusions A22
- References A24

B Prior-Exploiting Direction-of-Arrival Algorithms for Partially Uncorrelated Source Signals B1

- 1 Introduction B1
- 2 Problem description B3
- 3 POWDER B4
 - 3.1 Review of method B4
 - 3.2 Implementation for uniform linear array B6
 - 3.3 Statistical properties of the residual B7
- 4 Cramér-Rao bound B8
- 5 Performance analysis B10
 - 5.1 Consistency of estimates B10
 - 5.2 Asymptotic distribution B11
- 6 Numerical examples B12
 - 6.1 Performance comparison to existing state-of-art B13
 - 6.2 Asymptotic variance validity B14
 - 6.3 Robustness to modeling errors B14
- 7 Conclusions B16

Appendix.A	B17
Appendix.B.	B19
Appendix.C.	B20
Appendix.D.	B24
Appendix.E.	B26
Appendix.F.	B28
References	B29
C Robust Prior-based Direction of Arrival Estimation	C1
1 Introduction	C1
2 Problem formulation	C2
3 DOA estimator	C3
3.1 Exact prior knowledge	C5
3.2 Uncertainty in prior knowledge	C6
3.3 Estimator implementation	C7
4 Cramér-Rao bound	C8
5 Simulations	C9
6 Conclusions	C10
References	C11
D Subspace-based Frequency Estimation Utilizing Prior Information	D1
1 Introduction	D1
2 Problem formulation	D2
3 Estimator implementation	D2
4 Simulations	D5
5 Broken rotor bar diagnosis	D6
6 Conclusions	D8
References	D9
E An ESPRIT-based parameter estimator for spectroscopic data	E1
1 Introduction and data model	E1
2 The echo train ESPRIT algorithm	E3
3 Derivation of the Cramér-Rao bound	E5
4 Numerical examples	E8
5 Conclusions	E9
References	E10
F On Kronecker and Linearly Structured Covariance Matrix Estimation	F1
1 Introduction	F1
2 Problem formulation	F3
3 Kronecker-structured ML-estimation	F4
4 ML for persymmetric Kronecker model	F5
5 Toeplitz structured factor matrices	F7

5.1	Consistency of the factor matrix estimation	F7
5.2	EXIP-estimator	F9
5.3	SNIFF - structured non-iterative flip-flop algorithm	F12
6	Performance analysis of the SNIFF estimates	F14
6.1	Consistency	F14
6.2	Asymptotic variance	F14
7	Numerical simulations	F16
8	Conclusion	F19
	Appendix.A	F19
	Appendix.B.	F20
	Appendix.C	F24
	Appendix.D	F24
	Appendix.E.	F26
	References	F29

Nomenclature

Abbreviations and Acronyms

CRB:	Cramér-Rao Bound
DOA:	Direction of Arrival
ESPRIT:	Estimation of Signal Parameters via Rotational Invariance Techniques
ET-ESP:	Echo Train ESPRIT
FIM:	Fisher Information Matrix
FF:	Flip-Flop
MUSIC:	MUltiple SIgnal Classification
NIFF:	Non-Iterative Flip-Flop
NQR:	Nuclear Quadrupole Resonance
PLEDGE:	Prior knowLEDGE
POWDER:	Prior Orthogonally Weighted Direction Estimator
PS:	Persymmetric
PSL:	Pulse Spin-Locking
RMSE:	Root Mean Square Error
SAP:	Sensor Array Processing
SNIFF:	Structured Non-Iterative Flip-Flop
SNR:	Signal to Noise Ratio
ULA:	Uniform Linear Array

Notations

In this thesis vectors are denoted by lowercase, bold symbols: \mathbf{x} , $\boldsymbol{\theta}$, etc. Matrices are denoted by uppercase bold symbols: \mathbf{X} , $\boldsymbol{\Theta}$, etc.

$\mathbf{X} \in \mathbb{C}^{m \times n}$	\mathbf{X} is an $m \times n$ matrix with complex valued elements
$\mathbf{X} \in \mathbb{R}^{m \times n}$	\mathbf{X} is an $m \times n$ matrix with real valued elements
\mathbf{X}^* , \mathbf{X}^H	Conjugate transpose of \mathbf{X}
\mathbf{X}^c	Complex conjugate of \mathbf{X}
\mathbf{X}^T	Transpose of \mathbf{X}
\mathbf{X}^\dagger	Moore-Penrose pseudo inverse of \mathbf{X}
$E[\cdot]$	Expected value of random variable
\triangleq	“defined by”
\otimes	Kronecker-product: $\mathbf{A} \otimes \mathbf{B} \triangleq \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$
$\text{Tr}(\cdot)$	The trace function, i.e. the sum of the diagonal elements of a matrix
$ \cdot $	The determinant function
$\text{vec}(\cdot)$	Vectorization operator, stacking the columns of a matrix into a vector
$\text{diag}(\mathbf{x})$	The diagonal matrix with the elements of the vector \mathbf{x} on the diagonal
$\arg \min_x f(x)$	The value of x that minimizes $f(x)$
\mathbf{I}_n	The identity matrix of dimension $n \times n$; the subscript will often be omitted
$\mathbf{0}_{m \times n}$	The matrix of all zeros of dimension $m \times n$; the subscript will often be omitted
p.d.	Positive definite, also denoted $\mathbf{X} > \mathbf{0}$
$\mathcal{N}(\mathbf{x}, \mathbf{X})$	The (complex) normal distribution with mean \mathbf{x} and covariance \mathbf{X}

Part I

Introduction

Introduction

This thesis addresses a number of problems all related to parameter estimation in sensor array processing. The unifying presumption is that some of the parameters are known even before measurements are acquired. We thus study how to improve the estimation of the unknown parameters by incorporating the knowledge of the known parameters.

A straight-forward example of a *parameter estimation problem* that most people are familiar with is the following: what is the speed of the car you are driving? One might take this information for granted, but as will be shown, this is a perfect example of a parameter estimation problem.

Traditionally, one would solve this problem by measuring the rotation speed, or angular velocity, of the axle shaft connected to one of the wheels. One could expect that

$$s_0(t) = \Omega_0(t)R, \quad (1)$$

where $s_0(t)$ is the *parameter* of interest and the true speed of the wheel (and hence, the car) at time t , $\Omega_0(t)$ is the true angular velocity of the shaft at time t , and R is the (known) wheel radius. Thus, since our measurement device typically would be mounted on the wheel shaft, we can consider

$$\Omega_0(t) = \frac{1}{R}s_0(t) \quad (2)$$

to be our ideal observation. However, since no measurement device is perfect, the *signal* we expect to observe can be modeled as

$$\Omega(t) = \frac{1}{R}s_0(t) + n(t), \quad (3)$$

where $n(t)$ is a random noise term. Typically, there will be some assumptions on $n(t)$, and if we exploit these assumptions properly we can maximize the accuracy with which we can *estimate* $s_0(t)$ from $\Omega(t)$.

Further assumptions could include that $s_0(t) = s_0$, i.e. *constant* speed, during some time period $t = 1, \dots, N$; then, N *samples* of (3) could be used to estimate s_0 with higher accuracy.

There are infinite possibilities in choosing an estimator; however, the particular estimator we design will produce some estimate of s_0 : denote this estimate

\hat{s}_0 . An estimate \hat{s}_0 might be *unbiased*; popularly speaking, this means that given sufficiently many observations of the signal in question, $\hat{s}_0 = s_0$. Obviously this is an attractive property. See, e.g., [1] for a thorough treatment of this, and related, concepts.

If we only consider unbiased estimators of the parameter in question, we can rank the estimators' performance according to the *variance* of their estimates. We desire to have an estimator giving as low variance as possible. Remarkably, under quite mild assumptions, it is possible to find a bound on the *lowest possible* variance that *any* unbiased estimator of a certain parameter can have. One such bound is the Cramér-Rao Bound (CRB), [2].

With these preliminaries, we have the work laid out:

1. Choose a parameter estimation problem and determine the relevant CRB.
2. Design an unbiased estimator.
3. Show that its variance is lower than other existing estimators.
4. Show that the estimator gives the best possible accuracy by comparing it to the CRB.

As it turns out, 4) might very well not hold, or might be a too daunting challenge to accomplish. Then, as long as 3) can be established with sufficient rigor, 4) might potentially be irrelevant.

This general approach is applied to the conceptually more demanding estimation problems studied in the papers included in the thesis.

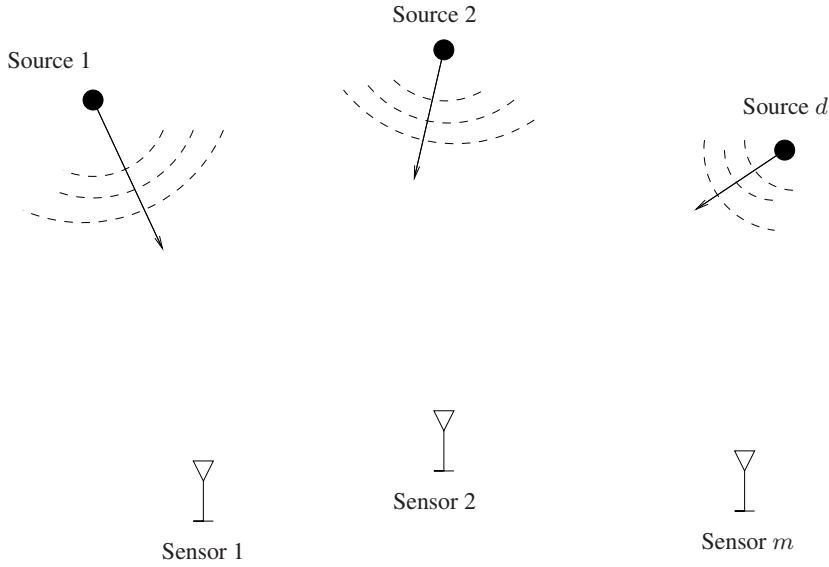


Figure 1: General SAP scenario, with d distinct sources each radiating waves, and m sensors forming the receive array.

1 Sensor array processing

This thesis mainly deals with the subject area Sensor Array Processing (SAP), and the parts that do not are at least treated in a mathematically similar manner. A general SAP scenario is depicted in Fig. 1, in which d narrow-band waves (for example, acoustic or electromagnetic) are impinging on an m -element sensor array located in the far field of all of the sources. The goal of SAP, in the context of this work, is to extract meaningful information from the impinging waves, or rather from the data acquired by sampling the sensor outputs after receive filtering.

All sensing involves the creation of noise; this fact has to be accounted for. In this thesis, the noise is modeled as additive on each sensor output. We will further explore the particular assumptions on the noise later.

The above mentioned narrow-band wave property implies that the time it takes for the (payload) signal modulated onto each wave to traverse the (maximum) length of the array is small in comparison to changes in the modulated signal transmitted by each source. This also means that the bandwidth of the payload is small in comparison to the carrier frequency. Furthermore, the far-field assumption stipulates that the waves can be assumed to be plane when reaching the array.

Given the above assumptions we can model the received baseband data after processing according to

$$\mathbf{y}(t) = \mathbf{A}(\bar{\boldsymbol{\theta}})\mathbf{s}(t) + \mathbf{n}(t); \quad (4)$$

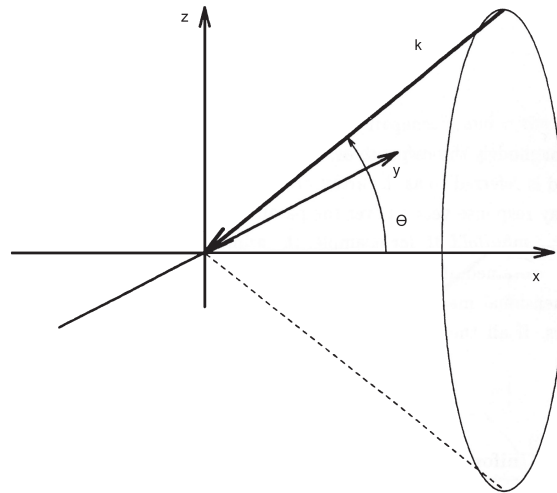


Figure 2: Definition of cone angle θ .

here, $\mathbf{y}(t) = [y_1(t) \ y_2(t) \ \dots \ y_m(t)]^T \in \mathbb{C}^{m \times 1}$ is the vector of the complex valued sensor outputs at time t , $\mathbf{A}(\bar{\boldsymbol{\theta}}) \in \mathbb{C}^{m \times d}$ is termed the *array steering matrix*, which for a given array geometry and sensor response is uniquely determined by the directions of arrival (DOAs) $\bar{\boldsymbol{\theta}}$ (the use of the somewhat unorthodox $\bar{\boldsymbol{\theta}}$ is because we want to reserve the symbol $\boldsymbol{\theta}$ for later use; see Section 2.2). Further, similarly to $\mathbf{y}(t)$, complex valued $\mathbf{s}(t) \in \mathbb{C}^{d \times 1}$ and $\mathbf{n}(t) \in \mathbb{C}^{m \times 1}$ are created by stacking the source signals and sensor noise, respectively. See, e.g., [3] for more on the complex signal representation and [4] for more on the data model.

In this thesis we let the DOA vector $\bar{\boldsymbol{\theta}} = [\bar{\theta}_1 \ \dots \ \bar{\theta}_d]^T$; the direction to each source is thus characterized by a single angle. There are two ways of viewing such a scenario; either the sources and the sensor array share a common plane. Then one angle is sufficient to uniquely determine the direction to each source. Or, if a common plane is not shared, then one angle places each source on a unique cone; see Fig. 2.

The number of signals d might not be known a-priori, and would in such a scenario also have to be estimated. However, from this point on, we assume that d is known; see e.g. [5], [6], [7] for methods of estimating d .

We further use the *stochastic* signal model, in contrast to the *deterministic* model; see e.g. [8] for elaboration on this subject. Under the stochastic signal model considered in this work, the source signal and noise are both assumed to be independent, identically distributed (i.i.d.), zero mean, circularly symmetric

complex Gaussian random processes with second order moments given by

$$\mathbb{E}[\mathbf{s}(t_1)\mathbf{s}^*(t_2)] = \mathbf{P}\delta_{t_1,t_2}, \quad \mathbb{E}[\mathbf{s}(t_1)\mathbf{s}^T(t_2)] = \mathbf{0}, \quad (5)$$

$$\mathbb{E}[\mathbf{n}(t_1)\mathbf{n}^*(t_2)] = \sigma^2\mathbf{I}\delta_{t_1,t_2}, \quad \mathbb{E}[\mathbf{n}(t_1)\mathbf{n}^T(t_2)] = \mathbf{0}, \quad (6)$$

where the Kronecker-delta is defined by $\delta_{t_1,t_2} = 1$ for $t_1 = t_2$ and 0 otherwise. We assume that we know the noise to be spatially white (see e.g. [9] for the treatment of colored noise with known covariance, and [10] for noise of unknown color), but the variance σ^2 is unknown. Thus, the received signal $\mathbf{y}(t)$ is a realization of an i.i.d., zero mean, circularly symmetric complex Gaussian random process, characterized by

$$\mathbb{E}[\mathbf{y}(t_1)\mathbf{y}^*(t_2)] = \mathbf{R}\delta_{t_1,t_2} \quad (7)$$

$$\mathbb{E}[\mathbf{y}(t_1)\mathbf{y}^T(t_2)] = \mathbf{0}. \quad (8)$$

Further, with \mathbf{y} given by (4), we have that

$$\mathbf{R} = \mathbf{A}(\bar{\boldsymbol{\theta}})\mathbf{P}\mathbf{A}^*(\bar{\boldsymbol{\theta}}) + \sigma^2\mathbf{I}. \quad (9)$$

We cannot observe (7) or (9) directly; rather, from N snapshots of $\mathbf{y}(t)$, we form an estimate according to

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{k=1}^N \mathbf{y}(k)\mathbf{y}^*(k). \quad (10)$$

1.1 Performance bounds in SAP

As mentioned earlier in the introductory example, we can determine the lowest achievable variance for any unbiased estimator of some parameters in a given problem. The Cramér-Rao Bound provides such a bound, and we will here show one way of finding it based on the data-model and assumptions of Section 1. See, e.g., [1] for additional background on the theory, and [11] for more on the methodology used below.

If observing N snapshots of $\mathbf{y}(t)$ under the assumptions in Section 1, the (k, l) -th element of the Fisher Information Matrix for a specified *parameter vector* $\boldsymbol{\alpha}$ is given by Bangs' formula [12]

$$\text{FIM}_{k,l} = N \text{Tr} \left(\frac{\partial \mathbf{R}}{\partial \alpha_k} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \alpha_l} \mathbf{R}^{-1} \right); \quad (11)$$

the elements of $\boldsymbol{\alpha}$ are the unknown parameters of the model. One can write (11) in matrix form as

$$\frac{1}{N} \text{FIM} = \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\alpha}^T} \right)^* (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}) \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\alpha}^T} \right), \quad (12)$$

where $\mathbf{r} = \text{vec}(\mathbf{R})$.

Thus, with no known parameters in (9), the parameter vector

$$\boldsymbol{\alpha} = [\boldsymbol{\theta}^T \quad \boldsymbol{\varrho}^T \quad \sigma^2]^T \quad (13)$$

is obtained, in which $\boldsymbol{\theta} = [\theta_1 \quad \dots \quad \theta_d]^T$ is the vector of *unknown* DOAs, $\boldsymbol{\varrho}$ is the vector made from $\{\mathbf{P}_{ii}\}$ and $\{\text{Re}(\mathbf{P}_{ij}), \text{Im}(\mathbf{P}_{ij}); i > j\}$, and σ^2 is the noise variance. The reason to choose the elements of $\boldsymbol{\varrho}$ as above is that \mathbf{P} is hermitian, and that the choice of real parameters facilitate later derivations.

In contrast to the low-rank model (9), in some applications (e.g., Paper F/ [13]) a viable option is to use the (unique) elements of \mathbf{R} as parameters in $\boldsymbol{\alpha}$; see e.g. [14] for a derivation of the CRB for such a scenario.

For $\boldsymbol{\alpha}$ as in (13) we proceed as follows. Note that

$$\mathbf{r} = (\mathbf{A}^c \otimes \mathbf{A}) \text{vec}(\mathbf{P}) + \sigma^2 \text{vec}(\mathbf{I}), \quad (14)$$

where the superscript c denotes complex conjugation, and the identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ for any matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of matching dimensions has been used. We can then introduce the following notation:

$$\begin{aligned} \mathbf{W}^{1/2} \left[\begin{array}{c|c|c} \frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^T} & \frac{\partial \mathbf{r}}{\partial \boldsymbol{\varrho}^T} & \frac{\partial \mathbf{r}}{\partial \sigma^2} \end{array} \right] \\ \triangleq \left[\mathbf{G} \mid \mathbf{V} \mid \mathbf{u} \right], \end{aligned} \quad (15)$$

where $\mathbf{W}^{1/2} = \mathbf{R}^{-T/2} \otimes \mathbf{R}^{-1/2}$. If one additionally defines

$$\left[\mathbf{V} \mid \mathbf{u} \right] \triangleq \Delta, \quad (16)$$

we can, using (15) and (16), re-write (12) as

$$\frac{1}{N} \text{FIM} = \begin{bmatrix} \mathbf{G}^* \\ \Delta^* \end{bmatrix} \left[\mathbf{G} \mid \Delta \right]. \quad (17)$$

The rationale for re-writing (12) as in (17) is the following: in Section 2, we are interested in closed-form expression for the CRB for the unknown DOAs $\boldsymbol{\theta}$ while the other parameters in $\boldsymbol{\alpha}$ are simply nuisance parameters which we do not explicitly desire to estimate. Hence, from (17) a standard result on partitioned matrix inversion gives that

$$\begin{aligned} \frac{1}{N} \text{CRB}_{\boldsymbol{\theta}}^{-1} &= \mathbf{G}^* \mathbf{G} - \mathbf{G}^* \Delta (\Delta^* \Delta)^{-1} \Delta^* \mathbf{G} \\ &= \mathbf{G}^* \boldsymbol{\Pi}_{\Delta}^{\perp} \mathbf{G}. \end{aligned} \quad (18)$$

Given the particular form of prior knowledge explored in Paper A/ [15] and Paper B/ [16], the solution to (18) will take different forms; for further details see the respective papers.

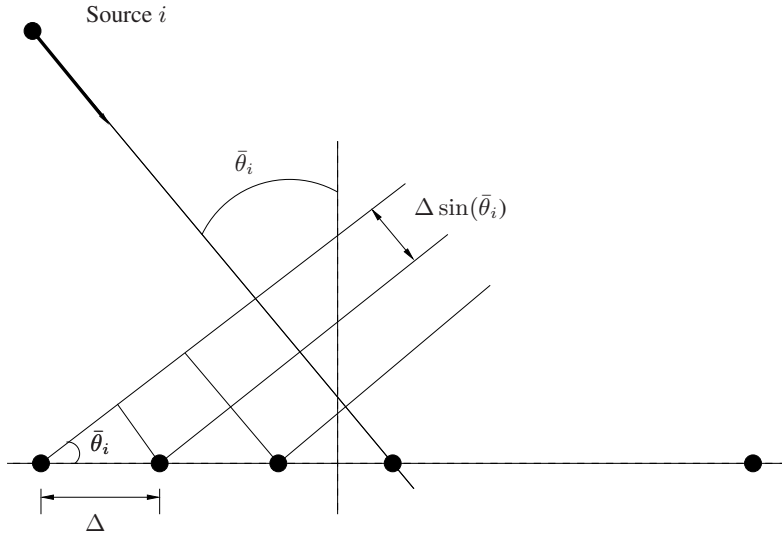


Figure 3: Schematic depiction of ULA and distance/time delay between sensors

2 Direction of arrival estimation

Paper A/ [15], Paper B/ [16], and Paper C/ [17] are all dedicated to Direction of Arrival (DOA) estimation problems. In these works, we assume that the sensors and the sources share a common plane in space — thus it is possible to characterize the direction to each source by a single angle. The objective is to estimate the vector $\boldsymbol{\theta} = [\bar{\theta}_1 \ \dots \ \bar{\theta}_d]^T$, where $\bar{\theta}_i$ is the angle to the i th source.

With the assumption that the array is unambiguous and its geometry known, for each given array response $\mathbf{A}(\boldsymbol{\theta})$, there is a unique parameterizing $\boldsymbol{\theta}$. From (9), we can access $\mathbf{A}(\boldsymbol{\theta})$ and hence the DOAs.

We can decompose \mathbf{R} in (9) as

$$\mathbf{R} = \sum_{k=1}^m \lambda_k \mathbf{u}_k \mathbf{u}_k^*, \quad (19)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ are the ordered eigenvalues, and \mathbf{u}_k the corresponding eigenvectors. The rank of \mathbf{P} is equal to $d' \leq d$ (a rank deficiency in \mathbf{P} signifies that some of the sources' signals are coherent), and we thus have that $\lambda_{d'+1} = \lambda_{d'+2} = \dots = \lambda_m = \sigma^2$. Denote the space spanned by $\mathbf{A}(\bar{\boldsymbol{\theta}})\mathbf{P}$ as the *signal subspace*; then, by collecting the eigenvectors corresponding to the largest d' eigenvalues in \mathbf{E}_s , we can note that \mathbf{E}_s spans this signal subspace. Similarly, the remaining $m-d'$ eigenvectors collected in \mathbf{E}_n span the *noise subspace*. Accordingly, we can then write (7) as

$$\mathbf{R} = \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^* + \mathbf{E}_n \boldsymbol{\Lambda}_n \mathbf{E}_n^*, \quad (20)$$

where $\mathbf{A}_s = \text{diag}(\lambda_1, \dots, \lambda_{d'})$, and $\mathbf{A}_n = \text{diag}(\lambda_{d'+1}, \dots, \lambda_m) = \sigma^2 \mathbf{I}$.

The rationale for (20) is of course that we do not have access to the true \mathbf{R} , but rather have to estimate it from data; accordingly, we observe N samples (*snapshots*) of the array output $\mathbf{y}(t)$, and then perform the eigen-decomposition of $\widehat{\mathbf{R}}$ (10), collecting the eigenvectors and -values as in (20). Then,

$$\widehat{\mathbf{R}} = \widehat{\mathbf{E}}_s \widehat{\mathbf{\Lambda}}_s \widehat{\mathbf{E}}_s^* + \widehat{\mathbf{E}}_n \widehat{\mathbf{\Lambda}}_n \widehat{\mathbf{E}}_n^*. \quad (21)$$

In order to estimate $\bar{\boldsymbol{\theta}}$ from $\widehat{\mathbf{E}}_s$ we note the following. If we define a matrix $\mathbf{B}(\bar{\boldsymbol{\theta}}) \in \mathbb{C}^{m \times m-d}$ that spans the null-space of $\mathbf{A}(\bar{\boldsymbol{\theta}})$, i.e. such that

$$\mathbf{B}^*(\bar{\boldsymbol{\theta}})\mathbf{A}(\bar{\boldsymbol{\theta}}) = \mathbf{0}, \quad (22)$$

we also have that $\mathbf{B}^*(\bar{\boldsymbol{\theta}})\mathbf{E}_s(\bar{\boldsymbol{\theta}}) = \mathbf{0}$. Hence we can form the criterion function associated with Weighted Subspace Fitting, [18]:

$$V_{\text{WSF}} = \text{vec}^* \left(\mathbf{B}^*(\bar{\boldsymbol{\theta}})\widehat{\mathbf{E}}_s \right) \mathbf{W} \text{vec} \left(\mathbf{B}^*(\bar{\boldsymbol{\theta}})\widehat{\mathbf{E}}_s \right), \quad (23)$$

where $\text{vec}(\cdot)$ stacks the columns of a matrix on top of each other, and $\mathbf{W} > \mathbf{0}$ is a weighting matrix.

The WSF-estimates are then found according to

$$\hat{\bar{\boldsymbol{\theta}}}_{\text{WSF}} = \arg \min_{\bar{\boldsymbol{\theta}}} V_{\text{WSF}}; \quad (24)$$

the weighting matrix \mathbf{W} in (23) is chosen as to minimize the asymptotic (in N) variance of the estimates $\hat{\bar{\boldsymbol{\theta}}}_{\text{WSF}}$.

2.1 Uniform linear array at receiver

This thesis focuses on when the receiving array is a so-called Uniform Linear Array (ULA) — this is an array where the m identical sensors are equi-spaced on a line, schematically depicted in Fig. 3. One advantage of the ULA formulation is that it allows significant computational savings when trying to estimate the DOAs. In Paper A/ [15] and Paper C/ [17], we have also specifically exploited the ULA structure to attain especially accurate results.

For a ULA, we can write

$$\mathbf{A}(\bar{\boldsymbol{\theta}}) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_d} \\ \vdots & \vdots & \dots & \vdots \\ e^{j(m-1)\omega_1} & e^{j(m-1)\omega_2} & \dots & e^{j(m-1)\omega_d} \end{bmatrix}, \quad (25)$$

where $j^2 = -1$, $\omega_i = -\pi \sin(\bar{\theta}_i)$ and the intra-sensor spacing of the ULA $\Delta = \frac{\lambda}{2}$, with λ the wave-length of the carrier signal. Then, the matrix \mathbf{B} in (22) can be

written as [19]

$$\mathbf{B}^T = \begin{bmatrix} b_0 & b_1 & \dots & b_d & \mathbf{0} \\ & \ddots & \ddots & & \ddots \\ \mathbf{0} & & b_0 & b_1 & \dots & b_d \end{bmatrix}, \quad (26)$$

where b_i are given as the coefficients of the polynomial

$$b_0 \prod_{i=1}^d (z - e^{-j\pi \sin(\bar{\theta}_i)}) \triangleq b_0 z^d + b_1 z^{d-1} + \dots + b_d. \quad (27)$$

The DOAs $\bar{\theta}$ parameterize $\mathbf{b} = [b_0 \dots b_d]^T$ through (27). We can hence minimize, e.g., (24) with respect to \mathbf{b} (since $\text{vec}(\mathbf{B}) = \mathbf{\Psi}\mathbf{b}$, with $\mathbf{\Psi}$ a selection matrix), and find $\hat{\bar{\theta}}$ from the minimizing $\hat{\mathbf{b}}$; see e.g. [19], [20] regarding such minimizations.

2.2 Prior information

In certain scenarios some of the parameters are known even before any data is acquired. Then, for a given data size N one can intuitively expect that if fewer parameters are to be estimated, a higher accuracy can be obtained in the ones that are to be estimated. This is indeed true in many scenarios, and can easily be confirmed by a comparison of the relevant CRBs for the sought parameters. It is then desirable to be able to include these known parameters in the functional form of the final estimator.

For example, as already mentioned, the number of sources d are considered to be known. This quantity could also be estimated; if estimated, there might be estimation errors. Such errors will impact the performance of e.g. the estimator shown in Section 2; the error would e.g. materialize as a dimensional error of $\hat{\mathbf{E}}_s$ in (23), which is likely to cause performance degradations in the estimate of θ , in addition to over- (or under-)estimating the number of actual sources.

This thesis essentially studies two types of prior information in DOA applications; the case when some of the DOAs of $\bar{\theta}$ are known, and the case when there is also information on the correlation state between (some of) the source signals. We first detail one way of incorporating known directional information, and then investigate two different types of correlation information in conjunction with known DOAs. Finally, we also look at the scenario when there is some uncertainty in directions considered to be known.

Known source directions

In the particular DOA problems that are under study in this work, it is assumed that there is prior information on some of the source locations: we assume that we know d_k of the angles in $\bar{\theta}$ — without loss of generality, it is then possible to

partition $\bar{\boldsymbol{\theta}}^T = [\boldsymbol{\theta}^T \quad \boldsymbol{\vartheta}^T]$, where we let $\boldsymbol{\theta}$ represent the unknown directions, and $\boldsymbol{\vartheta}$ the known ones. The objective is to estimate the $d_u = d - d_k$ angles $\boldsymbol{\theta}$. The idea of incorporating such information has been extensively investigated; some examples include [21], [22], and [23].

Another way of adding this information to the direction finding problem was first introduced in [24], and detailed properly in [15]; this method is denoted PLEDGE, described below. Such an approach utilizes the fact that the minimization variables of, e.g., the criterion function (23) are polynomial coefficients. Then, if some of the DOAs are known this corresponds to knowing some of the roots of the polynomial (27). Based on the known roots we can factor them from the unknown polynomial. Note that no information is lost or corrupted in this way: since we know some of the DOAs, which implies knowing some of the roots of the polynomial (27), what we in effect are doing is to constrain the polynomial to have some of its roots at certain known locations. The estimates of the remaining unknown roots will then be presumably more accurate.

Thus we write (27) as

$$b_0 \prod_{i=1}^d (z - e^{-j\pi \sin(\bar{\theta}_i)}) = P_k(z)P_u(z), \quad (28)$$

in which

$$P_u(z) = \tilde{b}_0 \prod_{i=1}^{d_u} (z - e^{-j\pi \sin(\bar{\theta}_i)}) = \tilde{b}_0 z^{d_u} + \dots + \tilde{b}_{d_u} \quad (29)$$

and

$$P_k(z) = c_0 \prod_{i=d_u+1}^d (z - e^{-j\pi \sin(\bar{\theta}_i)}) = c_0 z^{d_k} + \dots + c_{d_k}. \quad (30)$$

Here, $P_k(z)$ is the polynomial with d_k zeros corresponding to the known DOAs whereas $P_u(z)$ has $d_u = d - d_k$ zeros corresponding to the unknown DOAs. We use (28)-(30) to write

$$\mathbf{b} = \mathbf{C}\tilde{\mathbf{b}}, \quad (31)$$

where

$$\tilde{\mathbf{b}} = [\tilde{b}_0 \quad \tilde{b}_1 \quad \dots \quad \tilde{b}_{d_u}]^T \quad (32)$$

and

$$\mathbf{C}^T = \begin{bmatrix} c_0 & c_1 & \dots & c_{d_k} & & \mathbf{0} \\ & \ddots & \ddots & & \ddots & \\ \mathbf{0} & & c_0 & c_1 & \dots & c_{d_k} \end{bmatrix}, \quad (33)$$

with $\mathbf{C} \in \mathbb{C}^{(d+1) \times (d_u+1)}$.

In Paper A/ [15] it is described how to implement this framework. In Fig. 4, we investigate the general case when some DOAs are known; we display the performance of “MODE”, [25], which is not utilizing any of the above-mentioned

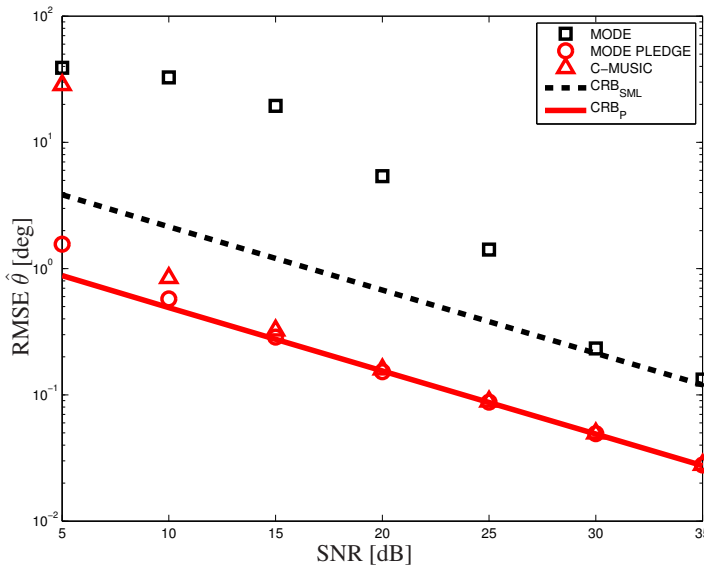


Figure 4: Sources located at $\bar{\boldsymbol{\theta}} = [10^\circ \ 15^\circ \ 12^\circ]^\text{T} \triangleq [\theta_1 \ \theta_2 \ \vartheta]^\text{T}$. RMSE of $\hat{\theta}_1$. Number of snapshots $N = 1000$, $m = 10$, equipowered sources, 1000 MC-realizations. θ_2 uncorrelated with the other (mutually coherent) sources.

prior knowledge; “MODE PLEDGE” [24] (and additionally in Paper A/ [15]), and “C-MUSIC” [22] which is exploiting the knowledge of $\boldsymbol{\vartheta}$, but in different ways. Additionally, the performance bounds are “CRB_{SML}”, corresponding to no prior knowledge and “CRB_P”, corresponding to knowledge of $\boldsymbol{\vartheta}$. \mathbf{P} in (5) can, for $d = 3$ sources, be expressed as

$$\mathbf{P} = \begin{bmatrix} p_1 & \rho_{12}\sqrt{p_1p_2} & \rho_{13}\sqrt{p_1p_3} \\ \rho_{12}^c\sqrt{p_1p_2} & p_2 & \rho_{23}\sqrt{p_2p_3} \\ \rho_{13}^c\sqrt{p_1p_3} & \rho_{23}^c\sqrt{p_2p_3} & p_3 \end{bmatrix}, \quad (34)$$

and for the scenario studied in Fig. 4, we have that $\rho_{12} = \rho_{23} = 0$; $\rho_{13} = \exp(-j\pi/12)$; ρ_{13} is chosen as to maximize the difficulty of the estimation scenario.

It can be seen from Fig. 4 that MODE PLEDGE is significantly more accurate than MODE. MODE PLEDGE is in the studied scenario also converging sooner (in SNR-terms) to its ultimate accuracy bound than C-MUSIC. Note that, due to the prior knowledge, the coherence between two of the signals is essentially nulled; thus C-MUSIC can be successfully used in this scenario, even though MUSIC in general breaks down when the signals are coherent.

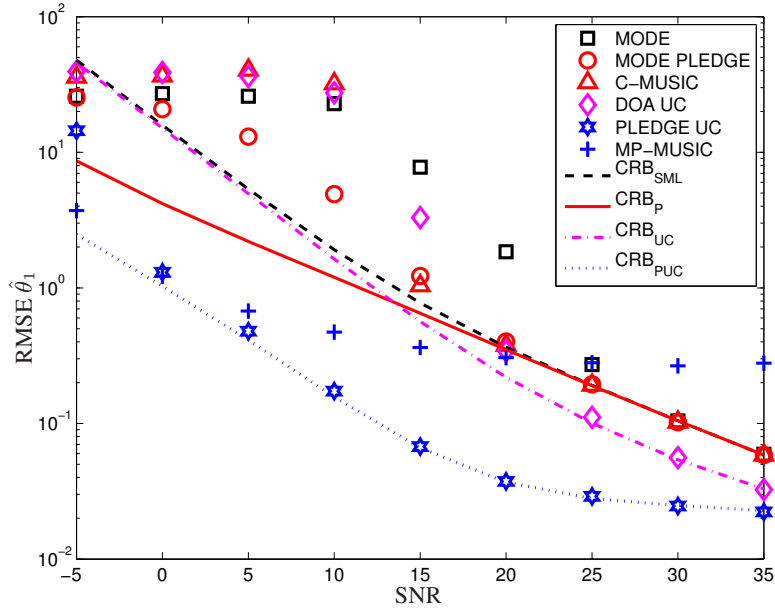


Figure 5: Showing the RMSE (for $\hat{\theta}_1$) of numerous estimators, detailed in the text, for the scenario in which the source signals are uncorrelated. Also displaying CRBs for the different assumed scenarios. $\theta = [10^\circ \ 15^\circ]^\text{T}$, $\vartheta = [12^\circ \ 25^\circ]^\text{T}$. Equipowered sources, $m = 10$, and averages of 1000 independent Monte Carlo realizations.

Other forms of prior knowledge

It turns out that combining several types of prior knowledge on the source signals is especially advantageous; this was first investigated in [26], and later in Paper A/ [15]. In these publications, it was assumed that in addition to knowing ϑ , there is the additional knowledge that the source signals are perfectly uncorrelated, i.e. $\mathbf{P} = \text{diag}([p_1 \ p_2 \ \dots \ p_d])$, where p_i represents the power of the signal from direction θ_i . Methods that explicitly exploit that the sources are uncorrelated are detailed in [27], [28], and thus in [26] the method in [28] was modified along the lines of the PLEDGE framework described in Section 2.2.

In Fig. 5, we compare several estimators in order to highlight the performance gained by exploiting the prior information: the estimators not already introduced are “DOA UC” [28] exploiting that \mathbf{P} is diagonal; and “PLEDGE UC” [26] and “P-MUSIC” [24], exploiting both forms of prior knowledge. Since each estimator is exploiting some assumptions (i.e., considering different parameters to be unknown), we show the CRB corresponding to these assumptions; bounds not previously detailed are “CRB_{UC}”, corresponding to \mathbf{P} known to be diagonal, and

“CRB_{PUC}”, corresponding to both the knowledge of \mathbf{P} being diagonal and knowing ϑ . As can be seen in Fig. 5, in the studied scenario (with uncorrelated sources) the mere knowledge of some DOAs in itself gives only a marginal accuracy increase (“MODE PLEDGE”/“C-MUSIC”) as compared to not exploiting any prior information (“MODE”). However, when properly accounting for the signals being uncorrelated, a very dramatic accuracy gain is displayed (“PLEDGE UC”).

A different method of exploiting combined prior information is further investigated in [29] and Paper B/ [16]; the method developed and analyzed in these publications is in some aspects similar to [22], but where the latter only is concerned about prior DOA knowledge, the former also assumes the source correlation matrix to be block diagonal, i.e.

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_u & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_k \end{bmatrix}, \quad (35)$$

with \mathbf{P}_u and \mathbf{P}_k denoting the correlation matrices of the unknown and known source signals, respectively. To see how this information can be especially beneficial, we note the following (where we additionally use the notation $\mathbf{A}_u = \mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{A}_k = \mathbf{A}(\vartheta)$): based on (4),

$$\begin{aligned} \mathbf{R} - \sigma^2 \mathbf{I} &= \mathbf{A}(\bar{\boldsymbol{\theta}}) \mathbf{P} \mathbf{A}^*(\bar{\boldsymbol{\theta}}) \\ &= \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* + \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^*. \end{aligned} \quad (36)$$

Now, since ϑ is known, we can multiply (36) by the projector onto the null space of \mathbf{A}_k , $\mathbf{\Pi}_{\mathbf{A}_k}^\perp = \mathbf{I} - \mathbf{A}_k (\mathbf{A}_k^* \mathbf{A}_k)^{-1} \mathbf{A}_k^*$, which nullifies the second term in (36):

$$(\mathbf{R} - \sigma^2 \mathbf{I}) \mathbf{\Pi}_{\mathbf{A}_k}^\perp = \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* \mathbf{\Pi}_{\mathbf{A}_k}^\perp. \quad (37)$$

Thus, similarly to (22), we can parameterize the minimization with respect to only the unknown DOAs by a matrix $\mathbf{B}_u \triangleq \mathbf{B}(\boldsymbol{\theta})$ according to

$$\mathbf{B}_u^* (\mathbf{R} - \sigma^2 \mathbf{I}) \mathbf{\Pi}_{\mathbf{A}_k}^\perp. \quad (38)$$

As in Section 2, \mathbf{R} (and σ^2) in (38) have to be estimated from data; the resulting WSF-related estimator was first presented in [29], and in Paper B/ [16] we properly derived and investigated its statistical properties. We denote the estimator POWDER (Prior Orthogonally Weighted Direction Estimator).

As it turns out, a very large fraction of the performance increase seen in Fig. 5 (and in [26] and Paper A/ [15]) can be traced to the assumption in (35); thus POWDER in practice gives almost the same accuracy as the methods exploiting the stricter, uncorrelated, assumptions (see Paper B/ [16] for more details). In addition, POWDER is also applicable to a larger set of scenarios since there are no requirements on \mathbf{P}_u or \mathbf{P}_k , and it shows better finite sample performance than the PLEDGE-based methods in some simulated scenarios. In Fig. 6, we explore the scenario when (35) is satisfied, but \mathbf{P}_u and \mathbf{P}_k are not diagonal, ruling out the

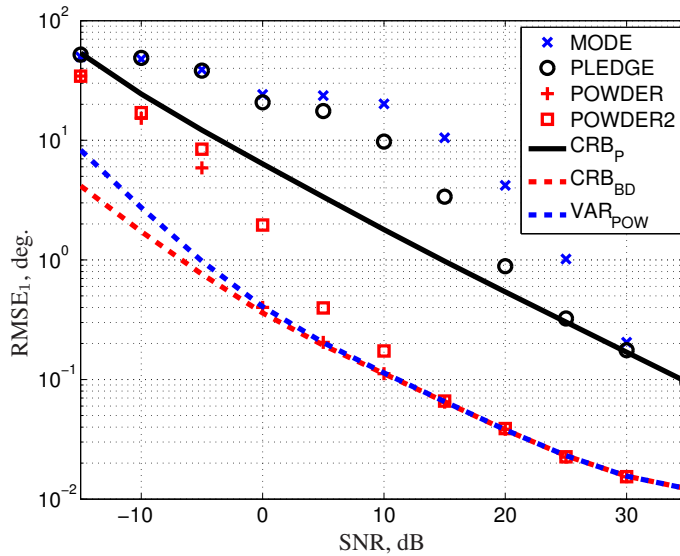


Figure 6: Comparison of POWDER to existing state-of-the-art methods. Normalized RMSE, averaged over 1000 MC realizations, along with the derived theoretical bounds; $\rho_k = \rho_u = 0.9$, and $N = 1000$. $\theta = [10^\circ \ 15^\circ]^\top$, and $\vartheta = [12^\circ \ 20^\circ]^\top$.

usage of e.g. PLEDGE UC. In addition to the above mentioned estimators, we have also included “POWDER2” which is a slightly modified version of POWDER (see Paper B/ [16] for more details), as well as the theoretical variance of POWDER (“VAR_{POW}”), along with the CRB (“CRB_{BD}”), given the assumption of block-diagonal \mathbf{P} and known ϑ . In Fig. 6 it can be seen that POWDER significantly outperforms PLEDGE (and MODE).

Modeling uncertainty in known source-directions

In Paper C/ [17], we study the scenario when we allow for some uncertainty in the known directions. A plausible such scenario is e.g. when an estimate of a location is provided to some form of fusion center that is also processing an array response. The “known” DOAs ϑ are modeled as random variables,

$$\vartheta \sim \mathcal{N}(\bar{\vartheta}, \Theta), \quad (39)$$

independent from the source signals $\mathbf{s}(t)$ and noise $\mathbf{n}(t)$.

The methodology in Paper C/ [17] is based on “Generalized WSF”, [30], where it is proven that the uncertainty in ϑ can be optimally countered by modifying the

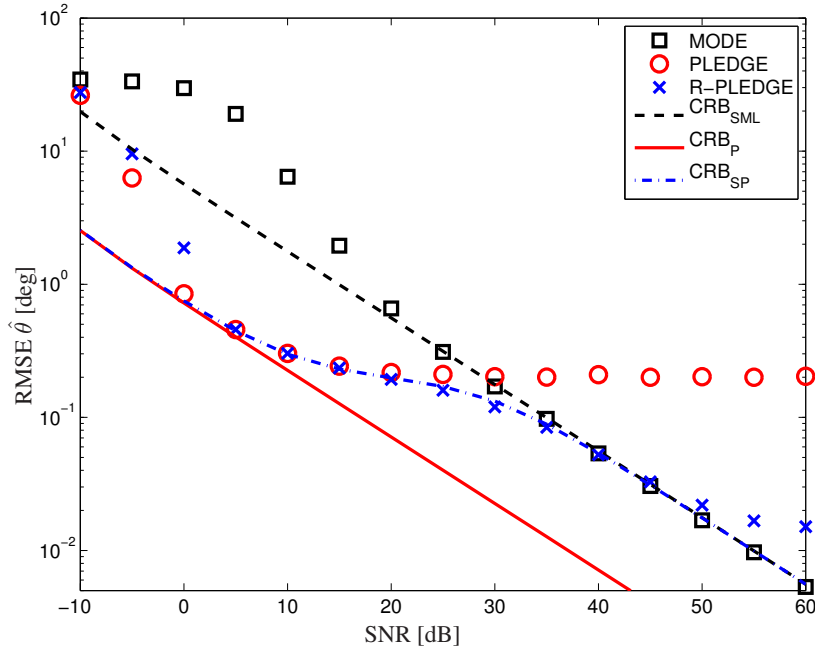


Figure 7: RMSE of 3 estimators and CRBs, averages of 4000 MC realizations; $N = 100$, $\theta = 15^\circ$, $\bar{\vartheta} = 12^\circ$, $\Theta = 0.2^2$, $m = 10$.

weighting matrix \mathbf{W} in (23), for “small” Θ . With the modified weighting matrix, the optimization problem is solved as in Paper A/ [15]. An example of the result is given in Fig. 7, where we simulate a scenario with two sources; the unknown is located at $\theta = 15^\circ$ and the known source is, for each Monte Carlo simulation, a realization of the random variable $\vartheta \sim \mathcal{N}(\bar{\vartheta}, \Theta)$ with $\bar{\vartheta} = 12^\circ$ and $\Theta = 0.2^2[\text{deg}^2]$. The sources are perfectly correlated which gives $\mathbf{P} = 10^{\text{SNR}/10} \begin{bmatrix} 1 & \rho \\ \rho^c & 1 \end{bmatrix}$, and we choose $\rho = \exp(-j0.72)$ in order to create a difficult estimation scenario.

In addition to the previously mentioned methods (note that “MODE PLEDGE” assumes $\vartheta = \bar{\vartheta}$, i.e. deterministic prior knowledge), we show the method proposed in Paper C/ [17], denoted “R-PLEDGE” (“Robust PLEDGE”). The accuracy bound “ CRB_{SP} ”, is the CRB taking the stated uncertainty into account, corresponding to the de-facto scenario. This bound is derived in Paper C/ [17]. The bound “ CRB_{P} ” is (in this case erroneously) assuming that $\vartheta = \bar{\vartheta}$.

In Fig. 7 it is seen that the proposed method R-PLEDGE and PLEDGE are about as accurate when the noise-induced estimation error is larger than the variance of the known source. However, as the SNR increases the known source variance becomes predominant and PLEDGE is not able to benefit from the higher

SNR due to the error in the prior. R-PLEDGE is seen to transition to the accuracy of the no-prior estimator, MODE, in this case, and can thus deliver as accurate estimates as is possible. For very high SNRs, however, R-PLEDGE begins to suffer from the error in the prior, and it is seen that its estimates are worse than those of MODE.

2.3 Conclusions and future work

In the DOA-related contribution of the thesis, the benefits of including known prior information already in the design of the estimator have been examined; we have shown that it is possible to achieve significantly more accurate results than if this information is neglected. Further, the devised prior-exploiting estimators are not significantly more computationally burdensome than the estimators that do not exploit such information. In Paper B/ [16], we additionally showed that for the known-source case, some assumptions on the receiving array can be relaxed.

Except for in Paper C/ [17], we have only studied deterministic prior information. However, the methodology of Paper C/ [17] is straight-forward to implement in the other studied scenarios. The full Bayesian treatment, as in e.g. the related problem analyzed in [31], is also a candidate for further studies.

3 Frequency estimation

Frequency estimation is one of the ubiquitous topics in signal processing: there are countless direct applications ranging from industrial machine diagnosis [32] to supply-current monitoring in smart grids [33], as well as mathematically analogous problems such as blood velocity estimation [34] and explosives detection [35]. Over decades, many algorithms have been proposed to solve such general problem, e.g. [19, 36, 37]. Recently there has been a renewed interest in the subject with the development of sparse estimation methods [38–40].

The general problem is to estimate the complex exponentials in the signal

$$y_0(t) = \sum_{k=1}^d \alpha_k e^{s_k t}, \quad t = 0, 1, \dots, N-1, \quad (40)$$

where α_k is the (complex) amplitude, and the exponential $s_k = \beta_k + j\omega_k$ contains the damping coefficient and angular frequency, respectively. Additionally, the amplitude can be written $\alpha_k = \bar{\alpha}_k e^{j\phi_k}$, where $\bar{\alpha}_k > 0$ is the real-valued amplitude of sinusoid k , and $\phi_k \in [0, 2\pi)$ its phase at $t = 0$; one may model ϕ_k as a random variable with a uniform distribution over its permissible values. Usually, ϕ_k is a nuisance parameter.

We cannot observe (40) directly and thus instead access

$$y(t) = y_0(t) + n(t), \quad (41)$$

where $n(t)$ is some additive noise corrupting the signal. Typically, we assume $n(t)$ to be a complex white circularly symmetric Gaussian i.i.d. process with variance σ^2 , i.e. $n(t) \sim \mathcal{N}(0, \sigma^2)$.

3.1 Sinusoidal estimation with prior information

We first study the estimation scenario when it is known that $\beta_k = 0, \quad \forall k$. A number of different methods have been proposed to tackle the sinusoids-in-noise estimation problem; we focus on the eigenanalysis-based methods such as [37, 41, 42]. The ability to incorporate prior knowledge of certain frequencies into the estimation of the remaining frequencies is not considered in the previous references. In [22, 43] some different approaches of doing so are evaluated. Our approach is to use the Prior knowLEDGE (PLEDGE) framework of [24] in two existing frequency estimators [25, 37, 42], and we also show the performance gains achieved by doing so. In addition, we apply the proposed estimation method to experimental data for broken rotor bar detection in an asynchronous induction motor; the objective of that analysis is to estimate sideband frequencies to the network frequency in order to diagnose the machine [44].

If we let

$$\mathbf{x}(t) = [\alpha_1 e^{j\omega_1 t} \quad \dots \quad \alpha_d e^{j\omega_d t}]^T, \quad (42)$$

we can reformulate the problem (41) into

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{n}(t), \quad t = 0, 1, \dots, N - m, \quad (43)$$

in which we have defined $\mathbf{y}(t) = [y(t) \ \dots \ y(t + m - 1)]^T$,

$$\mathbf{A} = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\omega_1} & \dots & e^{j\omega_d} \\ \vdots & & \vdots \\ e^{j(m-1)\omega_1} & \dots & e^{j(m-1)\omega_d} \end{bmatrix}, \quad (44)$$

m is a user parameter, and $\mathbf{n}(t)$ is defined similarly to $\mathbf{y}(t)$. Equation (43) is the standard matrix formulation for one-dimensional data, see e.g. [42]. From (42) we have that

$$\mathbf{P} \triangleq \mathbb{E}[\mathbf{x}(t)\mathbf{x}^*(t)] = \begin{bmatrix} \bar{\alpha}_1^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \bar{\alpha}_d^2 \end{bmatrix}, \quad (45)$$

for random i.i.d. phases ϕ_k . Then, from (43), one can find

$$\mathbf{R} \triangleq \mathbb{E}[\mathbf{y}(t)\mathbf{y}^*(t)] = \mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2\mathbf{I} \quad (46)$$

which is the data covariance matrix, cf. (7). Similarly, \mathbf{A} in (44) is nearly identical to (25). Thus, techniques used in SAP can be exploited, if adequately modified, to the frequency estimation problem.

The method devised in Paper D/ [45] is a straight-forward combination of [37] and PLEDGE, as described in Section 2.2 but tailored for the frequency estimation scenario. See Paper D/ [45] for explicit details. In Fig. 8, we investigate the performance of a number of different frequency estimators: “MODE” as in [25]; “ESPRIT” (using a forward-backward averaging of data), see e.g. [37]; “Markov” [37], [42]; and “MODE” and “Markov” modified according to Section 2.2, referred to as “MODE PLEDGE” and “Markov PLEDGE”. In the examined scenario, the vector length $\mathbf{y}(t)$ in (43) is varied by the parameter m ; the frequency content of the simulated signal is given by $\boldsymbol{\omega} = [2\pi 0.5 \ 2\pi 0.52 \ 2\pi 0.56]^T$, where $\omega_2 = 2\pi 0.52$ is considered known a-priori. We show the root mean squared error (RMSE) of the estimates corresponding to $\omega_1 = 2\pi 0.5$. For any value of m , Markov PLEDGE produces more accurate results than the other methods. Since MODE is not specifically tailored for frequency estimation, it is not surprising that it gives worse results; however, MODE PLEDGE is more accurate than the methods not utilizing the prior information.

While the ESPRIT method performs as well as the unmodified Markov method, it is not compatible with the PLEDGE concept and cannot inherently exploit the prior knowledge. It is otherwise an attractive estimator due to its lower computational complexity.

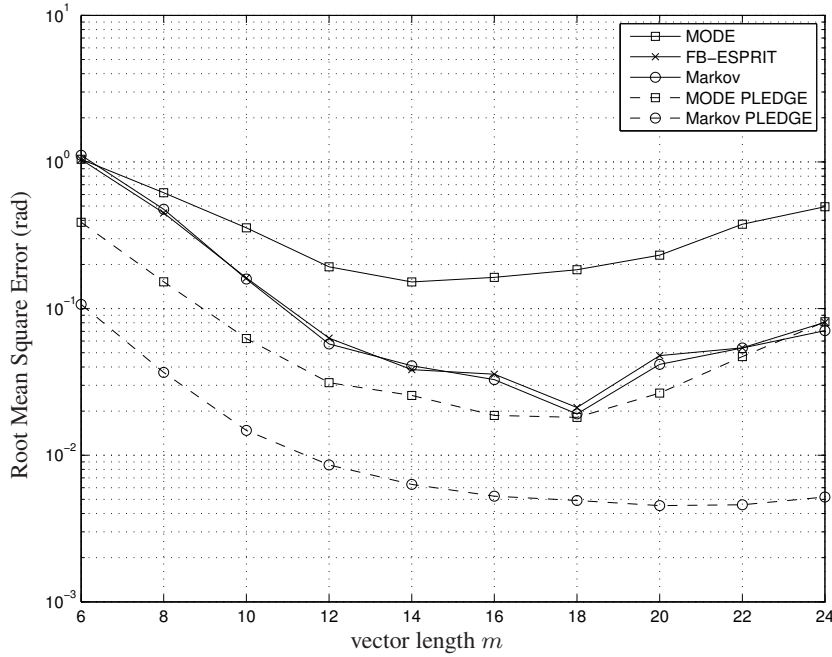


Figure 8: RMSE of estimates corresponding to $\omega_1 = 2\pi 0.5$, vs. m .

3.2 Pulse spin-locking data sequence estimation

Even though much work has been done on the estimation of damped and undamped sinusoids, there are only a few algorithms dealing with structured data models able to fit data produced by magnetic and quadrupolar resonance techniques. Such measurements are often resulting from the use of pulse spin-locking (PSL) sequences, which will then induce a fine structure into the signals. The PSL sequence consists of a preparatory pulse and a train of refocusing pulses, where the time between two consecutive refocusing pulses is 2τ , illustrated in Fig. 9. As discussed in [46, 47], the signal resulting from a PSL excitation can be well modeled as

$$y_{m,t} = x_{m,t} + w_{m,t}, \quad (47)$$

where $m = 0, \dots, M-1$ denotes the echo number, and $t = t_0, \dots, t_{N-1}$ the local time within each echo, where $t = 0$ denotes the center of the current pulse, and where we assume uniform sampling intervals within each echo. Moreover, $w_{m,t}$ is an additive circular symmetric white Gaussian i.i.d. noise with variance σ^2 , and

$$x_{m,t} = \sum_{k=1}^K \alpha_k \exp(i\omega_k t - \beta_k |t - \tau| - (t + 2\tau m)\eta_k), \quad (48)$$

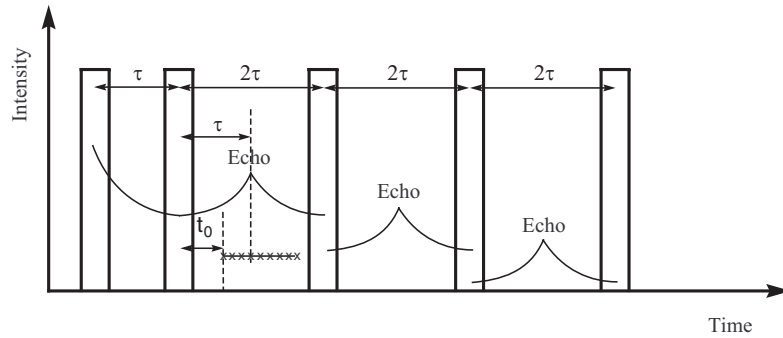


Figure 9: The PSL sequence.

where α_k , $\omega_k = 2\pi f_k$, β_k , and η_k denote the (complex) amplitude, frequency, damping coefficient (with respect to the *current* transmitted pulse), and the *compound*, or echo train, damping coefficient for the k th spectral line, respectively. Additionally, τ is a design parameter (due to operator choice of pulse repetition interval) and is thus known. The important differences between the models (48) and (41) are the following: in (48), an extra multiplicative damping term η_k is present. Further, measurements are essentially obtained for $t < 0$, meaning that the damping term β_k also acts as amplification, as seen in Fig. 9.

Common approaches for estimating the parameters for this form of data is to first sum the echoes, thereby destroying the finer details resulting from the echo train decay (and causing a bias in the estimates), and then use classical approaches such as the periodogram or the matrix pencil [48, 49]. Another approach to estimate all the unknown parameters in (48) was taken in [46]; however this method requires a gradient or grid-based search which in addition to careful initialization is also computationally demanding.

The algorithm developed in Paper E/ [50], termed the Echo Train ESPRIT (ET-ESP), requires no prior knowledge of typical parameter values, needing only knowledge of the number of present spectral lines K , a number that is generally known in these applications. In addition to the algorithm, the corresponding Cramér-Rao lower bound (CRB) for the problem is also derived, and the performance of the proposed algorithm is examined using numerical simulations.

Again, this method is also somewhat involved; in this introduction we will outline the first steps of the algorithm.

Let (48) be expressed as

$$x_{m,t} = \sum_{k=1}^K c_{m,k} z_k^t \quad (49)$$

with

$$c_{m,k} \triangleq \begin{cases} \alpha_k \exp(-\beta_k \tau) \cdot \exp(-2\eta_k \tau m) & \text{for } t < \tau \\ \alpha_k \exp(\beta_k \tau) \cdot \exp(-2\eta_k \tau m) & \text{for } t \geq \tau \end{cases}, \quad (50)$$

$$z_k \triangleq \begin{cases} \exp(i\omega_k) \cdot \exp(\beta_k - \eta_k) & \text{for } t < \tau \\ \exp(i\omega_k) \cdot \exp(-\beta_k - \eta_k) & \text{for } t \geq \tau \end{cases}; \quad (51)$$

note that z_k is not a function of m , as all echoes share the same poles. Reminiscent of [51], the noise-free data for each echo m is then stacked into the (Hankel) matrix

$$\mathbf{X}_m = \begin{bmatrix} x_{m,t_0} & x_{m,t_1} & \cdots & x_{m,t_{L'-1}} \\ x_{m,t_1} & x_{m,t_2} & \cdots & x_{m,t_{L'}} \\ \vdots & \vdots & & \vdots \\ x_{m,t_{L-1}} & \cdots & \cdots & x_{m,t_{N-1}} \end{bmatrix} \in \mathbb{C}^{L \times L'} \quad (52)$$

where $L' = N - L + 1$. This (noise-free) echo data matrix may then be collected into

$$\mathbf{X} = [\mathbf{X}_0 \quad \cdots \quad \mathbf{X}_{M-1}] \quad (53)$$

$$= \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \quad (54)$$

where the components of (54) are defined from the SVD of (53). Stacking the (noise corrupted) received data (47) similarly to (52) and (53), and then performing the SVD yields

$$\mathbf{Y} = \hat{\mathbf{U}} \hat{\mathbf{\Sigma}} \hat{\mathbf{V}}^* + \mathbf{W}, \quad (55)$$

where $\hat{\mathbf{\Sigma}}$ denotes the matrix formed from the K largest singular values, while $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ denote the matrices formed by the corresponding singular vectors. The residual term \mathbf{W} contains the noise.

Based on $\hat{\mathbf{U}}$ one can estimate the poles z_k using the ESPRIT-method [52]; the remaining parameters are estimated by least squares — see Paper E/ [50] for the complete treatment.

The algorithm was tested on simulated NQR data, formed as to mimic the response signal from the explosive TNT when excited using a PSL sequence. Such signals can be well modeled as a sum of four damped sinusoidal signals, with parameters as detailed in Paper E/ [50]. Based on the typical setup examined in [35], $N = 256$ measurements are used, for $M = 32$ echoes, with $\tau = 164$ and $t_0 = 36$, where the last two parameters are normalized with the sampling frequency and are therefore unit-less. The algorithm was evaluated using the normalized root mean squared error (NRMSE), defined as:

$$\text{NRMSE} = \sqrt{\frac{1}{P} \sum_{p=1}^P \left(\frac{\hat{x}_p}{x} - 1 \right)^2}, \quad (56)$$

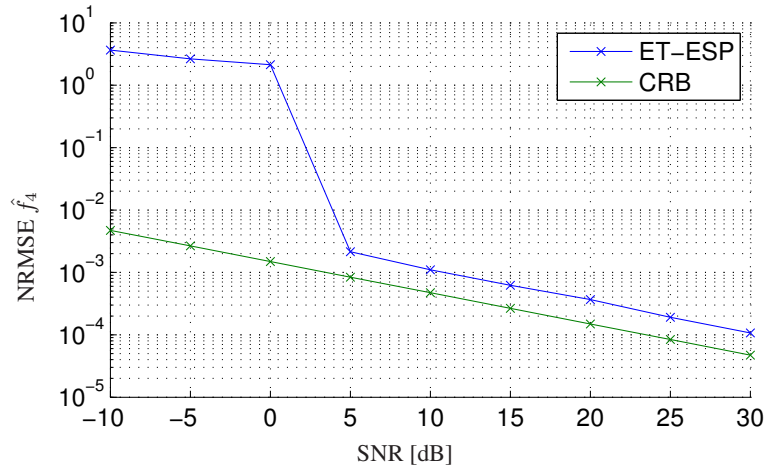


Figure 10: NRMSE for estimates of f_4 and the CRB.

where x denotes the true parameter value and \hat{x}_p the estimate of this parameter.

Fig. 10 shows the results from $P = 500$ Monte-Carlo simulations for the fourth spectral line (the performance for the other lines is similar). As is common for ESPRIT-based estimators, it can be noted that the ET-ESP estimate does not fully reach the CRB and is therefore not statistically efficient. However, the difference is very small down to $\text{SNR} = 0$ dB, before which the estimation error becomes very large. For $\text{SNR} = 5$ dB, the estimation error for the frequency f_4 is about 0.2%.

3.3 Conclusions and future work

Similarly to the DOA case in Section 2, we have explored various types of prior information. For the line-frequency estimation, it was shown in Paper D/ [45] that while the approach is beneficial, the accuracy increase is not as pronounced as in e.g. Paper A/ [15]. During the development of Paper D/ [45], it was realized that both the Markov-estimator, [37], and a related estimator, [42], could be improved; thus the creation of a WSF-like frequency estimator with better small-sample properties is a potential topic for further research.

In Paper E/ [50] we examined an existing data-model, and reformulated it in a novel way; we then tailored one of the classical SAP-estimators to this problem. We noted that the computational burden of the estimation process could be significantly reduced, while still giving sufficient accuracy. However, the ESPRIT-algorithm (which was the one tailored to the PSL-sequence data) is known to be sub-optimal in the SAP-scenario; thus, it appears plausible that if instead performing the parameter estimation in a proper WSF-like framework, it would be possible to obtain (asymptotical) ML-like performance while still keeping computational

complexity modest. This merits further detailed studies.

4 Kronecker structured covariance matrix estimation

In certain applications, e.g. [53], [54], [55], the objective of an estimation problem can be the covariance matrix \mathbf{R}_0 itself. This is explored in Paper F/ [13], where the covariance matrix of interest is modeled as a Kronecker product of two matrices of smaller dimensions, according to

$$\mathbf{R}_0 = \mathbf{A}_0 \otimes \mathbf{B}_0. \quad (57)$$

In practice, \mathbf{R}_0 is typically estimated from data, where \mathbf{A}_0 and \mathbf{B}_0 are *factor matrices* of dimensions $m \times m$, and $n \times n$, respectively, upon which further assumptions can be placed. In the herein studied scenario, \mathbf{A}_0 and \mathbf{B}_0 are positive definite (p.d.) and Toeplitz. When (57) is satisfied, it is highly desirable to tailor the estimator of \mathbf{R}_0 to this particular model in order to enhance accuracy and minimize computational costs: if \mathbf{A}_0 and \mathbf{B}_0 are hermitian, the number of (real valued) parameters in (57) are $m^2 + n^2$, whereas if estimating \mathbf{R}_0 simply under the hermitian constraint, $m^2 n^2$ parameters must be estimated.

Applications of (57) are numerous in the literature: in some instances of wireless communication, the model is applicable for the statistics of the channel between the transmitter and receiver, [56], [57]. When estimating covariance matrices from EEG/MEG data, the same model can be used, [58], [59]. In the separable statistical models [55], [60], (1) is the defining assumption. Recently, models where sparsity can be exploited in the estimation of \mathbf{R}_0 or its inverse have been proposed [61]. Especially when $m, n \gg 1$, sparsity has the potential to dramatically reduce the number of parameters to be estimated [62]; in this thesis, however, the factor matrices are non-sparse, and the dimensions m, n are fixed and finite.

In some of these applications, there is additional structure in the problem: each of the factor matrices in the Kronecker product might have a certain structure. For example, a factor matrix is Toeplitz if it is the covariance matrix of a stationary stochastic process [63]. In the EEG/MEG example, the temporal random process is stationary [58], [64], and thus the corresponding factor matrix is Toeplitz. In the MIMO communications example, when using uniform linear arrays at transmitter and receiver, the factor matrices are Toeplitz [65]. Similarly, if the arrays are merely symmetric around the respective array center point, we obtain persymmetric (PS) factor matrices (note that a Toeplitz structured matrix is per definition also PS). For further examples of structured factor matrices in signal processing applications, see [66]. From the above discussion, we conclude that it is desirable to design estimators that can benefit from such structure.

4.1 Accounting for Kronecker structure with persymmetric factor matrices

Based on the observation of (10) an ML-like estimator can be developed by finding the most probable candidate factor matrices \mathbf{A} and \mathbf{B} of (57). The estimator minimizes the negative log-likelihood function of the problem [14, 53, 55]; this function (up to a multiplicative constant) can be written

$$l(\mathbf{A}, \mathbf{B}) = n \log |\mathbf{A}| + m \log |\mathbf{B}| + \text{Tr} \left(\widehat{\mathbf{R}}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) \right), \quad (58)$$

if only including parameter dependent terms; $|\cdot|$ is the determinant function, and $\text{Tr}(\cdot)$ is the trace-operator.

If observing i.i.d. data $\{\mathbf{x}(t)\}_{t=1}^N$ drawn from $\mathcal{N}(\mathbf{0}, \mathbf{X})$, with \mathbf{X} unknown, it is well-known (see, e.g., [67]) that $\log |\mathbf{X}| + \text{Tr} \left(\widehat{\mathbf{R}}\mathbf{X}^{-1} \right)$ is minimized by $\mathbf{X} = \widehat{\mathbf{R}}$ if $\widehat{\mathbf{R}}$ is (p.d.). Thus, if we write

$$\text{Tr} \left(\widehat{\mathbf{R}}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) \right) = \text{Tr} \left(\sum_{k=1}^m \sum_{l=1}^m \widehat{\mathbf{R}}^{kl}(\mathbf{A}^{-1})_{lk} \mathbf{B}^{-1} \right), \quad (59)$$

where $\widehat{\mathbf{R}}^{kl}$ denotes the (k, l) th block of size $n \times n$ in $\widehat{\mathbf{R}}$, and $(\mathbf{A}^{-1})_{lk}$ is element l, k of \mathbf{A}^{-1} , it can be seen that (58) is minimized by

$$\widehat{\mathbf{B}} = \frac{1}{m} \sum_{k=1}^m \sum_{l=1}^m \widehat{\mathbf{R}}^{kl}(\mathbf{A}^{-1})_{lk} \quad (60)$$

for a fixed \mathbf{A} . A similar reasoning can be used to estimate \mathbf{A}_0 for a fixed \mathbf{B} , and thus a cyclic minimization algorithm (“flip-flop”) can be developed.

An objective of the algorithm is to incorporate the prior knowledge of PS factor matrices; that \mathbf{A}_0 and \mathbf{B}_0 are PS is equivalent to that

$$\mathbf{A}_0 = \mathbf{J}_m \mathbf{A}_0^T \mathbf{J}_m, \quad (61)$$

$$\mathbf{B}_0 = \mathbf{J}_n \mathbf{B}_0^T \mathbf{J}_n, \quad (62)$$

where

$$\mathbf{J}_k = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & & & \vdots \\ 1 & \dots & 0 & 0 \end{bmatrix}, \quad (63)$$

with the subscript k denoting the dimension of the square matrix. This constraint translates to solving the optimization problem

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}} l(\mathbf{A}, \mathbf{B}) \\ \text{s.t. } & \mathbf{A} \in \mathcal{P}_m, \mathbf{B} \in \mathcal{P}_n \end{aligned} \quad (64)$$

where $l(\mathbf{A}, \mathbf{B})$ is given by (4), and \mathcal{P}_m denotes the set of complex, persymmetric, hermitian p.d. matrices of dimension $m \times m$. Intuitively, the PS property means that a matrix is symmetric along the anti-diagonal; together with the fact that \mathbf{A}_0 and \mathbf{B}_0 are hermitian, it can be realized that the number of parameters to be estimated are significantly reduced as compared to an unstructured scenario.

If the sample covariance matrix $\hat{\mathbf{R}}$ is replaced by the Forward-Backward sample covariance matrix, $\hat{\mathbf{R}}_{\text{PS}} \triangleq \frac{1}{2}(\hat{\mathbf{R}} + \mathbf{J}\hat{\mathbf{R}}^T\mathbf{J})$ [68], the flip-flop algorithm (also known as alternating optimizations) described above solves (64), [13, 69]. Using $\hat{\mathbf{R}}_{\text{PS}}$ in lieu of $\hat{\mathbf{R}}$ thus gives estimates of \mathbf{A}_0 and \mathbf{B}_0 with the desired PS structure.

Additionally, it can be shown that the estimates from the flip-flop algorithm are asymptotically statistically efficient already after one complete cycle of the algorithm, [14]. The following algorithm is thus arrived at:

1. Choose an initial guess for \mathbf{A} , e.g. $\mathbf{A}_{\text{INIT}} = \mathbf{I}$.
2. Find \mathbf{B} based on $\hat{\mathbf{R}}_{\text{PS}}$ and \mathbf{A}_{INIT} , i.e. $\hat{\mathbf{B}}_{(1)} = \hat{\mathbf{B}}(\mathbf{A}_{\text{INIT}})$.
3. Find $\hat{\mathbf{A}}_{(2)}$ based on the previous estimate $\hat{\mathbf{B}}_{(1)}$ and $\hat{\mathbf{R}}_{\text{PS}}$.
4. Find $\hat{\mathbf{B}}_{(3)}$ based on $\hat{\mathbf{A}}_{(2)}$ and $\hat{\mathbf{R}}_{\text{PS}}$.
5. Finally, find the estimate $\hat{\mathbf{R}}_{\text{K}} = \left(\hat{\mathbf{A}}_{(2)} \otimes \hat{\mathbf{B}}_{(3)}\right)$ with the prescribed Kronecker-structure and PS structured factor matrices.

4.2 Accounting for Toeplitz structured factor matrices

A Toeplitz matrix is by construction also PS, which makes the method presented in Section 4.1 applicable; however, the exact Toeplitz structure will not be reproduced using such an approach. Hence, it is desirable to devise a method that produces factor matrices of the exact correct structure; it can be expected that estimates of \mathbf{R}_0 is more accurate for factor matrices of the correct structure.

We want to solve the optimization problem

$$\begin{aligned} \min_{\substack{\mathbf{A}, \mathbf{B} \\ \text{s.t. } \mathbf{A} \in \mathcal{T}_m, \mathbf{B} \in \mathcal{T}_n}} l(\mathbf{A}, \mathbf{B}), \end{aligned} \quad (65)$$

with $l(\mathbf{A}, \mathbf{B})$ given by (58) and where \mathcal{T}_m denotes the set of hermitian Toeplitz p.d. matrices of size $m \times m$. No algebraic solution is known to exist to (65); hence, we use the Extended Invariance Principle (EXIP) [70]. The cardinal idea of EXIP is the relaxation of the studied optimization problem such that it can be solved exactly, and a subsequent fit of the relaxed estimate to the restricted parameter set of the original optimization problem.

Applying this idea to the current context, we can relax (65) to (64); by doing so the number of available parameters is increased and an exact ML solution can be found to the relaxed problem. The PS estimate thus found is then considered to

be a noisy measurement of a Toeplitz-structured matrix; the reduced parameter set of this Toeplitz matrix is then estimated through weighted least squares [71].

To formalize the above discussion, we introduce some notation. Let $\hat{\mathbf{a}}_{(2)} = \text{vec}(\hat{\mathbf{A}}_{(2)})$, and $\hat{\mathbf{b}}_{(3)} = \text{vec}(\hat{\mathbf{B}}_{(3)})$ be the vectorized PS-structured estimates from the flip-flop algorithm, and let $\mathbf{a}_0 = \text{vec}(\mathbf{A}_0) \triangleq \mathbf{P}_A \boldsymbol{\theta}_{A_0}$, where $\boldsymbol{\theta}_{A_0}$ denotes the (real valued) Toeplitz parameters of \mathbf{A}_0 , and \mathbf{P}_A is the selection matrix mapping these parameters to the elements of \mathbf{a}_0 . We define \mathbf{b}_0 , \mathbf{P}_B , and $\boldsymbol{\theta}_{B_0}$ similarly, mutatis mutandis. Accordingly, we thus write

$$\boldsymbol{\epsilon} = \begin{bmatrix} \hat{\mathbf{a}}_{(2)} \\ \hat{\mathbf{b}}_{(3)} \end{bmatrix} - \begin{bmatrix} \mathbf{P}_A & \mathbf{0}_{m^2 \times n_B} \\ \mathbf{0}_{n^2 \times n_A} & \mathbf{P}_B \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_A \\ \boldsymbol{\theta}_B \end{bmatrix}, \quad (66)$$

where $\boldsymbol{\epsilon}$ is the error in the estimates of the true Toeplitz parameters. Rewriting (66), we have

$$\boldsymbol{\epsilon} = \hat{\boldsymbol{\eta}} - \mathbf{P}_\eta \boldsymbol{\theta}_\eta, \quad (67)$$

where

$$\hat{\boldsymbol{\eta}} \triangleq \begin{bmatrix} \hat{\mathbf{a}}_{(2)}^T & \hat{\mathbf{b}}_{(3)}^T \end{bmatrix}^T, \quad (68)$$

$$\mathbf{P}_\eta \triangleq \begin{bmatrix} \mathbf{P}_A & \mathbf{0}_{m^2 \times n_B} \\ \mathbf{0}_{n^2 \times n_A} & \mathbf{P}_B \end{bmatrix}, \quad (69)$$

$$\boldsymbol{\theta}_\eta \triangleq \begin{bmatrix} \boldsymbol{\theta}_A^T & \boldsymbol{\theta}_B^T \end{bmatrix}^T, \quad (70)$$

and $n_A \triangleq \dim(\boldsymbol{\theta}_A) = 2m - 1$, $n_B \triangleq \dim(\boldsymbol{\theta}_B) = 2n - 1$ are the number of parameters in the respective Toeplitz matrices. Based on the model (67), the weighted least squares estimate of $\boldsymbol{\theta}_\eta$ is found according to [71]

$$\hat{\boldsymbol{\theta}}_\eta = (\mathbf{P}_\eta^* \mathbf{W} \mathbf{P}_\eta)^{-1} \mathbf{P}_\eta^* \mathbf{W} \hat{\boldsymbol{\eta}}, \quad (71)$$

where the optimal choice of weighting matrix \mathbf{W} is $\mathbf{W}_{\text{OPT}} \triangleq (\mathbf{C}_{\hat{\boldsymbol{\eta}}} + \mathbf{P}_\eta \mathbf{P}_\eta^*)^\dagger$, [71] (here, $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$ is the asymptotic covariance of $\hat{\boldsymbol{\eta}}$).

The expression for $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$ involves the computation and manipulation of large matrices; a computationally less demanding method was proposed in [72] and in [13] the method was shown to give asymptotically statistically efficient estimates as well. It was proposed to estimate $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_B$ independently of each other, according to

$$\hat{\boldsymbol{\theta}}_{A_H} = (\mathbf{P}_A^* \mathbf{W}_{\hat{\mathbf{a}},H} \mathbf{P}_A)^{-1} (\mathbf{P}_A^* \mathbf{W}_{\hat{\mathbf{a}},H} \hat{\mathbf{a}}_{(2)}), \quad (72)$$

with $\mathbf{W}_{\hat{\mathbf{a}},H}$ a consistent estimate of $(\mathbf{A}_*^{-T} \otimes \mathbf{A}_*^{-1})$. The final factor matrix estimate is then found from $\text{vec}(\hat{\mathbf{A}}_{T,H}) = \mathbf{P}_A \hat{\boldsymbol{\theta}}_{A_H}$. The estimate $\text{vec}(\hat{\mathbf{B}}_{T,H})$ is found in the same way, mutatis mutandis.

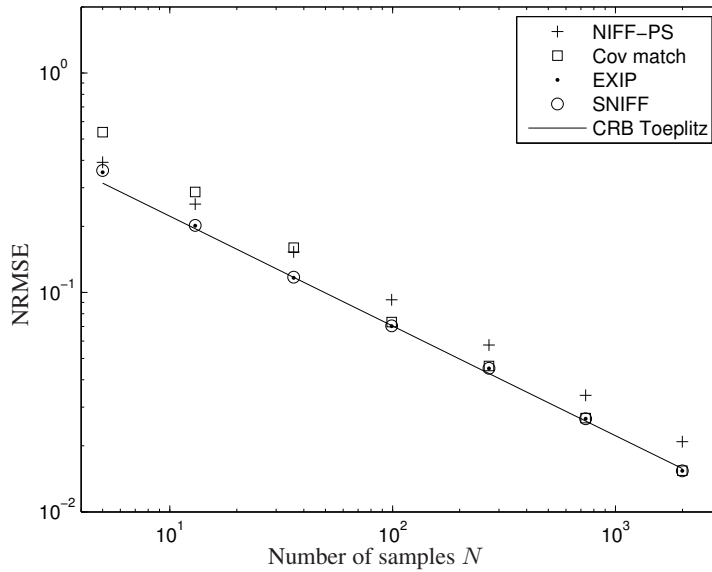


Figure 11: Normalized RMSE of four estimators as a function of number of samples. \mathbf{A}_0 is 4×4 , \mathbf{B}_0 is 4×4 ; both are Toeplitz structured. Averages taken over $L = 100$ Monte Carlo simulations for each sample size.

In Fig. 11 we compare the RMSE of 4 estimators of \mathbf{R}_0 , given that \mathbf{A}_0 and \mathbf{B}_0 are Toeplitz. “NIFF-PS” is the flip-flop estimator, accounting for the PS-structure of the matrices; “Cov match” from [14]; “EXIP”, where the Toeplitz parameters are found from (71); and finally “SNIFF”, based on (72). The estimates of \mathbf{R} given by the proposed WLS-based methods are not equivalent; however, as seen in Fig. 11, they both produce estimates with the same average accuracy. And, the variance of the estimates from both methods are equivalent to the relevant CRB, [13, 14], which is also attained by the covariance matching method (albeit at larger sample sizes).

5 Contributions

The thesis is based on the following papers:

- [A] Petter Wirfält, Guillaume Bouleux, Magnus Jansson, and Petre Stoica, “Optimal Prior Knowledge-Based Direction of Arrival Estimation”, in *IET Signal Processing*, vol. 6, no. 8, pp. 731–742, Oct. 2012.
- [B] Petter Wirfält and Magnus Jansson, “Robust Prior-based Direction of Arrival Estimation”, in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP’12)*, pp. 81–84, Ann Arbor, MI, USA, Aug. 2012.
- [C] Petter Wirfält and Magnus Jansson, “Prior-Exploiting Direction-of-Arrival Algorithms for Partially Uncorrelated Source Signals”, submitted to *Signal Processing*, Nov. 2013.
- [D] Petter Wirfält, Guillaume Bouleux, Magnus Jansson, and Petre Stoica, “Subspace-based Frequency Estimation Utilizing Prior Information”, in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP’11)*, pp. 533–536, Nice, France, Jun. 2011.
- [E] Erik Gudmundson, Petter Wirfält, Andreas Jakobsson, and Magnus Jansson, “An ESPRIT-based parameter estimator for spectroscopic data”, in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP’12)*, pp. 77–80, Ann Arbor, MI, USA, Aug. 2012.
- [F] Petter Wirfält and Magnus Jansson, “On Kronecker and Linearly Structured Covariance Matrix Estimation”, submitted to *IEEE Transactions on Signal Processing*, Jun. 2013.

Summary of the contributions by the author to papers A-F:

- [A] Wrote the paper; applied PLEDGE to the estimator originally tailored for the uncorrelated scenario; performed the CRB-analysis (the second of which was novel); performed the MC-simulations and real-data analysis.
- [B] Wrote the paper; tailored the existing GWSF-estimator to the case under study; performed CRB analysis; performed MC-simulations.
- [C] Wrote the paper; invented and developed the estimator; performed the theoretical analysis of the estimator; performed the CRB analysis; performed MC-simulations.
- [D] Wrote the paper; applied PLEDGE to the Markov-estimator; performed MC-simulations; applied estimators to real data.

- [E] Main responsible for estimator-section in paper; co-invented the estimator with principal author (approx. equal contribution).
- [F] Wrote the paper; invented and developed SNIFF, developed the EXIP-estimator; performed theoretical analysis; performed MC-simulations.

In addition to papers A-F, the following papers have also been (co)-authored by the thesis author:

- [i] Magnus Jansson, Petter Wirfält, Karl Werner, and Björn Ottersten, “ML Estimation of Covariance Matrices with Kronecker and Per-symmetric Structure”, in *IEEE Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop (DSP/SPE)*, pp. 298–301, Marco Island, FL, USA, Jan. 2009.
- [ii] Petter Wirfält and Magnus Jansson, “On Toeplitz and Kronecker Structure Covariance Matrix Estimation”, in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM’10)*, pp. 185–188, Jerusalem, Israel, Oct. 2010.
- [iii] Petter Wirfält, Magnus Jansson, Guillaume Bouleux, and Petre Stoica, “Prior knowledge-based direction of arrival estimation”, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2540–2543, Prague, Czech Republic, May 2011.
- [iv] Dave Zachariah, Petter Wirfält, Magnus Jansson, and Saikat Chatterjee, “Line spectrum estimation with probabilistic priors”, in *Signal Processing*, vol. 93, no. 11, pp. 2969–2974, 2013.
- [v] Petter Wirfält, Magnus Jansson, and Guillaume Bouleux, “Prior-Exploiting Direction-of-Arrival Algorithms for Partially Uncorrelated Source Signals”, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3972–3976, Vancouver, Canada, May 2013.

References

- [1] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1998.
- [2] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946.
- [3] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part III, Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Wiley, 2004.
- [4] H. Krim and M. Viberg, “Two Decades of Array Signal Processing Research,” *IEEE Signal Process. Mag.*, pp. 67–94, Jul. 1996.
- [5] J.-J. Fuchs, “Estimation of the number of signals in the presence of unknown correlated sensor noise,” *IEEE Trans. Signal Process.*, vol. 40, no. 5, pp. 1053–1061, 1992.
- [6] M. Wax, “Detection and localization of multiple sources via the stochastic signals model,” *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2450–2456, Nov. 1991.
- [7] M. Viberg, B. Ottersten, and T. Kailath, “Detection and estimation in sensor arrays using weighted subspace fitting,” *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.
- [8] P. Stoica and A. Nehorai, “Performance study of conditional and unconditional direction-of-arrival estimation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.
- [9] B. Ottersten, M. Viberg, and T. Kailath, “Analysis of subspace fitting and ml techniques for parameter estimation from sensor array data,” *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 590–600, Mar. 1992.
- [10] K. Werner and M. Jansson, “Doa estimation and detection in colored noise using additional noise-only data,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5309–5322, Nov. 2007.
- [11] P. Stoica, E. Larsson, and A. Gershman, “The stochastic CRB for array processing: a textbook derivation,” *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.
- [12] W. J. Bangs, “Array Processing with Generalized Beamformers,” Ph.D. dissertation, Yale University, New Haven, CT, 1971.
- [13] P. Wirfält and M. Jansson, “On kronecker and linearly structured covariance matrix estimation,” *IEEE Trans. Signal Process.*, submitted 2013.
- [14] K. Werner, M. Jansson, and P. Stoica, “On estimation of covariance matrices with Kronecker product structure,” *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 478–491, Feb. 2008.
- [15] P. Wirfält, G. Bouleux, M. Jansson, and P. Stoica, “Optimal prior knowledge-based direction of arrival estimation,” *IET Signal Process.*, vol. 6, no. 8, pp. 731–742, Oct. 2012.
- [16] P. Wirfält and M. Jansson, “Prior-exploiting direction-of-arrival algorithms for partially uncorrelated source signals,” *Signal Process.*, submitted 2013.
- [17] ———, “Robust prior-based direction of arrival estimation,” in *IEEE Stat. Signal Process. Workshop*, Aug. 2012, pp. 81–84.

- [18] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. on Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.
- [19] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. on Acoust., Speech and Signal Process.*, vol. 34, no. 5, pp. 1081–1089, Oct. 1986.
- [20] P. Stoica and K. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proc. Part F Radar and Signal Process.*, vol. 137, no. 1, pp. 19–26, Feb. 1990.
- [21] R. DeGroat, E. Dowling, and D. Linebarger, "The constrained MUSIC problem," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1445–1449, Mar. 1993.
- [22] D. Linebarger, R. DeGroat, E. Dowling, P. Stoica, and G. Fudge, "Incorporating a priori information into MUSIC-algorithms and analysis," *Signal Process.*, vol. 46, no. 1, pp. 85–104, 1995.
- [23] P. Stoica, P. Händel, and A. Nehorai, "Improved sequential MUSIC," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 4, pp. 1230–1239, Oct. 1995.
- [24] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior knowledge-based DOA estimation method," in *17th European Signal Process. Conf.*, Glasgow, UK, Aug. 2009, pp. 869–873.
- [25] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.
- [26] P. Wirfält, M. Jansson, G. Bouleux, and P. Stoica, "Prior knowledge-based direction of arrival estimation," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Prague, Czech Republic, May 2011, pp. 2540–2543.
- [27] M. Jansson, B. Göransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945–956, Apr. 1999.
- [28] M. Jansson and B. Ottersten, "Structured covariance matrix estimation: A parametric approach," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 3172–75.
- [29] P. Wirfält, M. Jansson, and G. Bouleux, "Prior-exploiting direction-of-arrival algorithm for partially uncorrelated source signals," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Vancouver, Canada, May 2013, pp. 3972–3976.
- [30] M. Jansson, A. Swindlehurst, and B. Ottersten, "Weighted subspace fitting for general array error models," *IEEE Trans. on Signal Process.*, vol. 46, no. 9, pp. 2484–2498, Sep. 1998.
- [31] D. Zachariah, P. Wirfält, M. Jansson, and S. Chatterjee, "Line spectrum estimation with probabilistic priors," *Signal Process.*, vol. 93, no. 11, pp. 2969–2974, 2013.
- [32] F. Cupertino, E. de Vanna, L. Salvatore, and S. Stasi, "Analysis techniques for detection of IM broken rotor bars after supply disconnection," *IEEE Trans. Ind. Appl.*, vol. 40, no. 2, pp. 526–533, Mar.–Apr. 2004.
- [33] Y. Xia, S. Douglas, and D. Mandic, "Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 44–54, 2012.

- [34] J. Jensen, *Estimation of Blood Velocities Using Ultrasound: A Signal Processing Approach*. Cambridge University Press, 1996.
- [35] S. D. Somasundaram, "Advanced Signal Processing Algorithms Based on Novel Nuclear Quadrupole Resonance Models for the Detection of Explosives," Ph.D. dissertation, King's College London, London, United Kingdom, 2007.
- [36] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, vol. 8, Boston, Massachusetts, USA, Apr. 1983, pp. 336–339.
- [37] A. Eriksson, P. Stoica, and T. Söderström, "Markov-based eigenanalysis method for frequency estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 3, pp. 586–594, Mar. 1994.
- [38] C. Austin, R. Moses, J. Ash, and E. Ertin, "On the relation between sparse reconstruction and parameter estimation with model order selection," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 560–570, June 2010.
- [39] P. Stoica, P. Babu, and J. Li, "Spice: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.
- [40] F. Gran, A. Jakobsson, and J. Jensen, "Adaptive spectral doppler estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 4, pp. 700–714, April 2009.
- [41] R. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. of RADC Spectrum Estimation Workshop*, Griffiss AFB, New York, USA, Oct. 1979, pp. 243–258.
- [42] M. Kristensson, M. Jansson, and B. Ottersten, "Further results and insights on subspace based sinusoidal frequency estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2962–2974, Dec. 2001.
- [43] E. Dowling, R. DeGroat, and D. Linebarger, "Exponential parameter estimation in the presence of known components and noise," *IEEE Trans. Antennas Propag.*, vol. 42, no. 5, pp. 590–599, May 1994.
- [44] J. Jung, J. Lee, and B. Kwon, "Online diagnosis of induction motors using MCSA," *IEEE Trans. Ind. Electron.*, vol. 53, no. 6, pp. 1842–1852, Dec. 2006.
- [45] P. Wirfält, G. Bouleux, M. Jansson, and P. Stoica, "Subspace-based frequency estimation utilizing prior information," in *IEEE Stat. Signal Process. Workshop*, Nice, France, Jun. 2011, pp. 533–536.
- [46] A. Jakobsson, M. Mossberg, M. Rowe, and J. A. S. Smith, "Exploiting Temperature Dependency in the Detection of NQR Signals," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1610–1616, May 2006.
- [47] A. Gregorovič and T. Apih, "TNT detection with 14N NQR: Multipulse sequences and matched filter," *J. Magn. Reson.*, vol. 198, no. 2, pp. 215–221, June 2009.
- [48] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, N.J.: Prentice Hall, 2005.
- [49] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

- [50] E. Gudmundson, P. Wirfalt, A. Jakobsson, and M. Jansson, "An ESPRIT-based parameter estimator for spectroscopic data," in *IEEE Stat. Signal Process. Workshop*, Aug. 2012, pp. 77–80.
- [51] H. Chen, S. V. Huffel, A. van den Boom, and P. van den Bosch, "Extended HTLS methods for parameter estimation of multiple data sets modeled as sums of exponentials," in *13th Int. Conf. Digit. Signal Process.*, vol. 2, Santorini, 1997, pp. 1035–1038.
- [52] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [53] P. Dutilleul, "The mle algorithm for the matrix normal distribution," *J. Stat. Comp. and Simulation*, vol. 64, no. 2, pp. 105–123, 1999.
- [54] A. Dawid, "Some matrix-variate distribution theory: Notational considerations and a bayesian application," *Biometrika*, vol. 68, no. 1, pp. 265–274, Apr. 1981.
- [55] N. Lu and D. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Stat. and Probability Lett.*, vol. 73, no. 4, pp. 449–457, 2005.
- [56] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic mimo radio channel model with experimental validation," *IEEE J. on Select. Areas in Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.
- [57] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.
- [58] J. de Munck, H. Huizenga, L. Waldorp, and R. Heethaar, "Estimating stationary dipoles from meg/eeg data contaminated with spatially and temporally correlated background noise," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1565–1572, Jul. 2002.
- [59] P. Strobach, "Low rank detection of multichannel gaussian signals using a constrained inverse," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, vol. 4, 1994, pp. 245–248.
- [60] N. Lu and D. Zimmerman, "On likelihood-based inference for a separable covariance matrix," Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa, Technical Report 337, 2004.
- [61] G. Allen and R. Tibshirani, "Transposable regularized covariance models with an application to missing data imputation," *Ann. Appl. Stat.*, vol. 4, no. 2, pp. 764–790, 2010.
- [62] T. Tsiligkaridis, A. Hero, and S. Zhou, "On convergence of kronecker graphical lasso algorithms," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1743–1755, Apr. 2013.
- [63] P. Stoica and R. Moses, *Spectral analysis of signals*. Prentice Hall, 2005.
- [64] J. McEwen and G. Anderson, "Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity," *IEEE Trans. Biomed. Eng.*, vol. BME-22, no. 5, pp. 361–369, Sep. 1975.
- [65] A. Molisch, "A generic model for MIMO wireless propagation channels in macro- and microcells," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 61–71, Jan. 2004.

-
- [66] T. Barton and D. Fuhrmann, "Covariance structures for multidimensional data," *Multidimensional Syst. and Signal Process.*, vol. 4, no. 2, pp. 111–123, Apr. 1993.
- [67] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc., 1958.
- [68] M. Jansson and P. Stoica, "Forward-only and forward-backward sample covariances – A comparative study," *Signal Process.*, vol. 77, no. 3, pp. 235–245, Sep. 1999.
- [69] M. Jansson, P. Wirfält, K. Werner, and B. Ottersten, "ML estimation of covariance matrices with Kronecker and persymmetric structure," in *IEEE Digital Signal Process. Workshop and 5th IEEE Signal Process. Educ. Workshop, (DSP/SPE)*, Jan. 2009, pp. 298–301.
- [70] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Signal Process.*, vol. 17, no. 4, pp. 383–387, Aug. 1989.
- [71] T. Söderström and P. Stoica, *System Identification*. Prentice Hall International (UK) Ltd, 1989.
- [72] P. Wirfält and M. Jansson, "On toeplitz and kronecker structured covariance matrix estimation," in *2010 IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, Jerusalem, Israel, Oct. 2010, pp. 185–188.

Part II

Included papers

Paper A

Optimal Prior Knowledge-Based Direction of Arrival Estimation

Petter Wirfält, Guillaume Bouleux, Magnus Jansson, and Petre Stoica

Published in
IET Signal Process., vol. 6, no. 8, pp. 731–742, Oct. 2012.

©2012 The Institution of Engineering and Technology
The layout has been revised

Optimal Prior Knowledge-Based Direction of Arrival Estimation

Petter Wirfält, Guillaume Bouleux, Magnus Jansson and Petre Stoica

Abstract

In certain applications involving direction of arrival (DOA) estimation we may have a priori information on some of the DOAs. This information could refer to a target known to be present at a certain position, or to a reflection. In this paper we investigate a methodology for array processing that exploits the information on the known DOAs for estimating the unknown DOAs as accurately as possible. We present algorithms that can efficiently handle the case of both correlated and uncorrelated sources when the receiver is a uniform linear array. We find a major improvement in estimator accuracy in feasible scenarios, and we compare the estimator performance to the corresponding theoretical stochastic Cramér-Rao Bounds (CRBs) as well as to the performance of other methods capable of exploiting such prior knowledge. In addition, we apply the investigated estimators to real data from an ultra-sound array.

Index Terms—Antenna arrays; array signal processing; direction of arrival estimation; Cramer-Rao bound

1 Introduction

In some direction of arrival (DOA) estimation applications there exists a priori information that is not exploited in the traditional DOA finding algorithms such as [1], [2], and [3]. For example, in a stationary RADAR scenario there can be a reflection that is always received. The DOA associated with this reflection is of no interest to the observer, since it is already known. However its presence might degrade the estimation of the unknown DOAs of interest.

A methodologically related problem is the frequency estimation, for example the diagnosis of rotating machines in industrial environments. The rotational behavior of such machines is well-known and hence some peaks of the related frequency spectrum are known, e.g. the power grid supply frequency [4]. Estimating

this supply frequency provides no information about the equipment; rather, the known spectral peaks might obscure weaker interesting peaks.

The important feature of this prior information is that it is hidden in the received data; regular algorithms will simply treat the known directions or frequencies as part of the set of unknown ones. It would then be advantageous if prior knowledge on some DOAs or frequencies could be used in order to increase the accuracy when estimating the remaining unknown parameters.

For angle estimation this has been done in various ways. The MUSIC method [1], [5], is a so-called subspace method in which the subspaces spanned by the desired signals and the noise, respectively, are separated. One approach, then, is to use the prior information to decrease the dimension of the space spanned by the signals such that this space only contains the unknown signals. This dimension decrease, or deflation, of the signal subspace can be viewed geometrically as an orthogonal projection onto a lower dimensional subspace. This type of approach is suggested in, e.g., [6], [7], and [8]: the space spanned by the array data is projected onto the orthogonal complement of the subspace containing the known DOAs, thus removing the subspace spanned by the known directions from the array manifold. Then, it is shown that the unknown directions are more easily found from the reduced-dimension data set.

A similar approach is to use oblique projections, see [9] for an introduction to this concept. In this case, instead of projecting orthogonally onto a certain subspace, a projection along the subspace that is to be removed is performed. This results in a different solution as compared to projecting orthogonally since the projection is along a different path. The benefit of this approach is that the entire contribution of the concerned subspace is cancelled — in the orthogonal projection case, only the portion perpendicular to the desired subspace is cancelled. Thus, if the two subspaces are not orthogonal, the oblique projection can be superior to the orthogonal one. However, the oblique projector may increase the contribution of noise terms; see [9] for more details on this subject.

Comparisons of these two approaches applied to MUSIC can be found in [10], where some theoretical bounds are studied as well. By studying the Cramér-Rao Bound (CRB) for the case when prior information is available, it can be seen that the previously mentioned projection based methods are not performing optimally in the sense that they are not achieving the theoretically best achievable accuracy. This should not be surprising — in [11] and [12] the asymptotic properties of MUSIC were studied, and it was concluded that MUSIC is not efficient in the case when there is correlation between the source signals. On the other hand in the case of uncorrelated sources, MUSIC is efficient in the sense that it asymptotically achieves the CRB. However, in [13] it was shown that a tighter CRB exists when the sources are known to be uncorrelated; from the perspective of asymptotic accuracy, MUSIC is thus suboptimal in this scenario as well. The conclusion is that we cannot achieve asymptotically efficient estimates with MUSIC whether the sources are correlated or not, if we properly account for the model-imposed limits on accuracy.

Motivated by the above observation, the focus of this paper is on incorporating prior DOA knowledge into methods that are asymptotically statistically efficient. We are studying a scenario with a Uniform Linear Array (ULA) at the receiver, i.e. a sensor array consisting of identical omnidirectional sensors with uniform inter-sensor separation. We investigate two ULA DOA methods ([2], [3]), which are asymptotically efficient in the case of correlated and uncorrelated impinging signals, respectively. Both of these methods find their estimates through polynomial rooting, and this is where the approach [14] denoted PLEDGE (as in Prior knowLEDGE) advocated in this paper can be used: we convert the knowledge of some DOAs into knowledge of the corresponding roots of the polynomial to be factored. Doing so allows us to refine the estimates of the other unknown roots, and hence of the unknown DOAs. An important fact is that we incorporate the known information in a statistically optimal way.

As will be shown, both DOA estimation methods show significant accuracy improvements when modified to utilize known DOAs: the signal-to-noise-ratio (SNR) required to achieve the theoretical performance limit is reduced, and in addition this limit is improved as predicted by the respective CRB. When the sources are uncorrelated, the exploitation of prior DOA information as in [14] does not enhance the method in [2] as much as when the sources are correlated; this motivates the application of PLEDGE to the method in [3], which is usable only when the sources are uncorrelated. Being able to exploit the uncorrelatedness of the signals in conjunction with prior DOA knowledge allows greater accuracy increases. The realizable gain was illustrated in [15] and is further investigated in this paper. To accurately judge the performance of the two PLEDGE-based methods we compare them to other methods capable of exploiting prior DOA knowledge, [7] and [10].

The paper is structured as follows. In Section 2 we define the scenario and in Section 3 we review the two DOA finding algorithms [2] and [3]. In Section 4, we discuss different methods of incorporating prior DOA-knowledge: the approach introduced in [14] and advocated in this paper, and two other methods ([7] and [10]) for comparison. In Section 5 we derive the CRBs for the studied scenarios, and in Section 6 we perform numerical simulations to evaluate the potential of the proposed methodology. Finally, we apply the estimators to real data in Section 7.

The PLEDGE-concept was previously introduced in [14], and it was examined for uncorrelated source signals in [15]; the contribution of this paper is a more thorough examination of the potentially significant accuracy gains that result from using prior information, comparing the algorithms to existing methods, and applying them to real data. We also derive the corresponding accuracy bounds in a more compact manner.

2 Data model

Consider d narrow-band, plane waves sensed by an m sensor ULA. We assume that d_k of the impinging waves originate from a priori known directions; hence, define

the angles of arrival as $\bar{\boldsymbol{\theta}} \triangleq [\boldsymbol{\theta}^T \ \boldsymbol{\vartheta}^T]^T \triangleq [\theta_1 \ \dots \ \theta_{d_u} \ \vartheta_1 \ \dots \ \vartheta_{d_k}]^T$, where θ_i denotes the unknown direction i and ϑ_j denotes the known direction j ; henceforth, the subscripts u and k will denote unknown and known quantities, respectively. The angular measurements are referenced to 0° at array broadside, T is the transposition operation, and $d_u = d - d_k$.

Assume that N array response snapshots are recorded, with $t = 1, \dots, N$, according to the data model

$$\mathbf{y}(t) = \mathbf{A}(\bar{\boldsymbol{\theta}})\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{A}(\bar{\boldsymbol{\theta}}) = [\mathbf{a}(\bar{\theta}_1), \dots, \mathbf{a}(\bar{\theta}_d)] \in \mathbb{C}^{m \times d}$ is the array response matrix and $\mathbf{a}(\bar{\theta}_i) \in \mathbb{C}^{m \times 1}$ defined as

$$\mathbf{a}(\bar{\theta}_i) = [1 \quad e^{j\phi_i} \quad \dots \quad e^{j(m-1)\phi_i}]^T \quad (2)$$

is the steering vector corresponding to the i -th angle, viz.

$$\phi_i = -2\pi\Delta \sin(\bar{\theta}_i), \quad (3)$$

with Δ being the ULA inter-sensor spacing in wavelengths. Further, in (1), $\mathbf{s}(t) \in \mathbb{C}^{d \times 1}$ and $\mathbf{n}(t) \in \mathbb{C}^{m \times 1}$ are the source signal and additive noise vectors at time t , respectively, and they are assumed to be mutually independent zero-mean circular Gaussian random sequences with second-order moments given by

$$\mathbb{E}[\mathbf{s}(t)\mathbf{s}^*(\tau)] = \mathbf{P}\delta(t, \tau) \quad (4)$$

$$\mathbb{E}[\mathbf{n}(t)\mathbf{n}^*(\tau)] = \sigma^2\mathbf{I}\delta(t, \tau). \quad (5)$$

Here, $*$ denotes conjugate transposition and $\delta(t, \tau) = 1$ for $t = \tau$, and 0 otherwise. Without loss of generality, we can additionally partition $\mathbf{A}(\bar{\boldsymbol{\theta}}) \triangleq [\mathbf{A}_u(\boldsymbol{\theta}) \ \mathbf{A}_k(\boldsymbol{\vartheta})]$ with the subscripts added for clarity. Further, let

$$\mathbf{R} = \mathbb{E}[\mathbf{y}(t)\mathbf{y}^*(t)] \quad (6)$$

be the data covariance matrix, and define

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}(t)\mathbf{y}^*(t) \quad (7)$$

as a sample version thereof. The superscript $\hat{\cdot}$ denotes a quantity estimated from data.

3 DOA estimation

This section describes the two algorithms to which we apply the PLEDGE framework [14]. None of these methods can exploit the prior information $\boldsymbol{\vartheta}$; hence these

algorithms will estimate all the elements of the vector $\bar{\theta}$. Both algorithms are Maximum Likelihood (ML) based methods that minimize a subspace projection, and they effectively reduce (in the ULA case) the DOA finding problem to polynomial rooting. These two methods should be used when the signal correlation is unknown [2] and, respectively, when the sources are known to be uncorrelated [3]. In these cases the methods can be shown to be asymptotically statistically efficient (see [2], [3] and references therein).

Let

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\mathbf{\Lambda}}_n \hat{\mathbf{E}}_n^* \quad (8)$$

be the eigendecomposition of (7), where $\hat{\mathbf{\Lambda}}_s$ is a diagonal matrix comprising the d' largest eigenvalues of $\hat{\mathbf{R}}$, with d' being the rank of \mathbf{P} , and $\hat{\mathbf{E}}_s$ comprises the associated eigenvectors. Similarly, the $m - d'$ smallest eigenvalues are collected in the diagonal matrix $\hat{\mathbf{\Lambda}}_n$ and $\hat{\mathbf{E}}_n$ contains the associated eigenvectors.

3.1 General signal correlation

Here there is no assumed structure for \mathbf{P} . To review the method of [2], in [16] the authors define a polynomial, based on ϕ_1, \dots, ϕ_d that correspond to the true DOAs, by

$$b_0 z^d + b_1 z^{d-1} + \dots + b_d = b_0 \prod_{i=1}^d (z - e^{j\phi_i}). \quad (9)$$

All the roots of (9) are by definition on the unit-circle, and thus the coefficients can be chosen to be conjugate symmetric: $b_i = b_{d-i}^*$, $i = 0, 1, \dots, d$ [16]. While the conjugate symmetry does not guarantee unit-circle roots, when it comes to estimating the polynomial coefficients from data it is much simpler to enforce this constraint than actually constraining the roots to the unit circle.

Using the polynomial coefficients, define

$$\mathbf{B}^* = \begin{bmatrix} b_d & \dots & b_1 & b_0 & 0 & \dots & 0 \\ 0 & b_d & \dots & b_1 & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b_d & \dots & b_1 & b_0 \end{bmatrix} \quad (10)$$

with $\mathbf{B}^* \in \mathbb{C}^{(m-d) \times m}$. The MODE (Method of Direction Estimation) estimate is found by minimizing the function [2]

$$V_{\text{MODE}}(\mathbf{b}) = \text{Tr}[\mathbf{B} (\hat{\mathbf{B}}^* \hat{\mathbf{B}})^{-1} \mathbf{B}^* \hat{\mathbf{E}}_s \hat{\mathbf{\Lambda}} \hat{\mathbf{E}}_s^*], \quad (11)$$

where $\hat{\mathbf{\Lambda}} = \hat{\mathbf{\Lambda}}_s^{-1} (\hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I})^2$, $\hat{\sigma}^2 = \frac{1}{m-d'} \text{Tr}[\hat{\mathbf{\Lambda}}_n]$ is a consistent estimate of the noise power, and $\hat{\mathbf{B}}$ is constructed from an estimate of $\mathbf{b} = [b_0 \ b_1 \ \dots \ b_d]^T$ — typically, the algorithm is initialized with $\hat{\mathbf{B}}^* \hat{\mathbf{B}} = \mathbf{I}$, giving a coarse $\hat{\mathbf{b}}$, and

then it is iterated one more time. Due to the special structure of \mathbf{B} it can be seen that (11) is a quadratic function of the polynomial coefficients \mathbf{b} . Thus we can equivalently write (11) as

$$V_{\text{MODE}}(\mathbf{b}) = \|\mathbf{H}\mathbf{b}\|^2, \quad (12)$$

where \mathbf{H} is estimated from the data and a function of $\hat{\mathbf{b}}$. The exact structure of \mathbf{H} is somewhat involved and not in the primary scope of this article, but can be readily obtained from [16] along with the modifications given by [2].

Minimizing (12) with respect to \mathbf{b} subject to a non-triviality constraint (i.e. $\mathbf{b} \neq \mathbf{0}$) and the conjugate symmetry constraint previously mentioned (which is a simple least squares or eigen-decomposition problem) provides an estimate of the polynomial coefficients \mathbf{b} . Then, the angles of the roots of the corresponding polynomial (9) gives estimates of (3) from which the DOAs $\bar{\boldsymbol{\theta}}$ are estimated.

3.2 Uncorrelated sources

If we have the a priori information that the sources are uncorrelated we can devise a more accurate estimator. In what follows, we give a brief summary of a statistically efficient algorithm originally developed in [3]. This algorithm, which we will call ‘‘DOA UC’’ (where UC stands for uncorrelated), is based on minimizing the covariance-matching criterion

$$V_{\text{UC}}(\boldsymbol{\eta}) = (\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\eta}))^* \mathbf{W} (\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\eta})). \quad (13)$$

In (13) $\mathbf{r} = \text{vec}(\mathbf{R})$, with $\text{vec}(\cdot)$ stacking the columns of a matrix on top of each other, and $\hat{\mathbf{r}}$ is the sample estimate thereof (7). Also $\boldsymbol{\eta} = [\bar{\boldsymbol{\theta}}^T \quad \mathbf{p}^T \quad \sigma^2]^T$ parameterizes \mathbf{r} , with \mathbf{p} denoting the diagonal elements of (4), and \mathbf{W} is a suitably chosen weighting matrix. Following [3] and [17], $\mathbf{W} = \hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}^{-1}$ minimizes the asymptotic variance of the (asymptotically unbiased) estimate given by minimizing (13) (here, \otimes denotes the Kronecker product).

In [3] it is shown how to successively solve (13) for the parameters \mathbf{p} and σ^2 , producing closed form estimates $\hat{\mathbf{p}}$ and $\hat{\sigma}^2$. Introducing these estimates in (13), and using the same polynomial parametrization \mathbf{b} of the DOAs as in (9), (10), it can be shown that (13) can be written as

$$V_{\text{UC}}(\mathbf{b}) = \|\tilde{\mathbf{H}}\mathbf{b}\|^2, \quad (14)$$

where $\tilde{\mathbf{H}}$ is given in [3]. Similarly to \mathbf{H} in (12) $\tilde{\mathbf{H}}$ is a function of the data and the DOA and (14) can obviously — using some initialization for the DOA-dependent terms in $\tilde{\mathbf{H}}$ — be minimized, in the same manner as (12).

4 Using prior knowledge

4.1 PLEDGE

As already mentioned the approach of incorporating prior DOA information into the estimation is referred to as PLEDGE (Prior knowLEDGE) [14]. This approach utilizes the fact that the minimization variables of the criterion functions in (12) and (14) are polynomial coefficients. Knowing some of the DOAs is equivalent to knowing some of the roots of the polynomial in (9); we can then partition this polynomial in two factors: one fully known factor corresponding to the known roots, and one factor that is unknown but has fewer unknown terms than the original polynomial. Note that no information is lost in this way: we are only constraining the polynomial to have some of its roots at certain locations, defined by the known DOAs. The estimates of the remaining unknown roots will then presumably be more accurate.

Formulating this mathematically we re-write (9) as

$$b_0 \prod_{i=1}^d (z - e^{j\phi_i}) = P_k(z)P_u(z), \quad (15)$$

in which

$$P_u(z) = \tilde{b}_0 \prod_{i=1}^{d_u} (z - e^{j\phi_i}) = \tilde{b}_0 z^{d_u} + \dots + \tilde{b}_{d_u} \quad (16)$$

and

$$P_k(z) = c_0 \prod_{i=d_u+1}^d (z - e^{j\phi_i}) = c_0 z^{d_k} + \dots + c_{d_k}. \quad (17)$$

Here, $P_k(z)$ is the polynomial with d_k zeros corresponding to the known DOAs ϑ whereas $P_u(z)$ has $d_u = d - d_k$ zeros corresponding to the unknown DOAs θ . Note that the coefficients $\mathbf{c} = [c_0 \dots c_{d_k}]^T$ are given by ϑ via (3); thus the prior DOA knowledge defines \mathbf{c} . If (15) is expanded, it is obvious that the polynomial coefficients \mathbf{b} can be written as a convolution of $\tilde{\mathbf{b}} = [\tilde{b}_0 \tilde{b}_1 \dots \tilde{b}_{d_u}]^T$ and \mathbf{c} . Thus we can rewrite (15)-(17) in matrix form as

$$\mathbf{b} = \mathbf{C}\tilde{\mathbf{b}}, \quad (18)$$

if we introduce the (Toeplitz structured) convolution matrix

$$\mathbf{C}^T = \begin{bmatrix} c_0 & c_1 & \dots & c_{d_k} & 0 & \dots & 0 \\ 0 & c_0 & c_1 & \dots & c_{d_k} & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \dots & 0 & c_0 & c_1 & \dots & c_{d_k} \end{bmatrix} \in \mathbb{C}^{(d_u+1) \times (d+1)}, \quad (19)$$

the elements of which are given by \mathbf{c} .

4.2 PLEDGE: general case

Substituting (18) in (12), we obtain the PLEDGE criterion function

$$V_P(\tilde{\mathbf{b}}) = \|\mathbf{H}\mathbf{C}\tilde{\mathbf{b}}\|^2 \quad (20)$$

(hereafter we let the subscript P stand for ‘‘PLEDGE’’).

The method of finding the unknown DOAs consists of the following steps:

1. Form the \mathbf{C} matrix from the known DOAs.
2. Perform an eigendecomposition of $\hat{\mathbf{R}}$ and, using $(\hat{\mathbf{B}}^*\hat{\mathbf{B}})^{-1} = \mathbf{I}$ as an initial estimate, form \mathbf{H} as in [16] and [2].
3. Find an estimate of $\tilde{\mathbf{b}}$ by minimizing (20) subject to $\|\mathbf{b}\| = 1$ and the conjugate symmetry constraint. Use this estimate in (18) to form $(\hat{\mathbf{B}}^*\hat{\mathbf{B}})^{-1}$, and update \mathbf{H} .
4. Minimize (20) once again subject to the same constraints to find $\tilde{\mathbf{b}}$, then find $\phi_i, i = 1, \dots, d_u$, by rooting (16), and finally the sought DOAs through (3).

4.3 PLEDGE: uncorrelated sources

Substituting (18) in (14), we obtain the following PLEDGE criterion function

$$V_{PUC} = \|\tilde{\mathbf{H}}\mathbf{C}\tilde{\mathbf{b}}\|^2 \quad (21)$$

(the subscript PUC denotes ‘‘PLEDGE uncorrelated’’). This criterion is minimized in the same way as for (20), that is:

1. Form \mathbf{C} in (19) from the known DOAs
2. Form an initial DOA-independent $\tilde{\mathbf{H}}$ based on the sample data according to [3].
3. Find an initial estimate of $\tilde{\mathbf{b}}$ by minimizing (21) subject to $\|\mathbf{b}\| = 1$ and the conjugate symmetry constraint; use this initial DOA estimate to find a more accurate $\tilde{\mathbf{H}}$.
4. Minimize (21), once again, subject to the same constraints to find $\tilde{\mathbf{b}}$, then obtain $\phi_i, i = 1, \dots, d_u$, by rooting (16), and finally the sought DOAs through (3).

4.4 Other methods

For comparison purposes, we also consider two prior-based methods that are not using the PLEDGE concept. These are C-MUSIC (as in Constrained MUSIC, see [7]), which is a method based on an orthogonal projection. We also look at a method based on an oblique projection; the method was introduced in [10] as P-MUSIC (short for Prior MUSIC), which was modified in [14] to remove an observed bias. We denote this method MP-MUSIC (Modified P-MUSIC). These MUSIC-based algorithms find the respective DOAs by a rooting approach, along the lines of root-MUSIC [18].

C-MUSIC

This algorithm is described in full in [7]. The main idea is to use an orthogonal projection that removes the known part of the signal subspace; this is done by pre- and post-multiplying the sample covariance matrix with a projection matrix. Accordingly, we apply a regular MUSIC-algorithm to

$$\hat{\mathbf{R}}_c = \mathbf{\Pi}_{\mathbf{A}_k}^\perp \hat{\mathbf{R}} \mathbf{\Pi}_{\mathbf{A}_k}^\perp \quad (22)$$

where $\mathbf{\Pi}_{\mathbf{X}}^\perp = \mathbf{I} - \mathbf{\Pi}_{\mathbf{X}}$ denotes the orthogonal projection onto the null-space of \mathbf{X}^* , $\mathbf{\Pi}_{\mathbf{X}} = \mathbf{X}\mathbf{X}^\dagger = \mathbf{X}(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*$ is the orthogonal projection onto the range space of \mathbf{X} , and \mathbf{A}_k is composed of the steering vectors corresponding to the known DOAs. The contribution of the known sources to $\hat{\mathbf{R}}$ has thus been cancelled, whereas the eigenvectors of $\hat{\mathbf{R}}_c$ still contain information on the unknown sources; see [7] for further details.

MP-MUSIC

In [10] a criterion function is proposed based on an oblique projection; it is then shown that this criterion condenses to a minimization only involving an orthogonal projection. It can be inferred from [10] that the resulting algorithm suffers from a bias, rendering it worse than other methods, e.g., C-MUSIC. Thus a corrective measure was suggested in [14], and here we denote the resulting algorithm MP-MUSIC; we below briefly investigate the consequences of this correction, which takes the form of removing the noise from the sample covariance matrix. Thus let

$$\mathbf{R} - \sigma^2\mathbf{I} = \mathbf{A}\mathbf{P}\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_u & \mathbf{A}_k \end{bmatrix} \begin{bmatrix} \mathbf{P}_u & \mathbf{P}_{uk} \\ \mathbf{P}_{uk}^* & \mathbf{P}_k \end{bmatrix} \begin{bmatrix} \mathbf{A}_u^* \\ \mathbf{A}_k^* \end{bmatrix} \quad (23)$$

in which \mathbf{A}_u contains the steering vectors corresponding to the unknown sources, \mathbf{P}_u and \mathbf{P}_k are the square signal covariance matrices related to the unknown and known emitters, respectively, and \mathbf{P}_{uk} is the cross-correlation matrix of the signals for the unknown and known emitters.

The obliquely projected criterion in [10] condensed to pre-multiplying the (now modified) sample covariance matrix with the same orthogonal projector as used in Section 4.4; hence, look at

$$\Pi_{\mathbf{A}_k}^\perp (\mathbf{R} - \sigma^2 \mathbf{I}) = \Pi_{\mathbf{A}_k}^\perp \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* + \Pi_{\mathbf{A}_k}^\perp \mathbf{A}_u \mathbf{P}_{uk} \mathbf{A}_k^*. \quad (24)$$

If in addition the known and unknown signals are uncorrelated to one another, i.e. $\mathbf{P}_{uk} = \mathbf{0}$, and we use estimated quantities in (24) then:

$$\Pi_{\mathbf{A}_k}^\perp (\hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{I}) = \Pi_{\mathbf{A}_k}^\perp \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* + \mathbf{E}_\epsilon \quad (25)$$

where \mathbf{E}_ϵ accounts for the error introduced by using estimated quantities. Using the right singular-vectors of (25) in a standard MUSIC algorithm then gives estimates of the unknown DOAs.

5 PLEDGE CRB

We now derive the CRB based on the prior-knowledge of some of the DOAs. We derive it both for the general case (which has been previously considered in [14]) and for the case when the sources are known to be uncorrelated.

The derivation starts out by following [19]. Let

$$\boldsymbol{\alpha} = [\boldsymbol{\theta}^T \quad \boldsymbol{\varrho}^T \quad \sigma^2]^T \quad (26)$$

be the vector of unknown parameters in the model. Here, $\boldsymbol{\theta}$ is again the vector of unknown DOAs, $\boldsymbol{\varrho}$ is the vector made from $\{\mathbf{P}_{ii}\}$ and $\{\text{Re}(\mathbf{P}_{ij}), \text{Im}(\mathbf{P}_{ij}); i > j\}$, and σ^2 is the noise variance. Under the assumptions in Section 2, the (k, l) th element of the Fisher Information Matrix for the parameter vector $\boldsymbol{\alpha}$ is given by

$$\text{FIM}_{k,l} = N \text{Tr} \left(\frac{\partial \mathbf{R}}{\partial \alpha_k} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \alpha_l} \mathbf{R}^{-1} \right). \quad (27)$$

This can be written in matrix form as

$$\frac{1}{N} \text{FIM} = \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\alpha}^T} \right)^* (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}) \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\alpha}^T} \right), \quad (28)$$

where

$$\mathbf{r} = \text{vec}(\mathbf{R}) = (\mathbf{A}^c \otimes \mathbf{A}) \text{vec}(\mathbf{P}) + \sigma^2 \text{vec}(\mathbf{I}), \quad (29)$$

and where the superscript c denotes complex conjugation. Introduce \mathbf{G} , \mathbf{V} , and \mathbf{u} via

$$\begin{aligned} \mathbf{W}^{1/2} \left[\begin{array}{c|c|c} \frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^T} & \frac{\partial \mathbf{r}}{\partial \boldsymbol{\varrho}^T} & \frac{\partial \mathbf{r}}{\partial \sigma^2} \end{array} \right] \\ \triangleq [\mathbf{G} \mid \mathbf{V} \mid \mathbf{u}], \end{aligned} \quad (30)$$

where $\mathbf{W}^{1/2} = \mathbf{R}^{-T/2} \otimes \mathbf{R}^{-1/2}$, and also let

$$\begin{bmatrix} \mathbf{V} & | & \mathbf{u} \end{bmatrix} \triangleq \Delta. \quad (31)$$

Then we can write (28) as

$$\frac{1}{N} \text{FIM} = \begin{bmatrix} \mathbf{G}^* \\ \Delta^* \end{bmatrix} \begin{bmatrix} \mathbf{G} & \Delta \end{bmatrix}. \quad (32)$$

Since we are interested in a bound on the angle estimates, we want an expression for the top-left block of the inverse of the partitioned matrix in (32). Using a standard result on partitioned matrix inversion gives

$$\begin{aligned} \frac{1}{N} \text{CRB}_P^{-1}(\boldsymbol{\theta}) &= \mathbf{G}^* \mathbf{G} - \mathbf{G}^* \Delta (\Delta^* \Delta)^{-1} \Delta^* \mathbf{G} \\ &= \mathbf{G}^* \Pi_{\Delta}^{\perp} \mathbf{G}. \end{aligned} \quad (33)$$

5.1 Correlated sources

Since only some of the DOAs are considered unknown, the \mathbf{G} in the present derivation differs from the one in [19]. This difference propagates to the final result and managing it is in essence the contribution of this section.

From (33), following [19], it can be shown that

$$\Pi_{\Delta}^{\perp} = \Pi_{\mathbf{V}}^{\perp} - \frac{\Pi_{\mathbf{V}}^{\perp} \mathbf{u} \mathbf{u}^* \Pi_{\mathbf{V}}^{\perp}}{\mathbf{u}^* \Pi_{\mathbf{V}}^{\perp} \mathbf{u}}. \quad (34)$$

By evaluating the derivatives in (30), and through some further algebra, one finds

$$\mathbf{u}^* \Pi_{\mathbf{V}}^{\perp} \mathbf{g}_k = \mathbf{0}, \quad (35)$$

where \mathbf{g}_k is the k th column of \mathbf{G} . This result allows us to rewrite the individual elements of (33) as

$$\frac{1}{N} [\text{CRB}_P^{-1}(\boldsymbol{\theta})]_{k,l} = \mathbf{g}_k^* \Pi_{\mathbf{V}}^{\perp} \mathbf{g}_l. \quad (36)$$

By further calculations we arrive at

$$\begin{aligned} &\frac{1}{N} [\text{CRB}_P^{-1}(\boldsymbol{\theta})]_{k,l} \\ &= \frac{2}{\sigma^2} \text{Re} \left((\mathbf{D}_u^* \Pi_{\mathbf{A}}^{\perp} \mathbf{D}_u)_{k,l} (\bar{\mathbf{P}}_u^* \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \bar{\mathbf{P}}_u)_{l,k} \right), \end{aligned} \quad (37)$$

where

$$\mathbf{D}_u = \begin{bmatrix} \frac{\partial \mathbf{a}(\theta_1)}{\partial \theta_1} & \dots & \frac{\partial \mathbf{a}(\theta_{d_u})}{\partial \theta_{d_u}} \end{bmatrix}, \quad (38)$$

and $\bar{\mathbf{P}}_u$ is the $d \times d_u$ sub-matrix of (4) associated with the unknown sources.

Finally, we can write (37) in matrix form as

$$\begin{aligned} & \text{CRB}_{\mathbf{P}}(\boldsymbol{\theta}) \\ &= \frac{\sigma^2}{2N} \left[\text{Re} \left((\mathbf{D}_u^* \boldsymbol{\Pi}_{\mathbf{A}}^\perp \mathbf{D}_u) \odot (\bar{\mathbf{P}}_u^* \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \bar{\mathbf{P}}_u)^T \right) \right]^{-1} \end{aligned} \quad (39)$$

where \odot is the Schur (element-wise) product.

5.2 Uncorrelated sources

Assuming the sources to be uncorrelated leads to a diagonal \mathbf{P} , which reduces the number of unknown parameters. We then have that $\boldsymbol{\varrho}$ in (26) is equal to the diagonal entries \mathbf{p} of (4), and the vector of unknown parameters becomes

$$\boldsymbol{\alpha}_{\text{UC}} = [\boldsymbol{\theta}^T \quad \mathbf{p}^T \quad \sigma^2]^T. \quad (40)$$

In this case the derivatives of (30) corresponding to $\boldsymbol{\alpha}_{\text{UC}}$ are given by:

$$\begin{aligned} \mathbf{G}_{\text{UC}} &= \mathbf{W}^{1/2} \frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^T} \\ &= \mathbf{W}^{1/2} [(\mathbf{D}_u^c \circ \mathbf{A}_u) + (\mathbf{A}_u^c \circ \mathbf{D}_u)] \mathbf{P}_u \\ &\triangleq \mathbf{W}^{1/2} \mathbf{D}_{\mathbf{P}}, \end{aligned} \quad (41)$$

$$\begin{aligned} \boldsymbol{\Delta}_{\text{UC}} &= \mathbf{W}^{1/2} \left[\begin{array}{c|c} \frac{\partial \mathbf{r}}{\partial \mathbf{p}^T} & \frac{\partial \mathbf{r}}{\partial \sigma^2} \end{array} \right] \\ &= \mathbf{W}^{1/2} [\mathbf{A}^c \circ \mathbf{A} \quad | \quad \text{vec}(\mathbf{I})] \\ &\triangleq \mathbf{W}^{1/2} \mathbf{D}_{\mathbf{n}}, \end{aligned} \quad (42)$$

where \circ is the Khatri-Rao product (column-wise Kronecker product), and \mathbf{P}_u is the square matrix with the unknown source powers on the diagonal. Using these results in (33) gives

$$\text{CRB}_{\text{PUC}}(\boldsymbol{\theta}) = \frac{1}{N} \left[\mathbf{D}_{\mathbf{P}}^* \mathbf{W}^{1/2} \boldsymbol{\Pi}_{\mathbf{W}^{1/2} \mathbf{D}_{\mathbf{n}}}^\perp \mathbf{W}^{1/2} \mathbf{D}_{\mathbf{P}} \right]^{-1}. \quad (43)$$

5.3 Comparison to no-prior bounds

Due to the model structures being nested, we immediately have, for sources of arbitrary correlation, that

$$\text{CRB}_{\text{SML}} \geq \text{CRB}_{\mathbf{P}}, \quad (44)$$

where the notation \geq implies that the difference $\text{CRB}_{\text{SML}} - \text{CRB}_{\mathbf{P}}$ is positive semi-definite. Here, CRB_{SML} denotes the bound when there is no prior information on the DOAs [19].

Similarly, if the sources are known to be uncorrelated, the relation

$$\text{CRB}_{\text{SML}} \geq \text{CRB}_{\text{UC}} \geq \text{CRB}_{\text{PUC}} \quad (45)$$

is true, where CRB_{UC} denotes the bound without prior DOA knowledge, but when the sources are known to be uncorrelated [13]. These relations follow from the general theory behind the CRB by noting the following. The unknown parameters for the tighter bounds represent a strict subset of the parameter vector for the wider bound; each successive reduction of the parameter set must produce a bound which is smaller than or equal to a bound obtained for the entire parameter vector. A more quantitative comparison of the different CRBs is quite involved and beyond the scope of this paper.

6 Simulated data examples

We perform Monte Carlo simulations to evaluate the performance of the studied estimators and compare the results to the standard MODE algorithm [2] and the DOA UC [3]. As previously stated, the latter methods are asymptotically optimal in the case of no prior DOA knowledge. For reference, we also include the prior-based methods mentioned in Section 4.4, from [10] and [7].

In each realization, N simulated snapshots of the data model (1) are used to form the sample covariance matrix $\hat{\mathbf{R}}$ in (7). For each snapshot, $\mathbf{s}(t)$ is a realization of a zero-mean temporally-white complex-Gaussian noise process with spatial covariance matrix \mathbf{P} . Likewise, $\mathbf{n}(t)$ is a realization of a similarly defined process but with spatial covariance matrix $\sigma^2\mathbf{I}$.

In the simulated scenarios, we vary the parameters of the model (1); we vary the number of sources d and their locations $\boldsymbol{\theta}$. We also choose \mathbf{P} as to study different properties of the estimators, e.g. SNR and correlation dependence. We restrict ourselves to studying DOAs that are closely spaced since this is the more difficult estimation scenario. In the case of widely spaced DOAs, traditional methods already produce accurate estimates; indeed, as evidenced in e.g. Fig. 3a, when the DOA separation grows large the prior information does not significantly enhance the accuracy of the estimators.

The estimated DOAs $\hat{\boldsymbol{\theta}}$ are compared to the true values $\boldsymbol{\theta}$, and for each Monte Carlo simulation the squared error is recorded. The performance measure in our simulations is the root-mean-square error, which for the unknown angle θ_i is computed as

$$\text{RMSE}_i = \sqrt{\frac{1}{L} \sum_{k=1}^L (\hat{\theta}_{i,k} - \theta_i)^2}, \quad i = 1, \dots, d_u, \quad (46)$$

where $\hat{\theta}_{i,k}$ denotes the k th estimate of the i th angle, and $L = 1000$ is the number of Monte Carlo realizations.

Since some of the algorithms are not designed to exploit any prior information, they will inherently attempt to estimate all d DOAs θ , including the known ones. Thus, when calculating the RMSEs of $\hat{\theta}$ for those algorithms, the d_k estimates closest to ϑ are associated with these known angles, and the remaining d_u angles are compared to the unknown angles θ .

Unless mentioned otherwise, we let all source signals have the same power (i.e., the diagonal elements of \mathbf{P} are the same); this is to simplify the scenarios such that the amount of parameters that are changed are kept limited. In most of the simulations below, the source SNR is given by the abscissa.

6.1 Uncorrelated sources

In the first simulation we investigate a scenario with closely spaced, uncorrelated sources located at $\bar{\theta} = [10^\circ \ 15^\circ \ 12^\circ \ 25^\circ]^\text{T} \triangleq [\theta_1 \ \theta_2 \ \vartheta_1 \ \vartheta_2]^\text{T}$; $\vartheta = [12^\circ \ 25^\circ]^\text{T}$ is assumed to be known a priori. The number of sensors is $m = 10$, and the number of snapshots $N = 1000$.

We compare the accuracy of all the estimators presented in the article: “MODE” from Section 3.1 and [2]; “MODE PLEDGE” as in Section 4.2 and [14]; “C-MUSIC” from Section 4.4 and [7]; “DOA UC” from Section 3.2 and [3]; “PLEDGE UC” as described in Section 4.3 (see also [15]); and “MP-MUSIC” as outlined in Section 4.4, [10], and [14]. In addition to the above methods, we also show the bounds corresponding to different amounts of prior knowledge: “CRB_{SML}” [19] corresponding to no prior knowledge; “CRB_P” which assumes some source DOAs to be known, see Section 5.1 and [14]; “CRB_{UC}” corresponding to the knowledge of sources being uncorrelated from [13]; and finally “CRB_{PUC}” which assumes knowledge of all sources being uncorrelated and some DOAs known, given by (43) in Section 5.2.

For uncorrelated sources, we have

$$\mathbf{P} = \text{diag}(p_1, \dots, p_d), \quad (47)$$

where $\text{diag}(\cdot)$ denotes the diagonal matrix with the arguments on the diagonal. For simplicity, $p_i = p_j$, $i, j = 1, \dots, d$, corresponding to all sources being equipowered; we also define $\text{SNR}_i = \frac{p_i}{\sigma^2}$.

The benefits of using all the available prior information are clearly illustrated in Fig. 1a — for a given accuracy, the required SNR is reduced by roughly 15 dB when comparing PLEDGE UC to DOA UC, and by 20 dB when comparing to MODE.

Note from the same figure that MP-MUSIC gives poor estimates; in particular, the RMSEs of its estimates are not decreasing with increasing SNR. The method exhibits such a saturation behavior in all scenarios. This may be due to the fact that the method relies heavily on the assumption that $\mathbf{P}_{uk} = 0$ (see (24)-(25)). This may well hold in theory but not in finite samples and in such cases the right singular vectors of (25) are perturbed by contributions from \mathbf{A}_k (cf. the last term in

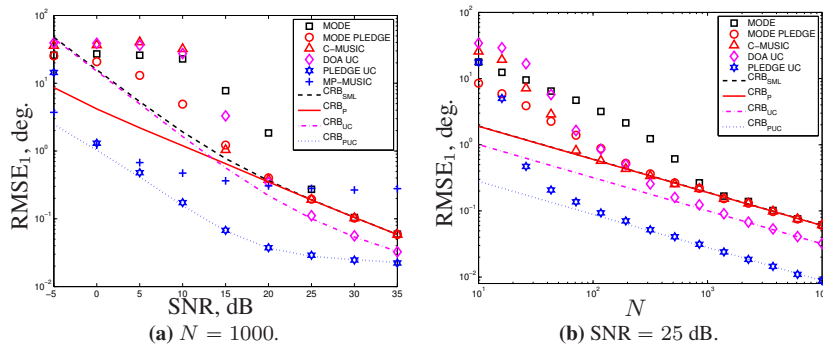


Figure 1: Uncorrelated sources: The plots show the RMSE (of $\hat{\theta}_1$) for the methods presented in the article, along with the CRBs for the different scenarios. $\theta = [10^\circ \ 15^\circ]^T$, $\vartheta = [12^\circ \ 25^\circ]^T$. Equipowered sources, $m = 10$, and average of 1000 independent Monte Carlo realizations.

(24)). In other scenarios (not shown), the threshold at which the RMSE saturates is quite low, and the method can then be quite accurate. However, due to the behavior shown in Fig. 1a we will not consider this method any further.

MODE PLEDGE, which does not exploit the prior information about the correlation state, exhibits a slightly lower resolution threshold than MODE. C-MUSIC is somewhat less accurate than MODE PLEDGE.

The method DOA UC, which uses the prior information that the sources are uncorrelated, behaves very similarly to MODE, except it reaches the uncorrelated, lower, CRB_{UC}; we thus see the benefits of utilizing the correlation state information in the estimator.

At high SNRs, the benefit of using the prior knowledge diminishes as is visible in the figure; when the SNR grows very large (not shown) all bounds converge to the same variance, and hence the estimators are equally accurate. Note however that, e.g., CRB_{PUC} always has a negative slope — a higher unknown-source SNR is always beneficial.

In Fig. 1b we repeat the simulation, but this time fix the SNR (for all sources) at 25 dB and vary the number of snapshots, N . PLEDGE UC is markedly superior to the other methods and can, as compared to MODE, decrease the number of samples required for a specific accuracy by roughly a factor of 50. Note that the simulated scenario is difficult for MODE: it requires about $N = 5000$ samples in order to be efficient. In addition, we see that C-MUSIC reaches the performance bound for slightly lower values of N than MODE PLEDGE.

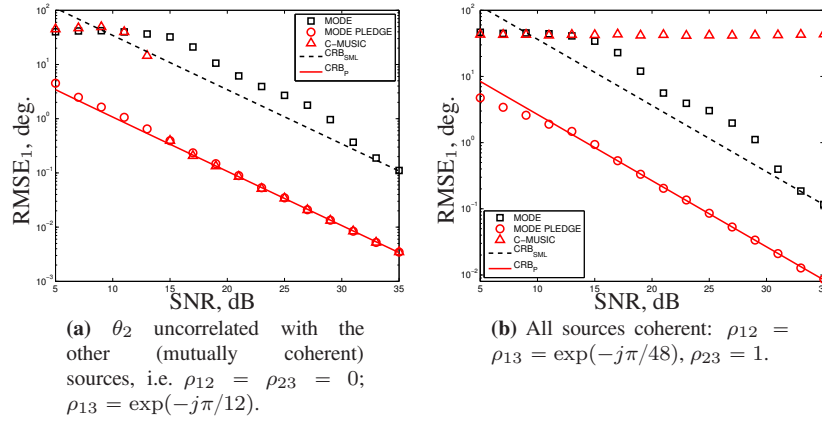


Figure 2: Unknown source correlation: sources located at $\bar{\theta} = [10^\circ \ 15^\circ \ 12^\circ]^\top \triangleq [\theta_1 \ \theta_2 \ \vartheta]^\top$. RMSE of $\hat{\theta}_1$. Number of snapshots $N = 1000$, $m = 6$, equipowered sources, 1000 Monte Carlo realizations. Non-zero correlation coefficients are chosen to maximize the difficulty of the estimation scenario. [Note that there is an error in the SNR-values of this figure; the actual SNR (in dB) is twice the displayed value.]

6.2 Unknown source correlation

In the second case we explore a situation with coherence between the sources; accordingly, the spatial covariance matrix of the emitted signals is

$$\mathbf{P} = \begin{bmatrix} p_1 & p_{12} & p_{13} \\ p_{12}^c & p_2 & p_{23} \\ p_{13}^c & p_{23}^c & p_3 \end{bmatrix} \quad (48)$$

where $p_{ij} = \rho_{ij}\sqrt{p_i p_j}$, with ρ_{ij} being the correlation coefficient between the sources i and j ($|\rho_{ij}| \leq 1$). The sources are now located at $\bar{\theta} = [10^\circ \ 15^\circ \ 12^\circ]^\top$ with $\vartheta = 12^\circ$ assumed to be known a priori. In (48), the entries are ordered as in $\bar{\theta}$; hence, for example, p_{12} corresponds to the correlation between the sources at 10 and 15 degrees. The correlation coefficient ρ_{13} in Fig. 2a and Fig. 2b is chosen to maximize the difficulty of this specific simulation scenario in the sense that the CRB_{SML} is large for every angle. The correlation is assumed unknown; we thus omit the estimators explicitly assuming the sources to be uncorrelated from the analysis of this scenario, along with the corresponding bounds. For simplicity we again assume the sources to be equipowered, giving $p_1 = p_2 = p_3$. In the interest of brevity, in Fig. 2 we only present the results for θ_1 ; the results for θ_2 are similar.

It can be seen from Fig. 2 that MODE PLEDGE outperforms the other meth-

ods, and that MODE PLEDGE is significantly more accurate than MODE in both cases. This can be contrasted to Fig. 1a, where the gain was not so pronounced. Hence it can be inferred that the major benefit of MODE PLEDGE is the removal of the correlation between known and unknown sources, rather than that the estimator has one less parameter to estimate.

It appears as if the correlation between the unknown sources and the known source has been cancelled through the exploitation of prior knowledge in C-MUSIC as well — in Fig. 2a C-MUSIC can accurately estimate the unknown source(s) even though $|\rho_{13}| = 1$. On the other hand, C-MUSIC completely fails in Fig. 2b as in that case there exists source correlation even after nullifying the known source — this is in agreement with the general behavior of the MUSIC-algorithm, which is well-known to fail when the sources are coherent. The fact that the correlation has indeed been cancelled can be realized by comparing (24) and (22); the parts of \mathbf{P} related to the known sources, including potential correlation with the unknown sources, vanish due to the orthogonal projectors operating on both sides of $\hat{\mathbf{R}}$.

As expected (and desired) MODE PLEDGE can handle both cases optimally. C-MUSIC can be seen in Fig. 2a to suffer from a higher threshold SNR than MODE PLEDGE; however, it passes quickly through the transition region to an asymptotic accuracy equivalent to that of MODE PLEDGE for larger SNR values. Additionally, the large decrease in $\text{CRB}_{\mathbf{P}}$ as compared to CRB_{SML} can be traced to the correlation cancellation; the estimates using prior knowledge are not perturbed by the coherence between the sources at 10 and 12 degrees.

Note also that the theoretical performance bound is attained by MODE PLEDGE already at low SNRs — however, the CRB at the lowest SNRs is too large to make the estimates meaningful (presumably due to the closely spaced sources).

6.3 Error in prior knowledge

Here we study the case when there is an error in the prior DOA knowledge. To circumvent some of the issues of DOA association in this case we consider only two sources located at $\theta = 10^\circ$ and $\vartheta = 12^\circ$, or $\vartheta = 11^\circ$ depending on the correlation state — when the sources are coherent (with $\rho = \exp(-j\pi/8)$), we increase the separation to simplify the scenario. In Fig. 3 we vary the assumed location of ϑ and study the impact on performance. As before, $m = 6$, $\text{SNR} = 25$ dB, and $N = 1000$.

Even though there are errors in the prior knowledge, the PLEDGE methods can still give enhanced estimates as long as the error is small when compared to the standard RMSE; this is seen in both Fig. 3a and Fig. 3b. Thus, since the studied methods are asymptotically efficient, it is possible to judge whether it is advantageous to utilize a prior DOA-knowledge, with a given accuracy, based on the CRB of the source to be estimated. A prior information with a smaller variance than the estimate's CRB gives a strict statistical performance increase (in the sense

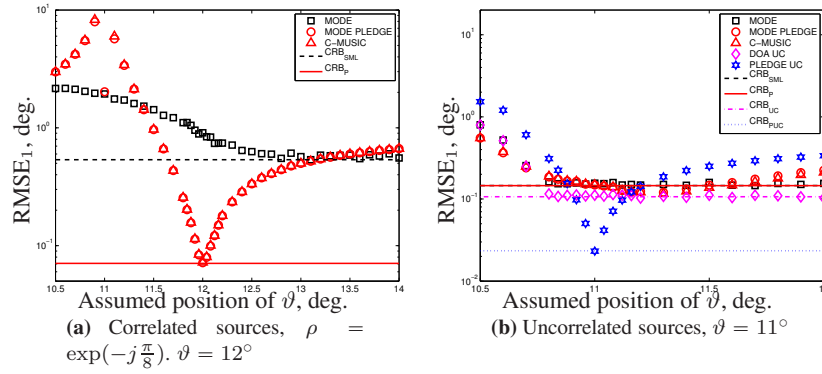


Figure 3: Error in prior knowledge: RMSE of the estimates of $\theta = 10^\circ$ when varying the assumed known location as shown. Equipowered sources with SNR = 25 dB, $m = 6$, $N = 1000$ and 1000 Monte Carlo realizations.

of a smaller variance) in the estimated DOA. As can be seen from e.g. Fig. 3a, in some cases errors larger than this bound can also translate into an increased estimator accuracy.

Note that in Fig. 3a MODE and in Fig. 3b DOA UC suffer when the known DOA is believed to be closer to the unknown one than it actually is — this is due to the angle association as explained in the fifth paragraph of Section 6.

6.4 Varying known-source properties

In this scenario we vary the separation between the known source and the unknown one, and also the known source power.

In Fig. 4a we vary the known DOA and, as can be seen, PLEDGE UC can offer dramatic accuracy increases. This method can correctly estimate and resolve the two sources in the studied case for very small separations.

In Fig. 4b, we vary the power of the known source. When the known source is significantly weaker than the unknown one, the methods not exploiting prior angular information have problems: they easily find the unknown source, but may associate it with the known one. The prior-based methods however do not suffer from this dramatic loss of accuracy. Once again, PLEDGE UC is significantly outperforming the other methods, and attains the accuracy bound for when there is only one source present (which we denoted CRB_{1S} in Fig. 4b). It can thus be seen that PLEDGE UC correctly accounts for the influence the known source has on the estimation problem; if the known source is too weak to affect the estimation of the unknown source, PLEDGE UC has in effect reduced the number of sources in the received data to only contain the unknown one. This is obviously a very desirable property for a prior-using method.

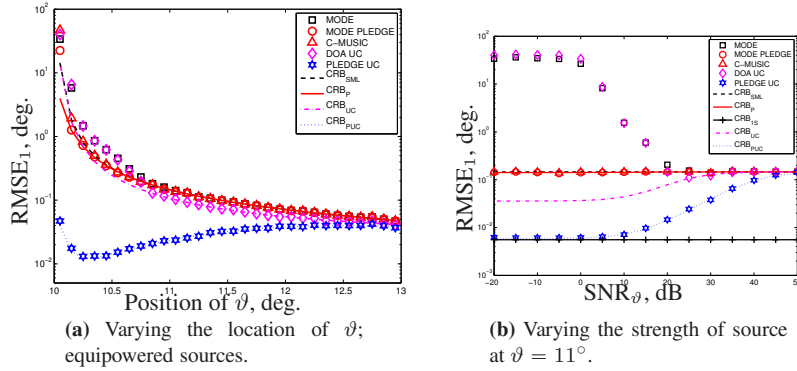


Figure 4: Varying known-source properties: RMSE of the estimates of $\theta = 10^\circ$ when varying the properties of the known source. Uncorrelated sources, $\text{SNR}_\theta = 25$ dB, $m = 6$, $N = 1000$ and 1000 Monte Carlo realizations.

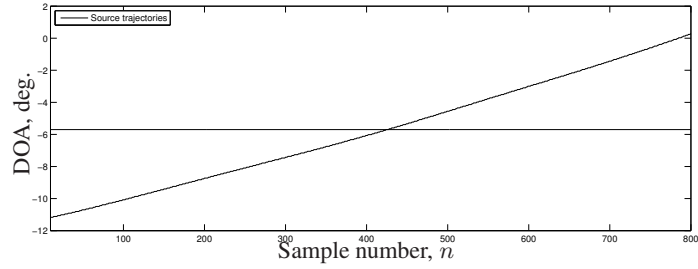
6.5 A note on the computational complexity of the algorithms

The algorithms investigated all showed execution times on the order of milliseconds on standard PCs. See Table 1 where we compare the mean time each algorithm needed to estimate the unknown parameters in the scenario shown in Fig. 1a. The methods are implemented in MATLAB to conceptually agree with their source publications; this has the effect that the implementations are not optimized for computational efficiency. Note that the algorithms scale differently with the scenario-specific parameters (i.e., m , d , etc). To illustrate this effect we also include simulation times for when $m = 20$ in Table 1.

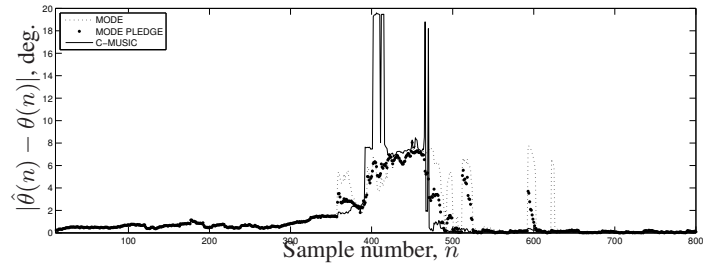
Table 1: Mean execution times, milliseconds.

	$m = 10$	$m = 20$
MODE:	0.8	2.2
MODE PLEDGE:	1.1	2.4
DOA UC:	2.1	8.5
PLEDGE UC:	2.2	8.6
C-MUSIC:	1.0	8.1
MP-MUSIC:	0.5	4.3

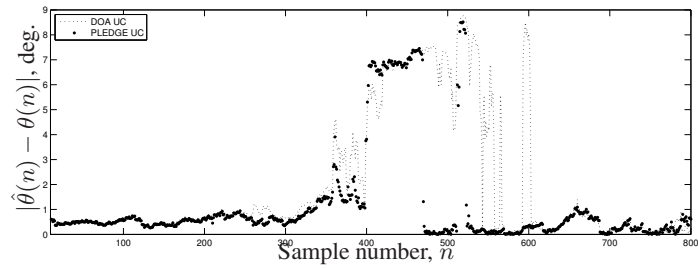
In Table 1 it can be seen that the PLEDGE methodology introduces a small computational burden; using PLEDGE in the MODE algorithm increased the execution time by about 0.3 ms or 40%, whereas for $m = 20$ the increase was about the same, 0.2 ms, which in this case translates to about 10%. The extra computa-



(a) True DOAs of the two sources, as a function of sample number. Dataset 4 from [20].



(b) Methods not exploiting correlation information.



(c) Methods exploiting the fact that the sources are uncorrelated.

Figure 5: Real-data scenario: (a) true positions, and (b), (c) absolute error of DOA estimates ($\lambda = 0.96$). Known source fixed at $\vartheta = -5.7^\circ$, as evident from (a).

tional time required when using PLEDGE in DOA UC was 0.1 ms in both cases which translates to an increase in computational burden of 5% and 1%, respectively.

7 Experimental data application

In this section we use experimental data from the University of Wyoming Source Tracking Array Testbed (UW STAT) [20] ("dataset 4") in order to test the estimators on actual data. The data are collected using a six sensor ultrasonic ULA, operating in a narrow-band setup, with two uncorrelated sources one of which is stationary and that we here consider known. The sources are each emitting 200 Hz bandlimited noise, and the inter-sensor spacing is 2.1 signal wavelengths. Such a spacing provides an unambiguous array response in the region of interest (approximately $\pm 15^\circ$), while enhancing resolution in this region compared to a six-sensor ULA with half-plane field of view. With a sample rate of 800 Hz, but a signal bandwidth of only 200 Hz, we in effect do not have independent samples. This means that the effective number of snapshots, as compared to a theoretical analysis, is reduced by a factor of 2.

The data-dependent input to all the studied algorithms is the covariance matrix $\hat{\mathbf{R}}$ in (7). A problem in this case is that each data vector from the sensors represents a certain DOA, since one target is non-stationary. Thus when forming $\hat{\mathbf{R}}$, data vectors representing different DOAs are combined into $\hat{\mathbf{R}}$. From this $\hat{\mathbf{R}}$, the algorithms will of course find a single DOA for each source. Hence, in order to achieve meaningful results the source movement, or the change in the true DOA, has to be small over the interval of snapshots used to form $\hat{\mathbf{R}}$. In order to take this fact into account a recursive update of $\hat{\mathbf{R}}$ is used. Accordingly, $\hat{\mathbf{R}}$ is updated as

$$\hat{\mathbf{R}}(n) = \lambda \hat{\mathbf{R}}(n-1) + (1-\lambda) \mathbf{y}(n) \mathbf{y}^*(n), \quad \hat{\mathbf{R}}(0) = \mathbf{0}. \quad (49)$$

Thus, at time instant n , the DOA vector $\bar{\boldsymbol{\theta}}(n)$ is estimated based on $\hat{\mathbf{R}}(n)$. In (49) the forgetting parameter λ is chosen based upon how many past samples $\hat{\mathbf{R}}(n)$ is desired to effectively contain, according to the approximative relation: $\lambda = 1 - \frac{2}{N}$. We use $\lambda = 0.96$, approximately corresponding to a rectangular window size of $N = 50$.

The true source trajectories as a function of time can be seen in Fig. 5a, and the absolute error $|\hat{\theta}(n) - \theta(n)|$ of the estimates in Fig. 5b and Fig. 5c, with Fig. 5b corresponding to the algorithms that do not exploit information on the source correlation, and Fig. 5c showing the ones utilizing this information. We plot the absolute error of the estimated angle at each time instant. We do not consider MP-MUSIC, as it was concluded in Section 6.1 that it is not robust.

In the beginning of the data ($n < 300$) all algorithms show a relatively small error. As we proceed in time the moving source is approaching the stationary one. The source separation decreases and we can see that the algorithms suffer in accuracy; at first, this manifests as an increasing bias, which shows up as a gradual increase in absolute error, but eventually the detection breaks down and no useful information is provided by the estimators.

In Fig. 5b, it can be seen that usage of the prior information through the MODE PLEDGE algorithm has a noticeable effect but is not as beneficial as one

might have hoped. Comparing to Fig. 1, this result is not so surprising: the region of substantial improvement is small for MODE PLEDGE when the sources are uncorrelated.

PLEDGE UC, as seen in Fig. 5c, improves the estimation accuracy in some parts of the data sample — note that while its detection breaks down at approximately the same angular separation as for the other algorithms, it resumes tracking sooner giving a larger region of reasonable DOA estimation.

C-MUSIC is about as accurate as MODE PLEDGE; at some time instances it is better, while worse at others.

When the source DOAs are rather close to each other, we see from the figures that the prior-based methods fail similarly to the conventional methods. The PLEDGE methods remove, or filters out, the source at the known direction; however as the unknown source comes too close to the known one, the data from the unknown source is filtered out as well. Thus we cannot circumvent the fact that too closely spaced sources are indistinguishable. In other words the PLEDGE method causes the unknown source SNR to decrease when it comes too close to the known source. This can be contrasted to the situation of the conventional methods, in which the received data transitions into an apparent scenario with a single source of double power in the direction of the merged sources.

By using the Empirical Root Mean Square Error, ERMSE, defined according to

$$\text{ERMSE} = \sqrt{\frac{1}{M} \sum_{k=1}^M (\hat{\theta}_k - \theta)^2} \quad (50)$$

with M being the number of measured DOAs, we can aggregate the results showed in Fig. 5. We thus find that MODE, MODE PLEDGE, and C-MUSIC gives ERMSE values of 2.7° , 2.2° , and 3.4° , respectively. The methods exploiting that the signals are uncorrelated, DOA UC and PLEDGE UC, gives ERMSE values of 3.3° and 2.3° , respectively. We can thus quantify the improvement in estimation accuracy given by the PLEDGE framework; however, in this particular case, the methods exploiting the uncorrelated nature of the source signals (i.e., DOA UC and PLEDGE UC) failed to improve the estimation, as compared to the methods that disregard this information. This is likely due to the real signals not satisfying the data model (1) — the data were acquired from moving sources and averaged according to (49), which is not what the methods were designed for. It is plausible that the UC methods, which are exploiting more of the structure in the received data, are more sensitive to such model imperfections.

8 Conclusions

We have closely examined a prior-knowledge-based framework that uses known positions to facilitate the estimation of unknown ones. We have shown how to exploit such prior DOA-knowledge in a manner that gives asymptotically optimal

estimates whether the impinging signals are correlated or not. Through numerical simulations we have shown that in most scenarios a significant accuracy increase can be expected; this benefit is especially pronounced in the region where the estimator transitions to its asymptotic accuracy.

We also applied the methodology to a real-world scenario with one known DOA and one unknown time-varying DOA. Even though the methods are designed to analyze data received from a stationary target, we saw improvements in the angular estimates.

References

- [1] R. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. of RADC Spectrum Estimation Workshop*, Griffiss AFB, New York, USA, Oct. 1979, pp. 243–258.
- [2] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.
- [3] M. Jansson and B. Ottersten, "Structured covariance matrix estimation: A parametric approach," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 3172–75.
- [4] P. Wirfält, G. Bouleux, M. Jansson, and P. Stoica, "Subspace-based frequency estimation utilizing prior information," in *IEEE Stat. Signal Process. Workshop*, Nice, France, Jun. 2011, pp. 533–536.
- [5] G. Biennu and L. Kopp, "Adaptivity to background noise spatial coherence for high resolution passive methods," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, vol. 5, Denver, Colorado, USA, Apr. 1980, pp. 307–310.
- [6] R. DeGroat, E. Dowling, and D. Linebarger, "The constrained MUSIC problem," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1445–1449, Mar. 1993.
- [7] D. Linebarger, R. DeGroat, E. Dowling, P. Stoica, and G. Fudge, "Incorporating a priori information into MUSIC-algorithms and analysis," *Signal Process.*, vol. 46, no. 1, pp. 85–104, 1995.
- [8] P. Stoica, P. Händel, and A. Nehorai, "Improved sequential MUSIC," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 4, pp. 1230–1239, Oct. 1995.
- [9] R. Behrens, L. Scharf, C. Inc, and C. Broomfield, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1413–1424, Jun. 1994.
- [10] R. Boyer and G. Bouleux, "Oblique projections for direction-of-arrival estimation with prior knowledge," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1374–1387, Apr. 2008.
- [11] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [12] —, "MUSIC, maximum likelihood, and Cramér-Rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.
- [13] M. Jansson, B. Göransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945–956, Apr. 1999.
- [14] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior knowledge-based DOA estimation method," in *17th European Signal Process. Conf.*, Glasgow, UK, Aug. 2009, pp. 869–873.
- [15] P. Wirfält, M. Jansson, G. Bouleux, and P. Stoica, "Prior knowledge-based direction of arrival estimation," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Prague, Czech Republic, May 2011, pp. 2540–2543.

-
- [16] P. Stoica and K. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proc. Part F Radar and Signal Process.*, vol. 137, no. 1, pp. 19–26, Feb. 1990.
 - [17] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–23, May 1999.
 - [18] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, vol. 8, Boston, Massachusetts, USA, Apr. 1983, pp. 336–339.
 - [19] P. Stoica, E. Larsson, and A. Gershman, "The stochastic CRB for array processing: a textbook derivation," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.
 - [20] J. Pierre, E. Scott, and M. Hays, "A sensor array testbed for source tracking algorithms," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Munich, Germany, Apr. 1997, pp. 3769–3772.

Paper B

Prior-Exploiting Direction-of-Arrival Algorithms for Partially Uncorrelated Source Signals

Petter Wirfält and Magnus Jansson

Submitted to
Signal Processing (nov. 2013)

Prior-Exploiting Direction-of-Arrival Algorithms for Partially Uncorrelated Source Signals

Petter Wirfält and Magnus Jansson

Abstract

In certain Direction-of-Arrival (DOA) estimation problems, there is information available regarding the source signals already prior to the actual estimation process. For example, a number of emitters whose DOAs are perfectly known may be present; additionally, there can be information on the correlation state of the transmitted signals. In such scenarios, it is desirable to exploit the prior information already in the estimator design such that the knowledge can benefit the estimation of the DOAs of the unknown sources. We present a novel way of exploiting such knowledge in this article, together with a more detailed analysis of a recent result in this field. Additionally we develop closed-form expressions for the asymptotic accuracy of the studied algorithms, and we compare the result to the relevant Cramér-Rao Bound. The realizable performance in the finite sample-case is studied through numerical Monte-Carlo simulations, from which one can conclude that the theoretically predicted accuracies are attained for modest sample sizes and comparatively low SNR, meaning that the studied algorithms significantly outperform current state-of-art.

Index Terms—Accuracy, Arrays, Covariance matrix, Direction of arrival estimation, Signal processing algorithms

1 Introduction

Direction-of-Arrival (DOA) estimation is a classical topic in signal processing and much work has been done in the area in the past decades. A passive array of sensors is receiving signals from a number of distinct sources, and the objective is to estimate the directions these signal are impinging from. In the seminal works [1] and [2], statistically efficient DOA estimation methods are presented. The underlying assumptions for the mentioned methods to be efficient were quite mild, notably,

i.i.d. spatially white sensor noise. Thus, the mild assumptions give estimators that are applicable to a wide range of scenarios.

However, many different scenarios exist in practice, and in some of these more restrictive assumptions can be made; for example, some of the source directions might be known *a-priori*. Exploiting such information in the design of the estimator can be expected to produce methods that are more accurate than [1] and [2], and that this is indeed possible has been shown numerous times, e.g. [3], [4], [5]. Another example of a more restrictive assumption is that it might be known that the source signals are spatially uncorrelated, e.g. [6], [7], and the combination of these two types of prior information was in [8] shown to be beneficial.

Recently, a new DOA algorithm denoted Prior Orthogonally Weighted Direction Estimator (POWDER), was proposed [9], where it was assumed that the known and the unknown DOAs are uncorrelated, but no assumptions were made on the correlation between the signals in the sets of known and unknown signals, respectively. Thus, the method of [9] is applicable in a wider set of scenarios than the one in [8], and it was also shown that for scenarios where both methods are applicable, the former possess better small-sample performance than the latter. Additionally, in other scenarios, POWDER significantly outperformed state-of-the-art methods [1], [5].

In this article we extend the work in [9] in the following ways: we devise an estimator denoted POWDER2 that accounts for the non-circular statistics of the POWDER residual; we derive closed form expressions for the asymptotic variances of POWDER and POWDER2, and show these to be identical for high Signal-to-Noise Ratios (SNRs); we derive the Cramér-Rao Bound (CRB) under the particular assumptions studied; we show that under some conditions both POWDER-methods attain the CRB; we conclude the article by, through numerical simulations, investigating the finite-sample, finite-SNR performance of the studied methods and the relation to the theoretically derived variances.

The article is structured as follows. In Section 2 we revisit the problem formulation given in [9], and in Section 3 we look at the theoretical derivation of POWDER; for reasons explained there we also introduce POWDER2. We derive an expression for the CRB in Section 4, and we also show how to simplify that expression for high SNR. In Section 5 we make a performance analysis of the POWDER-algorithms and we see that, asymptotically in N and for large SNR, the algorithms both obtain the CRB. We conclude the article in Section 6 by numerical Monte-Carlo (MC) experiments, in which we compare the performance of the investigated algorithms as well as the theoretically derived expressions for their accuracy.

We use the following notation: Boldface lowercase (uppercase) letters denote vectors (matrices). The operator \otimes denotes the Kronecker product, and the operator $\text{vec}(\mathbf{X})$ stacks the columns of the matrix \mathbf{X} into a vector. It can be verified that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ for matrices of matching dimensions. The superscripts T , c , and $*$ denote transpose, conjugate, and conjugate transpose, re-

spectively. By the symbol \triangleq we indicate a definition. If a quantity $X = O(x)$, then X/x is bounded as $x \rightarrow 0$, and if $X = o(x)$, then $X/x \rightarrow 0$ as $x \rightarrow 0$. We also use $O_p(x)$ and $o_p(x)$, which are the respective in-probability versions [10]. We denote the Frobenius norm of the matrix \mathbf{X} by $\|\mathbf{X}\|$. We define the projection matrix $\mathbf{\Pi}_{\mathbf{X}} \triangleq \mathbf{X}(\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^*$ for any full-rank matrix \mathbf{X} , and the orthogonal projector $\mathbf{\Pi}_{\mathbf{X}}^\perp \triangleq \mathbf{I} - \mathbf{\Pi}_{\mathbf{X}}$.

2 Problem description

Consider the narrow-band signal model (see e.g. [2])

$$\mathbf{y}(t) = \mathbf{A}(\bar{\boldsymbol{\theta}})\mathbf{x}(t) + \mathbf{n}(t), \quad t = 0, 1, \dots, N-1. \quad (1)$$

Here, the vector $\mathbf{y}(t) \in \mathbb{C}^{m \times 1}$ represents the sensor array output, and $\mathbf{x}(t) \in \mathbb{C}^{d \times 1}$ the signal samples, at time t . The matrix $\mathbf{A}(\bar{\boldsymbol{\theta}}) \in \mathbb{C}^{m \times d}$ is the array steering matrix, which is uniquely determined by the array geometry and the (assumed distinct) DOAs $\bar{\boldsymbol{\theta}}$ of the impinging signals (we reserve $\boldsymbol{\theta}$ for the *unknown* DOAs, see below). The dimensions m and d correspond to the number of sensors and source signals, respectively. Finally, $\mathbf{n}(t) \in \mathbb{C}^{m \times 1}$ represents the sensor noise. We model both the signal and the noise vectors as independent zero mean, i.i.d. circularly symmetric complex Gaussian random processes with spatial covariance matrices given by $\text{cov}(\mathbf{x}(t)) = \mathbf{P}$ and $\text{cov}(\mathbf{n}(t)) = \sigma^2 \mathbf{I}$, respectively. Using (1) and the definitions above, the sensor output covariance matrix is

$$\mathbf{R} \triangleq \text{E}[\mathbf{y}(t)\mathbf{y}^*(t)] = \mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2 \mathbf{I}. \quad (2)$$

The first key assumption in the current article, which delimits it from some well-known state of the art results in this field (e.g. [2]), is that we assume some of the signal directions to be known *a-priori*; hence we are only interested in estimating $d_u = d - d_k$ of the DOAs, where the subscripts u and k henceforth denote unknown and known, respectively. With that fact in mind we can, without loss of generality, write

$$\bar{\boldsymbol{\theta}}_0 = [\boldsymbol{\theta}_0^T \quad \boldsymbol{\vartheta}^T]^T; \quad (3)$$

$$\mathbf{A}(\bar{\boldsymbol{\theta}}_0) = [\mathbf{A}(\boldsymbol{\theta}_0) \quad \mathbf{A}(\boldsymbol{\vartheta})] \triangleq [\mathbf{A}_u \quad \mathbf{A}_k]; \quad (4)$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_u & \mathbf{P}_{uk} \\ \mathbf{P}_{uk}^* & \mathbf{P}_k \end{bmatrix} \quad (5)$$

where henceforth $\boldsymbol{\theta}_0$ and $\boldsymbol{\vartheta}$ denote the unknown and the known DOAs, respectively. We further distinguish between $\boldsymbol{\theta}_0$, representing the *true values* of the unknown DOAs, and $\boldsymbol{\theta}$, which is the *parametrization* of the unknown DOAs. In (5),

the subscripts denote the correlation states between the signals emanating from the unknown and known DOAs. The second defining assumption of this work is

$$\mathbf{P}_{uk} = \mathbf{0}; \quad (6)$$

hence we assume, and exploit, that there is no correlation between the signals from the known and unknown directions. There is no other work known to the authors which have studied this particular instance of the DOA problem; in the DOA scenarios studied in [6] and [11], it was assumed that the source signals were perfectly uncorrelated, i.e. diagonal \mathbf{P} — that assumption is a strict subset of the current assumption $\mathbf{P}_{uk} = \mathbf{0}$. Accordingly, we do not make any assumptions on \mathbf{P}_k or \mathbf{P}_u ; except, we need to know their respective rank and hence introduce the parameters $d'_u = \text{rank}(\mathbf{P}_u)$, $d'_k = \text{rank}(\mathbf{P}_k)$, and $d' = \text{rank}(\mathbf{P}) = d'_u + d'_k$. We also require

$$\begin{cases} m > 2d_u - d'_u \\ m \geq d \end{cases} \quad (7)$$

to ensure identifiability of θ ; we motivate (7) in Section 5.1.

3 POWDER

3.1 Review of method

Below we study the theoretical properties of the POWDER-method [9]. First we summarize the steps of the estimator. Start by creating the noise-free sensor-output covariance matrix,

$$\begin{aligned} \mathbf{R} - \sigma^2 \mathbf{I} &= \mathbf{A} \mathbf{P} \mathbf{A}^* \\ &= [\mathbf{A}_u \quad \mathbf{A}_k] \begin{bmatrix} \mathbf{P}_u & \mathbf{P}_{uk} \\ \mathbf{P}_{uk}^* & \mathbf{P}_k \end{bmatrix} \begin{bmatrix} \mathbf{A}_u^* \\ \mathbf{A}_k^* \end{bmatrix}. \end{aligned} \quad (8)$$

If then (some of) the contribution from the known directions is removed by multiplying with the orthogonal projector $\Pi_{\mathbf{A}_k}^\perp$ (such that $\mathbf{A}_k^* \Pi_{\mathbf{A}_k}^\perp = \mathbf{0}$) from the left, we get

$$\begin{aligned} (\mathbf{R} - \sigma^2 \mathbf{I}) \Pi_{\mathbf{A}_k}^\perp &= (\mathbf{A}_u \mathbf{P}_u + \mathbf{A}_k \mathbf{P}_{uk}^*) \mathbf{A}_u^* \Pi_{\mathbf{A}_k}^\perp \\ &= \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* \Pi_{\mathbf{A}_k}^\perp \end{aligned} \quad (9)$$

$$= \mathbf{U}_s \Sigma_s \mathbf{V}_s^*, \quad (10)$$

where (9) follows from that $\mathbf{P}_{uk} = \mathbf{0}$, and \mathbf{U}_s , Σ_s , and \mathbf{V}_s follow from the singular value decomposition of (9) (note that only d'_u singular values of Σ_s are non-zero). The subscript s denotes the signal subspace (corresponding to the unknown sources). In practice we do not have access to the true quantities (except,

by assumption, $\Pi_{\mathbf{A}_k}^\perp$) in (10); thus we estimate \mathbf{R} through

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}(t)\mathbf{y}^*(t). \quad (11)$$

To find $\hat{\sigma}^2$, we perform the eigenvalue decomposition (EVD) of (11) according to

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\mathbf{\Lambda}}_n \hat{\mathbf{E}}_n^*, \quad (12)$$

where $\hat{\mathbf{E}}_s$ contains the eigenvectors associated with the d' largest eigenvalues of $\hat{\mathbf{R}}$ and $\hat{\mathbf{\Lambda}}_s$ is the matrix containing these eigenvalues on the diagonal. Similarly, $\hat{\mathbf{E}}_n$ contains the remaining eigenvectors (and spans the noise subspace) and $\hat{\mathbf{\Lambda}}_n$ the associated eigenvalues. Now, we can find $\hat{\sigma}^2 = \frac{1}{m-d'} \text{Tr}(\hat{\mathbf{\Lambda}}_n)$. In practice the realization of (10) thus becomes:

$$\left(\hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{I}\right) \Pi_{\mathbf{A}_k}^\perp = \hat{\mathbf{U}}_s \hat{\mathbf{\Sigma}}_s \hat{\mathbf{V}}_s^* + \hat{\mathbf{U}}_n \hat{\mathbf{\Sigma}}_n \hat{\mathbf{V}}_n^*. \quad (13)$$

In (13), the singular vectors corresponding to the d'_u principal singular values are collected in $\hat{\mathbf{U}}_s$ and $\hat{\mathbf{V}}_s$, while the remaining $m - d'_u$ singular vectors are collected in $\hat{\mathbf{U}}_n$ and $\hat{\mathbf{V}}_n$.

Using (13) and the fact that $\hat{\mathbf{V}}_s$ and $\hat{\mathbf{V}}_n$ form an orthonormal basis,

$$\hat{\mathbf{U}}_s = \left(\hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{I}\right) \Pi_{\mathbf{A}_k}^\perp \hat{\mathbf{V}}_s \hat{\mathbf{\Sigma}}_s^{-1}. \quad (14)$$

We now introduce the matrix $\mathbf{B}(\boldsymbol{\theta}) \in \mathbb{C}^{m \times (m-d_u)}$, which spans the null space of \mathbf{A}_u^* (i.e. $\mathbf{B}^* \mathbf{A}_u = \mathbf{0}$), such that

$$\begin{aligned} \mathbf{B}^* \hat{\mathbf{U}}_s &= \mathbf{B}^* \left(\hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{I}\right) \Pi_{\mathbf{A}_k}^\perp \hat{\mathbf{V}}_s \hat{\mathbf{\Sigma}}_s^{-1} \\ &= \mathbf{B}^* \left(\tilde{\mathbf{R}} - \tilde{\sigma}^2 \mathbf{I}\right) \Pi_{\mathbf{A}_k}^\perp \mathbf{V}_s \mathbf{\Sigma}_s^{-1} + o_p(1/\sqrt{N}), \end{aligned} \quad (15)$$

since $\tilde{\mathbf{R}} \triangleq \hat{\mathbf{R}} - \mathbf{R} = O_p(1/\sqrt{N})$, $\tilde{\sigma}^2 \triangleq \hat{\sigma}^2 - \sigma^2 = O_p(1/\sqrt{N})$, $\mathbf{B}^*(\mathbf{R} - \sigma^2 \mathbf{I}) \Pi_{\mathbf{A}_k}^\perp = \mathbf{0}$, $\hat{\mathbf{V}}_s = \mathbf{V}_s + o_p(1)$, and $\hat{\mathbf{\Sigma}}_s = \mathbf{\Sigma}_s + o_p(1)$. Further, $\text{Span}(\mathbf{V}_s) \subseteq \text{Span}(\Pi_{\mathbf{A}_k}^\perp)$ and hence $\Pi_{\mathbf{A}_k}^\perp \mathbf{V}_s = \mathbf{V}_s$; then we can re-write (15) as

$$\mathbf{B}^* \hat{\mathbf{U}}_s = \mathbf{B}^* \left(\tilde{\mathbf{R}} - \tilde{\sigma}^2 \mathbf{I}\right) \mathbf{V}_s \mathbf{\Sigma}_s^{-1} + o_p(1/\sqrt{N}). \quad (16)$$

By forming the residual error vector

$$\boldsymbol{\epsilon} \triangleq \text{vec} \left(\hat{\mathbf{U}}_s^* \mathbf{B} \right), \quad (17)$$

a cost function can be created:

$$V_{\text{POW}_1}(\boldsymbol{\theta}) = \boldsymbol{\epsilon}^* \mathbf{W}_1 \boldsymbol{\epsilon}, \quad (18)$$

where $\mathbf{W}_1 > 0$ is a positive definite (p.d.) weighting matrix. In [9], estimates of $\boldsymbol{\theta}_0$ were found as

$$\hat{\boldsymbol{\theta}}_{\text{POW1}} = \arg \min_{\boldsymbol{\theta}} V_{\text{POW1}}(\boldsymbol{\theta}). \quad (19)$$

It is in (18) implicitly assumed that $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \mathbf{0}$, i.e. that $\boldsymbol{\epsilon}$ is circularly symmetric; this is not necessarily the case (please see Section 3.3), which implies that the target function (18), as chosen in [9], does not capture the full statistics of the problem. To circumvent this problem, we introduce the extended residual

$$\bar{\boldsymbol{\epsilon}} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}^c \end{bmatrix}, \quad (20)$$

with the corresponding cost function

$$V_{\text{POW2}}(\boldsymbol{\theta}) = \bar{\boldsymbol{\epsilon}}^* \mathbf{W}_2 \bar{\boldsymbol{\epsilon}}, \quad (21)$$

where $\mathbf{W}_2 > 0$ is a weighting matrix to be determined. The POWDER2 estimates of $\boldsymbol{\theta}_0$ are obtained as

$$\hat{\boldsymbol{\theta}}_{\text{POW2}} = \arg \min_{\boldsymbol{\theta}} V_{\text{POW2}}(\boldsymbol{\theta}). \quad (22)$$

3.2 Implementation for uniform linear array

When the receiving array is a uniform linear array (ULA) it is possible to rewrite the minimization problem in a particularly appealing form. This is due to that \mathbf{A}_u in this case has a Vandermonde-structure; we can write its l th column (for half wave-length intra sensor spacing) as

$$\mathbf{a}(\theta_l) = \begin{bmatrix} 1 \\ \exp(-j\pi \sin(\theta_l)) \\ \vdots \\ \exp(-j(m-1)\pi \sin(\theta_l)) \end{bmatrix}, \quad (23)$$

in which $j^2 = -1$ and $l = 1, \dots, d_u$. Then, the matrix \mathbf{B} can be written as [12]

$$\mathbf{B}^T = \begin{bmatrix} b_0 & b_1 & \dots & b_{d_u} & & \mathbf{0} \\ & \ddots & \ddots & & \ddots & \\ \mathbf{0} & & b_0 & b_1 & \dots & b_{d_u} \end{bmatrix}, \quad (24)$$

where b_i are given as the coefficients of the polynomial

$$b_0 \prod_{l=1}^{d_u} (z - e^{-j\pi \sin(\theta_l)}) \triangleq b_0 z^{d_u} + b_1 z^{d_u-1} + \dots + b_{d_u}. \quad (25)$$

We collect the polynomial coefficients of (25) according to $\mathbf{b} = [b_0 \ b_1 \ \dots \ b_{d_u}]^T$. Through the relation $\text{vec}(\mathbf{B}) = \mathbf{\Psi}\mathbf{b}$, with $\mathbf{\Psi}$ a selection matrix, we can minimize (18) (and (21)) with respect to \mathbf{b} . The DOAs are then estimated by rooting (25), with coefficients from $\hat{\mathbf{b}}$.

The minimization of $V_{\text{POW1}}(\boldsymbol{\theta})$ (or $V_{\text{POW2}}(\boldsymbol{\theta})$) with respect to \mathbf{b} can additionally be performed as an eigenvalue problem for \mathbf{b} . For some technical details regarding this minimization, please see [13].

3.3 Statistical properties of the residual

We now explore the statistical properties of the residual (17), and derive optimal choices of the weighting matrices \mathbf{W}_1 and \mathbf{W}_2 in (18) and (21).

THEOREM 1

Given the data model and assumptions in Section 2, the following expressions hold:

$$\mathbf{C}_\epsilon \triangleq \lim_{N \rightarrow \infty} N \text{E}[\epsilon\epsilon^*] = \mathbf{M}\mathbf{F}(\mathbf{R}^T \otimes \mathbf{R})\mathbf{F}^*\mathbf{M}^*, \quad (26)$$

$$\mathbf{C}_{\tilde{\epsilon}} \triangleq \lim_{N \rightarrow \infty} N \text{E}[\epsilon\epsilon^T] = \mathbf{M}\mathbf{F}(\mathbf{R}^T \otimes \mathbf{R})\mathbf{L}_{mm}\mathbf{F}^T\mathbf{M}^T, \quad (27)$$

where

$$\mathbf{M} \triangleq (\mathbf{B}^T \otimes \boldsymbol{\Sigma}_s^{-1}\mathbf{V}_s^*), \quad (28)$$

$$\mathbf{F} \triangleq \left(\mathbf{I}_m - \frac{1}{m-d'} \text{vec}(\mathbf{I}_m)\boldsymbol{\pi}^* \right), \quad (29)$$

$$\boldsymbol{\pi} \triangleq \text{vec}(\boldsymbol{\Pi}_{\mathbf{E}_s}^\perp), \quad (30)$$

and the permutation matrix \mathbf{L}_{mm} is defined such that $\mathbf{L}_{mm} \text{vec}(\mathbf{X}_{m \times m}) = \text{vec}(\mathbf{X}_{m \times m}^T)$. Further,

$$\mathbf{C}_\epsilon = \check{\mathbf{C}}_\epsilon + O(\sigma^4), \quad (31)$$

$$\mathbf{C}_{\tilde{\epsilon}} = O(\sigma^4), \quad (32)$$

in which

$$\check{\mathbf{C}}_\epsilon \triangleq (\mathbf{B}^T \mathbf{R}^T \mathbf{B}^c \otimes \boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{R} \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1}) > \mathbf{0}. \quad (33)$$

Proof: See Appendix A. ■

It is well known (see e.g. [14]) that the criteria (19) and (22) give minimum variance estimates (for the respective residuals) if the weighting matrices are chosen as (scaled versions of) the inverse covariance of the relevant residual; accordingly, we propose to choose \mathbf{W}_1 and \mathbf{W}_2 as

$$\mathbf{W}_1^{\text{OPT}} = \sigma^2 \mathbf{C}_\epsilon^{-1}, \quad (34)$$

and

$$\mathbf{W}_2^{\text{OPT}} = \sigma^2 \begin{bmatrix} \mathbf{C}_\epsilon & \mathbf{C}_{\bar{\epsilon}} \\ \mathbf{C}_{\bar{\epsilon}}^c & \mathbf{C}_\epsilon^T \end{bmatrix}^{-1}, \quad (35)$$

respectively, where the scaling σ^2 is introduced in order to keep the weighting matrices finite for small σ^2 .

Using the result of Theorem 1, we thus have that for high SNR,

$$\begin{aligned} V_{\text{POW2}}^{\text{OPT}} &= \sigma^2 \bar{\epsilon}^* \left(\begin{bmatrix} \check{\mathbf{C}}_\epsilon & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{C}}_\epsilon^T \end{bmatrix}^{-1} \right) \bar{\epsilon} \\ &= \sigma^2 \bar{\epsilon}^* \check{\mathbf{C}}_\epsilon^{-1} \bar{\epsilon} + \sigma^2 \epsilon^T \check{\mathbf{C}}_\epsilon^{-T} \epsilon^c \\ &= 2V_{\text{POW1}}^{\text{OPT}}, \end{aligned} \quad (36)$$

since $\check{\mathbf{C}}_\epsilon^{-1}$ is hermitian. Hence, we can see that the motivation for introducing $\hat{\theta}_{\text{POW2}}$ vanishes for high SNR.

We will explore the small sample, low SNR, properties of both estimators in Section 6.

From Theorem 1, it can be seen that $\mathbf{W}_1^{\text{OPT}}$ and $\mathbf{W}_2^{\text{OPT}}$, through \mathbf{M} , are functions of \mathbf{B} . Based on the consistency analysis in Section 5.1, it can however be seen that estimates of θ_0 are consistent for *any* p.d. choice of \mathbf{W}_1 , \mathbf{W}_2 ; hence initial estimates of θ_0 can be used to construct (consistent) estimates of $\mathbf{W}_1^{\text{OPT}}$ or $\mathbf{W}_2^{\text{OPT}}$, as appropriate.

Further, N is in practice finite; thus \mathbf{C}_ϵ and $\mathbf{C}_{\bar{\epsilon}}$ must be estimated from sample data. From an analysis similar to the one in (15), it can be deduced that $\widehat{\mathbf{W}}_1^{\text{OPT}} \triangleq \hat{\sigma}^2 \hat{\mathbf{C}}_\epsilon^{-1} = \mathbf{W}_1^{\text{OPT}} + o_p(1)$; hence, $\widehat{\mathbf{W}}_1^{\text{OPT}}$ is a consistent estimate of $\mathbf{W}_1^{\text{OPT}}$ and we can use $\widehat{\mathbf{W}}_1^{\text{OPT}}$ in (19) without impairing the asymptotic properties of the estimator. The same holds for $\widehat{\mathbf{W}}_2^{\text{OPT}}$.

4 Cramér-Rao bound

Since POWDER is designed to exploit a block-diagonal structure of the source covariance matrix \mathbf{P} , it is of interest to compare its accuracy with the theoretically best achievable performance under such assumptions. Let

$$\boldsymbol{\alpha} \triangleq [\boldsymbol{\theta}^T \quad \boldsymbol{\rho}^T \quad \sigma^2]^T \quad (37)$$

denote the vector of unknown parameters in the model; we thus have $\boldsymbol{\rho} = [\boldsymbol{\rho}_k^T \quad \boldsymbol{\rho}_u^T]^T$ with $\boldsymbol{\rho}_k$ being the vector made from $\{[\mathbf{P}_k]_{ii}\}$ and $\{\text{Re}([\mathbf{P}_k]_{ij}), \text{Im}([\mathbf{P}_k]_{ij}); i > j\}$, and $\boldsymbol{\rho}_u$ is made from \mathbf{P}_u in a similar fashion. The primary interest herein is an accuracy bound for the DOAs $\boldsymbol{\theta}$. Hence, a bound on the relevant subset of $\boldsymbol{\alpha}$ is found in the following theorem.

THEOREM 2

Given the assumptions in Section 2,

$$CRB_{\theta} = \frac{1}{N} \left(\mathbf{G}^* \mathbf{\Pi}_{\bar{\Delta}} \mathbf{G} \right)^{-1}, \quad (38)$$

where we by θ signify that the bound is only on the DOAs, but the entire α is unknown. In (38), we have

$$\mathbf{G} \triangleq \mathbf{C}_{\mathbf{r}}^{1/2} \frac{\partial \mathbf{r}}{\partial \theta^T}, \quad (39)$$

where $\mathbf{r} = \text{vec}(\mathbf{R})$, $\mathbf{C}_{\mathbf{r}}^{1/2} \triangleq \mathbf{R}^{-T/2} \otimes \mathbf{R}^{-1/2}$, and

$$\bar{\Delta} \triangleq \mathbf{C}_{\mathbf{r}}^{1/2} \left[(\mathbf{A}_k^c \otimes \mathbf{A}_k) \quad (\mathbf{A}_u^c \otimes \mathbf{A}_u) \quad \text{vec}(\mathbf{I}) \right]. \quad (40)$$

Proof: See Appendix B ■

It is possible to simplify (38) under more restrictive assumptions.

THEOREM 3

Given the assumptions in Section 2 and the additional condition $\lambda_{\min}^{-1}(\mathbf{P}_k) \rightarrow 0$, we have that

$$CRB_{\theta} = \overline{CRB}_{\theta} + O(\|\mathbf{P}_k^{-2}\|), \quad (41)$$

where

$$\overline{CRB}_{\theta} \triangleq \frac{1}{2N} \left[\text{Re} \left(\left(\mathbf{D}^* \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_{uw}} \mathbf{R}^{-1/2} \mathbf{D} \right) \odot \left(\mathbf{P}_u^T \mathbf{A}_{uw}^T \mathbf{A}_{uw}^c \mathbf{P}_u^c \right) \right) \right]^{-1}, \quad (42)$$

\mathbf{D} is given in (C.23), $\mathbf{A}_{uw} = \mathbf{R}^{-1/2} \mathbf{A}_u$, and \odot is the Hadamard (or Schur) product, i.e. element-wise multiplication.

Proof: See Appendix C. ■

As will be seen in Section 6, the additional assumption required in Theorem 3 is in practice not so restrictive; for example, the assumption of $\lambda_{\min}^{-1}(\mathbf{P}_k) \rightarrow 0$ may be valid already at moderate SNRs of the sources in the known directions.

5 Performance analysis

In this section we analyze the asymptotic properties of the estimator given by (19). The analysis of the estimator given by (22) is identical, and the results are very similar; additionally, we saw that, due to (36), the motivation for (22) vanishes for high SNR. Due to these reasons, we do not explicitly show the performance analysis of (22) in this article.

We now establish the consistency of the estimator and also find the asymptotic distribution of the errors in the estimates.

5.1 Consistency of estimates

The cost function (18) converges, asymptotically in N and uniformly in $\boldsymbol{\theta}$, with probability 1 to

$$\begin{aligned} V_\infty(\boldsymbol{\theta}) &= \lim_{N \rightarrow \infty} V_{\text{POW1}}(\boldsymbol{\theta}) \\ &= \text{vec}^*(\mathbf{U}_s^* \mathbf{B}(\boldsymbol{\theta})) \mathbf{W}_1 \text{vec}(\mathbf{U}_s^* \mathbf{B}(\boldsymbol{\theta})). \end{aligned} \quad (43)$$

Since \mathbf{W}_1 in (43) is p.d. (by definition), $V_\infty \geq 0$ with equality if and only if $\mathbf{U}_s^* \mathbf{B}(\boldsymbol{\theta}) = \mathbf{0}$. Now we establish that

$$V_\infty(\boldsymbol{\theta}) = 0 \implies \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad (44)$$

under the conditions in Section 2.

From [15], (44) follows for $m > 2d_u - d'_u$ if \mathbf{U}_s from (10) can be written

$$\mathbf{U}_s = \mathbf{A}_u(\boldsymbol{\theta}_0) \mathbf{T}, \quad (45)$$

for some full rank matrix $\mathbf{T} \in d_u \times d'_u$. Since \mathbf{U}_s is defined from the singular value decomposition of $\mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* \boldsymbol{\Pi}_{\mathbf{A}_k}^\perp$, it follows that

$$\text{rank}(\mathbf{T}) = \text{rank}(\mathbf{P}_u \mathbf{A}_u^* \boldsymbol{\Pi}_{\mathbf{A}_k}^\perp). \quad (46)$$

Using Frobenius' rank inequality (e.g., [16]),

$$\begin{aligned} \text{rank}(\mathbf{T}) &\geq \text{rank}(\mathbf{P}_u \mathbf{A}_u^*) + \text{rank}(\mathbf{A}_u^* \boldsymbol{\Pi}_{\mathbf{A}_k}^\perp) \\ &\quad - \text{rank}(\mathbf{A}_u^*). \end{aligned} \quad (47)$$

Now we note the following: if $m \geq d$, $\mathbf{A}(\bar{\boldsymbol{\theta}}_0) = [\mathbf{A}_u \quad \mathbf{A}_k]$ is full rank and then the Schur-complement of \mathbf{A}_k in $\mathbf{A}^* \mathbf{A}$, $\mathbf{A}_u^* \boldsymbol{\Pi}_{\mathbf{A}_k}^\perp \mathbf{A}_u > \mathbf{0}$; hence, $\mathbf{A}_u^* \boldsymbol{\Pi}_{\mathbf{A}_k}^\perp$ is of full rank d_u . Thus, for $m \geq d$, we have that

$$\text{rank}(\mathbf{T}) \geq d'_u + d_u - d_u, \quad (48)$$

or, equivalently, $\text{rank}(\mathbf{T}) = d'_u$. Hence, (44) is established; thus the estimates $\hat{\boldsymbol{\theta}}$ from (19) converges to $\boldsymbol{\theta}_0$ with probability 1 as $N \rightarrow \infty$.

5.2 Asymptotic distribution

To find the asymptotic variance of the estimates, we perform a Taylor expansion of $\frac{\partial V}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \triangleq V'(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$:

$$\mathbf{0} = V'(\hat{\boldsymbol{\theta}}) = V'(\boldsymbol{\theta}_0) + \frac{\partial V'}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|^2). \quad (49)$$

If we define

$$\mathbf{H} \triangleq \lim_{N \rightarrow \infty} \frac{\partial V'}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad (50)$$

and let $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, we have (for small errors $\tilde{\boldsymbol{\theta}}$, and asymptotically in N) from (49) that

$$\tilde{\boldsymbol{\theta}} = -\mathbf{H}^{-1} V'(\boldsymbol{\theta}_0). \quad (51)$$

Then we can, by defining

$$\mathbf{Q} \triangleq \lim_{N \rightarrow \infty} N \mathbf{E} [V'(\boldsymbol{\theta}_0) V'^*(\boldsymbol{\theta}_0)], \quad (52)$$

formulate the following theorem.

THEOREM 4

Given the assumptions of Section 2, the POWDER estimates asymptotically in N satisfy

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{POW}} - \boldsymbol{\theta}_0) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\text{POW}}), \quad (53)$$

where $\mathbf{C}_{\text{POW}} = \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1}$, and \mathbf{Q} and \mathbf{H} are given by (D.10) and (D.13), respectively.

Proof: The Gaussianity of (53) follows from the fact that $\sqrt{N} \text{vec}(\hat{\mathbf{R}} - \mathbf{R})$ is asymptotically Gaussian (due to the central limit theorem), together with (51), (D.9), and (A.3). That (53) is zero-mean follows from the consistency proven in Section 5.1. For the derivation of \mathbf{C}_{POW} , see Appendix D. ■

Due to (36), we have that the estimates from both estimators in Section 3, when using the optimal weighting matrices (34) and (35), respectively, have the asymptotic distribution given by Theorem 4 at high SNRs. In addition, when using said optimal weighting matrices, we can formulate a stronger result given by the following theorem.

THEOREM 5

Given the assumptions in Section 2 we have that, when using the optimal weighting matrices (34) (or (35), respectively),

$$\mathbf{C}_{\text{POW}} = N \cdot \overline{\text{CRB}}_{\boldsymbol{\theta}} + O(\sigma^4), \quad (54)$$

and $\overline{\text{CRB}}_{\boldsymbol{\theta}}$ is given by (42).

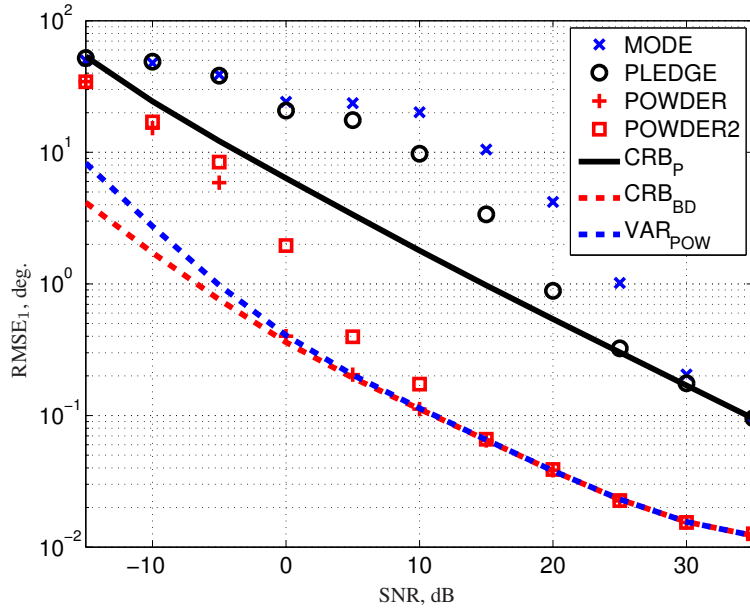


Figure 1: Comparison of POWDER to existing state-of-the-art methods. Normalized RMSE, averaged over 1000 MC realizations, along with the derived theoretical bounds; $\rho_k = \rho_u = 0.9$, and $N = 1000$.

Proof: See Appendix E ■

Thus we have shown that for large N and high SNR, the proposed method(s) are optimal in the sense that a higher accuracy is not obtainable. We will in Section 6 examine the applicability of the respective expressions (53) and (54) as predictors of estimator accuracy for finite N and SNR.

6 Numerical examples

In this section we investigate the practical accuracy of the investigated estimators through Monte-Carlo (MC) simulations, as well as the asymptotic accuracy results (53) and (54). In the MC-simulations, we generate pseudo-random Gaussian noise sequences $\mathbf{x}(t)$ and $\mathbf{n}(t)$ with the prescribed statistics. We have two unknown and two known sources: $\boldsymbol{\theta} = [10^\circ \ 15^\circ]^\top$, and $\boldsymbol{\vartheta} = [12^\circ \ 20^\circ]^\top$. In order to streamline the presentation, we control the source powers by the two parameters SNR_u and SNR_k , where the subscripts refer to unknown and known, respectively.

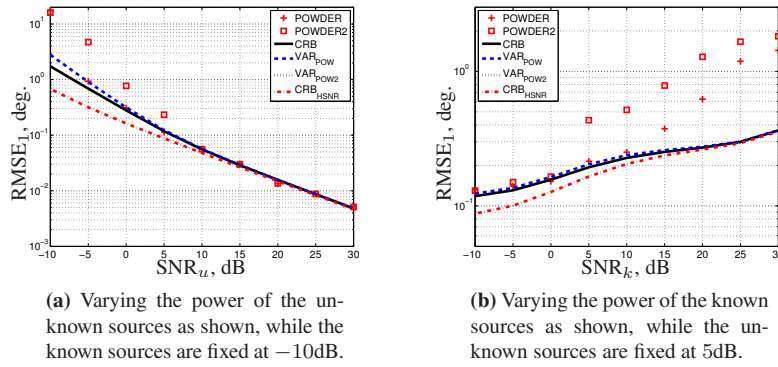


Figure 2: Scenario as in Fig. 1, except for SNR as detailed in the respective figure; averages of 1000 MC-simulations, $N = 1000$. Note that VAR_{POW2} is visually coinciding with the CRB.

Thus, we let the signals from each set of emitters be equipowered, according to

$$\mathbf{P}_u = \text{SNR}_u \begin{bmatrix} 1 & \rho_u \\ \rho_u^c & 1 \end{bmatrix}, \quad \mathbf{P}_k = \text{SNR}_k \begin{bmatrix} 1 & \rho_k \\ \rho_k^c & 1 \end{bmatrix}, \quad (55)$$

where ρ_k and ρ_u are parameters. We satisfy the assumption (6), except in the scenario depicted in Fig. 4. The receiving array is consistently a ULA; its number of sensors $m = 10$ unless otherwise specified.

To evaluate the practical performance of the investigated estimators, we use the Root-Mean-Square Error (RMSE) as performance metric, which we define as

$$\text{RMSE}_i = \sqrt{\frac{1}{L} \sum_{l=1}^L \left(\hat{\theta}_i^{(l)} - \theta_{0,i} \right)^2}, \quad (56)$$

where the superscript (l) denotes the l th MC-realization, and L is the number of MC-realizations.

6.1 Performance comparison to existing state-of-art

In Fig. 1 we compare simulated performance of both POWDER and POWDER2 to “MODE”, [1], (or, equivalently, WSF [2]) which is not using prior knowledge on source DOA or correlation, and “PLEDGE” [8], which exploits the knowledge of ϑ . Two CRBs are displayed: “CRB_P”, [8], assuming that ϑ is known, and “CRB_{BD}”, given by (38).

Here, all sources are of the same power, thus $\text{SNR}_u = \text{SNR}_k = \text{SNR}$. As can be seen, in the studied scenario PLEDGE is marginally better than MODE for low

SNR; at higher SNR, both methods are equivalent, and efficient. However, it is also seen that POWDER is consistently very much more accurate, and sooner (in terms of SNR) efficient. Hence, a significant decrease in SNR is possible for a given accuracy. It can be concluded that at high SNR, knowing ϑ (without correlation knowledge) becomes irrelevant as MODE and PLEDGE give the same accuracy. POWDER, however, seems to consistently benefit from its prior knowledge.

The figure is also showing the validity of the derived expression (53) (which in the studied scenario is equivalent for POWDER and POWDER2).

6.2 Asymptotic variance validity

In Fig. 2 we investigate the significance of the assumptions of Theorem 3, and study the empirical performance in comparison to the expressions (53), (54) and (38). In the figure, “CRB” denotes the exact expression (38), and “CRB_{H_{SNR}” (42).}

In Fig. 2a, we have fixed the known sources’ power at $\text{SNR}_k = -10\text{dB}$. We can see in the figure that for large SNR_u , the respective bounds converge even though SNR_k is fixed.

In Fig. 2b we study the opposite scenario; we fix $\text{SNR}_u = 5\text{dB}$, and vary SNR_k . We can thus see that while the estimation scenario is more difficult for high SNR_k , the assumptions of Theorem 3 are eventually satisfied and the bounds converge. However, for the higher SNR_k , SNR_u is too small and the estimators break down.

Additionally, in Fig. 3 we study the scenario when all sources are uncorrelated, and $\text{SNR}_u = \text{SNR}_k = \text{SNR}$; it is then possible to use methods relying on the information that \mathbf{P} is diagonal. Hence in this figure we also include “PLEDGE UC”, [11], which is tailored for the scenario with known sources and diagonal \mathbf{P} . Additionally, we include the corresponding bound, “CRB_{PUC}”, [8]. We observe several interesting phenomena in Fig. 3: even though there is a strict ordering in the sense that $\text{VAR}_{\text{POW}} > \text{CRB} > \text{CRB}_{\text{PUC}}$ for $\sigma^2 > 0$, the difference is in this scenario practically undistinguishable. Further, POWDER has better low-SNR properties than PLEDGE UC (and POWDER2), and outperforms PLEDGE UC in the studied scenario even though POWDER does not exploit the fact that \mathbf{P} is strictly diagonal. POWDER consistently transitions to the theoretical variance limit at lower SNRs than POWDER2 (whose theoretical limit has been omitted from this figure for brevity; for practical purposes, the two limits are identical in the studied scenario).

6.3 Robustness to modeling errors

In this simulation scenario we study the effects of violating the assumption that the signals from the known and unknown DOAs are uncorrelated. Thus we let $\mathbf{P}_{uk} = \rho_{uk} \mathbf{1}_2 \mathbf{1}_2^T$, where we vary the parameter ρ_{uk} according to Fig. 4, and $\mathbf{1}_2$ denotes the vector of all ones of length 2. The SNR is fixed at 15dB for both

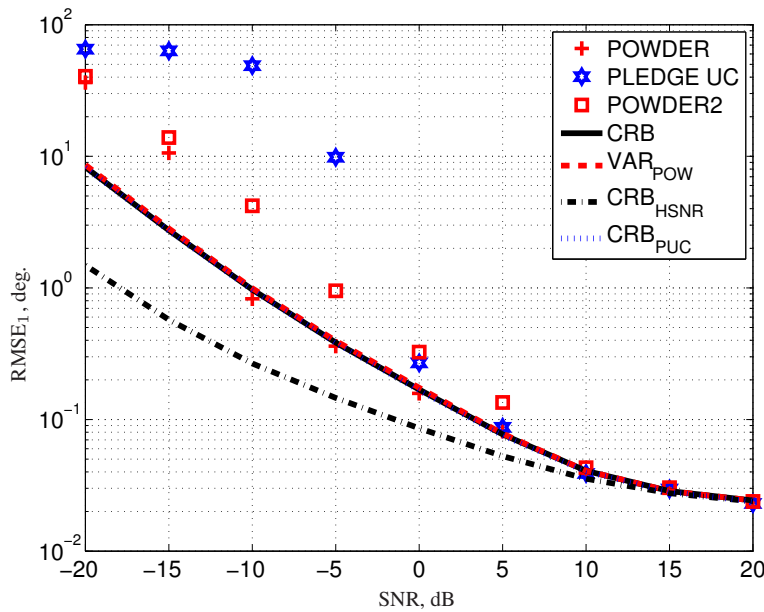


Figure 3: Comparison to existing state-of-the-art methods when $\rho_u = \rho_k = 0$. Normalized RMSE, averaged over 500 MC realizations, along with the derived theoretical bounds; $m = 15$ and $N = 1000$.

sets of DOAs, $\rho_u = \rho_k = 0.9$, and we plot against the number of obtained data samples N , in order to study finite sample effects. In addition to the RMSEs for the POWDER-methods for various choices of ρ_{uk} , we also show the CRB for the case $\mathbf{P}_{uk} = \mathbf{0}$, i.e. the ideal scenario the methods are designed for.

In Fig. 4 we can see that POWDER has better finite sample effects than POWDER2; in the studied scenario, it consistently attains the theoretical performance limit at smaller sample sizes N . Presumably, this effect is due to that the residual of POWDER2 is of larger dimension than that of POWDER; cf. (17) and (20).

From the figure, and by analyzing the individual MC-estimates, it can be seen that the modeling error introduces a bias in the estimates of the unknown DOAs. As ρ_{uk} grows larger, the methods increasingly suffer from not being able to distinguish between the known and unknown DOAs, cf. (10). Both methods appears to be fairly robust to the induced modeling error, in the sense that the estimators do not completely fail. However, as the error due to finite-samples decrease with increasing N , the error due to the induced bias limit the estimation accuracy.

The case of modeling errors in the known angles has been treated in e.g. [17]; a similar analysis of the herein studied estimator can be performed but is out of

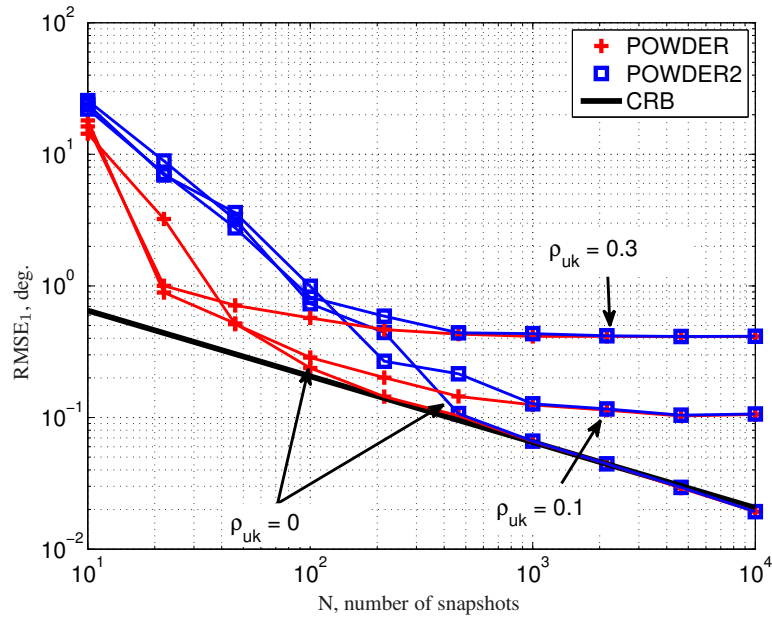


Figure 4: Showing the effects of violating the assumption $\mathbf{P}_{uk} = \mathbf{0}$, and the influence of finite sample effects. Examining three different values of the parameter ρ_{uk} . SNR fixed at 15dB, averages over 500 MC realizations.

scope of the present article.

7 Conclusions

In this article we have derived asymptotic expression for the accuracy of the recently proposed method [9] POWDER, and we have also extended the method into POWDER2. We have shown that, asymptotically in N and SNR, the methods are equivalent and statistically efficient. We also explored practical estimation results using Monte Carlo simulations, which showed the theoretically predicted results to be accurate, and we investigated robustness to modeling errors. Our simulations indicate that POWDER has better convergence properties for both SNR and N than POWDER2; we also showed that the estimators obtained their asymptotical (in N) accuracy already at moderate sample sizes and low SNR. Additionally, we saw a graceful convergence to the (proven) CRB-equivalent accuracy for moderate SNR for both methods.

Appendix A

First, derive an expression for $\mathbf{C}_\epsilon = \lim_{N \rightarrow \infty} N \mathbb{E} [\epsilon \epsilon^*]$. To do so, re-write (17) based on (16):

$$\begin{aligned} \epsilon &= \text{vec} \left(\hat{\mathbf{U}}_s^* \mathbf{B} \right) \\ &= (\mathbf{B}^T \otimes \boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^*) \text{vec} \left(\tilde{\mathbf{R}} - \tilde{\sigma}^2 \mathbf{I} \right) + o_p(1/\sqrt{N}) \end{aligned} \quad (\text{A.1})$$

From, e.g., [7]:

$$\begin{aligned} (m - d') \tilde{\sigma}^2 &= \text{Tr}(\hat{\boldsymbol{\Lambda}}_n) - \text{Tr}(\boldsymbol{\Lambda}_n) \\ &= \text{vec}^*(\mathbf{I}_m - \mathbf{E}_s \mathbf{E}_s^*) \text{vec}(\tilde{\mathbf{R}}) + o_p(1/\sqrt{N}) \\ &\triangleq \boldsymbol{\pi}^* \text{vec}(\tilde{\mathbf{R}}) + o_p(1/\sqrt{N}), \end{aligned} \quad (\text{A.2})$$

since $\mathbf{E}_s^* \mathbf{E}_s = \mathbf{I}_{d'}$, and $\boldsymbol{\pi}$ is defined in (30). Using (A.2) in (A.1), with the definitions (28) and (29), we have

$$\epsilon = \mathbf{M} \mathbf{F} \text{vec}(\tilde{\mathbf{R}}) + o_p(1/\sqrt{N}). \quad (\text{A.3})$$

From, e.g., [18], we have that

$$\mathbb{E}(\text{vec}(\tilde{\mathbf{R}}) \text{vec}^*(\tilde{\mathbf{R}})) = \frac{1}{N} (\mathbf{R}^T \otimes \mathbf{R}). \quad (\text{A.4})$$

Using (A.3) together with (A.4) gives

$$\lim_{N \rightarrow \infty} N \mathbb{E} [\epsilon \epsilon^*] = \mathbf{M} \mathbf{F} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{F}^* \mathbf{M}^*. \quad (\text{A.5})$$

Expand (A.5) to read

$$\begin{aligned} \mathbf{C}_\epsilon &= \left(\mathbf{M} - \frac{1}{m - d'} \mathbf{M} \mathbf{i} \boldsymbol{\pi}^* \right) (\mathbf{R}^T \otimes \mathbf{R}) \times \\ &\quad \left(\mathbf{M}^* - \frac{1}{m - d'} \boldsymbol{\pi} \mathbf{i}^* \mathbf{M}^* \right), \end{aligned} \quad (\text{A.6})$$

where $\mathbf{i} = \text{vec}(\mathbf{I}_m)$. Evaluating (A.6) we will have four terms. The first one is

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{M} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{M}^* \\ &= (\mathbf{B}^T \mathbf{R}^T \mathbf{B}^c \otimes \boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{R} \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1}), \end{aligned} \quad (\text{A.7})$$

which is guaranteed to be p.d. for $\sigma^2 > 0$ since the constituent matrices of each Kronecker factor then are p.d. The second term in (A.6) is given by

$$\mathbf{C}_2 = -\frac{1}{m - d'} \mathbf{M} \mathbf{i} \boldsymbol{\pi}^* (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{M}^*. \quad (\text{A.8})$$

Note that from (A.8)

$$\begin{aligned}\boldsymbol{\pi}^* (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{M}^* &= \text{vec}^* (\mathbf{R}(\mathbf{I} - \mathbf{E}_s \mathbf{E}_s^*) \mathbf{R}) \mathbf{M}^* \\ &= \text{vec}^* (\mathbf{E}_n \boldsymbol{\Lambda}_n^2 \mathbf{E}_n^*) \mathbf{M}^* \\ &= \sigma^4 \text{vec}^* (\boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{E}_n \mathbf{E}_n^* \mathbf{B}),\end{aligned}\quad (\text{A.9})$$

where the second equality follows from the EVD of \mathbf{R} and the third from (28). By noting that $\mathbf{M}\mathbf{i} = \text{vec}(\boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{B})$, we have that (A.8) evaluates to

$$\mathbf{C}_2 = -\frac{\sigma^4}{m-d'} \text{vec}(\boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{B}) \text{vec}^* (\boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{E}_n \mathbf{E}_n^* \mathbf{B}). \quad (\text{A.10})$$

Obviously, $\mathbf{C}_3 = \mathbf{C}_2^*$.

It is easily verified that $\boldsymbol{\pi}^* (\mathbf{R}^T \otimes \mathbf{R}) \boldsymbol{\pi} = \sigma^4 \text{Tr}(\mathbf{E}_n \mathbf{E}_n^*) = \sigma^4(m-d')$, which gives

$$\begin{aligned}\mathbf{C}_4 &= \frac{1}{(m-d')^2} \mathbf{M}\mathbf{i} \boldsymbol{\pi}^* (\mathbf{R}^T \otimes \mathbf{R}) \boldsymbol{\pi} \mathbf{i}^* \mathbf{M}^* \\ &= \frac{\sigma^4}{m-d'} \text{vec}(\boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{B}) \text{vec}^* (\boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{B}).\end{aligned}\quad (\text{A.11})$$

Hence we can write

$$\mathbf{C}_\epsilon = \mathbf{C}_1 + O(\sigma^4), \quad (\text{A.12})$$

which establishes (31).

To continue, we need the following Lemma:

LEMMA 1

For a permutation matrix \mathbf{L}_{mn} such that

$$\mathbf{L}_{mn} \text{vec}(\mathbf{X}_{m \times n}) = \text{vec}(\mathbf{X}_{m \times n}^T), \quad (\text{A.13})$$

it holds that

$$\mathbf{L}_{mn} (\mathbf{X}_{n \times d} \otimes \mathbf{Y}_{m \times p}) = (\mathbf{Y}_{m \times p} \otimes \mathbf{X}_{n \times d}) \mathbf{L}_{pd}. \quad (\text{A.14})$$

Proof: Write

$$\begin{aligned}\mathbf{L}_{mn} (\mathbf{X}_{n \times d} \otimes \mathbf{Y}_{m \times p}) \text{vec}(\mathbf{Z}_{p \times d}) &= \mathbf{L}_{mn} \text{vec}(\mathbf{Y} \mathbf{Z} \mathbf{X}^T) \\ &= \text{vec}(\mathbf{X} \mathbf{Z}^T \mathbf{Y}^T) \\ &= (\mathbf{Y} \otimes \mathbf{X}) \text{vec}(\mathbf{Z}^T) \\ &= (\mathbf{Y} \otimes \mathbf{X}) \mathbf{L}_{pd} \text{vec}(\mathbf{Z});\end{aligned}\quad (\text{A.15})$$

since \mathbf{Z} is arbitrary, (A.14) follows. \blacksquare

We will in the following omit the dimensional marker of these permutation matrices to simplify notation. Thus, if we write $\text{vec}^T(\tilde{\mathbf{R}}) = \text{vec}^*(\tilde{\mathbf{R}}^T) = \text{vec}^*(\tilde{\mathbf{R}})\mathbf{L}$ (since $\tilde{\mathbf{R}}$ is Hermitian), we have from (A.4) that

$$E(\text{vec}(\tilde{\mathbf{R}}) \text{vec}^T(\tilde{\mathbf{R}})) = N^{-1} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{L}; \quad (\text{A.16})$$

analogously to (A.5), we have

$$\mathbf{C}_{\tilde{\epsilon}} = \mathbf{M}\mathbf{F} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{L}\mathbf{F}^T\mathbf{M}^T \quad (\text{A.17})$$

$$= \mathbf{M} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{L}\mathbf{M}^T + O(\sigma^4), \quad (\text{A.18})$$

where the last equality follows from (29), (A.14), and (A.12). Now,

$$\mathbf{L}\mathbf{M}^T = \mathbf{L} (\mathbf{B} \otimes \mathbf{V}_s^c \boldsymbol{\Sigma}_s^{-1}) = (\mathbf{V}_s^c \boldsymbol{\Sigma}_s^{-1} \otimes \mathbf{B}) \mathbf{L}, \quad (\text{A.19})$$

again using (A.14). Using (A.19) in (A.18),

$$\begin{aligned} \mathbf{M} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{L}\mathbf{M}^T &= (\mathbf{B}^T \mathbf{R}^T \mathbf{V}_s^c \boldsymbol{\Sigma}_s^{-1} \otimes \boldsymbol{\Sigma}_s^{-1} \mathbf{V}_s^* \mathbf{R}\mathbf{B}) \mathbf{L} \\ &= O(\sigma^4), \end{aligned} \quad (\text{A.20})$$

since $\mathbf{V}_s^* \mathbf{R}\mathbf{B} = \sigma^2 \mathbf{V}_s^* \mathbf{B}$, which is due to $\mathbf{B}^* \mathbf{A}_u = \mathbf{0}$ and $\mathbf{V}_s = \boldsymbol{\Pi}_{\mathbf{A}_k}^\perp \mathbf{V}_s$. Hence,

$$\mathbf{C}_{\tilde{\epsilon}} = O(\sigma^4). \quad (\text{A.21})$$

Appendix B

This proof follows the methodology of [19] (see also [7]), in which the general scenario without prior knowledge of $\boldsymbol{\vartheta}$ or $\mathbf{P}_{uk} = \mathbf{0}$ was studied. The Fisher Information Matrix (FIM) for the parameter vector $\boldsymbol{\alpha}$ is given by

$$\frac{1}{N} \text{FIM} = \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\alpha}^T} \right)^* (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}) \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\alpha}^T} \right), \quad (\text{B.1})$$

where

$$\mathbf{r} = \text{vec}(\mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^*) + \text{vec}(\mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^*) + \sigma^2 \text{vec}(\mathbf{I}), \quad (\text{B.2})$$

due to \mathbf{P} being block-diagonal. If we let

$$[\mathbf{G} \mid \boldsymbol{\Delta}] \triangleq \mathbf{C}_r^{1/2} \left[\begin{array}{c|cc} \frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^T} & \frac{\partial \mathbf{r}}{\partial \boldsymbol{\rho}^T} & \frac{\partial \mathbf{r}}{\partial \sigma^2} \end{array} \right], \quad (\text{B.3})$$

we can, by comparing (B.3) to (B.1), write

$$\frac{1}{N} \text{FIM} = \begin{bmatrix} \mathbf{G}^* \\ \boldsymbol{\Delta}^* \end{bmatrix} [\mathbf{G} \mid \boldsymbol{\Delta}]. \quad (\text{B.4})$$

The respective elements of (B.3) are found as (for $i = 1, \dots, d_u$)

$$\frac{\partial \mathbf{r}}{\partial \theta_i} = (\mathbf{A}_u^c \mathbf{P}_u^c \otimes \mathbf{I}) \bar{\mathbf{d}}_i + (\mathbf{I} \otimes \mathbf{A}_u \mathbf{P}_u) \mathbf{L} \bar{\mathbf{d}}_i^c, \quad (\text{B.5})$$

where

$$\bar{\mathbf{d}}_i = \frac{\partial \text{vec}(\mathbf{A}_u)}{\partial \theta_i}, \quad (\text{B.6})$$

and \mathbf{L} follows from Lemma 1. Collecting $\bar{\mathbf{D}} = [\bar{\mathbf{d}}_1 \ \dots \ \bar{\mathbf{d}}_{d_u}]$, we have that

$$\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^T} = (\mathbf{A}_u^c \mathbf{P}_u^c \otimes \mathbf{I}) \bar{\mathbf{D}} + (\mathbf{I} \otimes \mathbf{A}_u \mathbf{P}_u) \mathbf{L}_1 \bar{\mathbf{D}}^c. \quad (\text{B.7})$$

Continuing with the other terms of (B.3),

$$\frac{\partial \mathbf{r}}{\partial \boldsymbol{\rho}^T} = [(\mathbf{A}_k^c \otimes \mathbf{A}_k) \mathbf{J}_k \quad (\mathbf{A}_u^c \otimes \mathbf{A}_u) \mathbf{J}_u], \quad (\text{B.8})$$

where \mathbf{J}_k is the selection matrix such that $\text{vec}(\mathbf{P}_k) = \mathbf{J}_k \boldsymbol{\rho}_k$, and \mathbf{J}_u is defined from $\text{vec}(\mathbf{P}_u) = \mathbf{J}_u \boldsymbol{\rho}_u$. Finally, $\partial \mathbf{r} / \partial \sigma^2 = \text{vec}(\mathbf{I})$.

Since the interest is in a lower bound for the DOAs $\boldsymbol{\theta}$ (and not a bound for all the parameters $\boldsymbol{\alpha}$), a standard result on block-matrix inversion on (B.4) gives

$$\text{CRB}_{\boldsymbol{\theta}} = \frac{1}{N} \left(\mathbf{G}^* \boldsymbol{\Pi}_{\bar{\Delta}}^{\perp} \mathbf{G} \right)^{-1}, \quad (\text{B.9})$$

where we have also exploited that since \mathbf{J}_k and \mathbf{J}_u are invertible, $\text{Span}(\Delta) = \text{Span}(\bar{\Delta})$ (and $\bar{\Delta}$ is defined in (40)).

Appendix C

Start by partitioning $\bar{\Delta} = [\bar{\mathbf{v}} \ \mathbf{r}_i]$, where we define

$$\begin{aligned} \bar{\mathbf{v}} &\triangleq [\bar{\mathbf{A}}_k \quad \bar{\mathbf{A}}_u], \\ \bar{\mathbf{A}}_k &\triangleq \mathbf{C}_r^{1/2} (\mathbf{A}_k^c \otimes \mathbf{A}_k), \\ \bar{\mathbf{A}}_u &\triangleq \mathbf{C}_r^{1/2} (\mathbf{A}_u^c \otimes \mathbf{A}_u), \\ \mathbf{r}_i &\triangleq \text{vec}(\mathbf{R}^{-1}). \end{aligned}$$

Then it is possible to decompose the projection operator $\boldsymbol{\Pi}_{\bar{\Delta}}^{\perp}$ (see e.g. [20]) according to

$$\boldsymbol{\Pi}_{\bar{\Delta}}^{\perp} = \boldsymbol{\Pi}_{\bar{\mathbf{v}}}^{\perp} - \frac{\boldsymbol{\Pi}_{\bar{\mathbf{v}}}^{\perp} \mathbf{r}_i \mathbf{r}_i^* \boldsymbol{\Pi}_{\bar{\mathbf{v}}}^{\perp}}{\mathbf{r}_i^* \boldsymbol{\Pi}_{\bar{\mathbf{v}}}^{\perp} \mathbf{r}_i}. \quad (\text{C.1})$$

Similarly, we decompose $\Pi_{\bar{\mathbf{v}}}$ in (C.1) according to

$$\Pi_{\bar{\mathbf{v}}} = \Pi_{\bar{\mathbf{A}}_u} + \Pi_{\bar{\mathbf{A}}_u}^\perp \bar{\mathbf{A}}_k \left(\bar{\mathbf{A}}_k^* \Pi_{\bar{\mathbf{A}}_u}^\perp \bar{\mathbf{A}}_k \right)^{-1} \bar{\mathbf{A}}_k^* \Pi_{\bar{\mathbf{A}}_u}^\perp. \quad (\text{C.2})$$

Introduce

$$\mathbf{R}_u \triangleq \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* + \sigma^2 \mathbf{I}; \quad (\text{C.3})$$

with the assumption that \mathbf{P}_k in (5) is p.d. (together with (6)), it is straight-forward to write the inverse of \mathbf{R} as

$$\mathbf{R}^{-1} = \mathbf{R}_u^{-1} - \mathbf{R}_u^{-1} \mathbf{A}_k \left(\mathbf{P}_k^{-1} + \mathbf{A}_k^* \mathbf{R}_u^{-1} \mathbf{A}_k \right)^{-1} \mathbf{A}_k^* \mathbf{R}_u^{-1}. \quad (\text{C.4})$$

If we then exploit that $\lambda_{\min}^{-1}(\mathbf{P}_k) \rightarrow 0$ (i.e., the signals from the known directions are strong), we immediately have from (C.4) that

$$\mathbf{A}_k^* \mathbf{R}^{-1} = O(\|\mathbf{P}_k^{-1}\|), \quad (\text{C.5})$$

due to the expansion

$$(\mathbf{X} + \mathbf{Y})^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1} \mathbf{Y} \mathbf{X}^{-1} + O(\|\mathbf{Y}^2\|) \quad (\text{C.6})$$

for small \mathbf{Y} . The result of (C.5) allows us to write

$$\bar{\mathbf{A}}_k^* \Pi_{\bar{\mathbf{A}}_u}^\perp \bar{\mathbf{A}}_k = \bar{\mathbf{A}}_k^* \bar{\mathbf{A}}_k + O(\|\mathbf{P}_k^{-4}\|), \quad (\text{C.7})$$

which in turn, with (C.6), lets us rewrite (C.2) as

$$\Pi_{\bar{\mathbf{v}}} = \Pi_{\bar{\mathbf{A}}_u} + \Pi_{\bar{\mathbf{A}}_u}^\perp \Pi_{\bar{\mathbf{A}}_k} \Pi_{\bar{\mathbf{A}}_u}^\perp + O(\|\mathbf{P}_k^{-2}\|). \quad (\text{C.8})$$

From (B.3),

$$\mathbf{G}^* \Pi_{\bar{\Delta}}^\perp \mathbf{G} = \left(\frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^\top} \right)^* \mathbf{C}_r^{*/2} \Pi_{\bar{\Delta}}^\perp \mathbf{C}_r^{1/2} \frac{\partial \mathbf{r}}{\partial \boldsymbol{\theta}^\top}; \quad (\text{C.9})$$

together with (C.1), we have that

$$\begin{aligned} \mathbf{C}_r^{*/2} \Pi_{\bar{\Delta}}^\perp \mathbf{C}_r^{1/2} = \\ \mathbf{C}_r^{*/2} \Pi_{\bar{\mathbf{v}}}^\perp \mathbf{C}_r^{1/2} - \mathbf{C}_r^{*/2} \frac{\Pi_{\bar{\mathbf{v}}}^\perp \mathbf{r}_i \mathbf{r}_i^* \Pi_{\bar{\mathbf{v}}}^\perp}{\mathbf{r}_i^* \Pi_{\bar{\mathbf{v}}}^\perp \mathbf{r}_i} \mathbf{C}_r^{1/2}. \end{aligned} \quad (\text{C.10})$$

Due to (C.8) and (C.5), we have that

$$\mathbf{C}_r^{*/2} \Pi_{\bar{\mathbf{v}}}^\perp \mathbf{C}_r^{1/2} = \mathbf{C}_r^{*/2} \Pi_{\bar{\mathbf{A}}_u}^\perp \mathbf{C}_r^{1/2} + O(\|\mathbf{P}_k^{-2}\|). \quad (\text{C.11})$$

By a similar argument, the second term in (C.10) becomes

$$\begin{aligned} \mathbf{C}_r^{*/2} \frac{\Pi_{\mathbf{v}}^\perp \mathbf{r}_i \mathbf{r}_i^* \Pi_{\mathbf{v}}^\perp}{\mathbf{r}_i^* \Pi_{\mathbf{v}}^\perp \mathbf{r}_i} \mathbf{C}_r^{1/2} &= \mathbf{C}_r^{*/2} \frac{\Pi_{\mathbf{A}_u}^\perp \mathbf{r}_i \mathbf{r}_i^* \Pi_{\mathbf{A}_u}^\perp}{\mathbf{r}_i^* \Pi_{\mathbf{A}_u}^\perp \mathbf{r}_i} \mathbf{C}_r^{1/2} \\ &+ O(\|\mathbf{P}_k^{-2}\|). \end{aligned} \quad (\text{C.12})$$

Using (C.11) and (C.12) in (C.9), together with (B.7), gives

$$\begin{aligned} \mathbf{G}^* \Pi_{\Delta}^\perp \mathbf{G} &= 2 \operatorname{Re} \left(\bar{\mathbf{D}}^* \left(\mathbf{P}_u^T \mathbf{A}_{uw}^T \otimes \mathbf{R}^{-1/2} \right) \right. \\ &\times \left(\Pi_{\mathbf{A}_u}^\perp - \frac{\Pi_{\mathbf{A}_u}^\perp \mathbf{r}_i \mathbf{r}_i^* \Pi_{\mathbf{A}_u}^\perp}{\mathbf{r}_i^* \Pi_{\mathbf{A}_u}^\perp \mathbf{r}_i} \right) \\ &\times \left[\left(\mathbf{A}_{uw}^c \mathbf{P}_u^c \otimes \mathbf{R}^{-1/2} \right) \bar{\mathbf{D}} \right. \\ &\left. \left. + \left(\mathbf{R}^{-T/2} \otimes \mathbf{A}_{uw} \mathbf{P}_u \right) \mathbf{L} \bar{\mathbf{D}}^c \right] \right) + O(\|\mathbf{P}_k^{-2}\|), \end{aligned} \quad (\text{C.13})$$

where we have introduced

$$\mathbf{A}_{uw} \triangleq \mathbf{R}^{-1/2} \mathbf{A}_u. \quad (\text{C.14})$$

Note that in (C.13),

$$\begin{aligned} &\left(\mathbf{P}_u^T \mathbf{A}_{uw}^T \otimes \mathbf{R}^{-1/2} \right) \left(\Pi_{\mathbf{A}_u}^\perp \mathbf{r}_i \right) \\ &= \left(\mathbf{P}_u^T \mathbf{A}_{uw}^T \otimes \mathbf{R}^{-1/2} \right) \left(\mathbf{r}_i - \left(\Pi_{\mathbf{A}_{uw}}^c \otimes \Pi_{\mathbf{A}_{uw}} \right) \mathbf{r}_i \right) \\ &= \operatorname{vec} \left(\mathbf{R}^{-1/2} \mathbf{R}^{-1} \mathbf{A}_{uw} \mathbf{P}_u \right) - \operatorname{vec} \left(\mathbf{R}^{-1/2} \Pi_{\mathbf{A}_{uw}} \mathbf{R}^{-1} \mathbf{A}_{uw} \mathbf{P}_u \right) \\ &= \operatorname{vec} \left(\mathbf{R}^{-1/2} \Pi_{\mathbf{A}_{uw}}^\perp \mathbf{R}^{-1} \mathbf{A}_{uw} \mathbf{P}_u \right) = O(\|\mathbf{P}_k^{-1}\|). \end{aligned} \quad (\text{C.15})$$

The last equality in (C.15) is proven in the following: we have that

$$\begin{aligned} \mathbf{R}^{-1} \mathbf{A}_{uw} &= \mathbf{R}^{-1/2} \mathbf{R}^{-1} \mathbf{A}_u \\ &= \mathbf{R}^{-1/2} \left(\mathbf{R}_u^{-1} \mathbf{A}_u - \mathbf{R}_u^{-1} \mathbf{A}_k \left(\mathbf{A}_k^* \mathbf{R}_u^{-1} \mathbf{A}_k \right)^{-1} \right. \\ &\quad \left. \times \mathbf{A}_k^* \mathbf{R}_u^{-1} \mathbf{A}_u \right) + O(\|\mathbf{P}_k^{-1}\|), \end{aligned} \quad (\text{C.16})$$

where the last equality follows from (C.4) and (C.5). Using

$$\mathbf{R}_u^{-1} = \sigma^{-2} \left(\mathbf{I} - \mathbf{A}_u \left(\mathbf{P}_u \mathbf{A}_u^* \mathbf{A}_u + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{P}_u \mathbf{A}_u^* \right), \quad (\text{C.17})$$

we can re-write the first term in (C.16):

$$\begin{aligned}\mathbf{R}^{-1/2}\mathbf{R}_u^{-1}\mathbf{A}_u &= \mathbf{R}^{-1/2}\mathbf{A}_u \left(\mathbf{P}_u\mathbf{A}_u^*\mathbf{A}_u + \sigma^2\mathbf{I}\right)^{-1} \\ &= \mathbf{A}_{uw} \left(\mathbf{P}_u\mathbf{A}_u^*\mathbf{A}_u + \sigma^2\mathbf{I}\right)^{-1}.\end{aligned}\quad (\text{C.18})$$

Define $\Xi \triangleq (\mathbf{A}_k^*\mathbf{R}_u^{-1}\mathbf{A}_k)^{-1}\mathbf{A}_k^*$ to shorten notation; then examine the second term in (C.16) and exploit (C.17):

$$\begin{aligned}&-\mathbf{R}^{-1/2}\mathbf{R}_u^{-1}\mathbf{A}_k\Xi\mathbf{R}_u^{-1}\mathbf{A}_u \\ &= -\sigma^{-2}\mathbf{R}^{-1/2}\left(\mathbf{I} - \mathbf{A}_u \left(\mathbf{P}_u\mathbf{A}_u^*\mathbf{A}_u + \sigma^2\mathbf{I}\right)^{-1}\mathbf{P}_u\mathbf{A}_u^*\right) \\ &\quad \times \mathbf{A}_k\Xi\mathbf{R}_u^{-1}\mathbf{A}_u \\ &= -\sigma^{-2}\mathbf{R}^{-1/2}\mathbf{A}_k\Xi\mathbf{R}_u^{-1}\mathbf{A}_u + \mathbf{A}_{uw}\Xi_2,\end{aligned}\quad (\text{C.19})$$

where $\Xi_2 \triangleq \sigma^{-2}\left(\mathbf{P}_u\mathbf{A}_u^*\mathbf{A}_u + \sigma^2\mathbf{I}\right)^{-1}\mathbf{P}_u\mathbf{A}_u^*\mathbf{A}_k\Xi\mathbf{R}_u^{-1}\mathbf{A}_u$. Using (C.18) and (C.19) in (C.16), we have

$$\mathbf{R}^{-1}\mathbf{A}_{uw} = \mathbf{A}_{uw}\Xi_3 - \mathbf{R}^{-1/2}\mathbf{A}_k\Xi_4, \quad (\text{C.20})$$

where $\Xi_3 \triangleq \Xi_2 + \left(\mathbf{P}_u\mathbf{A}_u^*\mathbf{A}_u + \sigma^2\mathbf{I}\right)^{-1}$ and $\Xi_4 \triangleq \sigma^{-2}\Xi\mathbf{R}_u^{-1}\mathbf{A}_u$. Now compare (C.20) to (C.15): the first term of (C.20) vanishes due to the orthogonal projector in (C.15). The second term inserted into (C.15) gives

$$\begin{aligned}&\text{vec}\left(\mathbf{R}^{-1/2}\Pi_{\mathbf{A}_{uw}}^\perp\mathbf{R}^{-1}\mathbf{A}_{uw}\mathbf{P}_u\right) \\ &= -\text{vec}\left(\mathbf{R}^{-1}\mathbf{A}_k\Xi_4\mathbf{P}_u - \mathbf{R}^{-1/2}\Pi_{\mathbf{A}_{uw}}\mathbf{R}^{-1/2}\mathbf{A}_k\Xi_4\mathbf{P}_u\right) \\ &= O(\|\mathbf{P}_k^{-1}\|),\end{aligned}\quad (\text{C.21})$$

due to (C.5).

Hence, we can write (C.13) as

$$\begin{aligned}\mathbf{G}^*\Pi_{\Delta}^\perp\mathbf{G} &= 2\text{Re}\left(\bar{\mathbf{D}}^*\left(\mathbf{P}_u^T\mathbf{A}_{uw}^T \otimes \mathbf{R}^{-1/2}\right)\right. \\ &\quad \times \Pi_{\mathbf{A}_u}^\perp\left[\left(\mathbf{A}_{uw}^c\mathbf{P}_u^c \otimes \mathbf{R}^{-1/2}\right)\bar{\mathbf{D}}\right. \\ &\quad \left.\left. + \left(\mathbf{R}^{-T/2} \otimes \mathbf{A}_{uw}\mathbf{P}_u\right)\mathbf{L}\bar{\mathbf{D}}^c\right]\right) + O(\|\mathbf{P}_k^{-2}\|) \\ &= 2\text{Re}\left(\bar{\mathbf{D}}^*\left[\mathbf{P}_u^T\mathbf{A}_{uw}^T\mathbf{A}_{uw}^c\mathbf{P}_u^c \otimes\right.\right. \\ &\quad \left.\left.\mathbf{R}^{-1/2}\Pi_{\mathbf{A}_{uw}}^\perp\mathbf{R}^{-1/2}\right]\bar{\mathbf{D}}\right) + O(\|\mathbf{P}_k^{-2}\|),\end{aligned}\quad (\text{C.22})$$

where the last equality follows due to $\mathbf{\Pi}_{\mathbf{A}_u}^\perp$. Additionally, if we let

$$\mathbf{D} = \begin{bmatrix} \frac{\partial \mathbf{a}(\theta_1)}{\partial \theta_1} & \frac{\partial \mathbf{a}(\theta_2)}{\partial \theta_2} & \cdots & \frac{\partial \mathbf{a}(\theta_{d_u})}{\partial \theta_{d_u}} \end{bmatrix}, \quad (\text{C.23})$$

and take the inverse of (C.22), we can write

$$\begin{aligned} \text{CRB}_\theta &= \frac{1}{2N} \left[\text{Re} \left(\left(\mathbf{D}^* \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_{uw}}^\perp \mathbf{R}^{-1/2} \mathbf{D} \right) \right. \right. \\ &\quad \left. \left. \odot \left(\mathbf{P}_u^T \mathbf{A}_{uw}^T \mathbf{A}_{uw}^c \mathbf{P}_u^c \right) \right) \right]^{-1} \\ &\quad + \frac{1}{N} O(\|\mathbf{P}_k^{-2}\|), \end{aligned} \quad (\text{C.24})$$

where \odot denotes element-wise multiplication.

Appendix D

V' in (51) is found by looking at the derivatives of V with respect to the individual elements θ_i of θ :

$$\begin{aligned} V'_i &= \frac{\partial V}{\partial \theta_i} = \boldsymbol{\epsilon}_i^* \mathbf{W} \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^* \mathbf{W} \boldsymbol{\epsilon}_i \\ &= \boldsymbol{\epsilon}_i^* \mathbf{W} \boldsymbol{\epsilon} + \boldsymbol{\epsilon}_i^T \mathbf{W}^T \boldsymbol{\epsilon}^c \triangleq \bar{\boldsymbol{\epsilon}}_i^* \bar{\mathbf{W}} \bar{\boldsymbol{\epsilon}}, \end{aligned} \quad (\text{D.1})$$

where

$$\boldsymbol{\epsilon}_i = \frac{\partial \boldsymbol{\epsilon}}{\partial \theta_i}, \quad \bar{\boldsymbol{\epsilon}}_i = \begin{bmatrix} \boldsymbol{\epsilon}_i \\ \boldsymbol{\epsilon}_i^c \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{W}} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^T \end{bmatrix}. \quad (\text{D.2})$$

From (17),

$$\frac{\partial \boldsymbol{\epsilon}}{\partial \theta_i} = \text{vec} \left(\hat{\mathbf{U}}_s^* \frac{\partial \mathbf{B}}{\partial \theta_i} \right) = \text{vec} \left(\mathbf{U}_s^* \frac{\partial \mathbf{B}}{\partial \theta_i} \right) + o_p(1). \quad (\text{D.3})$$

Note that

$$\mathbf{U}_s^* \frac{\partial \mathbf{B}}{\partial \theta_i} = \mathbf{T}^* \mathbf{A}_u^* \frac{\partial \mathbf{B}}{\partial \theta_i} = -\mathbf{T}^* \frac{\partial \mathbf{A}_u^*}{\partial \theta_i} \mathbf{B}, \quad (\text{D.4})$$

where \mathbf{T} is full rank and defined from

$$\mathbf{U}_s = \mathbf{A}_u \mathbf{T}, \quad (\text{D.5})$$

and the third equality follows from the fact that $\mathbf{B}^* \mathbf{A}_u = \mathbf{0}$. Using (D.4) together with (B.6), we see

$$\text{vec} \left(\mathbf{U}_s^* \frac{\partial \mathbf{B}}{\partial \theta_i} \right) = - (\mathbf{B}^T \otimes \mathbf{T}^*) \mathbf{L} \bar{\mathbf{d}}_i^c. \quad (\text{D.6})$$

Then we have

$$\frac{\partial \boldsymbol{\epsilon}}{\partial \boldsymbol{\theta}^T} = - (\mathbf{B}^T \otimes \mathbf{T}^*) \mathbf{L} \bar{\mathbf{D}}^c + o_p(1), \quad (\text{D.7})$$

and if we additionally introduce $\bar{\mathbf{K}} = \begin{bmatrix} \mathbf{K} \\ \mathbf{K}^c \end{bmatrix}$, where

$$\mathbf{K} \triangleq - (\mathbf{B}^T \otimes \mathbf{T}^*) \mathbf{L} \bar{\mathbf{D}}^c, \quad (\text{D.8})$$

(and hence \mathbf{k}_i , the i th column of \mathbf{K} , is given by (D.6)), we can compactly write

$$\frac{\partial V}{\partial \boldsymbol{\theta}} = \bar{\mathbf{K}}^* \bar{\mathbf{W}} \bar{\boldsymbol{\epsilon}} + o_p(1/\sqrt{N}). \quad (\text{D.9})$$

With the definition (52), we use (D.9) to state the expression for \mathbf{Q} :

$$\mathbf{Q} = \bar{\mathbf{K}}^* \bar{\mathbf{W}} \mathbf{C}_{\bar{\boldsymbol{\epsilon}}} \bar{\mathbf{W}}^* \bar{\mathbf{K}}. \quad (\text{D.10})$$

In (D.10), we have defined

$$\mathbf{C}_{\bar{\boldsymbol{\epsilon}}} \triangleq \begin{bmatrix} \mathbf{C}_{\boldsymbol{\epsilon}} & \mathbf{C}_{\bar{\boldsymbol{\epsilon}}} \\ \mathbf{C}_{\boldsymbol{\epsilon}}^c & \mathbf{C}_{\boldsymbol{\epsilon}}^T \end{bmatrix}, \quad (\text{D.11})$$

in which $\mathbf{C}_{\boldsymbol{\epsilon}}$ is given by (31), and $\mathbf{C}_{\bar{\boldsymbol{\epsilon}}}$ by (32).

From (50), and using the expressions (D.6) and (D.9),

$$\begin{aligned} H_{ij} &= \lim_{N \rightarrow \infty} \begin{bmatrix} \frac{\partial \mathbf{k}_i^*}{\partial \theta_j} & \frac{\partial \mathbf{k}_i^T}{\partial \theta_j} \end{bmatrix} \bar{\mathbf{W}} \boldsymbol{\epsilon} + \begin{bmatrix} \mathbf{k}_i^* & \mathbf{k}_i^T \end{bmatrix} \bar{\mathbf{W}} \begin{bmatrix} \mathbf{k}_j \\ \mathbf{k}_j^c \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{k}_i^* & \mathbf{k}_i^T \end{bmatrix} \bar{\mathbf{W}} \begin{bmatrix} \mathbf{k}_j \\ \mathbf{k}_j^c \end{bmatrix} \end{aligned} \quad (\text{D.12})$$

since $\lim_{N \rightarrow \infty} \boldsymbol{\epsilon} = \mathbf{0}$, and hence

$$\mathbf{H} = \bar{\mathbf{K}}^* \bar{\mathbf{W}} \bar{\mathbf{K}}. \quad (\text{D.13})$$

Appendix E

Define $\widetilde{\mathbf{W}}_1^{\text{OPT}} \triangleq \sigma^2 \check{\mathbf{C}}_\epsilon^{-1}$, such that $\widetilde{\mathbf{W}}_1^{\text{OPT}}$ exists for small σ^2 ; then, due to Theorem 1, we can write (34)

$$\mathbf{W}_1^{\text{OPT}} = \widetilde{\mathbf{W}}_1^{\text{OPT}} + O(\sigma^2) \quad (\text{E.1})$$

since $\check{\mathbf{C}}_\epsilon^{-1} = O(\sigma^{-2})$. Similarly,

$$\mathbf{W}_2^{\text{OPT}} = \widetilde{\mathbf{W}}_2^{\text{OPT}} + O(\sigma^2), \quad (\text{E.2})$$

where $\widetilde{\mathbf{W}}_2^{\text{OPT}} = \sigma^2 \begin{bmatrix} \check{\mathbf{C}}_\epsilon^{-1} & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{C}}_\epsilon^{-\text{T}} \end{bmatrix}$. Capitalizing on these facts, we can rewrite (D.13) according to

$$\sigma^2 \mathbf{H}_{\text{OPT}} = \bar{\mathbf{H}}_{\text{OPT}} + O(\sigma^2), \quad (\text{E.3})$$

where

$$\bar{\mathbf{H}}_{\text{OPT}} \triangleq \bar{\mathbf{K}}^* \widetilde{\mathbf{W}}_2^{\text{OPT}} \bar{\mathbf{K}}; \quad (\text{E.4})$$

note that (E.3) is valid for *both* target functions (18) and (21), if utilizing the respective optimal weighting matrix. From (E.4), it follows that

$$\begin{aligned} \bar{\mathbf{H}}_{\text{OPT}} &= \sigma^2 \mathbf{K}^* \check{\mathbf{C}}_\epsilon^{-1} \mathbf{K} + \sigma^2 \mathbf{K}^{\text{T}} \check{\mathbf{C}}_\epsilon^{-\text{c}} \mathbf{K}^{\text{c}} \\ &= 2\sigma^2 \text{Re} \left(\mathbf{K}^* \check{\mathbf{C}}_\epsilon^{-1} \mathbf{K} \right). \end{aligned} \quad (\text{E.5})$$

Using (D.8), (33), and (A.14), (E.5) can be simplified to read

$$\begin{aligned} \bar{\mathbf{H}}_{\text{OPT}} &= 2\sigma^2 \text{Re} \left(\bar{\mathbf{D}}^* \left[(\mathbf{T}^{\text{c}} \boldsymbol{\Sigma}_s^{\text{c}} (\mathbf{V}_s^{\text{T}} \mathbf{R}^{\text{T}} \mathbf{V}_s^{\text{c}})^{-1} \boldsymbol{\Sigma}_s^{\text{c}} \mathbf{T}^{\text{T}}) \right. \right. \\ &\quad \left. \left. \otimes (\mathbf{B} (\mathbf{B}^* \mathbf{R} \mathbf{B})^{-1} \mathbf{B}^*) \right] \bar{\mathbf{D}} \right). \end{aligned} \quad (\text{E.6})$$

From (45) and (10), $\mathbf{T} = \mathbf{A}_u^\dagger \mathbf{U}_s = \mathbf{P}_u \mathbf{A}_u^* \mathbf{V}_s \boldsymbol{\Sigma}_s^{-1}$; using this result together with the observation that $\mathbf{V}_s^* \mathbf{R} \mathbf{V}_s = \mathbf{V}_s^* \mathbf{R}_u \mathbf{V}_s$, we have that the first Kronecker factor in (E.6) can be written as

$$\begin{aligned} &\left(\mathbf{T} \boldsymbol{\Sigma}_s (\mathbf{V}_s^* \mathbf{R} \mathbf{V}_s)^{-1} \boldsymbol{\Sigma}_s \mathbf{T}^* \right)^{\text{c}} \\ &= \left(\mathbf{P}_u \mathbf{A}_u^* \mathbf{V}_s (\mathbf{V}_s^* \mathbf{R}_u \mathbf{V}_s)^{-1} \mathbf{V}_s^* \mathbf{P}_u \mathbf{A}_u \right)^{\text{c}}. \end{aligned} \quad (\text{E.7})$$

In (E.7),

$$(\mathbf{V}_s^* \mathbf{R}_u \mathbf{V}_s)^{-1} = (\mathbf{V}_s^* \mathbf{A}_u \mathbf{P}_u \mathbf{A}_u^* \mathbf{V}_s + \sigma^2 \mathbf{I})^{-1}. \quad (\text{E.8})$$

Note that we can decompose $\mathbf{P}_u = \mathbf{S}_u \mathbf{S}_u^*$, where $\mathbf{S}_u \in \mathbb{C}^{d_u \times d'_u}$ is full rank. From the definition of \mathbf{V}_s , (10), we have $\text{Span}(\mathbf{V}_s) = \text{Span}(\mathbf{\Pi}_{\mathbf{A}_k}^\perp \mathbf{A}_u \mathbf{S}_u)$. Thus, since $\mathbf{V}_s^* \mathbf{\Pi}_{\mathbf{A}_k}^\perp = \mathbf{V}_s^*$, $\mathbf{V}_s^* \mathbf{\Pi}_{\mathbf{A}_k}^\perp \mathbf{A}_u \mathbf{S}_u = \mathbf{V}_s^* \mathbf{A}_u \mathbf{S}_u$ is non-singular. Using these results and (E.8) in (E.7), we have

$$\begin{aligned} & \left(\mathbf{T} \boldsymbol{\Sigma}_s (\mathbf{V}_s^* \mathbf{R} \mathbf{V}_s)^{-1} \boldsymbol{\Sigma}_s \mathbf{T}^* \right)^c \\ &= (\mathbf{S}_u \mathbf{S}_u^*)^c + O(\sigma^2) = \mathbf{P}_u^c + O(\sigma^2). \end{aligned} \quad (\text{E.9})$$

By introducing $\mathbf{B}_w \triangleq \mathbf{R}^{1/2} \mathbf{B}$, we can rewrite the second Kronecker factor of (E.6) according to

$$\begin{aligned} \mathbf{B} (\mathbf{B}^* \mathbf{R} \mathbf{B})^{-1} \mathbf{B}^* &= \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{B}_w} \mathbf{R}^{-1/2} \\ &= \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_{uw}}^\perp \mathbf{R}^{-1/2}, \end{aligned} \quad (\text{E.10})$$

since \mathbf{B}_w spans the null-space of \mathbf{A}_{uw}^* .

Now using (E.9) and (E.10) in (E.6), and re-writing using (C.23),

$$\begin{aligned} \bar{\mathbf{H}}_{\text{OPT}} &= 2\sigma^2 \text{Re} \left(\left(\mathbf{D}^* \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_{uw}}^\perp \mathbf{R}^{-1/2} \mathbf{D} \right) \odot \mathbf{P}_u^c \right) \\ &\quad + O(\sigma^2). \end{aligned} \quad (\text{E.11})$$

Since $\mathbf{C}_{\text{POW}} = \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1}$, we have from the above derivations and (D.10) that

$$\begin{aligned} \mathbf{C}_{\text{POW}} &= \sigma^2 (\sigma^2 \mathbf{H}_{\text{OPT}})^{-1} = \sigma^2 (\bar{\mathbf{H}}_{\text{OPT}} + O(\sigma^2))^{-1} \\ &= \frac{1}{2} \left[\text{Re} \left(\left(\mathbf{D}^* \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_{uw}}^\perp \mathbf{R}^{-1/2} \mathbf{D} \right) \odot \mathbf{P}_u^c \right) \right]^{-1} \\ &\quad + O(\sigma^4). \end{aligned} \quad (\text{E.12})$$

We now examine the second factor of the Schur-product in (42), $\mathbf{P}_u \mathbf{A}_u^* \mathbf{R}^{-1} \mathbf{A}_u \mathbf{P}_u$. Let $\mathbf{P}_k = \mathbf{S}_k \mathbf{S}_k^*$, with $\mathbf{S}_k \in \mathbb{C}^{d_k \times d'_k}$ full rank. Then we can write

$$\mathbf{R} = [\mathbf{A}_u \mathbf{S}_u \quad \mathbf{A}_k \mathbf{S}_k] \begin{bmatrix} \mathbf{S}_u^* \mathbf{A}_u^* \\ \mathbf{S}_k^* \mathbf{A}_k^* \end{bmatrix} + \sigma^2 \mathbf{I}. \quad (\text{E.13})$$

Let $\bar{\mathbf{A}} = [\mathbf{A}_u \mathbf{S}_u \quad \mathbf{A}_k \mathbf{S}_k]$; then, as in (F.1), we have that

$$\sigma^2 \mathbf{R}^{-1} = \mathbf{\Pi}_{\bar{\mathbf{A}}}^\perp + \sigma^2 (\bar{\mathbf{A}}^\dagger)^* \bar{\mathbf{A}}^\dagger + O(\sigma^4). \quad (\text{E.14})$$

Using (E.14),

$$\begin{aligned}
\mathbf{P}_u \mathbf{A}_u^* \mathbf{R}^{-1} \mathbf{A}_u \mathbf{P}_u &= \frac{1}{\sigma^2} \mathbf{S}_u \mathbf{S}_u^* \mathbf{A}_u^* \left(\mathbf{\Pi}_{\mathbf{A}}^\perp + \sigma^2 (\bar{\mathbf{A}}^\dagger)^* \bar{\mathbf{A}}^\dagger \right) \mathbf{A}_u \mathbf{S}_u \mathbf{S}_u^* \\
&\quad + O(\sigma^2) \\
&= \mathbf{S}_u \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{S}_u^* + O(\sigma^2) \\
&= \mathbf{P}_u + O(\sigma^2),
\end{aligned} \tag{E.15}$$

which gives

$$\begin{aligned}
\overline{\text{CRB}}_\theta &= \frac{1}{2N} \left[\text{Re} \left(\left(\mathbf{D}^* \mathbf{R}^{-1/2} \mathbf{\Pi}_{\mathbf{A}_{uw}}^\perp \mathbf{R}^{-1/2} \mathbf{D} \right) \odot \mathbf{P}_u^c \right) \right]^{-1} \\
&\quad + O(\sigma^4).
\end{aligned} \tag{E.16}$$

Comparing (E.16) to (E.12), we have that

$$\mathbf{C}_{\text{POW}} = N \cdot \overline{\text{CRB}}_\theta + O(\sigma^4); \tag{E.17}$$

hence, both POWDER methods are efficient for small σ^2 .

Appendix F

For small $\sigma^2 \mathbf{P}^{-1}$ we have that

$$\begin{aligned}
\sigma^2 \mathbf{R}^{-1} &= \mathbf{I} - \mathbf{A} (\mathbf{A}^* \mathbf{A} + \sigma^2 \mathbf{P}^{-1})^{-1} \mathbf{A}^* \\
&= \mathbf{\Pi}_{\mathbf{A}}^\perp + \sigma^2 (\mathbf{A}^\dagger)^* \mathbf{P}^{-1} \mathbf{A}^\dagger + O(\sigma^4 \|\mathbf{P}^{-2}\|).
\end{aligned} \tag{F.1}$$

References

- [1] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.
- [2] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. on Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.
- [3] R.D. DeGroat, E.M. Dowling, and D.A. Linebarger, "The constrained MUSIC problem," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1445–1449, Mar. 1993.
- [4] D.A. Linebarger, R.D. DeGroat, E.M. Dowling, P. Stoica, and G.L. Fudge, "Incorporating a priori information into MUSIC-algorithms and analysis," *Signal Process.*, vol. 46, no. 1, pp. 85–104, 1995.
- [5] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior knowledge-based DOA estimation method," in *17th European Signal Process. Conf.*, Glasgow, UK, Aug. 2009, pp. 869–873.
- [6] M. Jansson and B. Ottersten, "Structured covariance matrix estimation: A parametric approach," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 3172–75.
- [7] M. Jansson, B. Göransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945–956, Apr. 1999.
- [8] P. Wirfält, G. Bouleux, M. Jansson, and P. Stoica, "Optimal prior knowledge-based direction of arrival estimation," *IET Signal Process.*, vol. 6, no. 8, pp. 731–742, Oct. 2012.
- [9] P. Wirfält, M. Jansson, and G. Bouleux, "Prior-exploiting direction-of-arrival algorithm for partially uncorrelated source signals," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Vancouver, Canada, May 2013, pp. 3972–3976.
- [10] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate analysis*, Probability and mathematical statistics. Academic Press, 1979.
- [11] P. Wirfält, M. Jansson, G. Bouleux, and P. Stoica, "Prior knowledge-based direction of arrival estimation," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Prague, Czech Republic, May 2011, pp. 2540–2543.
- [12] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. on Acoust., Speech and Signal Process.*, vol. 34, no. 5, pp. 1081–1089, Oct. 1986.
- [13] P. Stoica and K. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proc. Part F Radar and Signal Process.*, vol. 137, no. 1, pp. 19–26, Feb. 1990.
- [14] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International (UK) Ltd, 1989.
- [15] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 996–1000, Jul. 1989.
- [16] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

- [17] P. Wirfält and M. Jansson, “Robust prior-based direction of arrival estimation,” in *IEEE Stat. Signal Process. Workshop*, Aug. 2012, pp. 81–84.
- [18] B. Ottersten, P. Stoica, and R. Roy, “Covariance matching estimation techniques for array signal processing applications,” *Digital Signal Process.*, vol. 8, no. 3, pp. 185–210, Jul. 1998.
- [19] P. Stoica, E.G. Larsson, and A.B. Gershman, “The stochastic CRB for array processing: a textbook derivation,” *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.
- [20] I. Ziskind and M. Wax, “Maximum likelihood localization of multiple sources by alternating projection,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

Paper C

Robust Prior-based Direction of Arrival Estimation

Petter Wirfält and Magnus Jansson

Published in
Proceedings of the IEEE Statistical Signal Processing Workshop (SSP'12)

©2012 IEEE
The layout has been revised

Robust Prior-based Direction of Arrival Estimation

Petter Wirfält and Magnus Jansson

Abstract

In certain Direction of Arrival (DOA) scenarios some of the sources are approximately known a priori. It is then desirable to be able to exploit this prior knowledge when estimating the DOAs of the unknown sources. In this paper we modify an estimator utilizing exact angular prior knowledge of some sources such that the estimator is able to exploit prior knowledge with some uncertainty. We derive the corresponding Cramér-Rao lower bound and present numerical results showing that the estimator can benefit from prior information, even when it is inaccurate.

Index Terms—Direction of arrival estimation; array signal processing; Bayesian estimation

1 Introduction

Direction of Arrival (DOA) estimation is one of the most widely researched areas in Signal Processing. In certain applications some of the received signals are coming from a known direction. These signals can for example be from an obstacle in the field of view (FOV) of the sensing array, or perhaps from a friendly target whose location is known.

Most DOA algorithms estimate all sources in the received signal. For such algorithms a known source among the unknown will be treated as another parameter to estimate, and as such will (in general) degrade the accuracy of the estimates of the other sources.

Several methods have been proposed to exploit the prior knowledge of DOAs in the array FOV. Some use orthogonal projections which in this context means that the received signal is orthogonally projected onto a subspace which does not contain the known signal direction(s) [1]. Recently, a new method of exploiting such knowledge, denoted PLEDGE, was introduced [2]. Especially in some scenarios, this method gave a dramatic increase in estimation accuracy [3].

However, both the above mentioned methods suffer from that they require the exact angular knowledge of the sources; no uncertainty is accounted for in the prior information. In some scenarios this is valid; the position of the known source is exact. In other scenarios it is more proper to model a known DOA as having a nominal value with some uncertainty. This could for example be the case when a friendly target is reporting its GPS-position; then usually a quantitative measure of the uncertainty is also provided.

In this paper we propose a method to accurately exploit such uncertain known sources. The method requires some nominally known DOAs and a measure of the uncertainty in said DOAs. As will be shown this translates into a method that adaptively values the prior information in proportion to the uncertainty; this allows the method to benefit from prior DOA information, even when that information is not absolutely accurate.

The idea in this paper is to model the uncertainty in the prior as an *error* in the receiving array. Seen from this perspective, it is possible to utilize results pertaining to methods designed to be robust to array errors [4], [5]. One such estimator, which in addition has the attractive property of being relatively straight-forward to realize for a uniform linear array, is [6], and we apply the general framework in that article to our specific case.

2 Problem formulation

We consider the case of d narrow-band, plane waves impinging on an m sensor uniform linear array. Of these d waves, d_k are emanating from sources whose positions are known with some uncertainty. We can formulate a model for the received data according to

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, 2, \dots, N, \quad (1)$$

where $\mathbf{A}(\boldsymbol{\theta})$ is the array steering matrix parameterized by the vector of DOAs $\boldsymbol{\theta} = [\boldsymbol{\theta}_u^T \quad \boldsymbol{\theta}_k^T]^T$, where the subscripts u and k denote unknown and known, respectively, and T is the transpose operation. In addition, $\mathbf{s}(t)$ is the source signal waveforms sampled at time t , and $\mathbf{n}(t)$ the sensor noise. The i :th column of \mathbf{A} is defined as

$$\mathbf{a}(\theta_i) = [1 \quad e^{j\phi_i} \quad \dots \quad e^{j(m-1)\phi_i}]^T, \quad i = 1, \dots, d, \quad (2)$$

where

$$\phi_i = -2\pi\Delta \sin(\theta_i) \quad (3)$$

and Δ is the sensor separation in wavelengths. The complex-Gaussian, zero-mean signal and noise sequences $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are characterized by their second order statistics, which are assumed to be

$$E[\mathbf{s}(t)\mathbf{s}^*(\tau)] = \mathbf{P}\delta(t, \tau), \quad (4)$$

$$E[\mathbf{n}(t)\mathbf{n}^*(\tau)] = \sigma^2\mathbf{I}_m\delta(t, \tau), \quad (5)$$

where $*$ denotes conjugate transpose and \mathbf{I}_m denotes the identity matrix of size $m \times m$. For coherent sources, the rank of \mathbf{P} is less than d and accordingly we let $d' = \text{rank}(\mathbf{P})$.

The contribution of this paper is the design of an estimator that can take some uncertainty in the known DOAs into account. Accordingly, we model the known DOAs as Gaussian random variables independent of $\mathbf{s}(t)$ and $\mathbf{n}(t)$ defined by their mean $\mathbb{E}(\boldsymbol{\theta}_k) = \bar{\boldsymbol{\theta}}_k$ and their covariance

$$\mathbb{E}[(\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}}_k)(\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}}_k)^T] = \boldsymbol{\Theta}. \quad (6)$$

We assume we have access to the prior knowledge of the mean position $\bar{\boldsymbol{\theta}}_k$ and the covariance $\boldsymbol{\Theta}$; the idea is that the estimator should in some sense weight the knowledge of the known DOA according to this covariance. In the derivation of the proposed robust estimator and in the CRB analysis we will assume that $\boldsymbol{\Theta}$ is small, in the sense that the error in the unknown DOAs due to the error in the prior information is of the same order of magnitude as that due to finite sample effects of the noise. See [6] for more comments on this assumption.

3 DOA estimator

We begin by looking at a general description of a DOA-estimator; we will then introduce the concept of exact knowledge of some DOAs, and finally modify that estimator such that we can account for the stated uncertainty in the known DOAs.

Some additional statistics is needed in order to accomplish this. Let

$$\mathbb{E}[\mathbf{x}(t)\mathbf{x}^*(t)] \triangleq \mathbf{R} = \mathbf{A}_0\mathbf{P}\mathbf{A}_0^* + \sigma^2\mathbf{I}_m, \quad (7)$$

where $\mathbf{A}_0 = \mathbf{A}(\bar{\boldsymbol{\theta}}_u, \bar{\boldsymbol{\theta}}_k)$ and $\bar{\boldsymbol{\theta}}_u$ denotes the true positions of the unknowns sources. Performing an eigendecomposition of (7) gives

$$\mathbf{R} = \mathbf{E}_s\boldsymbol{\Lambda}_s\mathbf{E}_s^* + \mathbf{E}_n\boldsymbol{\Lambda}_n\mathbf{E}_n^*, \quad (8)$$

where $\mathbf{E}_s \in \mathbb{C}^{m \times d'}$ contains the d' principal eigenvectors of \mathbf{R} and spans what is referred to as the signal subspace, while \mathbf{E}_n containing the remaining $m - d'$ eigenvectors is spanning the noise subspace. In addition, $\boldsymbol{\Lambda}_s$ contains the associated signal eigenvalues, and $\boldsymbol{\Lambda}_n = \sigma^2\mathbf{I}_{m-d'}$. It can be shown that $\mathbf{E}_s = \mathbf{A}_0\mathbf{T}$, where $\mathbf{T} \in \mathbb{C}^{d' \times d'}$ is a full-rank matrix. Based on this fact we introduce a matrix $\mathbf{B}(\boldsymbol{\theta})$ which spans the null space of $\mathbf{A}(\boldsymbol{\theta})$; thus $\mathbf{B}^*(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) = \mathbf{0}$ which immediately gives $\mathbf{B}^*(\boldsymbol{\theta}_0)\mathbf{E}_s = \mathbf{0}$. In the case of a ULA at the receiving antenna, we can write

$$\mathbf{B}^* = \begin{bmatrix} b_d & \dots & b_1 & b_0 & 0 & \dots & 0 \\ 0 & b_d & \dots & b_1 & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b_d & \dots & b_1 & b_0 \end{bmatrix} \quad (9)$$

where the coefficients b_i are defined by the polynomial having its roots at the DOAs $\boldsymbol{\theta}$ according to

$$b_0 z^d + b_1 z^{d-1} + \dots + b_d = b_0 \prod_{i=1}^d (z - e^{j\phi_i}), \quad (10)$$

with ϕ_i defined in (3).

However, \mathbf{E}_s is not immediately accessible to us, and must be estimated from data. This is preferably done using

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) \mathbf{x}^*(t) \quad (11)$$

and then, similarly to (8),

$$\hat{\mathbf{R}} = \hat{\mathbf{E}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{E}}_s^* + \hat{\mathbf{E}}_n \hat{\mathbf{\Lambda}}_n \hat{\mathbf{E}}_n^*. \quad (12)$$

We can thus estimate $\boldsymbol{\theta}$ by minimizing a suitable norm of the residual

$$\boldsymbol{\epsilon} = \text{vec}(\mathbf{B}^*(\boldsymbol{\theta}) \hat{\mathbf{E}}_s), \quad (13)$$

where $\text{vec}(\cdot)$ is the vectorization operator which stacks the columns of a matrix on top of each other. Due to finite sample effects, (13) will in general not be equal to zero even for the true parameter values. In order to correctly minimize this residual, we must ensure that both the real and the imaginary parts are minimized. To that end and following [6], we introduce the notation

$$\hat{\mathbf{e}}_s = \begin{bmatrix} \text{vec}(\hat{\mathbf{E}}_s) \\ \text{vec}(\hat{\mathbf{E}}_s^c) \end{bmatrix} \quad (14)$$

$$\underline{\mathbf{A}} = \begin{bmatrix} (\mathbf{I}_{d'} \otimes \mathbf{A}) & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{d'} \otimes \mathbf{A}^c) \end{bmatrix} \quad (15)$$

$$\underline{\mathbf{B}} = \begin{bmatrix} (\mathbf{I}_{d'} \otimes \mathbf{B}) & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{d'} \otimes \mathbf{B}^c) \end{bmatrix} \quad (16)$$

$$\underline{\boldsymbol{\epsilon}} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}^c \end{bmatrix} = \underline{\mathbf{B}}^* \hat{\mathbf{e}}_s, \quad (17)$$

where the superscript c denotes complex conjugate, and \otimes denotes the Kronecker product. With this definition of $\underline{\boldsymbol{\epsilon}}$ we can formulate a minimization criteria that can be used to estimate the DOAs:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}} V(\boldsymbol{\theta}), \\ V(\boldsymbol{\theta}) &= \underline{\boldsymbol{\epsilon}}^*(\boldsymbol{\theta}) \mathbf{W} \underline{\boldsymbol{\epsilon}}(\boldsymbol{\theta}). \end{aligned} \quad (18)$$

In the expression (18), \mathbf{W} refers to a positive definite weighting matrix chosen such that the asymptotic variance of the estimates $\hat{\boldsymbol{\theta}}$ are minimized. Note that the optimal choice realizes the asymptotically statistically efficient algorithms in [7] and [8].

With (9) we can rewrite (13) according to

$$\begin{aligned}\boldsymbol{\epsilon} &= \text{vec}(\mathbf{B}^*(\boldsymbol{\theta})\hat{\mathbf{E}}_s) = \left(\hat{\mathbf{E}}_s^T \otimes \mathbf{I}_{m-d}\right) \text{vec}(\mathbf{B}^*) \\ &= \left(\hat{\mathbf{E}}_s^T \otimes \mathbf{I}_{m-d}\right) \boldsymbol{\Phi}_{\mathbf{B}\mathbf{b}} \mathbf{b} \triangleq \mathbf{Q}\mathbf{b},\end{aligned}\quad (19)$$

where we have utilized that $\text{vec}(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3) = (\mathbf{A}_3^T \otimes \mathbf{A}_1) \text{vec} \mathbf{A}_2$ for arbitrary matrices of suitable dimensions, introduced $\boldsymbol{\Phi}_{\mathbf{B}\mathbf{b}}$ as the selection matrix mapping the elements of $\mathbf{b} = [b_0 \ b_1 \ \dots \ b_d]^T$ to \mathbf{B}^* and defined $\mathbf{Q} \triangleq \left(\hat{\mathbf{E}}_s^T \otimes \mathbf{I}_{m-d}\right) \boldsymbol{\Phi}_{\mathbf{B}\mathbf{b}}$. Using (19) in (18), we get the expression

$$V(\boldsymbol{\theta}) = \underline{\mathbf{b}}^* \underline{\mathbf{Q}}^* \mathbf{W} \underline{\mathbf{Q}} \underline{\mathbf{b}}, \quad (20)$$

where

$$\underline{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^c \end{bmatrix}, \quad (21)$$

$$\underline{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ \mathbf{b}^c \end{bmatrix}. \quad (22)$$

The reason to rewrite in this manner is that the minimization of (18) is greatly simplified when parameterized by \mathbf{b} as in (20); through (10) there is a simple mapping between the elements of \mathbf{b} and the DOAs.

Remark: Note that the roots of (10) are constrained to the unit circle. To enforce that constraint when solving (20) significantly complicates the minimization procedure. We can relax this constraint to \mathbf{b} being conjugate symmetric (i.e., $b_i = b_{d-i}^c, i = 0, \dots, d$) and still obtain good accuracy. This has the advantage that the conjugate symmetry constraint can be algebraically enforced, allowing an unconstrained minimization of a slightly different expression than (20). Also note that to avoid the trivial solution $\mathbf{b} = \mathbf{0}$ a norm constraint should be enforced, i.e. $\|\mathbf{b}\| = 1$ where $\|\cdot\|$ is the euclidian norm, or by constraining a specific element of \mathbf{b} , e.g. $\text{Re}(b_0) = 1$. Such constraints can also be enforced without complicating the minimization step. See, e.g., [6] for more and exact details on this minimization.

3.1 Exact prior knowledge

When some of the DOAs $\boldsymbol{\theta}_k$ are known exactly these known roots can be factored from the polynomial (10) according to the PLEDGE framework [2]

$$b_0 \prod_{i=1}^d (z - e^{j\phi_i}) = P_k(\boldsymbol{\theta}_k) P_u(\boldsymbol{\theta}_u). \quad (23)$$

Let the coefficients of the known polynomial P_k constitute the set $\{c_i\}_{i=0}^{d_k}$, and use them to form the matrix $\mathbf{C} \in \mathbb{C}^{(d+1) \times (d_u+1)}$ similarly to (9), (10):

$$\mathbf{C}^T = \begin{bmatrix} c_0 & c_1 & \dots & c_{d_k} & \mathbf{0} \\ & \ddots & \ddots & & \ddots \\ \mathbf{0} & & c_0 & c_1 & \dots & c_{d_k} \end{bmatrix}. \quad (24)$$

We collect the coefficients of the unknown polynomial in $\tilde{\mathbf{b}} = [\tilde{b}_0 \ \tilde{b}_1 \ \dots \ \tilde{b}_{d_u}]^T$. We can then write the relation (23) in matrix notation as

$$\mathbf{b} = \mathbf{C}\tilde{\mathbf{b}}. \quad (25)$$

With (25) it is immediate to modify (22) and (20) to take the known sources into account, viz.:

$$V(\theta_u) = \tilde{\mathbf{b}}^* \tilde{\mathbf{Q}}^* \mathbf{W} \tilde{\mathbf{Q}} \tilde{\mathbf{b}}, \quad (26)$$

where

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^c \mathbf{C}^c \end{bmatrix} \quad (27)$$

$$\tilde{\mathbf{b}} = \begin{bmatrix} \tilde{\mathbf{b}} \\ \tilde{\mathbf{b}}^c \end{bmatrix}. \quad (28)$$

3.2 Uncertainty in prior knowledge

We now turn to the scenario of primary interest to this article, namely the one in which we know some target locations with a given uncertainty. This translates into there being some uncertainty in the \mathbf{C} -matrix.

In [6], a scenario with a criterion function as given by (26) is analyzed in the context of there being some uncertainty in the \mathbf{A} matrix; more exactly, some of the controlling array parameters are allowed to be random according to $\boldsymbol{\rho} \sim \mathcal{N}(\boldsymbol{\rho}_0, \boldsymbol{\Omega})$. In the context of the scenario studied in this article, this translates into there being uncertainty in the elements of the columns of $\mathbf{A}(\boldsymbol{\theta})$ representing the known angles. Hence, we can exploit the solution proposed to the problem given in [6] in order to correctly account for the uncertainty in the known DOAs.

Thus, in order to minimize the criterion function (26), we follow [6] and choose $\mathbf{W} = \mathbf{C}_\epsilon^{-1}$, where

$$\mathbf{C}_\epsilon = \lim_{N \rightarrow \infty} N E[\underline{\epsilon}\underline{\epsilon}^*], \quad (29)$$

which can be written

$$\mathbf{C}_\epsilon = \mathbf{L} + \mathbf{G}\mathbf{G}^* \quad (30)$$

where

$$\underline{\mathbf{L}} = \begin{bmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}^c \end{bmatrix}, \quad \mathbf{L} = \left(\sigma^2 \tilde{\mathbf{\Lambda}}^{-2} \mathbf{\Lambda}_s \otimes \mathbf{B}^* \mathbf{B} \right), \quad (31)$$

$$\underline{\mathbf{G}} = \begin{bmatrix} \mathbf{G} \\ \mathbf{G}^c \end{bmatrix}, \quad \mathbf{G} = (\mathbf{T}^T \otimes \mathbf{B}^*) \bar{\mathbf{D}}_k \bar{\mathbf{\Theta}}^{1/2}. \quad (32)$$

Here $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}_s - \sigma^2 \mathbf{I}_{d'}$, $\mathbf{T} = \mathbf{A}_0^\dagger \mathbf{E}_s$, where \dagger denotes the Moore-Penrose Pseudo-inverse, $\bar{\mathbf{\Theta}}^{1/2}$ is a symmetric square root of $(N \cdot \mathbf{\Theta})$, and $\bar{\mathbf{D}}_k$ is defined as

$$\bar{\mathbf{D}}_k = \begin{bmatrix} \frac{\partial \text{vec}(\mathbf{A}(\boldsymbol{\theta}))}{\partial \theta_{k_1}} & \dots & \frac{\partial \text{vec}(\mathbf{A}(\boldsymbol{\theta}))}{\partial \theta_{k_{d_k}}} \end{bmatrix} \quad (33)$$

evaluated at $\bar{\boldsymbol{\theta}}_k$.

It is additionally shown in [6] that the estimates $\hat{\boldsymbol{\theta}}_u$ given by the minimization of (26) with \mathbf{W} chosen according to (30) are asymptotically statistically efficient when using estimated quantities in the expressions (31) and (32).

3.3 Estimator implementation

We now succinctly state the steps of the proposed estimator following the analysis in Section 3.

The method of finding the unknown DOAs is performed according to the following sequence:

1. Form \mathbf{C} and $\bar{\mathbf{D}}_k$ using the nominally known positions $\bar{\boldsymbol{\theta}}_k$.
2. Perform an eigendecomposition of $\hat{\mathbf{R}}$, (12), to find $\hat{\mathbf{E}}_s$ and $\hat{\mathbf{\Lambda}}_s$; use $\hat{\mathbf{E}}_s$ to form \mathbf{Q} ($\Phi_{\mathbf{B}\mathbf{b}}$ can be formed offline), and create $\hat{\mathbf{Q}}$ according to (27).
3. Estimate $\hat{\sigma}^2 = \frac{1}{m-d'}(\text{Tr}(\hat{\mathbf{R}}) - \text{Tr}(\hat{\mathbf{\Lambda}}_s))$, and $\hat{\tilde{\mathbf{\Lambda}}} = \hat{\mathbf{\Lambda}}_s - \hat{\sigma}^2 \mathbf{I}$ to create an estimate of \mathbf{L} in (31).
4. Use $(\hat{\mathbf{B}}^* \hat{\mathbf{B}})^{-1} = \mathbf{I}$ and $\hat{\mathbf{G}} = \mathbf{0}$ as initial estimates, form $\hat{\mathbf{W}} = \hat{\mathbf{C}}_{\underline{\mathbf{e}}}^{-1}$ according to (30); then minimize (26) to find $\hat{\mathbf{b}}_{\text{INIT}}$, an initial estimate of $\hat{\mathbf{b}}$.
5. Use $\hat{\mathbf{b}}_{\text{INIT}}$ with (25) to form $(\hat{\mathbf{B}}^* \hat{\mathbf{B}})^{-1}$ and, additionally, with (10) and (3) to form $\hat{\mathbf{T}} = \mathbf{A}^\dagger(\hat{\boldsymbol{\theta}}) \hat{\mathbf{E}}_s$. This gives improved estimates of $\hat{\mathbf{L}}$ and $\hat{\mathbf{G}}$, which gives a more accurate $\hat{\mathbf{W}}$.
6. Minimize (26) once again to find $\hat{\mathbf{b}}$, then ϕ_i , $i = 1, \dots, d_u$, by rooting (10), and finally the sought DOAs through (3).

The minimization of the function (26) should be constrained such that \mathbf{b} is conjugate symmetric and such that $\|\mathbf{b}\| = 1$. Please see e.g. [6] for suitable ways of enforcing these constraints with the given criterion function.

4 Cramér-Rao bound

From [6] we have that the estimates $\hat{\theta}_u$ given by the method summarized in Section 3.3 are asymptotically unbiased and that their asymptotic variance is equal to the Cramér-Rao Bound (CRB). We now desire to find that bound.

Since we are only concerned with the DOA estimation we note from, e.g., [8] that the bound for the estimation of a *deterministic*¹ θ is given by

$$\mathbf{C}_\theta = \frac{\sigma^2}{2N} \operatorname{Re} \left((\mathbf{D}^* \Pi_{\mathbf{A}}^\perp \mathbf{D}) \odot (\mathbf{P}^* \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \mathbf{P})^T \right)^{-1}, \quad (34)$$

where

$$\mathbf{D} = \begin{bmatrix} \frac{\partial \mathbf{a}(\theta_1)}{\partial \theta_1} & \dots & \frac{\partial \mathbf{a}(\theta_d)}{\partial \theta_d} \end{bmatrix}, \quad (35)$$

and $\Pi_{\mathbf{A}}^\perp = \mathbf{I}_m - \mathbf{A}(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$. Without loss of generality, we partition $\theta = [\bar{\theta}_u^T \ \bar{\theta}_k^T]^T$, simultaneously partitioning (34) into a block matrix. Following [4], [5] we can, under the assumption that θ_k is close to $\bar{\theta}_k$, modify the lower right block of \mathbf{C}_θ to take the stochastic nature of θ_k into account; doing so we arrive at

$$\tilde{\mathbf{C}}_\theta = \frac{\sigma^2}{2N} \begin{bmatrix} \mathbf{C}_P & \mathbf{F} \\ \mathbf{F}^T & \mathbf{\Gamma} \end{bmatrix}^{-1}, \quad (36)$$

in which

$$\mathbf{C}_P = \operatorname{Re} \left((\mathbf{D}_u^* \Pi_{\mathbf{A}}^\perp \mathbf{D}_u) \odot (\bar{\mathbf{P}}_u^* \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \bar{\mathbf{P}}_u)^T \right) \quad (37)$$

$$\mathbf{F} = \operatorname{Re} \left((\mathbf{D}_u^* \Pi_{\mathbf{A}}^\perp \mathbf{D}_k) \odot (\bar{\mathbf{P}}_k^* \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \bar{\mathbf{P}}_u)^T \right) \quad (38)$$

$$\mathbf{\Gamma} = \operatorname{Re} \left((\mathbf{D}_k^* \Pi_{\mathbf{A}}^\perp \mathbf{D}_k) \odot (\bar{\mathbf{P}}_k^* \mathbf{A}^* \mathbf{R}^{-1} \mathbf{A} \bar{\mathbf{P}}_k)^T \right) + \frac{\sigma^2}{2N} \mathbf{\Theta}^{-1}. \quad (39)$$

In these equations, the subscript u and k on \mathbf{D} signifies that the derivatives in (35) are only with respect to the unknown and known sources, respectively, while $\bar{\mathbf{P}}_u$ denotes the sub-matrix of \mathbf{P} containing the *columns* associated with the unknown sources, and $\bar{\mathbf{P}}_k$ defined analogously. Thus, using a standard result on block-matrix inversion, the CRB for the unknown DOAs is found as

$$\operatorname{CRB}(\theta_u) = \frac{\sigma^2}{2N} (\mathbf{C}_P - \mathbf{F} \mathbf{\Gamma}^{-1} \mathbf{F}^T)^{-1}. \quad (40)$$

We can also note that $\frac{\sigma^2}{2N} \mathbf{C}_P^{-1}$ is the CRB of the estimates $\hat{\theta}_u$ when there is no uncertainty in the prior; see [2].

¹Not to be confused with the deterministic CRB.

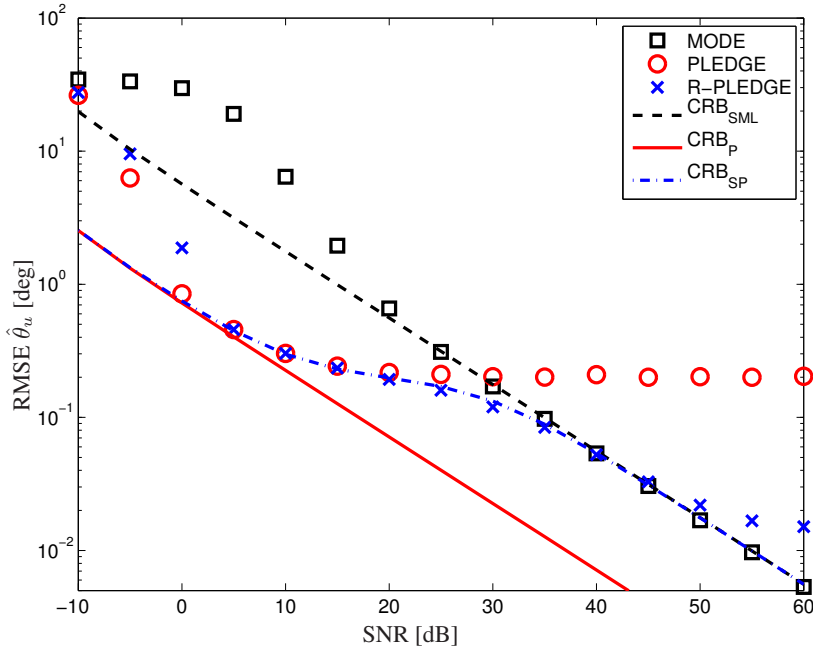


Figure 1: RMSE of 3 estimators and CRBs, averages of 4000 MC realizations; $N = 100$, $\bar{\theta}_u = 15^\circ$, $\bar{\theta}_k = 12^\circ$, $\Theta = 0.2^2$, $m = 10$.

5 Simulations

To study the empirical performance offered by the method in Section 3, we perform numerical Monte Carlo simulations. We simulate a scenario with two sources; the unknown is located at $\bar{\theta}_u = 15^\circ$ and the known source is, for each Monte Carlo simulation, a realization of the random variable $\theta_k \sim \mathcal{N}(\bar{\theta}_k, \Theta)$ with $\bar{\theta}_k = 12^\circ$ and $\Theta = 0.2^2 [\text{deg}^2]$. The sources are perfectly correlated which gives $\mathbf{P} = 10^{\text{SNR}/10} \begin{bmatrix} 1 & \rho \\ \rho^c & 1 \end{bmatrix}$. We choose the correlation phase $\rho = \exp(-j0.72)$ in order to create a difficult estimation scenario.

We compare the accuracy of three estimators in terms of the root-mean square error (RMSE) along with the respective CRB: MODE/WSF [7], [8] which is optimal in the case of no prior information; PLEDGE [2] which is optimal in the case of exact prior information; and the method proposed in this paper, hereby termed R-PLEDGE (“Robust PLEDGE”). The assumptions for the respective CRBs are: no prior information (denoted “CRB_{SML}”) [8]; exact prior information (“CRB_P”) [2], (37); and stochastic prior information (“CRB_{SP}”), (40).

As can be seen in Fig. 1 the proposed method R-PLEDGE and PLEDGE are about as accurate when the noise induced estimation error is larger than the vari-

ance of the known source. However, as the SNR increases the known source variance becomes predominant and PLEDGE is not able to benefit from the higher SNR due to the error in the prior. R-PLEDGE is seen to transition to the accuracy of the no-prior estimator, MODE, in this case, and can thus deliver as accurate estimates as is possible. For very high SNRs, however, R-PLEDGE begins to suffer from the error in the prior, and it is seen that its estimates are worse than those of MODE. However, the feasible region in which the prior information is usable has been significantly enhanced; a slightly higher threshold seems to be the only negative effects of R-PLEDGE as compared to PLEDGE. Hence R-PLEDGE generalizes PLEDGE to efficiently handle the case of uncertain prior knowledge.

6 Conclusions

In this paper we have developed a DOA estimator that can exploit prior information on some DOAs, even if the uncertainty in this knowledge is of the same order of magnitude as the expected unaided estimation accuracy. We have also developed an expression for the CRB of the problem under consideration, and this bound was seen to be attainable at moderate SNRs.

References

- [1] R. DeGroat, E. Dowling, and D. Linebarger, "The constrained MUSIC problem," *IEEE Trans. Signal Process.*, vol. 41, no. 3, pp. 1445–1449, Mar. 1993.
- [2] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior knowledge-based DOA estimation method," in *17th European Signal Process. Conf.*, Glasgow, UK, Aug. 2009, pp. 869–873.
- [3] P. Wirfält, M. Jansson, G. Bouleux, and P. Stoica, "Prior knowledge-based direction of arrival estimation," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Prague, Czech Republic, May 2011, pp. 2540–2543.
- [4] B. Wahlberg, B. Ottersten, and M. Viberg, "Robust signal parameter estimation in the presence of array perturbations," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Toronto, Canada, Apr. 1991, pp. 3277–3280 vol.5.
- [5] M. Viberg and A. Swindlehurst, "A bayesian approach to auto-calibration for parametric array signal processing," *IEEE Trans. on Signal Process.*, vol. 42, no. 12, pp. 3495–3507, Dec. 1994.
- [6] M. Jansson, A. Swindlehurst, and B. Ottersten, "Weighted subspace fitting for general array error models," *IEEE Trans. on Signal Process.*, vol. 46, no. 9, pp. 2484–2498, Sep. 1998.
- [7] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.
- [8] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. on Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.

Paper D

Subspace-based Frequency Estimation Utilizing Prior Information

Petter Wirfält, Guillaume Bouleux, Magnus Jansson, and Petre Stoica

Published in
Proceedings of the IEEE Statistical Signal Processing Workshop (SSP'11)

©2011 IEEE
The layout has been revised

Subspace-based Frequency Estimation Utilizing Prior Information

Petter Wirfält, Guillaume Bouleux, Magnus Jansson, and Petre Stoica

Abstract

In certain frequency estimation applications one or more of the underlying frequencies are known. For example, in rotary machines the known frequency may be a strong network frequency masking important closely spaced frequencies. Being able to include this information in the design of the estimator can be expected to improve the performance when estimating such closely spaced frequencies. We present a framework to include such prior information in a class of subspace-based estimators. Through Monte Carlo simulations and real-data applications we show the usefulness of our approach.

Index Terms—Frequency estimation; Parameter estimation; Rotating machines; Failure analysis

1 Introduction

Frequency estimation is an important tool in a number of signal processing applications, including radar, sonar, communications, etc. Another application that the current article will look more closely at is performance monitoring of rotating mechanical systems, and especially the diagnosis of faults in the motors of such systems. In all of these situations we might have *a priori* known frequencies in the signals we analyze, which in the case of an electric motor might correspond to the supply frequency, known gear frequencies, etc. The crucial point is that these frequencies do not carry any information that we are interested in but might in fact degrade the estimation of the unknown, interesting, frequencies.

A number of different methods have been proposed to tackle the general sinusoids-in-noise estimation problem, but in this article we focus on eigenanalysis-based methods such as [1], [2]. Not considered in the previous references is the ability to incorporate prior knowledge of certain frequencies into the

estimation of the remaining frequencies. In [3], [4] some different approaches of doing so are evaluated. Some of the approaches considered there use orthogonal projections, which we explicitly want to avoid due to the inherent information loss.

The diagnosis of rotating machines, and in particular broken rotor bars in induction motors, is a widely researched area, see, e.g. [5] for an overview of recent developments. Subspace techniques (MUSIC [6]) has previously been used in [7], [8], but DFT-methods still appear to be the prevalent diagnostic tool in the literature [5]. Recently a state-space model capturing physical parameters of the motor has been proposed in [9], where a motor failure can be detected and diagnosed from estimated model parameters.

In this paper we use the Prior knowLEDGE (PLEDGE) framework of [10] in two existing frequency estimators [1], [2], [11], and show the performance gains achieved by doing so. In addition, we apply the proposed estimation method to real data for broken rotor bar detection in an asynchronous induction motor; the objective of that analysis is to estimate sideband frequencies to the network frequency in order to diagnose the machine [12].

2 Problem formulation

The frequency estimation problem can be described as follows. We observe N samples of a scalar valued complex signal $y(t)$ according to

$$y(t) = \sum_{k=1}^d \alpha_k e^{j(\omega_k t + \phi_k)} + n(t), \quad t = 0, \dots, N-1, \quad (1)$$

in which, for sinusoid k , $\alpha_k > 0$ is a real valued amplitude, $\omega_k \in [0, 2\pi)$ is the angular frequency and $\phi_k \in [0, 2\pi)$ is the phase at $t = 0$. Further, α_k and ω_k are deterministic parameters, while ϕ_k is assumed random with a uniform distribution over its possible values. The number of sinusoids, d , present in the signal is assumed known, and $n(t)$ is a zero mean circular Gaussian random sequence independent of ϕ_k for all k , with non-zero second order moment given by

$$\mathbb{E}[n(t)n^c(\tau)] = \sigma^2 \delta(t, \tau), \quad (2)$$

where the superscript c denotes the complex conjugate and $\delta(t, \tau) = 1$ for $t = \tau$ and 0 otherwise. We additionally assume that a subset of $\{\omega_i\}_{i=1}^d$ is known *a priori*.

3 Estimator implementation

In this section we present the general idea of the estimators later on evaluated. For more thorough explanations see [1], [11].

If we let

$$\mathbf{x}(t) = [\alpha_1 e^{j(\omega_1 t + \phi_1)} \quad \dots \quad \alpha_d e^{j(\omega_d t + \phi_d)}]^T, \quad (3)$$

where the superscript T denotes the transpose, we can reformulate the problem (1) into

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{n}(t), \quad t = 0, \dots, N - m \quad (4)$$

in which we have defined $\mathbf{y}(t) = [y(t) \quad \dots \quad y(t + m - 1)]^T$,

$$\mathbf{A} = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\omega_1} & \dots & e^{j\omega_d} \\ \vdots & & \vdots \\ e^{j(m-1)\omega_1} & \dots & e^{j(m-1)\omega_d} \end{bmatrix}, \quad (5)$$

m is a user parameter, and $\mathbf{n}(t)$ is defined similarly to $\mathbf{y}(t)$. Equation (4) is the standard matrix formulation for one-dimensional data, see e.g. [2]. From (3) it is clear that

$$\mathbf{P} \triangleq \mathbb{E}[\mathbf{x}(t)\mathbf{x}^*(t)] = \begin{bmatrix} \alpha_1^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \alpha_d^2 \end{bmatrix}, \quad (6)$$

where the superscript $*$ denotes the conjugate transpose. Then, from (4), we find that

$$\mathbf{R} \triangleq \mathbb{E}[\mathbf{y}(t)\mathbf{y}^*(t)] = \mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2\mathbf{I} \quad (7)$$

which is the data covariance matrix. Next, we perform an eigendecomposition of \mathbf{R} and let

$$\mathbf{E}_s = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \dots \quad \mathbf{s}_d], \quad (8)$$

in which $\{\mathbf{s}_i\}_{i=1}^d$ correspond to the eigenvectors associated with the d largest eigenvalues $\{\lambda_i\}_{i=1}^d$ of \mathbf{R} . With this definition of \mathbf{E}_s we can use (7) to write

$$\mathbf{A}\mathbf{P}\mathbf{A}^*\mathbf{E}_s = \mathbf{E}_s\tilde{\mathbf{\Lambda}} \quad (9)$$

where

$$\tilde{\mathbf{\Lambda}} = \begin{bmatrix} \lambda_1 - \sigma^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_d - \sigma^2 \end{bmatrix}. \quad (10)$$

Now we define

$$\boldsymbol{\beta}_k = \left[\underbrace{0 \quad \dots \quad 0}_{k-1} \quad b_0 \quad \dots \quad b_d \quad \underbrace{0 \quad \dots \quad 0}_{m-d-k} \right]^T \quad (11)$$

in which $\{b_i\}$ are the polynomial coefficients of

$$b_0 z^d + b_1 z^{d-1} + \dots + b_d = b_0 \prod_{i=1}^d (z - e^{j\omega_i}), \quad (12)$$

the roots of which correspond to the frequencies in (1). Defining $\{b_i\}$ this way, it follows from (11) that $\mathbf{A}^* \boldsymbol{\beta}_k = \mathbf{0}$, $k = 1, \dots, m - d$, and from (9), and the fact that \mathbf{A} , \mathbf{P} , and \mathbf{E}_s are all full rank, that $\mathbf{E}_s^* \boldsymbol{\beta}_k = \mathbf{0}$. This last equality can explicitly be written for all k as (see e.g. [1])

$$\mathbf{F}\mathbf{b} \triangleq \begin{bmatrix} s_{1,1}^c & \cdots & s_{1,d+1}^c \\ \vdots & & \vdots \\ s_{d,1}^c & \cdots & s_{d,d+1}^c \\ s_{1,2}^c & \cdots & s_{1,d+2}^c \\ \vdots & & \vdots \\ s_{1,m-d}^c & \cdots & s_{1,m}^c \\ \vdots & & \vdots \\ & & \vdots \\ s_{d,m-d}^c & \cdots & s_{d,m}^c \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_d \end{bmatrix} = \mathbf{0} \quad (13)$$

where $s_{i,j}$ is the j th element in \mathbf{s}_i , and $\mathbf{b} = [b_0 \ \cdots \ b_d]^T$. In practice we can only use an estimate of (7),

$$\hat{\mathbf{R}} = \frac{1}{N - m + 1} \sum_{t=0}^{N-m} \mathbf{y}(t) \mathbf{y}^*(t). \quad (14)$$

We then find an estimate $\hat{\mathbf{E}}_s$ of \mathbf{E}_s in (8) from $\hat{\mathbf{R}}$, and hence an estimate $\hat{\mathbf{F}}$ of \mathbf{F} ; we can then write (13) estimated from data as

$$\hat{\mathbf{F}}\mathbf{b} = \boldsymbol{\mu}, \quad (15)$$

where we have accounted for the error in $\hat{\mathbf{F}}$ by introducing the residual vector $\boldsymbol{\mu}$. In [1], it is shown how to exploit the structure of \mathbf{F} and \mathbf{b} in (13) to arrive at an estimator of \mathbf{b} which is asymptotically efficient. Having found \mathbf{b} , one can find the sought frequencies from (12).

The above estimator finds $\boldsymbol{\omega}$ through the polynomial coefficients \mathbf{b} . The main idea of [10] was that a prior knowledge of some of the frequencies $\boldsymbol{\omega}$ is equivalent to knowing some factors of (12). This knowledge can easily be exploited by factoring the polynomial (12) according to

$$b_0 \prod_{i=1}^d (z - e^{j\omega_i}) = P_u(z) P_k(z), \quad (16)$$

where $P_k(z) = c_0 \prod_{i=1}^{d_k} (z - e^{j\omega_{k,i}})$ is defined by the known frequencies $\{\omega_{k,i}\}_{i=1}^{d_k}$, and d_k is the number of known frequencies. Reverting to vector notation, we can write

$$\mathbf{b} = \mathbf{C}\tilde{\mathbf{b}} \quad (17)$$

where $\tilde{\mathbf{b}}$ is the vector containing the polynomial coefficients of $P_u(z)$ corresponding to the unknown frequencies in (1), and \mathbf{C} is the matrix defined by

$$\mathbf{C}^T = \begin{bmatrix} c_0 & c_1 & \dots & c_{d_k} & & \mathbf{0} \\ & \ddots & \ddots & & \ddots & \\ \mathbf{0} & & c_0 & c_1 & \dots & c_{d_k} \end{bmatrix} \quad (18)$$

in which the coefficients $\{c_i\}_{i=0}^{d_k}$ are the coefficients of $P_k(z)$. Using (17) we rewrite (15) as

$$\hat{\mathbf{F}}\mathbf{C}\tilde{\mathbf{b}} = \boldsymbol{\mu}. \quad (19)$$

In a certain sense what we have done is to filter the data $\hat{\mathbf{F}}$ by \mathbf{C} in order to find a, presumably, more accurate $\tilde{\mathbf{b}}$ which only depends on the unknown frequencies. Thus we have integrated the prior frequency knowledge into the data.

Note that re-writing (15) as in (19) is not limited to the estimator developed in [1]. All estimators that use such a polynomial rooting can be expected to benefit from using the proposed PLEDGE framework: MODE [11] is such an example, where we have an expression on the form of (15) but with a different definition of \mathbf{F} ; in [11] the data has been weighted in a specific way as to achieve asymptotical efficiency in the related direction of arrival-estimation scenario. We examine the extension of MODE as well in the following sections.

4 Simulations

To study the practical performance improvements offered by the method in Section 3 we perform numerical Monte Carlo simulations in which $\boldsymbol{\omega} = [2\pi 0.5 \quad 2\pi 0.52 \quad 2\pi 0.56]^T$, where we consider $2\pi 0.52$ to be known. The sinusoids are equally strong with an $\text{SNR} = 20\text{dB}$, which we define as $\text{SNR}_k = \alpha_k^2/\sigma^2$ for the k th sinusoid. $N = 30$ samples according to (1) are created, on which the respective algorithms (see below) are allowed to operate. We repeat the simulations 10000 times.

The implemented methods are: MODE as in [11]; ESPRIT (using a forward-backward averaging of data), see e.g. [1]; Markov [1], [2]; and MODE and Markov modified according to Section 3 in this paper, referred to as MODE PLEDGE and Markov PLEDGE.

In Fig. 1 we show the root mean squared error (RMSE) of the estimates corresponding to $\omega_1 = 2\pi 0.5$ as a function of m , the vector length in (4). The estimates pertaining to the second unknown frequency follow the same pattern. For any value of m , Markov PLEDGE produces significantly more accurate results than the other methods. Since MODE is not specifically tailored for frequency estimation, it is not surprising that it gives worse results; however, MODE PLEDGE is more accurate than the methods not utilizing the prior information.

While the Esprit method is performing as well as the unmodified Markov method, it is not compatible with the PLEDGE concept as described in this paper,

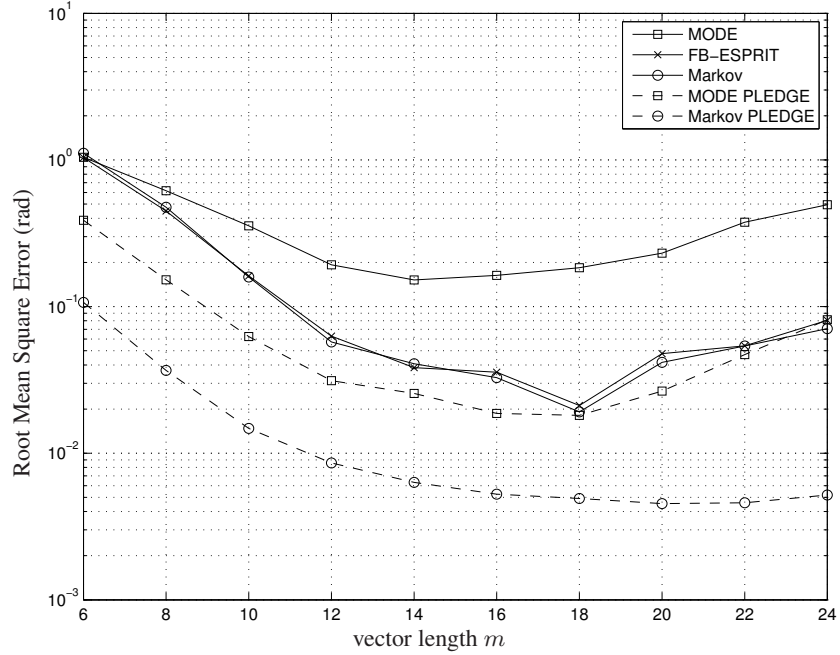


Figure 1: RMSE of estimates corresponding to $\omega_1 = 2\pi 0.5$, vs. m .

and we cannot inherently exploit the prior knowledge. Due to its lower computational complexity, it is otherwise an attractive estimator.

5 Broken rotor bar diagnosis

We also apply the methodology described in this paper to real data acquired from the current signal extracted from an induction motor. In order to interpret the results, we give a brief background to the problem; see e.g. [13], [9] for more details.

Here we are interested in defects of rotor bars, which can be diagnosed by the slip speed of the motor. The slip speed is a measure relating the rotation speed of the motor to the current fluctuations in the stator; the stator is driving the motor rotation. The slip speed is defined by

$$s = \frac{v_{\text{stat}} - v_{\text{rot}}}{v_{\text{stat}}}, \quad (20)$$

expressed in percent, where v_{rot} is the actual speed and v_{stat} the oscillation frequency in the stator. For industrial induction motors, the slip speed is typically between 1 and 2 %.

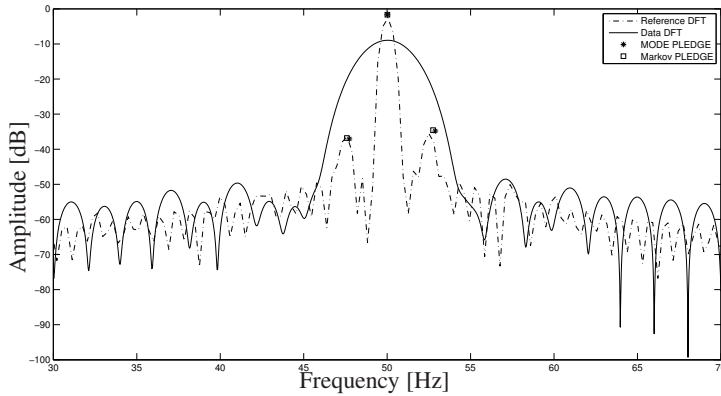


Figure 2: Estimated frequency content of the acquired motor current signal – proposed methods and “Data DFT” are using $N = 5000$ samples sampled at 10 kHz, with $m = 62$, $d = 5$. The “Reference DFT” is using the entire data set consisting of $1 \cdot 10^5$ samples corresponding to 10 seconds. Comparing the proposed methods to the reference, we can see that the proposed methods successfully identify the frequency peaks of interest.

A useful indicator of broken rotor bars is based on sideband components [12], [14] around the network frequency f_{NET} . These sidebands are located in the current spectrum at frequencies $f_B = (1 \pm 2ks)f_{\text{NET}}$, where k assumes integer values. Since f_{NET} is significantly stronger than all other frequencies in the spectrum, it can be expected to hamper the estimation of closely spaced frequencies; the idea then is to utilize the knowledge of f_{NET} , in order to detect f_B and thus to find s .

Comparing to, e.g., [7] the methods described herein produce frequency estimates without any associated amplitude estimate. We can however find estimates of the (complex) amplitudes based on (4), with $m = N$, according to

$$\hat{\mathbf{x}} = \mathbf{A}_N^\dagger \mathbf{y}_N \quad (21)$$

where the subscript N denotes that we include *all* N data-points, i.e. $\mathbf{y}_N = [y(0) \ \cdots \ y(N-1)]^T$ and \mathbf{A}_N defined accordingly, cf. (5). Additionally, in (21) \mathbf{A}_N^\dagger denotes the Moore-Penrose pseudo-inverse of \mathbf{A}_N . The amplitude corresponding to the estimated frequency ω_i can thus be found as $\alpha_i = |\hat{x}_i|$.

The data analyzed consisted of $1 \cdot 10^5$ samples acquired with a sampling frequency of 10 kHz. For such data lengths, the DFT, with proper windowing, can be expected to resolve and correctly diagnose the side band frequencies. Thus we can use the DFT of the full data set as a benchmark to compare against.

A more interesting scenario is when fewer data points are available; hence, we use $5 \cdot 10^3$ data points, which we then down sample by a factor of 40 producing

$N = 125$ samples which we analyze with the methods proposed in Section 3. The down sampling serves two purposes: the subspace methods are more expensive to implement for large m , and it also narrows the frequency band within which frequency estimates are found. Choosing the largest eigenvalues of $\hat{\mathbf{R}}$ suggests $d = 5$. Further, $m = 62$ (roughly $N/2$) is chosen.

Fig. 2 shows several plots: "Reference DFT" is the DFT spectrum based on the entire dataset, which can be seen as a reference to compare against. The proposed methods, utilizing the lesser amount of data, are also shown together with the DFT spectrum based on the reduced number of samples, "Data DFT".

We can see in Fig. 2 that the estimates produced by the proposed methods correspond to peaks in the reference spectrum. These estimated peaks match the theoretical expected values at $f = (1 \pm ks)f_{\text{NET}}$ for $k = 1$, which are the strongest side band frequencies. Thus, both methods accurately detect the side band components, which allows an estimate of the slip speed s . We can also see that for the investigated data length, detection using the DFT-based method is impossible due to the inherent low resolution of that approach.

In the studied case, it seems MODE PLEDGE gives about as accurate predictions as Markov PLEDGE. Further studies are needed in order to verify the theoretical accuracy-increase of Fig. 1.

6 Conclusions

In this paper we have shown how to incorporate prior information into an existing class of subspace-based frequency estimators. In the simulated data cases, significant gains can be achieved through using *a priori* known frequency information. We also investigated the estimator performance on real data and showed that the amount of data needed to estimate relevant parameters based on frequency spectra is reduced.

References

- [1] A. Eriksson, P. Stoica, and T. Söderström, "Markov-based eigenanalysis method for frequency estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 3, pp. 586–594, Mar. 1994.
- [2] M. Kristensson, M. Jansson, and B. Ottersten, "Further results and insights on subspace based sinusoidal frequency estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2962–2974, Dec. 2001.
- [3] E. Dowling, R. DeGroat, and D. Linebarger, "Exponential parameter estimation in the presence of known components and noise," *IEEE Trans. on Antennas and Propagation*, vol. 42, no. 5, pp. 590–599, May 1994.
- [4] D. Linebarger, R. DeGroat, E. Dowling, P. Stoica, and G. Fudge, "Incorporating a priori information into MUSIC-algorithms and analysis," *Signal Process.*, vol. 46, no. 1, pp. 85–104, 1995.
- [5] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4109–4126, Dec. 2008.
- [6] R. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. of RADC Spectrum Estimation Workshop*, Griffiss AFB, New York, USA, Oct. 1979, pp. 243–258.
- [7] S. Kia, H. Henao, and G.-A. Capolino, "A high-resolution frequency estimation method for three-phase induction machine fault detection," *IEEE Trans. Ind. Electron.*, vol. 54, no. 4, pp. 2305–2314, Aug. 2007.
- [8] F. Cupertino, E. de Vanna, L. Salvatore, and S. Stasi, "Analysis techniques for detection of IM broken rotor bars after supply disconnection," *IEEE Trans. Ind. Appl.*, vol. 40, no. 2, pp. 526–533, Mar.–Apr. 2004.
- [9] S. Bachir, S. Tnani, J.-C. Trigeassou, and G. Champenois, "Diagnosis by parameter estimation of stator and rotor faults occurring in induction machines," *IEEE Trans. Ind. Electron.*, vol. 53, no. 3, pp. 963–973, Jun. 2006.
- [10] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior knowledge-based DOA estimation method," in *17th European Signal Process. Conf.*, Glasgow, UK, Aug. 2009, pp. 869–873.
- [11] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.
- [12] J. Jung, J. Lee, and B. Kwon, "Online diagnosis of induction motors using MCSA," *IEEE Trans. Ind. Electron.*, vol. 53, no. 6, pp. 1842–1852, Dec. 2006.
- [13] M. Benbouzid, "A Review of Induction Motors Signature Analysis as a Medium for Faults Detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 984–994, Oct. 2000.
- [14] W. Deleroi, "Der Stabbruch im Käfigläufer eines Asynchronmotors," *Electrical Engineering (Archiv für Elektrotechnik)*, vol. 67, pp. 91–99, 1984.

Paper E

An ESPRIT-based parameter estimator for spectroscopic data

Erik Gudmundson, Petter Wirfält, Andreas Jakobsson, and Magnus Jansson

Published in
Proceedings of the IEEE Statistical Signal Processing Workshop (SSP'12)

©2012 IEEE
The layout has been revised

An ESPRIT-based parameter estimator for spectroscopic data

Erik Gudmundson, Petter Wirfält, Andreas Jakobsson, and Magnus Jansson

Abstract

The pulse spin-locking sequence is a common excitation sequence for magnetic resonance and nuclear quadrupole resonance signals, with the resulting measurement data being well modeled as a train of exponentially damped sinusoids. In this paper, we derive an ESPRIT-based estimator for such signals, together with the corresponding Cramér-Rao lower bound. The proposed estimator is computationally efficient and only requires prior knowledge of the number of spectral lines, which is in general available in the considered applications. Numerical simulations indicate that the proposed method is close to statistically efficient, and that it offers an attractive approach for initialization of existing statistically efficient gradient or search based techniques.

Index Terms—Parameter estimation; damped sinusoids; subspace techniques; multidimensional signal processing; NQR; NMR

1 Introduction and data model

Spectral estimation is a classical problem which has found application in a wide variety of fields, such as astronomy, medical imaging, radar, and spectroscopic techniques (for example nuclear magnetic resonance, NMR, and nuclear quadrupole resonance, NQR). Subspace-based estimators form an important class of spectral estimation methods that have proven to be useful for estimation of both damped and undamped sinusoidal signals (see, e.g., [1, 2]). Even though much work has been done on the estimation of damped and undamped sinusoids, there are only a few algorithms dealing with structured data models able to fit data produced by magnetic and quadrupolar resonance techniques. Such measurements are often resulting from the use of pulse spin-locking (PSL) sequences, which will then induce a fine structure into the signals. The PSL sequence consists of a preparatory

pulse and a train of refocusing pulses, where the time between two consecutive refocusing pulses is 2τ , as illustrated in Fig. 1. As discussed in [3, 4], the signal resulting from a PSL excitation can be well modeled as

$$y_{m,t} = x_{m,t} + w_{m,t}, \quad (1)$$

where $m = 0, \dots, M - 1$ denotes the echo number, and $t = t_0, \dots, t_{N-1}$ the local time within each echo, with $t = 0$ denoting the center of the current pulse, and where we assume uniform sampling intervals within each echo. Moreover, $w_{m,t}$ is an additive circular symmetric white Gaussian i.i.d. noise with variance σ^2 , and

$$x_{m,t} = \sum_{k=1}^K \alpha_k \exp(i\omega_k t - \beta_k |t - \tau| - (t + 2\tau m)\eta_k), \quad (2)$$

where α_k , $\omega_k = 2\pi f_k$, β_k , and η_k denote the (complex) amplitude, frequency, damping coefficient (with respect to the *current* transmitted pulse), and the *compound*, or echo train, damping coefficient for the k th spectral line, respectively; additionally, τ is a design parameter (due to operator choice of pulse repetition interval) and is thus known. Common approaches for estimating the parameters for this form of data is to sum the echoes, thereby destroying the finer details resulting from the echo train decay (and causing a bias in the estimates), and then use classical approaches such as the periodogram or the matrix pencil [1, 2]. An alternative approach was taken in [3], where a least-squares (LS) based algorithm for estimating all the unknown parameters in (2) was proposed. Such an estimate is formed using a gradient or grid-based search, but then requires a careful initialization of the various parameters, an often non-trivial task. As the search can often not be well initialized, this also implies that the resulting algorithm can be computationally demanding, especially when the data consists of several spectral lines. In this paper, we propose a computationally efficient ESPRIT-based algorithm, termed the echo train ESPRIT (ET-ESP), which requires no prior knowledge of typical parameter values, needing only knowledge of the number of present spectral lines, K , a number that is generally known in these applications. Furthermore, we introduce the corresponding Cramér-Rao lower bound (CRB) for the problem at hand and examine the performance of the proposed algorithm using numerical simulations.

Some words on the notation: hereafter, we denote by $\text{Re}\{x\}$, \mathbf{X}^T , \mathbf{X}^H , $\text{diag}\{\mathbf{x}\}$, and $[\mathbf{X}]_{l,k}$, the real part of x , the transpose of \mathbf{X} , the conjugate, or Hermitian, transpose of \mathbf{X} , the diagonal matrix with the vector \mathbf{x} along the diagonal, and the l, k -th element of \mathbf{X} , respectively.

2 The echo train ESPRIT algorithm

Let (2) be expressed as

$$x_{m,t} = \sum_{k=1}^K c_{m,k} z_k^t \quad (3)$$

with

$$c_{m,k} \triangleq \begin{cases} \alpha_k \exp(-\beta_k \tau) \cdot \exp(-2\eta_k \tau m) & \text{for } t < \tau \\ \alpha_k \exp(\beta_k \tau) \cdot \exp(-2\eta_k \tau m) & \text{for } t \geq \tau \end{cases}, \quad (4)$$

$$z_k \triangleq \begin{cases} \exp(i\omega_k) \cdot \exp(\beta_k - \eta_k) & \text{for } t < \tau; \\ \exp(i\omega_k) \cdot \exp(-\beta_k - \eta_k) & \text{for } t \geq \tau; \end{cases} \quad (5)$$

where it should be noted that z_k is not a function of m , as all echoes share the same poles. Reminiscent of [5], the noise-free data for each echo m is then stacked into the (Hankel) matrix

$$\mathbf{X}_m = \begin{bmatrix} x_{m,t_0} & x_{m,t_1} & \cdots & x_{m,t_{L'-1}} \\ x_{m,t_1} & x_{m,t_2} & \cdots & x_{m,t_{L'}} \\ \vdots & \vdots & & \vdots \\ x_{m,t_{L-1}} & \cdots & \cdots & x_{m,t_{N-1}} \end{bmatrix} \in \mathbb{C}^{L \times L'} \quad (6)$$

where $L' = N - L + 1$. This (noise-free) echo data matrix may then be collected, and partitioned, as

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}_0 \quad \cdots \quad \mathbf{X}_{M-1}] \\ &= \mathbf{S} [\mathbf{C}_0 \mathbf{T}^T \quad \cdots \quad \mathbf{C}_{M-1} \mathbf{T}^T] \end{aligned} \quad (7)$$

where $\mathbf{S} \in \mathbb{C}^{L \times K}$, $[\mathbf{S}]_{l,k} = z_k^{t_{l-1}}$, $\mathbf{T} \in \mathbb{C}^{L' \times K}$, $[\mathbf{T}]_{l',k} = z_k^{t_{l'-1}}$, $\mathbf{C}_m = \text{diag}\{[c_{m,1} \quad \cdots \quad c_{m,k}]\}$, and where $c_{m,k}$ is given by (4). Thus, \mathbf{S} and \mathbf{T} may be factored from each \mathbf{X}_m as in (7) due to z_k in (5) being independent of m . Forming the singular value decomposition (SVD) of \mathbf{X} , i.e.,

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H, \quad (8)$$

it may be noted by comparing (7) and (8) that \mathbf{S} and \mathbf{U} will span the same subspace. Regrettably, only $y_{m,t}$, i.e., the noise-corrupted measurements of (2) are available, instead necessitating the forming of \mathbf{Y}_m and \mathbf{Y} from $y_{m,t}$ similarly to (6) and (7).

Typically, magnetic resonance measurements may be obtained as either one or two-sided signals. For scenarios when measurements of both the expanding and the decaying part of the signal are available, so-called two-sided echoes, as is illustrated in Fig. 1, one may partition each echo into two parts, based on (4) and (5), such that one part is formed from $t < \tau$ and the other from $t \geq \tau$. Thus,

$$\mathbf{y}_m = [y_{m,t_0} \quad \cdots \quad y_{m,t_{N-1}}]^T \triangleq \begin{bmatrix} \mathbf{y}_m^{(+)} \\ \mathbf{y}_m^{(-)} \end{bmatrix}, \quad (9)$$

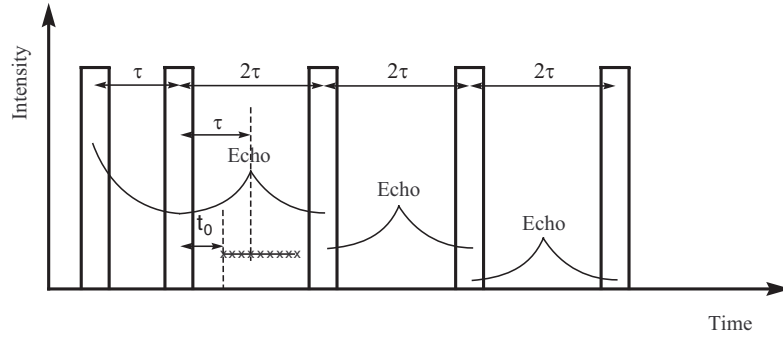


Figure 1: The PSL sequence.

where the superscripts $\mathbf{y}_m^{(+)}$ and $\mathbf{y}_m^{(-)}$ have been introduced to denote the expanding ($t < \tau$) and decaying ($t \geq \tau$) parts of \mathbf{y}_m , respectively. One then forms $\mathbf{Y}^{(+)}$ using only $\mathbf{y}_m^{(+)}$, and similarly for $\mathbf{Y}^{(-)}$. The following estimation is then performed independently for $\mathbf{Y}^{(+)}$ and $\mathbf{Y}^{(-)}$; accordingly, one gets two independent estimates for each parameter, i.e., $\hat{\gamma}_k^{(+)}$ and $\hat{\gamma}_k^{(-)}$. These are then combined to form the estimate $\hat{\gamma}_k = \frac{1}{2}\hat{\gamma}_k^{(+)} + \frac{1}{2}\hat{\gamma}_k^{(-)}$, where γ_k represents $\beta_k, \eta_k, \omega_k$, as appropriate.

Alternatively, for cases when only one-sided echoes are available, i.e., when one only obtains measurements for the decaying part $\mathbf{y}_m^{(-)}$, for $t \geq \tau$, the analysis is analogous, although with appropriate changes dictated by (4)-(6). In order to simplify the notation, we omit the superscript $(-)$ in the following.

Proceeding with either forms of measurements, let \mathbf{X}^\uparrow denote the operation of removing the bottom-most row of the matrix \mathbf{X} , and similarly let \mathbf{X}^\downarrow denote removal of the top-most row. Then, it is easily seen that $\mathbf{S}^\downarrow = \mathbf{S}^\uparrow \mathbf{Z}$, where

$$\mathbf{Z} = \text{diag} \{ [z_1, \dots, z_K] \}, \quad (10)$$

and hence $\mathbf{U}^\downarrow = \mathbf{U}^\uparrow \mathbf{\Omega}$, where $\mathbf{\Omega}$ and \mathbf{Z} are related by a similarity transformation and thus have the same eigenvalues. Using the measured data, the SVD of \mathbf{Y} is formed, yielding

$$\mathbf{Y} = \hat{\mathbf{U}} \hat{\mathbf{\Sigma}} \hat{\mathbf{V}}^H + \mathbf{W}, \quad (11)$$

where $\hat{\mathbf{\Sigma}}$ denotes the matrix formed from the K largest singular values and $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ denote the matrices formed by the corresponding singular vectors. The residual term, \mathbf{W} , contains the noise. The total-least squares (TLS) estimate $\hat{\mathbf{\Omega}}$ of $\mathbf{\Omega}$ may then be formed from

$$\hat{\mathbf{U}}^\downarrow = \hat{\mathbf{U}}^\uparrow \hat{\mathbf{\Omega}}, \quad (12)$$

and we may obtain estimates of the K poles $\{\hat{z}_k\}_{k=1}^K$ from the eigenvalues of $\hat{\mathbf{\Omega}}$. Using (5), we then find $\hat{\omega}_k = \angle \hat{z}_k$ and $\widehat{\beta_k + \eta_k} = -\log |\hat{z}_k|$. With the estimated

poles, one may then, for each echo m , write

$$\begin{bmatrix} \hat{z}_1^{t_0} & \cdots & \hat{z}_K^{t_0} \\ \hat{z}_1^{t_1} & \cdots & \hat{z}_K^{t_1} \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{t_{N-1}} & \cdots & \hat{z}_K^{t_{N-1}} \end{bmatrix} \begin{bmatrix} c_{m,1} \\ c_{m,2} \\ \vdots \\ c_{m,K} \end{bmatrix} = \begin{bmatrix} y_{m,t_0} \\ y_{m,t_1} \\ \vdots \\ y_{m,t_{N-1}} \end{bmatrix} \quad (13)$$

which forms a regular LS problem for $\{c_{m,k}\}_{k=1}^K$. Using (4), we simplify the notation by introducing $d_k \triangleq \alpha_k \exp(\beta_k \tau)$ and then, for each spectral line $k = 1, \dots, K$, one may form the following LS problem for the estimation of $\{\eta_k\}_{k=1}^K$

$$\begin{bmatrix} \log |\hat{c}_{0,k}| \\ \vdots \\ \log |\hat{c}_{M-1,k}| \end{bmatrix} = \begin{bmatrix} 1 & -2\tau \cdot 0 \\ \vdots & \vdots \\ 1 & -2\tau \cdot (M-1) \end{bmatrix} \begin{bmatrix} \log |d_k| \\ \eta_k \end{bmatrix}, \quad (14)$$

where $\{\hat{c}_{m,k}\}_{m=0}^{M-1}$ denote the LS solution to (13). The LS solution to (14) is readily found as

$$\hat{\eta}_k = \frac{3 \left[(M-1) \sum_{m=0}^{M-1} \log |\hat{c}_{m,k}| - 2 \sum_{m=0}^{M-1} m \log |\hat{c}_{m,k}| \right]}{\tau M (M-1) (M+1)}. \quad (15)$$

Using (5) and (15), one may then also estimate $\hat{\beta}_k$ as

$$\hat{\beta}_k = -(\hat{\eta}_k + \log |\hat{z}_k|). \quad (16)$$

Finally, given the estimates $\{\hat{\beta}_k, \hat{\omega}_k, \hat{\eta}_k\}_{k=1}^K$, an estimate of α_k , $k = 1, \dots, K$, may be formed from (1) using a maximum likelihood algorithm, which in this case coincides with the LS solution. Due to space limitations, the reader is referred to, e.g., [3, 6] for the details.

3 Derivation of the Cramér-Rao bound

We proceed to form the Cramér-Rao Bound (CRB) for the problem at hand, stacking the data from each measurement echo as

$$\mathbf{y} = \mathbf{x} + \mathbf{w}, \quad (17)$$

where

$$\mathbf{x} = [\mathbf{x}_0^T \cdots \mathbf{x}_{M-1}^T]^T \in \mathbb{C}^{NM \times 1}, \quad (18)$$

$$\mathbf{x}_m = [x_{m,t_0} \cdots x_{m,t_{N-1}}]^T \in \mathbb{C}^{N \times 1}, \quad (19)$$

$$\mathbf{w} = [w_{0,t_0} \cdots w_{M-1,t_{N-1}}]^T \in \mathbb{C}^{NM \times 1}. \quad (20)$$

In order to simplify the notation, let

$$\xi_k^{m,t} \triangleq \exp(i\omega_k t - \beta_k |t - \tau| - (t + 2\tau m)\eta_k), \quad (21)$$

so that

$$x_{m,t} = \sum_{k=1}^K \alpha_k \xi_k^{m,t}. \quad (22)$$

The CRB is given as $\text{CRB}(\boldsymbol{\gamma}) = [\text{FIM}(\boldsymbol{\gamma})]^{-1}$, where $\text{FIM}(\boldsymbol{\gamma})$ denotes the Fisher information matrix given the unknown (vector) parameters $\boldsymbol{\gamma}$, with

$$\boldsymbol{\gamma} = [a_1, \dots, a_K, b_1, \dots, b_K, \omega_1, \dots, \omega_K, \beta_1, \dots, \beta_K, \eta_1, \dots, \eta_K]^T, \quad (23)$$

where $a_k = |\alpha_k|$ and $b_k = \angle \alpha_k$. The CRB for a particular unknown γ_i is then obtained as the (i, i) -th element of the CRB matrix, i.e., $\text{CRB}(\gamma_i) = [\text{CRB}]_{ii}$. From the Slepian-Bang's formula [1], it is known that

$$\text{FIM}(\boldsymbol{\gamma}) = \frac{2}{\sigma^2} \text{Re} \left\{ \left(\frac{\partial \mathbf{x}}{\partial \boldsymbol{\gamma}} \right)^H \left(\frac{\partial \mathbf{x}}{\partial \boldsymbol{\gamma}} \right) \right\} \quad (24)$$

where the derivatives may be found as

$$\begin{aligned} \frac{\partial x_{m,t}}{\partial a_k} &= \exp(ib_k) \xi_k^{m,t}, & \frac{\partial x_{m,t}}{\partial \beta_k} &= -|t - \tau| \alpha_k \xi_k^{m,t}, \\ \frac{\partial x_{m,t}}{\partial b_k} &= i \alpha_k \xi_k^{m,t}, & \frac{\partial x_{m,t}}{\partial \eta_k} &= -(t + 2\tau m) \alpha_k \xi_k^{m,t}, \\ \frac{\partial x_{m,t}}{\partial \omega_k} &= it \alpha_k \xi_k^{m,t}. \end{aligned} \quad (25)$$

Reminiscent of the presentation in [7], these derivatives may be expressed on matrix form as $\partial \mathbf{x}_m / \partial \boldsymbol{\gamma} = \mathbf{Q}_m \mathbf{P}$, where

$$\begin{aligned} \mathbf{Q}_m &\triangleq \begin{bmatrix} \tilde{\boldsymbol{\Xi}}_m \boldsymbol{\Theta} & i \tilde{\boldsymbol{\Xi}}_m \boldsymbol{\Theta} & i \tilde{\tilde{\boldsymbol{\Xi}}}_m \boldsymbol{\Theta} & -\hat{\tilde{\boldsymbol{\Xi}}}_m \boldsymbol{\Theta} & -\check{\tilde{\boldsymbol{\Xi}}}_m \boldsymbol{\Theta} \end{bmatrix}, \\ \mathbf{P} &\triangleq \text{diag} \{ [\mathbf{I} \quad \boldsymbol{\Lambda} \quad \boldsymbol{\Lambda} \quad \boldsymbol{\Lambda} \quad \boldsymbol{\Lambda}] \} \in \mathbb{R}^{5K \times 5K}, \\ \tilde{\boldsymbol{\Xi}}_m &\triangleq \begin{bmatrix} \xi_1^{m,t_0} & \cdots & \xi_K^{m,t_0} \\ \vdots & \ddots & \vdots \\ \xi_1^{m,t_{N-1}} & \cdots & \xi_K^{m,t_{N-1}} \end{bmatrix}, \\ \tilde{\tilde{\boldsymbol{\Xi}}}_m &\triangleq \tilde{\mathbf{T}} \tilde{\boldsymbol{\Xi}}_m, \quad \hat{\tilde{\boldsymbol{\Xi}}}_m \triangleq \hat{\mathbf{T}} \tilde{\boldsymbol{\Xi}}_m, \quad \check{\tilde{\boldsymbol{\Xi}}}_m \triangleq \check{\mathbf{T}} \tilde{\boldsymbol{\Xi}}_m, \\ \boldsymbol{\Theta} &\triangleq \text{diag} \{ [e^{ib_1}, \dots, e^{ib_K}] \}, \\ \boldsymbol{\Lambda} &\triangleq \text{diag} \{ [a_1, \dots, a_K] \}, \\ \tilde{\mathbf{T}} &\triangleq \text{diag} \{ [t_0, \dots, t_{N-1}] \}, \\ \hat{\mathbf{T}} &\triangleq \text{diag} \{ [|t_0 - \tau|, \dots, |t_{N-1} - \tau|] \}, \\ \check{\mathbf{T}} &\triangleq \text{diag} \{ [(t_0 + 2\tau m), \dots, (t_{N-1} + 2\tau m)] \}. \end{aligned}$$

Stacking the derivatives from each echo m , yields $\partial \mathbf{x} / \partial \boldsymbol{\gamma} = \mathbf{Q} \mathbf{P}$, where

$$\mathbf{Q} = [\mathbf{Q}_0^T \quad \cdots \quad \mathbf{Q}_{M-1}^T]^T \in \mathbb{C}^{NM \times 5K}. \quad (26)$$

The FIM thus becomes

$$\text{FIM}(\boldsymbol{\gamma}) = \frac{2}{\sigma^2} \mathbf{P} \text{Re} \{ \mathbf{Q}^H \mathbf{Q} \} \mathbf{P} \quad (27)$$

implying that

$$\text{CRB}(\boldsymbol{\gamma}) = \frac{\sigma^2}{2} \mathbf{P}^{-1} [\text{Re} \{ \mathbf{Q}^H \mathbf{Q} \}]^{-1} \mathbf{P}^{-1}. \quad (28)$$

Letting

$$\boldsymbol{\Gamma} \triangleq [2 \text{Re} \{ \mathbf{Q}^H \mathbf{Q} \}]^{-1} \quad (29)$$

yields the further simplified expression for the sought CRBs

$$\text{CRB}(a_k) = [\boldsymbol{\Gamma}]_{kk} a_k^2 / \text{SNR}_k \quad (30)$$

$$\text{CRB}(b_k) = [\boldsymbol{\Gamma}]_{(k+K)(k+K)} / \text{SNR}_k \quad (31)$$

$$\text{CRB}(\omega_k) = [\boldsymbol{\Gamma}]_{(k+2K)(k+2K)} / \text{SNR}_k \quad (32)$$

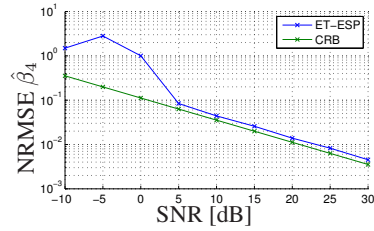
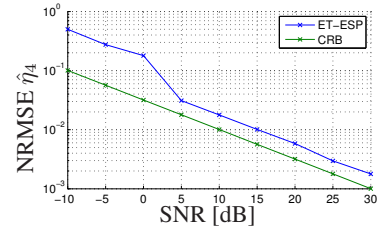
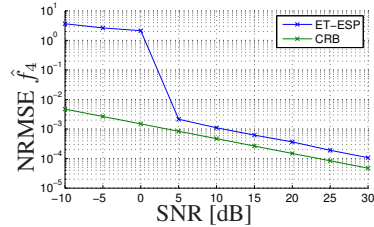
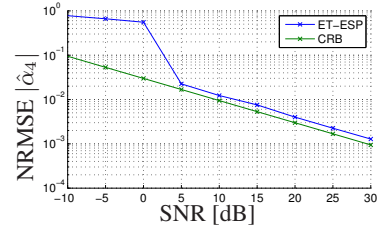
$$\text{CRB}(\beta_k) = [\boldsymbol{\Gamma}]_{(k+3K)(k+3K)} / \text{SNR}_k \quad (33)$$

$$\text{CRB}(\eta_k) = [\boldsymbol{\Gamma}]_{(k+4K)(k+4K)} / \text{SNR}_k, \quad (34)$$

where $\text{SNR}_k = a_k^2 / \sigma^2$.

Table 1: Parameters for simulated data

k	1	2	3	4
f_k (Hz)	0.0329	0.0122	0.0049	-0.0232
β_k (Hz)	0.0202	0.0077	0.0053	0.0035
η_k (10^{-3})	0.1811	0.2647	0.2130	0.2221
$ \alpha_k $	1.20	5.00	4.30	3.65
$\angle\alpha_k$ (rad)	0.4591	-2.8045	0.0661	-1.9922

**(a)** NRMSE for estimates of β_4 and the CRB as given in (33).**(b)** NRMSE for estimates of η_4 and the CRB as given in (34).**(c)** NRMSE for estimates of f_4 and the CRB as given in (32).**(d)** NRMSE for estimates of $|\alpha_4|$ and the CRB as given in (30).**Figure 2:** NRMSE given by the proposed estimator for different parameters, compared with the respective CRB. Here, the NRMSE is empirically evaluated over 500 Monte-Carlo realizations, for symmetric echoes, with $N = 256$ and $M = 32$.

4 Numerical examples

The proposed algorithm was evaluated using simulated NQR data, formed as to mimic the response signal from the explosive TNT when excited using a PSL sequence. Such signals can be well modeled as a sum of four damped sinusoidal signals, with the parameters as detailed in Table 1 (see [6] for further details on the signal and the relevant measurement setup). Based on the typical setup examined in [6], we use $N = 256$ measurements, for $M = 32$ echoes, with $\tau = 164$

and $t_0 = 36$, where the last two parameters are normalized with the sampling frequency and are therefore unit-less. The algorithm was evaluated using the normalized root mean squared error (NRMSE), defined as:

$$\text{NRMSE} = \sqrt{\frac{1}{P} \sum_{p=1}^P \left(\frac{\hat{x}_p}{x} - 1 \right)^2}, \quad (35)$$

where x denotes the true parameter value and \hat{x}_p the estimate of this parameter. The signal-to-noise ratio (SNR) is defined as $\sigma_s^2 \sigma^{-2}$, with σ_s^2 and σ^2 denoting the power of the noise-free signal and the noise variance, respectively. Moreover, we use $L = \tilde{N}/2$, where \tilde{N} denotes the number of samples of either the expanding or the decaying part of the echo. With the used τ , one obtains a symmetric echo [6], so that $L^{(+)} = L^{(-)} = 64$. Fig. 2 shows the results from $P = 500$ Monte-Carlo simulations for the fourth spectral line (the performance for the other lines was similar). As is common for ESPRIT-based estimators, it can be noted that the ET-ESP estimate does not fully reach the CRB and is therefore not statistically efficient. However, the difference is very small down to SNR = 0 dB (which corresponds to $\sigma^2 \approx 6$), before which the estimation error becomes very large. For SNR = 5 dB, the estimation error for the frequency f_4 is about 0.2%, whereas for the damping coefficient, β_4 , and the damping coefficient, η_4 , it is about 8% and 3%, respectively. The amplitude error $|\alpha_4|$ is about 2%.

5 Conclusions

In this paper, we have derived an ESPRIT-based estimator and the corresponding CRB for the data model detailing the typically damped sinusoidal signals obtained in magnetic resonance measurements when formed using PSL data sequences. The estimator is computationally efficient and only requires the number of sinusoids to be known, which is typically the case in the considered applications. Via Monte-Carlo simulations, we have shown that the algorithm is close to being statistically efficient for typical signal-to-noise ratios. The proposed method offers an attractive alternative as a standalone estimator or as an initial estimator for further refined estimates based on gradient or search-based techniques.

References

- [1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, N.J.: Prentice Hall, 2005.
- [2] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.
- [3] A. Jakobsson, M. Mossberg, M. Rowe, and J. A. S. Smith, "Exploiting Temperature Dependency in the Detection of NQR Signals," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1610–1616, May 2006.
- [4] A. Gregorovič and T. Apih, "TNT detection with 14N NQR: Multipulse sequences and matched filter," *J. Magn. Reson.*, vol. 198, no. 2, pp. 215–221, June 2009.
- [5] H. Chen, S. V. Huffel, A. van den Boom, and P. van den Bosch, "Extended HTLS methods for parameter estimation of multiple data sets modeled as sums of exponentials," in *13th Int. Conf. Digit. Signal Process.*, vol. 2, Santorini, 1997, pp. 1035–1038.
- [6] S. D. Somasundaram, "Advanced Signal Processing Algorithms Based on Novel Nuclear Quadrupole Resonance Models for the Detection of Explosives," Ph.D. dissertation, King's College London, London, United Kingdom, 2007.
- [7] Y. Yao and S. P. Pandit, "Cramér-Rao Lower Bound for a Damped Sinusoidal Process," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 878–885, April 1995.

Paper F

On Kronecker and Linearly Structured Covariance Matrix Estimation

Petter Wirfält and Magnus Jansson

Submitted to
IEEE Transactions on Signal Processing (Jun. 2013)

On Kronecker and Linearly Structured Covariance Matrix Estimation

Petter Wirfält and Magnus Jansson

Abstract

The estimation of covariance matrices is an integral part of numerous signal processing applications. In many scenarios, there exists prior knowledge on the structure of the true covariance matrix; e.g., it might be known that the matrix is Toeplitz in addition to hermitian. Given the available data and such prior structural knowledge, estimates using the known structure can be expected to be more accurate since more data per unknown parameter is available. In this work we study the case when a covariance matrix is known to be the Kronecker product of two factor matrices, and in addition the factor matrices are Toeplitz. We devise a two-step estimator to accurately solve this problem: the first step is a maximum likelihood (ML) based closed form estimator, which has previously been shown to give asymptotically (in the number of samples) efficient estimates when the relevant factor matrices are hermitian or persymmetric. The second step is a re-weighting of the estimates found in the first steps, such that the final estimate satisfies the desired Toeplitz structure. We derive the asymptotic distribution of the proposed two-step estimator and conclude that the estimator is asymptotically statistically efficient, and hence asymptotically ML. Through Monte Carlo simulations we further show that the estimator converges to the relevant Cramér-Rao lower bound for fewer samples than existing methods.

Index Terms—Parameter estimation, structured covariance estimation, signal processing algorithms, Kronecker model

1 Introduction

For certain statistical models [1], [2], [3] a covariance matrix can be modeled as a Kronecker product of two matrices of smaller dimensions, according to

$$\mathbf{R}_0 = \mathbf{A}_0 \otimes \mathbf{B}_0 \tag{1}$$

In applications, \mathbf{R}_0 is typically to be estimated from data, and \mathbf{A}_0 and \mathbf{B}_0 are *factor matrices* of dimensions $m \times m$, and $n \times n$, respectively, upon which further assumptions can be placed. Typically (and specifically, in this article), \mathbf{A}_0 and \mathbf{B}_0 are complex, hermitian, and positive definite (p.d.); additionally they may possess some linear structure (e.g., Toeplitz). When satisfying (1), it is highly desirable to tailor the estimator of \mathbf{R}_0 to this particular model in order to enhance accuracy and minimize computational costs: if \mathbf{A}_0 and \mathbf{B}_0 are complex hermitian, the number of real parameters in (1) are $m^2 + n^2$, whereas if estimating \mathbf{R}_0 simply under the hermitian constraint, $m^2 n^2$ real parameters must be estimated.

Applications of (1) are numerous in the literature: in some instances of wireless communication, the model is applicable for the statistics of the channel between the transmitter and receiver, [4], [5]. When estimating covariance matrices from EEG/MEG data, the same model can be used, [6], [7]. In the separable statistical models [3], [8], (1) is the defining assumption. Recently, models where sparsity can be exploited in the estimation of \mathbf{R}_0 or its inverse have been proposed [9]. Especially when $m, n \gg 1$, sparsity has the potential to dramatically reduce the number of parameters to be estimated [10]; in the current article, however, the factor matrices under consideration are non-sparse, and the dimensions m, n are fixed and finite.

In some of these applications, there is additional structure in the problem: each of the factor matrices in the Kronecker product might have a certain structure. For example, a factor matrix is Toeplitz if it is the covariance matrix of a stationary stochastic process [11]. In the EEG/MEG example, the temporal random process is stationary [6], [12], and thus the corresponding factor matrix is Toeplitz. In the MIMO communications example, when using uniform linear arrays at transmitter and receiver, the factor matrices are Toeplitz [13]. Similarly, if the arrays are merely symmetric around the respective array center point, we obtain persymmetric (PS) factor matrices (note that a Toeplitz structured matrix is per definition also PS). For further examples of structured factor matrices in signal processing applications, see [14]. Thus it is desirable to design estimators that can use such structure.

The idea in this paper is to revisit a maximum likelihood (ML) based iterative algorithm which is designed to find the factor matrices of a Kronecker product. The algorithm is called the flip-flop algorithm [1], [3], [15]. A nice property of the method is that it in only one iteration finds an asymptotically (in the number of samples) efficient estimate [15]. It also exhibits very good properties when the number of samples are limited and quickly produces estimates as good as the Cramér-Rao lower Bound (CRB) [15].

In [16], the flip-flop algorithm was enhanced to also capture PS structure in the factor matrices. However, the method cannot reproduce possible Toeplitz structure of the factor matrices it estimates. Since such a structure is prevalent in applications, it would be desirable to be able to capture this structure while retaining the good performance for small sample sizes. The “covariance matching”-algorithm developed in [15] is capable of capturing any such linear structure, but

shows worse estimation performance if compared to the algorithm in [16] when the number of samples are low.

The estimator developed in this article is shown to be asymptotically (in the number of samples, N) ML. We also see, through synthetic experiments in Section 7, that the estimator displays superior performance compared to that of the “covariance matching”-algorithm [15], when the number of available data samples is limited. Although the method is applicable to any linear structure in the factor matrices of (1), we from hereon restrict our analysis to the Toeplitz case due to its prevalence in applications (e.g., [6], [13], [17]). The results are easily generalizable to other linear structures.

The following notation will be used. Uppercase/lowercase bold-face letters, e.g. \mathbf{X}/\mathbf{x} , will denote matrices/vectors, respectively. In general, we will let $\mathbf{x} = \text{vec}(\mathbf{X})$, where the $\text{vec}(\cdot)$ operator stacks the columns of a matrix on top of each other. The symbol \simeq denotes that only dominating terms are retained, in the sense that $\mathbf{x}_1 \simeq \mathbf{x}_2$ is equivalent to $\mathbf{x}_1 = \mathbf{x}_2 + o_p(1/\sqrt{N})$. Here, $\xi_N = o_p(1/\sqrt{N})$ means that the sequence $\sqrt{N}\xi_N$ converges to zero in probability. We will also use the notation $\xi_N = O_p(1/\sqrt{N})$ to say that $\sqrt{N}\xi_N$ is bounded in probability.

Further, $\hat{\cdot}$ will denote an estimated quantity, T denotes transpose, c conjugate, H conjugate transpose, and † denotes the Moore-Penrose pseudo-inverse (see e.g. [18]).

2 Problem formulation

We observe measurements of a zero-mean, vector valued discrete time process $\mathbf{x}(t)$. Each observation is assumed to be drawn independently from a complex circularly symmetric Gaussian distribution having covariance matrix

$$\mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(\tau)] = \mathbf{R}_0\delta(t, \tau), \quad (2)$$

where $\delta(t, \tau)$ is the Kronecker-delta function, which is 1 for $t = \tau$ and 0 otherwise. The underlying process is satisfying the Kronecker model, such that \mathbf{R}_0 satisfies (1), with $\mathbf{A}_0 \in \mathcal{T}_m$, and $\mathbf{B}_0 \in \mathcal{T}_n$, where \mathcal{T}_m denotes the set of complex hermitian Toeplitz p.d. matrices of size $m \times m$.

Our objective is to estimate \mathbf{R}_0 given the data $\mathbf{x}(t), t = 1, 2, \dots, N$. We will see that this estimation will benefit from first finding estimates of \mathbf{A}_0 and \mathbf{B}_0 — however, it should be noted that \mathbf{A}_0 and \mathbf{B}_0 are unidentifiable based on the measurements of $\mathbf{x}(t)$. This is due to the fact that $\mathbf{R}_0 = \mathbf{A}_0 \otimes \mathbf{B}_0 = (\gamma\mathbf{A}_0) \otimes (\gamma^{-1}\mathbf{B}_0)$, with γ an arbitrary constant. This does not, as will be seen, prevent us from finding accurate estimates of \mathbf{R}_0 .

It is well known that the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t)\mathbf{x}^H(t) \quad (3)$$

is a sufficient statistic of the problem, [19]; thus the data only enters our estimators through (3).

3 Kronecker-structured ML-estimation

We now look at the likelihood function for the estimation problem at hand, and sketch how to obtain the ML estimator based on this function. The complete derivation was previously shown in [1], [3], [15]; the interested reader is referred to these references for further detail. We only impose the Kronecker structure on \mathbf{R}_0 , and, implicitly, that \mathbf{A}_0 and \mathbf{B}_0 are p.d.; the additional structure on the factor matrices is imposed in later sections.

The negative log-likelihood function of the problem, up to a scaling and with parameter independent terms omitted, can be written

$$l(\mathbf{A}, \mathbf{B}) = n \log |\mathbf{A}| + m \log |\mathbf{B}| + \text{Tr}\{\hat{\mathbf{R}}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\}. \quad (4)$$

Here, $|\cdot|$ is the determinant function, and $\text{Tr}(\cdot)$ is the trace-operator. It can be shown that (4) is minimized by

$$\hat{\mathbf{B}} = \frac{1}{m} \sum_{k=1}^m \sum_{l=1}^m \hat{\mathbf{R}}^{kl}(\mathbf{A}^{-1})_{lk} \quad (5)$$

for a fixed \mathbf{A} , where $\hat{\mathbf{R}}^{kl}$ denotes the k, l th block of size $n \times n$ in $\hat{\mathbf{R}}$, and $(\mathbf{A}^{-1})_{lk}$ is element l, k of \mathbf{A}^{-1} .

Likewise, the minimizer of (4) with respect to \mathbf{A} is

$$\hat{\mathbf{A}} = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \hat{\mathbf{R}}^{kl}(\mathbf{B}^{-1})_{lk}, \quad (6)$$

for a fixed \mathbf{B} . Here, $\hat{\mathbf{R}}^{kl}$ denotes the k, l th block of size $m \times m$ in $\hat{\mathbf{R}}$, which is defined as

$$\hat{\mathbf{R}} = \mathbf{L}_R \hat{\mathbf{R}}_R \mathbf{L}_R^T, \quad (7)$$

where \mathbf{L}_R is a permutation matrix defined such that

$$\mathbf{L}_R(\mathbf{X} \otimes \mathbf{Y}) = (\mathbf{Y} \otimes \mathbf{X})\mathbf{L}_R, \quad (8)$$

for any matrices \mathbf{X}, \mathbf{Y} of size $m \times m$ and $n \times n$, respectively. Following [1], it can be shown that $\hat{\mathbf{B}}$ and $\hat{\mathbf{A}}$, from (5) and (6), are p.d. almost surely (a.s.) [20], given that $N \geq \max(\frac{m}{n}, \frac{n}{m}) + 1$ and that \mathbf{A} and \mathbf{B} on the right hand sides of (5) and (6), respectively, are p.d.

With the sequential estimates given by (5) and (6), an iterative minimization algorithm (“flip-flop”) can be developed [1], [3], [15]. If the initial guess of \mathbf{A} in (5) is p.d., it was in [15] shown that the estimates given by the flip-flop algorithm are asymptotically efficient already after one iteration. The following algorithm is thus arrived at:

1. Choose an initial guess for \mathbf{A} , e.g. $\mathbf{A}_{\text{INIT}} = \mathbf{I}$.
2. Find \mathbf{B} based on \mathbf{A}_{INIT} , $\hat{\mathbf{B}}_{(1)} = \hat{\mathbf{B}}(\mathbf{A}_{\text{INIT}})$ using (5).
3. Find \mathbf{A} based on the previous estimate $\hat{\mathbf{B}}_{(1)}$ through (6), giving $\hat{\mathbf{A}}_{(2)}$.
4. Find $\hat{\mathbf{B}}_{(3)}$ based on $\hat{\mathbf{A}}_{(2)}$ again using (5).
5. Finally, find the estimate $\hat{\mathbf{R}}_{\text{K}} = \left(\hat{\mathbf{A}}_{(2)} \otimes \hat{\mathbf{B}}_{(3)} \right)$ with the prescribed Kronecker-structure.

Of course, in step 1 one could instead choose to initialize the algorithm with \mathbf{B}_{INIT} ; then the roles of \mathbf{A} and \mathbf{B} are simply interchanged in the algorithm. In the rest of the article, we will assume that the algorithm is initialized with \mathbf{A}_{INIT} . We will refer to this algorithm as the Non-Iterative Flip-Flop algorithm (NIFF) [15].

4 ML for persymmetric Kronecker model

In this section we look at finding the ML estimate of \mathbf{R}_0 under the additional assumption that the factor matrices are persymmetric (PS). That \mathbf{A}_0 and \mathbf{B}_0 are PS is equivalent to that (see, e.g., [21])

$$\mathbf{A}_0 = \mathbf{J}_m \mathbf{A}_0^T \mathbf{J}_m, \quad (9)$$

$$\mathbf{B}_0 = \mathbf{J}_n \mathbf{B}_0^T \mathbf{J}_n, \quad (10)$$

where

$$\mathbf{J}_k = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & & & \vdots \\ 1 & \dots & 0 & 0 \end{bmatrix}, \quad (11)$$

with the subscript k denoting the dimension of the square matrix. We will occasionally drop the index on \mathbf{J} when there is no risk for confusion. Clearly, a PS matrix is, in addition to being hermitian, symmetric along the anti-diagonal.

We will now show how to incorporate the PS structure in the ML-estimates given by the method in Section 3. The final results presented below can be found in [16], but no proof was given there.

The difference in the ML problem for the PS Kronecker model is that the criterion in (4) should be minimized with respect to the hermitian p.d. matrices \mathbf{A} and \mathbf{B} subject to the PS constraints (9)–(10). That is, we should find the minimizers of the optimization problem

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}} l(\mathbf{A}, \mathbf{B}) \\ & \text{s.t. } \mathbf{A} \in \mathcal{P}_m, \mathbf{B} \in \mathcal{P}_n \end{aligned} \quad (12)$$

where $l(\mathbf{A}, \mathbf{B})$ is given by (4), and \mathcal{P}_m denotes the set of complex, persymmetric, hermitian p.d. matrices of dimension $m \times m$. To find these minimizers we will rewrite the criterion function by using the following facts. With the constraints given by (12) we have

$$\mathbf{A}^{-1} = (\mathbf{J}\mathbf{A}^T\mathbf{J})^{-1} = \mathbf{J}\mathbf{A}^{-T}\mathbf{J}$$

since $\mathbf{J}^{-1} = \mathbf{J}$. This implies that

$$\begin{aligned} \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} &= \mathbf{J}_m \mathbf{A}^{-T} \mathbf{J}_m \otimes \mathbf{J}_n \mathbf{B}^{-T} \mathbf{J}_n \\ &= (\mathbf{J}_m \otimes \mathbf{J}_n) (\mathbf{A}^{-T} \otimes \mathbf{B}^{-T}) (\mathbf{J}_m \otimes \mathbf{J}_n) \\ &= \mathbf{J}_{mn} (\mathbf{A}^{-T} \otimes \mathbf{B}^{-T}) \mathbf{J}_{mn}, \end{aligned} \quad (13)$$

where we used that

$$\mathbf{X}\mathbf{Y} \otimes \mathbf{Z}\mathbf{W} = (\mathbf{X} \otimes \mathbf{Z})(\mathbf{Y} \otimes \mathbf{W}) \quad (14)$$

for arbitrary matrices \mathbf{X} , \mathbf{Y} , \mathbf{Z} , \mathbf{W} of suitable dimensions. Using (13) in the trace term in (4) we get

$$\begin{aligned} \text{Tr}\{\widehat{\mathbf{R}}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\} &= \text{Tr}\{\widehat{\mathbf{R}}(\mathbf{J}[\mathbf{A}^{-T} \otimes \mathbf{B}^{-T}]\mathbf{J})\} \\ &= \text{Tr}\{\mathbf{J}\widehat{\mathbf{R}}\mathbf{J}(\mathbf{A}^{-T} \otimes \mathbf{B}^{-T})\} \\ &= \text{Tr}\{\mathbf{J}\widehat{\mathbf{R}}^T\mathbf{J}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\}, \end{aligned} \quad (15)$$

where we used the facts that $\text{Tr}\{\mathbf{X}\mathbf{Y}\} = \text{Tr}\{\mathbf{Y}\mathbf{X}\}$ and $\text{Tr}\{\mathbf{X}\} = \text{Tr}\{\mathbf{X}^T\}$ for arbitrary matrices \mathbf{X} and \mathbf{Y} of appropriate dimensions.

We can now use these results in (12); with the above mentioned constraints on \mathbf{A} , \mathbf{B} , we use the equivalence in (15) to rewrite the criterion in (12) according to

$$\begin{aligned} l_{\text{PS}}(\mathbf{A}, \mathbf{B}) &= n \log |\mathbf{A}| + m \log |\mathbf{B}| \\ &\quad + \text{Tr}\{\widehat{\mathbf{R}}_{\text{PS}}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\}, \end{aligned} \quad (16)$$

where we have defined

$$\widehat{\mathbf{R}}_{\text{PS}} \triangleq \frac{1}{2}(\widehat{\mathbf{R}} + \mathbf{J}\widehat{\mathbf{R}}^T\mathbf{J}). \quad (17)$$

Note that (17) is PS by construction; the subscript PS on l in (16) denotes that we have implicitly incorporated the PS structure of \mathbf{A} and \mathbf{B} , and the p.d. constraint is enforced in the same way as in Section 3. See [21] for more details regarding the ‘‘Forward-Backward averaging’’ employed in (17). We can now formulate the following theorem:

THEOREM 1

The NIFF estimates of \mathbf{A} and \mathbf{B} are PS if the estimated sample covariance matrix used by the NIFF-algorithm is PS and the initial guess \mathbf{A}_{INIT} (or \mathbf{B}_{INIT}) is PS.

Proof: See Appendix A. ■

The conclusion from Theorem 1 is that we can solve the constrained problem (12) as easily as the unconstrained problem by only modifying the input to NIFF! Employing the modified sample covariance matrix $\widehat{\mathbf{R}}_{\text{PS}}$ we thus obtain an ML-estimate of \mathbf{R}_0 with the correct structure in the factor matrices. We will denote the resulting algorithm based on $\widehat{\mathbf{R}}_{\text{PS}}$ “NIFF-PS.” The benefit of this approach will be apparent in the simulations shown in Section 7.

5 Toeplitz structured factor matrices

Several relevant scenarios exist in which the factor matrices \mathbf{A}_0 and \mathbf{B}_0 are Toeplitz, [6], [13], [14]. A Toeplitz matrix is by construction also PS; thus we can expect the ML method presented in Section 4 (NIFF-PS) to give more accurate estimates than the unstructured method of Section 3 (NIFF). However, the exact Toeplitz structure will not be reproduced. Hence, it is desirable to devise a method that produces factor matrices of the correct structure; it can be expected that estimates of \mathbf{R}_0 is more accurate for factor matrices of the correct structure. The proposed method developed below is not ML, but asymptotically (in the number of snapshots, N), we show the method to be statistically efficient and hence to have the same asymptotic distribution as ML.

We want to solve the optimization problem

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{B}} \quad & l(\mathbf{A}, \mathbf{B}), \\ \text{s.t. } \mathbf{A} \in \mathcal{T}_m, \mathbf{B} \in \mathcal{T}_n \end{aligned} \quad (18)$$

with $l(\mathbf{A}, \mathbf{B})$ given by (4) and where \mathcal{T}_m , as previously, denotes the set of complex hermitian Toeplitz p.d. matrices of size $m \times m$.

The Toeplitz constraint makes the optimization problem (18) much harder to solve than (12); no algebraic solution as in Section 3 or Section 4 is known to exist. To proceed, we use the extended invariance principle (EXIP) [22]. The cardinal idea of EXIP is the relaxation of the studied optimization problem such that it can be solved exactly, and a subsequent fit of the relaxed estimate to the restricted parameter set of the original optimization problem.

The remainder of this section is structured as follows. First, we investigate the consistency of NIFF and NIFF-PS from Section 3 and Section 4. Then, we detail EXIP for the problem under study. Finally, we propose a computationally more efficient incarnation of EXIP.

5.1 Consistency of the factor matrix estimation

We noted earlier that we can only hope to estimate the factor matrices up to a scale factor. Thus, in order to find a meaningful definition of the error in our estimates of these matrices, we must establish what to compare the estimates to. It turns

out that the choice of initial guess in step 1 of the algorithm in Section 3 uniquely determines which matrices to use for comparison.

Following the notation of [15], we introduce some variables which facilitate this analysis as well as being useful in the continuation of the paper. Let

$$\begin{aligned} \mathbf{R}_B &= [\mathbf{r}^{11} \quad \mathbf{r}^{12} \quad \dots \quad \mathbf{r}^{1m} \quad \dots \quad \mathbf{r}^{mm}] \\ &= \text{vec}(\mathbf{B}_0) \text{vec}^H(\mathbf{A}_0) \triangleq \bar{\mathbf{f}}_0 \mathbf{a}_0^H, \end{aligned} \quad (19)$$

where $\mathbf{r}^{kl} \triangleq \text{vec}(\mathbf{R}^{kl})$ are the vectorized versions of the sub-matrices used in the NIFF-algorithm of Section 3, $\bar{\mathbf{f}}_0 \triangleq \text{vec}(\mathbf{B}_0)$, and $\mathbf{a}_0 \triangleq \text{vec}(\mathbf{A}_0)$. \mathbf{R}_B is related to \mathbf{R}_0 through

$$\text{vec}(\mathbf{R}_B) = \mathbf{L}_B \text{vec}(\mathbf{R}_0), \quad (20)$$

where \mathbf{L}_B is a permutation matrix satisfying $\mathbf{L}_B^T = \mathbf{L}_B^{-1}$. Similarly, we define

$$\mathbf{R}_A = [\bar{\mathbf{r}}^{11} \quad \bar{\mathbf{r}}^{12} \quad \dots \quad \bar{\mathbf{r}}^{nn}] = \text{vec}(\mathbf{A}_0) \text{vec}^H(\mathbf{B}_0), \quad (21)$$

where $\bar{\mathbf{r}}^{kl}$ is the vectorized version of the k, l th block of $\mathbf{L}_R \mathbf{R}_0 \mathbf{L}_R^T$, cf. (7), and

$$\text{vec}(\mathbf{R}_A) = \mathbf{L}_A \text{vec}(\mathbf{R}_0). \quad (22)$$

With these preliminaries, we can analyze the estimates produced by the NIFF-algorithm presented above. If we rewrite (5) in vector notation we get [15]

$$\hat{\mathbf{f}} \triangleq \text{vec}(\hat{\mathbf{B}}) = \frac{1}{m} \hat{\mathbf{R}}_B \text{vec}(\mathbf{A}^{-1}), \quad (23)$$

where $\hat{\mathbf{R}}_B$ is the sample version of (19). Completing the first step of the NIFF-algorithm, we find

$$\hat{\mathbf{f}}_{(1)} = \frac{1}{m} \hat{\mathbf{R}}_B \text{vec}(\mathbf{A}_{\text{INIT}}^{-1}) = \frac{1}{m} \hat{\mathbf{R}}_B \mathbf{a}_{\text{INIT}}^i, \quad (24)$$

where henceforth subscripts $(1), (2), \dots$ denote the step in which the scripted estimate is found in the NIFF-algorithm, and the superscript i denote inverse, in the sense $\mathbf{x}^i = \text{vec}(\mathbf{X}^{-1})$. If we let the number of observed snapshots $N \rightarrow \infty$, $\hat{\mathbf{R}}$ converges almost surely (a.s.) to \mathbf{R}_0 and thus $\hat{\mathbf{R}}_B \xrightarrow{\text{a.s.}} \mathbf{R}_B$. Hence,

$$\begin{aligned} \hat{\mathbf{f}}_{(1)} &\xrightarrow[N \rightarrow \infty]{\text{a.s.}} \frac{1}{m} \mathbf{R}_B \mathbf{a}_{\text{INIT}}^i = \frac{1}{m} \bar{\mathbf{f}}_0 \mathbf{a}_0^H \mathbf{a}_{\text{INIT}}^i \\ &= \frac{\text{Tr}(\mathbf{A}_0 \mathbf{A}_{\text{INIT}}^{-1})}{m} \bar{\mathbf{f}}_0 \triangleq \bar{\mathbf{f}}_*, \end{aligned} \quad (25)$$

where we have defined $\bar{\mathbf{f}}_*$ as the scaled version of $\bar{\mathbf{f}}_0$ that the algorithm will converge to.

Similarly, the estimate (6) found in the second step of the NIFF-algorithm converges according to

$$\begin{aligned}\hat{\mathbf{a}}_{(2)} &= \frac{1}{n} \hat{\mathbf{R}}_A \text{vec}(\hat{\mathbf{B}}_{(1)}^{-1}) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \frac{1}{n} \mathbf{a}_0^H \bar{\mathbf{f}}_0^i \bar{\mathbf{f}}_0^i \\ &= \frac{m}{\text{Tr}(\mathbf{A}_0 \mathbf{A}_{\text{INIT}}^{-1})} \mathbf{a}_0 \triangleq \mathbf{a}_*.\end{aligned}\quad (26)$$

In the third step we analogously have

$$\begin{aligned}\hat{\mathbf{f}}_{(3)} &= \frac{1}{m} \hat{\mathbf{R}}_B \text{vec}(\hat{\mathbf{A}}_{(2)}^{-1}) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \frac{1}{m} \bar{\mathbf{f}}_0 \mathbf{a}_0^H \mathbf{a}_0^i \\ &= \frac{\text{Tr}(\mathbf{A}_0 \mathbf{A}_{\text{INIT}}^{-1})}{m} \bar{\mathbf{f}}_0 = \bar{\mathbf{f}}_*.\end{aligned}\quad (27)$$

Thus, in the ensuing analysis we can consider \mathbf{a}_* and $\bar{\mathbf{f}}_*$ to be the *true values* of the factor matrices; of course, the original model still holds — $\mathbf{R}_0 = \mathbf{A}_* \otimes \mathbf{B}_*$. The above limits are additionally independent of the “forward-backward averaging” introduced in Section 4, since $\hat{\mathbf{R}}_{\text{PS}} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbf{R}_0$ as well.

5.2 EXIP-estimator

The EXIP [22] estimation procedure is specifically designed to simplify difficult ML-problems, such as (18). The procedure entails that the (difficult) optimization problem under consideration is relaxed, such that an exact ML-estimator can be used to find consistent estimates of the parameters of the relaxed problem. In the current context, then, we relax (18) to (12); by doing so the number of available parameters is increased, and an exact ML estimator of the parameters of the relaxed problem can be found by NIFF-PS, as shown in Section 4. Additionally, as shown in Section 5.1, that estimator is consistent.

Then, EXIP dictates that the so-found estimates, belonging to a subspace of larger dimension than that of the sought parameters, are considered to be noisy measurements of the parameters of the original optimization problem. The desired parameters are then found through a Markov-like (weighted least squares) optimization approach, which gives asymptotically (in the number of samples, N) ML estimates of the sought parameters.

Note that another viable approach is to relax (18) to (4), i.e. using NIFF, and then perform the Markov-like optimization. According to the theory of EXIP, the asymptotic performance of the thus found estimate is the same as if using NIFF-PS in the first step. Presumably, exploiting NIFF-PS (as appropriate in the problem under study) would lead to better finite-sample performance.

Applied to the problem (18), since $\mathcal{T}_m \subset \mathcal{P}_m$, the estimated parameters of the relaxed problem (12), $\hat{\mathbf{a}}_{(2)}$ and $\hat{\mathbf{f}}_{(3)}$, can then be viewed as noisy measurements of the parameters of $\mathbf{A}_* \in \mathcal{T}_m$ and $\mathbf{B}_* \in \mathcal{T}_n$, respectively. Explicitly, the Toeplitz

parameters are given from $\mathbf{a}_* \triangleq \mathbf{P}_A \boldsymbol{\theta}_{A_*}$ and $\bar{\mathbf{f}}_* \triangleq \mathbf{P}_B \boldsymbol{\theta}_{B_*}$, respectively, in which $\boldsymbol{\theta}_{A_*}$ ($\boldsymbol{\theta}_{B_*}$) denotes the true, real, Toeplitz parameters of \mathbf{A}_* (\mathbf{B}_*), and \mathbf{P}_A (\mathbf{P}_B) is the selection matrix mapping these parameters to the elements of \mathbf{a}_* ($\bar{\mathbf{f}}_*$).

The (unconstrained) EXIP optimization problem then becomes

$$\min_{\boldsymbol{\theta}_\eta} (\hat{\boldsymbol{\eta}} - \mathbf{P}_\eta \boldsymbol{\theta}_\eta)^H \mathbf{W}_{\text{EXIP}} (\hat{\boldsymbol{\eta}} - \mathbf{P}_\eta \boldsymbol{\theta}_\eta), \quad (28)$$

in which

$$\hat{\boldsymbol{\eta}} \triangleq \begin{bmatrix} \hat{\mathbf{a}}_{(2)} \\ \hat{\mathbf{f}}_{(3)} \end{bmatrix}, \quad (29)$$

$$\mathbf{P}_\eta \triangleq \begin{bmatrix} \mathbf{P}_A & \mathbf{0}_{m^2 \times n_B} \\ \mathbf{0}_{n^2 \times n_A} & \mathbf{P}_B \end{bmatrix}, \quad (30)$$

$$\boldsymbol{\theta}_\eta \triangleq \begin{bmatrix} \boldsymbol{\theta}_A \\ \boldsymbol{\theta}_B \end{bmatrix}, \quad (31)$$

\mathbf{W}_{EXIP} is a to-be-determined weighting matrix, and $n_A \triangleq \dim(\boldsymbol{\theta}_A) = 2m - 1$, $n_B \triangleq \dim(\boldsymbol{\theta}_B) = 2n - 1$ are the number of real parameters in the respective complex Toeplitz matrices. Note that (28) is valid irrespectively of whether NIFF or NIFF-PS is used to find the initial estimates; the entries of (29), however, changes accordingly. The solution to (28) is given by the weighted least squares estimate [23]

$$\hat{\boldsymbol{\theta}}_\eta = (\mathbf{P}_\eta^H \mathbf{W}_{\text{EXIP}} \mathbf{P}_\eta)^{-1} \mathbf{P}_\eta^H \mathbf{W}_{\text{EXIP}} \hat{\boldsymbol{\eta}}. \quad (32)$$

The optimal choice, in terms of asymptotic (in N) variance of $\hat{\boldsymbol{\theta}}_\eta$, of the matrix \mathbf{W}_{EXIP} is dependent on the covariance of $\hat{\boldsymbol{\eta}}$, which we denote by

$$\begin{aligned} \mathbf{C}_{\hat{\boldsymbol{\eta}}} &\triangleq \lim_{N \rightarrow \infty} N \text{E} ((\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^H) \\ &= \lim_{N \rightarrow \infty} N \text{E} \begin{bmatrix} \tilde{\mathbf{a}}_{(2)} \tilde{\mathbf{a}}_{(2)}^H & \tilde{\mathbf{a}}_{(2)} \tilde{\mathbf{f}}_{(3)}^H \\ \tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H & \tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{f}}_{(3)}^H \end{bmatrix}. \end{aligned} \quad (33)$$

From [23], we have that if $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$ is p.d., \mathbf{W}_{EXIP} should be chosen as the inverse of $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$. However, below we show that $\mathbf{C}_{\hat{\boldsymbol{\eta}}}$ is singular in the current problem. In such cases,

$$\mathbf{W}_{\text{EXIP}} \triangleq (\mathbf{C}_{\hat{\boldsymbol{\eta}}} + \mathbf{P}_\eta \mathbf{P}_\eta^*)^\dagger \quad (34)$$

is the optimal choice of weighting matrix [23]. We now find explicit expressions for the respective elements of (33), for both NIFF and NIFF-PS.

THEOREM 2

Given the data model in Section 2, but relaxing the assumptions on \mathbf{A}_0 and \mathbf{B}_0 such that they are hermitian p.d., the estimates from the NIFF algorithm have the covariances:

$$\begin{aligned} \lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{a}}_{(2)} \tilde{\mathbf{a}}_{(2)}^H \right) &= \frac{1}{n} \left[(\mathbf{A}_*^T \otimes \mathbf{A}_*) + \frac{\text{Tr}(\mathbf{A}_I^2)}{m^2} \mathbf{a}_* \mathbf{a}_*^H \right. \\ &\quad \left. - \frac{1}{m} (\text{vec}(\mathbf{A}_I \mathbf{A}_*) \mathbf{a}_*^H \right. \\ &\quad \left. + \mathbf{a}_* \text{vec}^H(\mathbf{A}_I \mathbf{A}_*)) \right]; \end{aligned} \quad (35)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{f}}_{(3)}^H \right) &= \frac{1}{m} \left[(\mathbf{B}_*^T \otimes \mathbf{B}_*) \right. \\ &\quad \left. + \frac{\text{Tr}(\mathbf{A}_I^2) - m \bar{\mathbf{f}}_* \bar{\mathbf{f}}_*^H}{nm} \bar{\mathbf{f}}_* \bar{\mathbf{f}}_*^H \right]; \end{aligned} \quad (36)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H \right) &= \frac{1}{mn} \left[\bar{\mathbf{f}}_* \text{vec}^H(\mathbf{A}_I \mathbf{A}_*) \right. \\ &\quad \left. - \frac{\text{Tr}(\mathbf{A}_I^2)}{m} \bar{\mathbf{f}}_* \mathbf{a}_*^H \right], \end{aligned} \quad (37)$$

where $\mathbf{A}_I = \mathbf{A}_* \mathbf{A}_{\text{INIT}}^{-1}$. While (35) is rank deficient, (36) is full rank.

Proof: See Appendix B. ■

If using the method detailed in Section 4, and the modified sample covariance matrix $\tilde{\mathbf{R}}_{\text{PS}}$ (17), the covariances of the estimates as given in Theorem 2 should change accordingly. We state the thus obtained covariances in the following theorem:

THEOREM 3

Given the data model in Section 2, but relaxing the assumptions on \mathbf{A}_0 and \mathbf{B}_0 such that they are hermitian p.d., the estimates from the NIFF-PS algorithm have the covariances:

$$\lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{a}}_{(2)} \tilde{\mathbf{a}}_{(2)}^H \right) = \Xi_{\mathbf{a}} \mathbf{T}_{\text{PS}} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{T}_{\text{PS}}^H \Xi_{\mathbf{a}}^H; \quad (38)$$

$$\lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{f}}_{(3)}^H \right) = \Xi_{\mathbf{f}} \mathbf{T}_{\text{PS}} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{T}_{\text{PS}}^H \Xi_{\mathbf{f}}^H; \quad (39)$$

$$\lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H \right) = \Xi_{\mathbf{f}} \mathbf{T}_{\text{PS}} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{T}_{\text{PS}}^H \Xi_{\mathbf{a}}^H, \quad (40)$$

in which

$$\mathbf{\Xi}_{\mathbf{a}} \triangleq \frac{1}{n} \left[((\bar{\mathbf{f}}_*^i)^T \otimes \mathbf{I}_{m^2}) \mathbf{L}_A - \frac{1}{m} \mathbf{a}_* (\bar{\mathbf{f}}_*^i)^H ((\mathbf{a}_{\text{INIT}}^i)^T \otimes \mathbf{I}_{n^2}) \mathbf{L}_B \right], \quad (41)$$

$$\mathbf{T}_{\text{PS}} \triangleq \frac{1}{2} [\mathbf{I} + (\mathbf{J} \otimes \mathbf{J}) \mathbf{L}_T]; \quad (42)$$

$$\begin{aligned} \mathbf{\Xi}_{\bar{\mathbf{f}}} \triangleq & \frac{1}{m} ((\mathbf{a}_*^i)^T \otimes \mathbf{I}_{n^2}) \mathbf{L}_B \\ & - \frac{1}{mn} \bar{\mathbf{f}}_* (\mathbf{a}_*^i)^H ((\bar{\mathbf{f}}_*^i)^T \otimes \mathbf{I}_{m^2}) \mathbf{L}_A + \\ & + \frac{1}{mn} \bar{\mathbf{f}}_* (\bar{\mathbf{f}}_*^i)^H ((\mathbf{a}_{\text{INIT}}^i)^T \otimes \mathbf{I}_{n^2}) \mathbf{L}_B. \end{aligned} \quad (43)$$

Proof: See Appendix C. ■

Using (35)-(37), or (38)-(40) as appropriate, in (33) we have that (32) gives asymptotically (in N) efficient estimates of $\hat{\boldsymbol{\theta}}_\eta$, for \mathbf{W}_{EXIP} according to (34). Explicitly, the asymptotic variance of $\hat{\boldsymbol{\theta}}_\eta$ is [23]

$$\text{cov}(\hat{\boldsymbol{\theta}}_\eta) = (\mathbf{P}_\eta^H \mathbf{W}_{\text{EXIP}} \mathbf{P}_\eta)^{-1} - \mathbf{I}. \quad (44)$$

Toeplitz-structured estimates of \mathbf{A} and \mathbf{B} are then readily obtained as

$$\begin{bmatrix} \text{vec}(\hat{\mathbf{A}}_\eta) \\ \text{vec}(\hat{\mathbf{B}}_\eta) \end{bmatrix} = \mathbf{P}_\eta \hat{\boldsymbol{\theta}}_\eta, \quad (45)$$

and the asymptotic covariance of the estimate of \mathbf{R}_0 given $\hat{\boldsymbol{\theta}}_\eta$ is

$$\mathbf{C}_{\text{S},\eta} = \frac{1}{N} \mathbf{K} \left[(\mathbf{P}_\eta^H \mathbf{W}_{\text{EXIP}} \mathbf{P}_\eta)^{-1} - \mathbf{I} \right] \mathbf{K}^H, \quad (46)$$

in which

$$\mathbf{K} \triangleq \mathbf{L}_B^T (\mathbf{P}_A^c \otimes \mathbf{P}_B) [(\mathbf{I}_{n_A} \otimes \boldsymbol{\theta}_{B_*}) \quad (\boldsymbol{\theta}_{A_*} \otimes \mathbf{I}_{n_B})], \quad (47)$$

and \mathbf{L}_B is given in (20). According to the EXIP theory [22], [23], (46) coincides with the CRB (the explicit form of which we state in (60)). The veracity of this statement is easily verifiable numerically.

5.3 SNIFF - structured non-iterative flip-flop algorithm

In contrast to the above (joint) estimation of $\boldsymbol{\theta}_\eta^T = [\boldsymbol{\theta}_A^T \quad \boldsymbol{\theta}_B^T]$, it was in [24] instead proposed to estimate the respective parameter vectors independently of

each other:

$$\hat{\boldsymbol{\theta}}_{\mathbf{A}_H} = (\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}},H} \mathbf{P}_A)^{-1} (\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}},H} \hat{\mathbf{a}}_{(2)}), \quad (48)$$

$$\hat{\boldsymbol{\theta}}_{\mathbf{B}_H} = (\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}},H} \mathbf{P}_B)^{-1} (\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}},H} \hat{\mathbf{f}}_{(3)}), \quad (49)$$

with $\mathbf{W}_{\hat{\mathbf{a}},H}$ and $\mathbf{W}_{\hat{\mathbf{f}},H}$ weighting matrices, and then forming:

$$\text{vec}(\hat{\mathbf{A}}_{T,H}) \triangleq \mathbf{P}_A \hat{\boldsymbol{\theta}}_{\mathbf{A}_H}, \quad (50)$$

$$\text{vec}(\hat{\mathbf{B}}_{T,H}) \triangleq \mathbf{P}_B \hat{\boldsymbol{\theta}}_{\mathbf{B}_H}. \quad (51)$$

It was additionally in [24] proposed to use consistent estimates of the weighting matrices

$$\mathbf{W}_{\hat{\mathbf{a}},H} = (\mathbf{A}_*^{-T} \otimes \mathbf{A}_*^{-1}), \quad (52)$$

$$\mathbf{W}_{\hat{\mathbf{f}},H} = (\mathbf{B}_*^{-T} \otimes \mathbf{B}_*^{-1}). \quad (53)$$

The inverses in (48) and (49) are guaranteed to exist as long as $\mathbf{W}_{\hat{\mathbf{a}},H}$, and $\mathbf{W}_{\hat{\mathbf{f}},H}$, respectively, are p.d. since \mathbf{P}_A and \mathbf{P}_B are full rank. $\mathbf{W}_{\hat{\mathbf{a}},H}$ and $\mathbf{W}_{\hat{\mathbf{f}},H}$ are guaranteed to be p.d. as long as the estimates used in (52) and (53) are p.d., which they are a.s. for $N \geq \max(\frac{m}{n}, \frac{n}{m}) + 1$, cf. Section 3 and [1].

As evident from Theorem 2 and Theorem 3, the expression for the weighting matrix in (32) is quite involved; this can be contrasted to the simple expressions (52) and (53). In Section 6, we show that this computationally simpler (due to the reduced dimensions of the involved matrices), but conceptually ad-hoc, estimation process also gives asymptotically efficient estimates.

Summarizing, we propose to estimate \mathbf{R}_0 according to:

1. Form $\hat{\mathbf{R}}_{PS}$ (17) from sample covariance matrix $\hat{\mathbf{R}}$ (3) or data vectors $\mathbf{x}(t)$.
2. Apply the NIFF-algorithm to $\hat{\mathbf{R}}_{PS}$ (i.e., use NIFF-PS) to get $\hat{\mathbf{A}}_{(2)}$ and $\hat{\mathbf{B}}_{(3)}$.
3. Use (48)–(51) to find the Toeplitz structured estimates $\hat{\mathbf{A}}_{T,H}$ and $\hat{\mathbf{B}}_{T,H}$, with $\hat{\mathbf{A}}_{(2)}$ and $\hat{\mathbf{B}}_{(3)}$ replacing \mathbf{A}_* and \mathbf{B}_* , respectively, in (52)–(53).
4. Form the sought estimate as $\hat{\mathbf{R}}_S = \hat{\mathbf{A}}_{T,H} \otimes \hat{\mathbf{B}}_{T,H}$.

We will refer to the above algorithm as the Structured Non-Iterative Flip-Flop algorithm (SNIFF).

Note that the methodology above is applicable to the scenario in which, e.g., \mathbf{A}_0 is Toeplitz, and \mathbf{B}_0 is PS. In order to find accurate estimates of \mathbf{R}_0 in this case, one would retain the estimate of \mathbf{B}_0 from NIFF-PS, and employ the above algorithm to only $\hat{\mathbf{A}}_{(2)}$ — the estimate of the true covariance matrix would then be $\hat{\mathbf{R}}_0 = \hat{\mathbf{A}}_{T,H} \otimes \hat{\mathbf{B}}_{(3)}$.

6 Performance analysis of the SNIFF estimates

In this section we prove the asymptotic efficiency of the estimates given by the SNIFF estimator.

6.1 Consistency

In Section 5.1 we have shown that the estimates from NIFF converge almost surely to \mathbf{A}_* and \mathbf{B}_* , respectively. From (48)–(51), we then immediately have that $\text{vec}(\hat{\mathbf{A}}_{T,H}) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \mathbf{a}_*$, and $\text{vec}(\hat{\mathbf{B}}_{T,H}) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \bar{\mathbf{f}}_*$. Hence, the estimates given by SNIFF are (strongly) consistent.

6.2 Asymptotic variance

Let $\hat{\mathbf{R}}_S$ denote the estimate of \mathbf{R}_0 given by SNIFF. Then

$$\tilde{\mathbf{r}}_S \triangleq \text{vec}(\hat{\mathbf{R}}_S) - \text{vec}(\mathbf{R}_0), \quad (54)$$

which we write, based on a first-order approximation and using the notation $\tilde{\mathbf{a}}_T \triangleq \text{vec}(\hat{\mathbf{A}}_{T,H} - \mathbf{A}_*)$, $\tilde{\mathbf{f}}_T \triangleq \text{vec}(\hat{\mathbf{B}}_{T,H} - \mathbf{B}_*)$,

$$\begin{aligned} \tilde{\mathbf{r}}_S &\simeq \mathbf{L}_B^T \text{vec} \left(\tilde{\mathbf{f}}_T \mathbf{a}_*^H + \bar{\mathbf{f}}_* \tilde{\mathbf{a}}_T^H \right) \\ &= \mathbf{L}_B^T \left[\text{vec} \left(\mathbf{P}_B \tilde{\boldsymbol{\theta}}_{B_H} \boldsymbol{\theta}_{A_*}^T \mathbf{P}_A^H \right) + \text{vec} \left(\mathbf{P}_B \boldsymbol{\theta}_{B_*} \tilde{\boldsymbol{\theta}}_{A_H}^T \mathbf{P}_A^H \right) \right], \end{aligned} \quad (55)$$

due to the factor matrix estimates being Toeplitz. Also note that $\tilde{\boldsymbol{\theta}}_{A_H}$ and $\tilde{\boldsymbol{\theta}}_{B_H}$ denote the error in the *Toeplitz* parameters. Continuing from (55),

$$\begin{aligned} \tilde{\mathbf{r}}_S &\simeq \mathbf{L}_B^T \left[(\mathbf{P}_A^c \boldsymbol{\theta}_{A_*} \otimes \mathbf{P}_B) \tilde{\boldsymbol{\theta}}_{B_H} + (\mathbf{P}_A^c \otimes \mathbf{P}_B \boldsymbol{\theta}_{B_*}) \tilde{\boldsymbol{\theta}}_{A_H} \right] \\ &= \mathbf{L}_B^T (\mathbf{P}_A^c \otimes \mathbf{P}_B) \\ &\quad \times \left[(\boldsymbol{\theta}_{A_*} \otimes \mathbf{I}_{n_B}) \tilde{\boldsymbol{\theta}}_{B_H} + (\mathbf{I}_{n_A} \otimes \boldsymbol{\theta}_{B_*}) \tilde{\boldsymbol{\theta}}_{A_H} \right] \\ &= \mathbf{K} \begin{bmatrix} \tilde{\boldsymbol{\theta}}_{A_H} \\ \tilde{\boldsymbol{\theta}}_{B_H} \end{bmatrix}. \end{aligned} \quad (56)$$

Hence, we have the variance of the estimator as

$$\begin{aligned} \mathbf{C}_S &\triangleq \lim_{N \rightarrow \infty} N \mathbf{E} [\tilde{\mathbf{r}}_S \tilde{\mathbf{r}}_S^H] \\ &= \lim_{N \rightarrow \infty} N \mathbf{K} \mathbf{E} \left[\begin{bmatrix} \tilde{\boldsymbol{\theta}}_{A_H} \\ \tilde{\boldsymbol{\theta}}_{B_H} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\theta}}_{A_H}^T & \tilde{\boldsymbol{\theta}}_{B_H}^T \end{bmatrix} \right] \mathbf{K}^H \\ &= \mathbf{K} \mathbf{C}_\theta \mathbf{K}^H, \end{aligned} \quad (57)$$

where

$$\mathbf{C}_\theta \triangleq \lim_{N \rightarrow \infty} N \mathbb{E} \begin{bmatrix} \tilde{\boldsymbol{\theta}}_{\mathbf{A}_H} \tilde{\boldsymbol{\theta}}_{\mathbf{A}_H}^T & \tilde{\boldsymbol{\theta}}_{\mathbf{A}_H} \tilde{\boldsymbol{\theta}}_{\mathbf{B}_H}^T \\ \tilde{\boldsymbol{\theta}}_{\mathbf{B}_H} \tilde{\boldsymbol{\theta}}_{\mathbf{A}_H}^T & \tilde{\boldsymbol{\theta}}_{\mathbf{B}_H} \tilde{\boldsymbol{\theta}}_{\mathbf{B}_H}^T \end{bmatrix}. \quad (58)$$

Explicitly, the different blocks of (58) are found, through (48), according to (we illustrate by examining the lower-left block; the others follow, mutatis mutandis):

$$\begin{aligned} \lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\boldsymbol{\theta}}_{\mathbf{B}_H} \tilde{\boldsymbol{\theta}}_{\mathbf{A}_H}^T] &= \left(\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, H} \mathbf{P}_B \right)^{-1} \mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, H} \\ &\quad \times \lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H] \\ &\quad \times \mathbf{W}_{\hat{\mathbf{a}}, H} \mathbf{P}_A \left(\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}}, H} \mathbf{P}_A \right)^{-1}. \end{aligned} \quad (59)$$

We observe that \mathbf{C}_θ is invariant to the forward-backward averaging of Section 4:

LEMMA 1

\mathbf{C}_θ does not depend on whether its constituting blocks, as e.g. given by (59), are found based on $\{\hat{\mathbf{a}}_{(2)}, \hat{\mathbf{f}}_{(3)}\}$ or $\{\hat{\mathbf{a}}_{\text{PS}(2)}, \hat{\mathbf{f}}_{\text{PS}(3)}\}$.

Proof: See Appendix D. ■

The minimum achievable variance of any unbiased estimator of \mathbf{R}_0 is given by the Cramér-Rao Bound. This bound is in [15] shown to be

$$\begin{aligned} \mathbf{C}_{\text{CRB}} &= \frac{1}{N} \mathbf{K} \left[\mathbf{K}^H (\mathbf{R}_0^{-T} \otimes \mathbf{R}_0^{-1}) \mathbf{K} \right]^\dagger \mathbf{K}^H \\ &\triangleq \mathbf{K}(\text{FIM})^\dagger \mathbf{K}^H, \end{aligned} \quad (60)$$

where FIM is the Fisher information matrix for the data under the assumption that the factor matrices are Toeplitz. We can now formulate the following theorem:

THEOREM 4

Given the data model in Section 2, the estimates given by the SNIFF algorithm (given in Section 5.3) are asymptotically statistically efficient.

Proof: In Appendix E it is shown that $\mathbf{K} \mathbf{C}_\theta \mathbf{K}^H - \mathbf{K}(\text{FIM})^\dagger \mathbf{K}^H = \mathbf{0}$. ■

Note that, from (B.13), (48), and (50), the estimates of the factor matrices given by SNIFF are asymptotically Gaussian; together with asymptotic consistency and efficiency, SNIFF is thus asymptotically (in N) ML. Hence, SNIFF can be used to obtain asymptotically ML estimates of \mathbf{R}_0^{-1} , as studied in e.g. [9], as well.

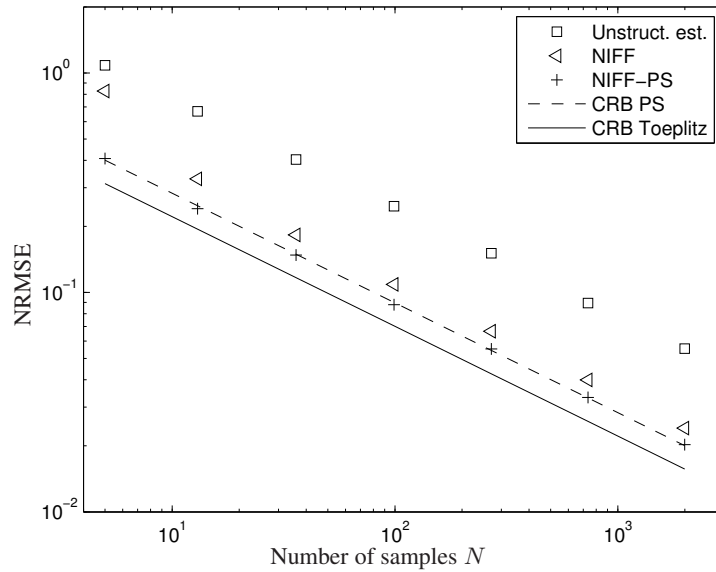


Figure 1: Normalized RMSE of three estimators; Unstructured (sample average) estimator, NIFF and NIFF-PS as a function of number of samples. \mathbf{A}_0 is 4×4 , \mathbf{B}_0 is 4×4 ; both are Toeplitz structured. Averages taken over $L = 100$ Monte Carlo simulations for each sample size.

7 Numerical simulations

We perform numerical simulations in order to evaluate the performance of the methods proposed in Section 4 and Section 5. We compare these results to those of previously published estimators, and also to the Cramér Rao Bound [15].

In our simulations we generate random complex positive definite hermitian Toeplitz structured matrices \mathbf{A}_0 and \mathbf{B}_0 , and then find \mathbf{R}_0 from these. \mathbf{R}_0 is kept fixed during the different Monte Carlo simulations. In each MC-simulation iteration we generate N samples from a complex Gaussian distribution with covariance \mathbf{R}_0 , and from these samples we find $\hat{\mathbf{R}}$, (3). This is the sufficient statistic available to the respective algorithms. The different estimators show similar relative performance and convergence behavior for different realizations of \mathbf{R}_0 as the ones presented.

Our performance measure is the normalized root-mean square error (NRMSE)

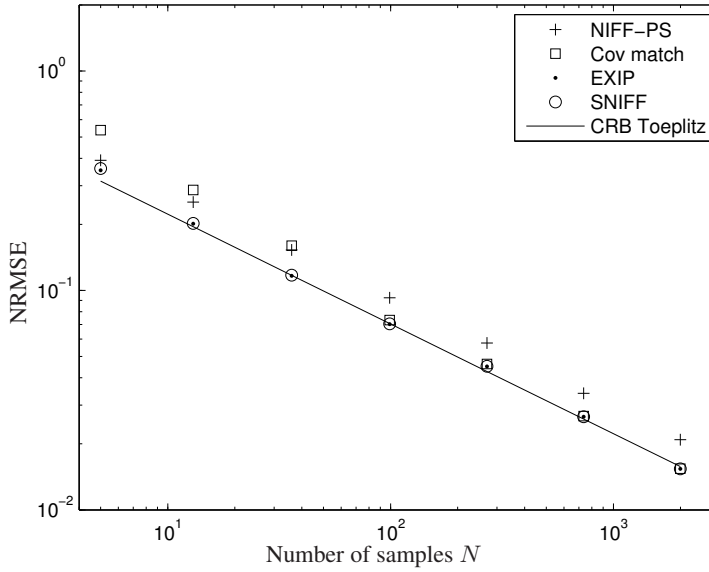


Figure 2: Normalized RMSE of four estimators as a function of number of samples. \mathbf{A}_0 is 4×4 , \mathbf{B}_0 is 4×4 ; both are Toeplitz structured. Averages taken over $L = 100$ Monte Carlo simulations for each sample size.

of the estimates, according to

$$\text{NRMSE} = \sqrt{\frac{1}{L} \sum_{k=1}^L \frac{\|\hat{\mathbf{R}}_k - \mathbf{R}_0\|_F^2}{\|\mathbf{R}_0\|_F^2}}, \quad (61)$$

where L is the number of Monte Carlo simulations, \mathbf{R}_0 is the true covariance matrix, $\hat{\mathbf{R}}_k$ is the estimate from Monte Carlo simulation k and $\|\cdot\|_F^2$ is the Frobenius norm according to $\|\mathbf{A}\|_F^2 = \sum_{i,j} |\mathbf{A}_{ij}|^2$.

In Fig. 1 we compare three estimators: the unstructured, sample average estimator (3); the Non-Iterative Flip-Flop (“NIFF”, see [15] and Section 3); and finally the improved NIFF-PS, introduced in [16] and further explored in Section 4, which exploits the persymmetry of the factor matrices.

As can be expected, we see that estimators utilizing more of the structure of the estimation problem performs better than those using less. It should also be noted that NIFF-PS is reaching its asymptotical accuracy already at few samples. Though, this bound is not the optimal bound for the current problem, since the method is not guaranteed to fully reproduce the existing Toeplitz structure. For comparison, the CRB corresponding to persymmetric factor matrices has also been plotted in Fig. 1, and the method is seen to attain this.

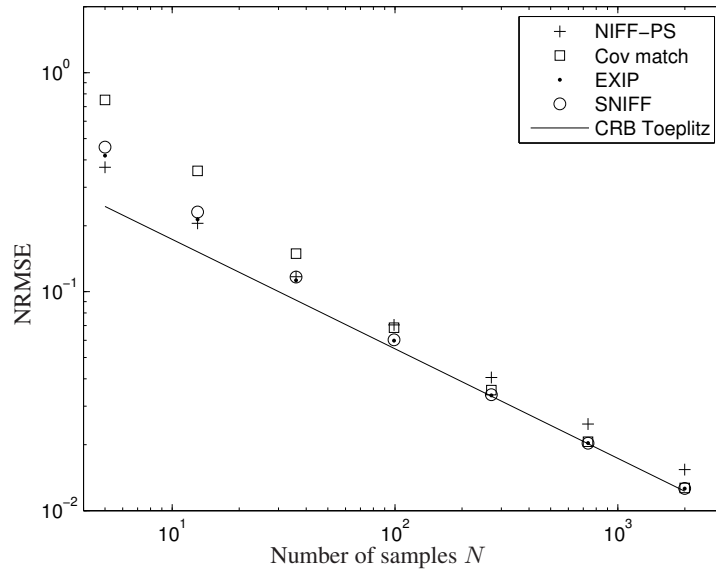


Figure 3: Normalized RMSE of four estimators as a function of number of samples. \mathbf{A}_0 is 8×8 , \mathbf{B}_0 is 8×8 ; both are Toeplitz structured. Averages taken over $L = 100$ Monte Carlo simulations for each sample size.

In Fig. 2 we repeat the estimation scenario, but with a different set of estimators: NIFF-PS from the previous figure for comparison; the covariance matching method [15]; the EXIP-estimator from Section 5.2; and finally, SNIFF, as proposed in Section 5.3. Note that the latter three are all asymptotically statistically efficient.

In the figure it is seen that the proposed method SNIFF is performing significantly better than the covariance matching approach of [15] at lower sample numbers; SNIFF and EXIP are apparently both in the asymptotic region already when $N \geq 10$. We can also see that SNIFF is as accurate as the EXIP-estimator. As predicted, SNIFF, EXIP, and covariance matching all attain the theoretical performance limit as the number of samples grow. We can also see that the NIFF-PS algorithm is better than the covariance matching one for small sample sizes, but as we increase the amount of data the former is surpassed in accuracy by the efficient estimator, as it is not able to take the Toeplitz structure into account.

In Fig. 3, we have increased the matrix sizes to create a more difficult estimation scenario; there we can see that the covariance-based estimators requires more data in order to attain the CRB. Additionally, when the number of samples is limited, SNIFF and EXIP suffers due to that the estimated weighting matrices used in

these algorithms are poor estimates of the actual ones. As the number of samples grows, this problem disappears.

Due to the finite number of MC realizations, there is variability in the NRMSEs. For $m = n = 4$ and $L = 100$ MC realizations, the variance of the MC induced error for each NRMSE is on the order of $\frac{0.1}{\sqrt{N}}\%$ of each shown NRMSE. This translates to that allowing for a 2 standard deviation outlier would still only entail a shift in the figures corresponding to approximately half the height of the markers.

8 Conclusion

In this article we have proven that a recently proposed method [24] is asymptotically, in N , ML, and we have more explicitly motivated the algorithm developed in [16]. Thus, both investigated estimators are statistically efficient in the respective estimation scenario they are designed for.

In addition, in numerous Monte-Carlo simulations, we have seen that the proposed method SNIFF is obtaining the asymptotic performance bound at smaller sample sizes than the previous state-of-the-art method [15].

Appendix A

Below we provide a proof of Theorem 1: Note that

$$\text{Tr}\{\hat{\mathbf{R}}(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})\} = \text{Tr}\left\{\sum_{k=1}^m \sum_{l=1}^m \hat{\mathbf{R}}^{kl}(\mathbf{A}^{-1})_{lk} \mathbf{B}^{-1}\right\}. \quad (\text{A.1})$$

Rewriting the last term in (16) as in (A.1), we find

$$\hat{\mathbf{B}} = \frac{1}{m} \sum_{k=1}^m \sum_{l=1}^m \hat{\mathbf{R}}_{\text{PS}}^{kl}(\mathbf{A}^{-1})_{lk}. \quad (\text{A.2})$$

We then have that

$$\begin{aligned} \hat{\mathbf{B}}_{i,j} &\equiv \mathbf{e}_i^T \hat{\mathbf{B}} \mathbf{e}_j \\ &= \frac{1}{m} \sum_{k=1}^m \sum_{l=1}^m \mathbf{e}_i^T \hat{\mathbf{R}}_{\text{PS}}^{kl} \mathbf{e}_j (\mathbf{A}^{-1})_{lk}, \end{aligned} \quad (\text{A.3})$$

where \mathbf{e}_i is the i th column of \mathbf{I}_m . If we introduce

$$\mathbf{C}^{i,j} = (\mathbf{I}_m \otimes \mathbf{e}_i^T) \hat{\mathbf{R}}_{\text{PS}} (\mathbf{I}_m \otimes \mathbf{e}_j), \quad (\text{A.4})$$

it is possible to rewrite (A.3) as

$$\begin{aligned}\widehat{\mathbf{B}}_{i,j} &= \frac{1}{m} \text{Tr} [\mathbf{C}^{i,j} \mathbf{A}^{-1}] = \frac{1}{m} \text{Tr} [\mathbf{C}^{i,j} \mathbf{J}_m \mathbf{A}^{-\text{T}} \mathbf{J}_m] \\ &= \frac{1}{m} \text{Tr} [\mathbf{J}_m \mathbf{C}^{i,j} \mathbf{J}_m \mathbf{A}^{-\text{T}}],\end{aligned}\tag{A.5}$$

where the persymmetry of \mathbf{A} and the properties of the trace-operator have been used. Looking at $\mathbf{J}_m \mathbf{C}^{i,j} \mathbf{J}_m$ from (A.5),

$$\begin{aligned}\mathbf{J}_m \mathbf{C}^{i,j} \mathbf{J}_m &= (\mathbf{J}_m \otimes \mathbf{e}_i^{\text{T}}) \widehat{\mathbf{R}}_{\text{PS}} (\mathbf{J}_m \otimes \mathbf{e}_j) \\ &= (\mathbf{I}_m \otimes \mathbf{e}_i^{\text{T}} \mathbf{J}_n) \mathbf{J}_{mn} \widehat{\mathbf{R}}_{\text{PS}} \mathbf{J}_{mn} (\mathbf{I}_m \otimes \mathbf{J}_n \mathbf{e}_j) \\ &= (\mathbf{I}_m \otimes \mathbf{e}_{n-i+1}^{\text{T}}) \widehat{\mathbf{R}}_{\text{PS}}^{\text{T}} (\mathbf{I}_m \otimes \mathbf{e}_{n-j+1}) \\ &= (\mathbf{C}^{n-j+1, n-i+1})^{\text{T}},\end{aligned}\tag{A.6}$$

where we have made extensive use of (14). Now, using the result of (A.6) in (A.5) gives

$$\begin{aligned}\widehat{\mathbf{B}}_{i,j} &= \frac{1}{m} \text{Tr} [(\mathbf{C}^{n-j+1, n-i+1})^{\text{T}} \mathbf{A}^{-\text{T}}] \\ &= \frac{1}{m} \text{Tr} [(\mathbf{C}^{n-j+1, n-i+1}) \mathbf{A}^{-1}] \\ &= \widehat{\mathbf{B}}_{n-j+1, n-i+1},\end{aligned}\tag{A.7}$$

which is the elementwise definition of $\widehat{\mathbf{B}}$ being PS. The analysis is equivalent for $\widehat{\mathbf{A}}$, mutatis mutandis.

Appendix B

We begin by defining the error in the estimates; according to the analysis in Section 5.1, let

$$\begin{aligned}\tilde{\mathbf{f}}_{(1)} &= \widehat{\mathbf{f}}_{(1)} - \bar{\mathbf{f}}_* = \frac{1}{m} [\widehat{\mathbf{R}}_{\text{B}} - \mathbf{R}_{\text{B}}] \mathbf{a}_{\text{INIT}}^i \\ &\triangleq \frac{1}{m} \tilde{\mathbf{R}}_{\text{B}} \mathbf{a}_{\text{INIT}}^i.\end{aligned}\tag{B.1}$$

Here, and henceforth, we have used the notation $\tilde{\cdot}$ to denote an error in an estimated quantity, according to

$$\tilde{\mathbf{X}} \triangleq \widehat{\mathbf{X}} - \mathbf{X}\tag{B.2}$$

for any quantity \mathbf{X} . Next, note that we can write

$$\begin{aligned}\tilde{\mathbf{R}}_{\text{B}} \mathbf{a}_{\text{INIT}}^i &= ((\mathbf{a}_{\text{INIT}}^i)^{\text{T}} \otimes \mathbf{I}_{n^2}) \text{vec}(\tilde{\mathbf{R}}_{\text{B}}) \\ &= ((\mathbf{a}_{\text{INIT}}^i)^{\text{T}} \otimes \mathbf{I}_{n^2}) \mathbf{L}_{\text{B}} \text{vec}(\tilde{\mathbf{R}}) = \mathbf{T}_{\text{B}1} \tilde{\mathbf{r}},\end{aligned}\tag{B.3}$$

where $\tilde{\mathbf{R}}$ follows from (B.2), the identity

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B}), \quad (\text{B.4})$$

which is valid for any matrices \mathbf{A} , \mathbf{B} and \mathbf{C} of suitable dimensions, has been used, and we defined $\mathbf{T}_{B1} \triangleq ((\mathbf{a}_{\text{INIT}}^i)^T \otimes \mathbf{I}_{n^2}) \mathbf{L}_B$. The relation (B.4) will be used extensively throughout the current section. With the result of (B.3) we can find

$$\begin{aligned} \text{cov}(\hat{\mathbf{f}}_{(1)}) &= \text{E} \left[\tilde{\mathbf{f}}_{(1)} \tilde{\mathbf{f}}_{(1)}^H \right] \\ &= \frac{1}{m^2} \mathbf{T}_{B1} \text{E} [\tilde{\mathbf{r}} \tilde{\mathbf{r}}^H] \mathbf{T}_{B1}^H. \end{aligned} \quad (\text{B.5})$$

Using the well-known (see, e.g., [25]) result that

$$\text{E} \left[\text{vec}(\tilde{\mathbf{R}}) \text{vec}^H(\tilde{\mathbf{R}}) \right] = \frac{1}{N} (\mathbf{R}^T \otimes \mathbf{R}), \quad (\text{B.6})$$

we find

$$\text{cov}(\hat{\mathbf{f}}_{(1)}) = \frac{1}{Nm^2} \mathbf{T}_{B1} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{T}_{B1}^H, \quad (\text{B.7})$$

which can be simplified to

$$\text{cov}(\hat{\mathbf{f}}_{(1)}) = \frac{\text{Tr}(\mathbf{A}_I^2)}{Nm^2} (\mathbf{B}_*^T \otimes \mathbf{B}_*), \quad (\text{B.8})$$

where $\mathbf{A}_I = \mathbf{A}_* \mathbf{A}_{\text{INIT}}^{-1}$. Proceeding to the covariance of the error in the second estimate, we find (since $\tilde{\mathbf{r}} = O_p(1/\sqrt{N})$)

$$\begin{aligned} \tilde{\mathbf{a}}_{(2)} &= \hat{\mathbf{a}}_{(2)} - \mathbf{a}_* = \frac{1}{n} \hat{\mathbf{R}}_A \hat{\mathbf{f}}_{(1)}^i - \frac{1}{n} \mathbf{R}_A \bar{\mathbf{f}}_*^i \\ &\simeq \frac{1}{n} \tilde{\mathbf{R}}_A \bar{\mathbf{f}}_*^i - \frac{1}{n} \mathbf{R}_A \text{vec}(\mathbf{B}_*^{-1} \tilde{\mathbf{B}}_{(1)} \mathbf{B}_*^{-1}), \end{aligned} \quad (\text{B.9})$$

using the first order approximation

$$(\mathbf{B}_* + \tilde{\mathbf{B}}_{(1)})^{-1} \simeq \mathbf{B}_*^{-1} - \mathbf{B}_*^{-1} \tilde{\mathbf{B}}_{(1)} \mathbf{B}_*^{-1}. \quad (\text{B.10})$$

The first term in (B.9) can be written

$$\begin{aligned} \tilde{\mathbf{R}}_A \bar{\mathbf{f}}_*^i &= ((\bar{\mathbf{f}}_*^i)^T \otimes \mathbf{I}_{m^2}) \text{vec}(\tilde{\mathbf{R}}_A) \\ &= ((\bar{\mathbf{f}}_*^i)^T \otimes \mathbf{I}_{m^2}) \mathbf{L}_A \tilde{\mathbf{r}} \\ &= \mathbf{T}_{A2} \tilde{\mathbf{r}}, \end{aligned} \quad (\text{B.11})$$

where $\mathbf{T}_{A2} \triangleq ((\tilde{\mathbf{f}}_*^i)^T \otimes \mathbf{I}_{m^2}) \mathbf{L}_A$. We can also simplify the second term in (B.9) according to

$$\begin{aligned}
 \mathbf{R}_A \text{vec}(\mathbf{B}_*^{-1} \tilde{\mathbf{B}}_{(1)} \mathbf{B}_*^{-1}) &= \\
 &= \mathbf{R}_A (\mathbf{B}_*^{-T} \otimes \mathbf{B}_*^{-1}) \tilde{\mathbf{f}}_{(1)} \\
 &= \mathbf{a}_* \tilde{\mathbf{f}}_*^H (\mathbf{B}_*^{-T} \otimes \mathbf{B}_*^{-1}) \tilde{\mathbf{f}}_{(1)} \\
 &= \mathbf{a}_* \text{vec}^H(\mathbf{B}_*^{-1}) \tilde{\mathbf{f}}_{(1)} \\
 &= \frac{1}{m} \mathbf{a}_* (\tilde{\mathbf{f}}_*^i)^H \tilde{\mathbf{R}}_B \mathbf{a}_*^i \\
 &= \frac{1}{m} \mathbf{a}_* (\tilde{\mathbf{f}}_*^i)^H \mathbf{T}_{B1} \tilde{\mathbf{r}}. \tag{B.12}
 \end{aligned}$$

With (B.11) and (B.12), we can write (B.9) as

$$\begin{aligned}
 \tilde{\mathbf{a}}_{(2)} &\simeq \frac{1}{n} \mathbf{T}_{A2} \tilde{\mathbf{r}} - \frac{1}{mn} \mathbf{a}_* (\tilde{\mathbf{f}}_*^i)^H \mathbf{T}_{B1} \tilde{\mathbf{r}} \\
 &= \tilde{\mathbf{\Xi}}_2 \tilde{\mathbf{r}}, \tag{B.13}
 \end{aligned}$$

if we define

$$\tilde{\mathbf{\Xi}}_2 \triangleq \frac{1}{n} \left[\mathbf{T}_{A2} - \frac{1}{m} \mathbf{a}_* (\tilde{\mathbf{f}}_*^i)^H \mathbf{T}_{B1} \right]. \tag{B.14}$$

Thus,

$$\lim_{N \rightarrow \infty} N \text{cov}(\hat{\mathbf{a}}_{(2)}) = \tilde{\mathbf{\Xi}}_2 (\mathbf{R}^T \otimes \mathbf{R}) \tilde{\mathbf{\Xi}}_2^H, \tag{B.15}$$

which can be simplified to

$$\begin{aligned}
 \lim_{N \rightarrow \infty} N \text{cov}(\hat{\mathbf{a}}_{(2)}) &= \frac{1}{n} \left[(\mathbf{A}_*^T \otimes \mathbf{A}_*) + \frac{\text{Tr}(\mathbf{A}_I^2)}{m^2} \mathbf{a}_* \mathbf{a}_*^H \right. \\
 &\quad \left. - \frac{1}{m} (\text{vec}(\mathbf{A}_I \mathbf{A}_*) \mathbf{a}_*^H \right. \\
 &\quad \left. + \mathbf{a}_* \text{vec}^H(\mathbf{A}_I \mathbf{A}_*)) \right]. \tag{B.16}
 \end{aligned}$$

By noting that, per definition, $\text{Tr}(\mathbf{A}_I) = m$, it is easily established that $\text{vec}(\mathbf{A}_{\text{INIT}}^{-1})$ spans the null-space of (B.16), and thus that $\lim_{N \rightarrow \infty} N \text{cov}(\hat{\mathbf{a}}_{(2)})$ is singular.

Continuing with the third iterate in NIFF, we see that

$$\begin{aligned}
 \tilde{\mathbf{f}}_{(3)} &= \hat{\mathbf{f}}_{(3)} - \tilde{\mathbf{f}}_* = \frac{1}{m} \hat{\mathbf{R}}_B \text{vec} \left((\hat{\mathbf{A}}_{(2)})^{-1} \right) - \frac{1}{m} \mathbf{R}_B \mathbf{a}_*^i \\
 &\stackrel{(a)}{\simeq} \frac{1}{m} \tilde{\mathbf{R}}_B \mathbf{a}_*^i - \frac{1}{m} \mathbf{R}_B \text{vec}(\mathbf{A}_*^{-1} \tilde{\mathbf{A}}_{(2)} \mathbf{A}_*^{-1}) \\
 &= \frac{1}{m} \tilde{\mathbf{R}}_B \mathbf{a}_*^i - \frac{1}{m} \mathbf{R}_B (\mathbf{A}_*^{-T} \otimes \mathbf{A}_*^{-1}) \tilde{\mathbf{a}}_{(2)} \\
 &= \frac{1}{m} \tilde{\mathbf{R}}_B \mathbf{a}_*^i - \frac{1}{m} \tilde{\mathbf{f}}_* (\mathbf{a}_*^i)^H \tilde{\mathbf{a}}_{(2)}, \tag{B.17}
 \end{aligned}$$

where (a) follows from retaining the dominant error terms, analogously to (B.10). Similarly to (B.3) and (B.11) we rewrite the first term in (B.17) as

$$\tilde{\mathbf{R}}_{\mathbf{B}_*} \mathbf{a}_*^i = \left((\mathbf{a}_*^i)^\top \otimes \mathbf{I}_{n^2} \right) \mathbf{L}_{\mathbf{B}} \tilde{\mathbf{r}} = \mathbf{T}_{\mathbf{B}_3} \tilde{\mathbf{r}}, \quad (\text{B.18})$$

and insert this together with (B.13) in (B.17) to get

$$\begin{aligned} \tilde{\mathbf{f}}_{(3)} &\simeq \frac{1}{m} \mathbf{T}_{\mathbf{B}_3} \tilde{\mathbf{r}} - \frac{1}{m} \bar{\mathbf{f}}_* (\mathbf{a}_*^i)^H \boldsymbol{\Xi}_2 \tilde{\mathbf{r}} \\ &\triangleq \boldsymbol{\Xi}_3 \tilde{\mathbf{r}}, \end{aligned} \quad (\text{B.19})$$

where

$$\boldsymbol{\Xi}_3 \triangleq \frac{1}{m} \left[\mathbf{T}_{\mathbf{B}_3} - \bar{\mathbf{f}}_* (\mathbf{a}_*^i)^H \boldsymbol{\Xi}_2 \right]. \quad (\text{B.20})$$

Thus,

$$\lim_{N \rightarrow \infty} N \text{cov}(\hat{\mathbf{f}}_{(3)}) = \boldsymbol{\Xi}_3 (\mathbf{R}^\top \otimes \mathbf{R}) \boldsymbol{\Xi}_3^H, \quad (\text{B.21})$$

which can be simplified to read

$$\begin{aligned} \lim_{N \rightarrow \infty} N \text{cov}(\hat{\mathbf{f}}_{(3)}) &= \frac{1}{m} \left[(\mathbf{B}_*^\top \otimes \mathbf{B}_*) \right. \\ &\quad \left. + \frac{\text{Tr}(\mathbf{A}_1^2) - m}{nm} \bar{\mathbf{f}}_* \bar{\mathbf{f}}_*^H \right]. \end{aligned} \quad (\text{B.22})$$

Using the fact that the trace of a product of two matrices is an inner product, we have from the Cauchy-Schwarz inequality (see e.g. [26]) that $\text{Tr}(\mathbf{A}_1^2) \text{Tr}(\mathbf{I}_m) \geq \text{Tr}(\mathbf{A}_1)^2$, implying that $\text{Tr}(\mathbf{A}_1^2) - m \geq 0$ since $\text{Tr}(\mathbf{A}_1) = m$. Then (B.22) is p.d., since \mathbf{B}_* is p.d.

The covariance between $\tilde{\mathbf{f}}_3$ and $\tilde{\mathbf{a}}_2$ contains several terms:

$$\begin{aligned} \lim_{N \rightarrow \infty} N \text{E}(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H) &= \frac{1}{m} \left[\frac{1}{n} \mathbf{T}_{\mathbf{B}_3} (\mathbf{R}^\top \otimes \mathbf{R}) \mathbf{T}_{\mathbf{A}_2}^H \right. \\ &\quad \left. - \frac{1}{mn} \mathbf{T}_{\mathbf{B}_3} (\mathbf{R}^\top \otimes \mathbf{R}) \mathbf{T}_{\mathbf{B}_1}^H \bar{\mathbf{f}}_*^i \mathbf{a}_*^H \right. \\ &\quad \left. - \bar{\mathbf{f}}_* (\mathbf{a}_*^i)^H \lim_{N \rightarrow \infty} N \text{cov}(\tilde{\mathbf{a}}_{(2)}) \right]. \end{aligned} \quad (\text{B.23})$$

The first term in (B.23) can be found as

$$\frac{1}{mn} \mathbf{T}_{\mathbf{B}_3} (\mathbf{R}^\top \otimes \mathbf{R}) \mathbf{T}_{\mathbf{A}_2}^H = \frac{1}{mn} \bar{\mathbf{f}}_* \mathbf{a}_*^H. \quad (\text{B.24})$$

One finds the second term as

$$-\frac{1}{m^2 n} \mathbf{T}_{\mathbf{B}_3} (\mathbf{R}^\top \otimes \mathbf{R}) \mathbf{T}_{\mathbf{B}_1}^H \bar{\mathbf{f}}_*^i \mathbf{a}_*^H = -\frac{1}{mn} \bar{\mathbf{f}}_* \mathbf{a}_*^H, \quad (\text{B.25})$$

which together with (B.24) gives

$$\begin{aligned}
 \lim_{N \rightarrow \infty} N E(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H) &= \lim_{N \rightarrow \infty} -\frac{N}{m} \bar{\mathbf{f}}_*(\mathbf{a}_*^i)^H \text{cov}(\tilde{\mathbf{a}}_2) \\
 &= -\frac{1}{mn} \bar{\mathbf{f}}_*(\mathbf{a}_*^i)^H \left[(\mathbf{A}_*^T \otimes \mathbf{A}_*) + \frac{\text{Tr}(\mathbf{A}_I^2)}{m^2} \mathbf{a}_* \mathbf{a}_*^H \right. \\
 &\quad \left. - \frac{1}{m} (\text{vec}(\mathbf{A}_I \mathbf{A}_*) \mathbf{a}_*^H + \mathbf{a}_* \text{vec}^H(\mathbf{A}_I \mathbf{A}_*)) \right] \\
 &= \frac{1}{mn} \bar{\mathbf{f}}_* \left[\text{vec}^H(\mathbf{A}_I \mathbf{A}_*) - \frac{\text{Tr}(\mathbf{A}_I^2)}{m} \mathbf{a}_*^H \right]. \tag{B.26}
 \end{aligned}$$

Appendix C

When using the FB-averaging, we have that

$$\hat{\mathbf{R}}_{\text{PS}} = \mathbf{R}_{\text{PS}} + \tilde{\mathbf{R}}_{\text{PS}} = \mathbf{R}_0 + \tilde{\mathbf{R}}_{\text{PS}}, \tag{C.1}$$

since \mathbf{R}_0 is PS. It is straight-forward to rewrite (C.1) to find

$$\tilde{\mathbf{R}}_{\text{PS}} = \frac{1}{2} (\tilde{\mathbf{R}} + \mathbf{J} \tilde{\mathbf{R}}^T \mathbf{J}), \tag{C.2}$$

where $\tilde{\mathbf{R}} = \hat{\mathbf{R}} - \mathbf{R}_0$. Vectorizing (C.2) then gives

$$\begin{aligned}
 \tilde{\mathbf{r}}_{\text{PS}} &= \frac{1}{2} \text{vec} (\tilde{\mathbf{R}} + \mathbf{J} \tilde{\mathbf{R}}^T \mathbf{J}) = \frac{1}{2} (\tilde{\mathbf{r}} + (\mathbf{J} \otimes \mathbf{J}) \text{vec}(\tilde{\mathbf{R}}^T)) \\
 &= \frac{1}{2} (\mathbf{I} + (\mathbf{J} \otimes \mathbf{J}) \mathbf{L}_T) \tilde{\mathbf{r}} \triangleq \mathbf{T}_{\text{PS}} \tilde{\mathbf{r}}, \tag{C.3}
 \end{aligned}$$

in which the permutation matrix \mathbf{L}_T is defined such that $\mathbf{L}_T \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}^T)$ for any \mathbf{X} , and $\mathbf{T}_{\text{PS}} \triangleq \frac{1}{2} [\mathbf{I} + (\mathbf{J} \otimes \mathbf{J}) \mathbf{L}_T]$. Thus, we can modify (B.6) to take the PS structure into account by replacing $(\mathbf{R}^T \otimes \mathbf{R})$ with $\mathbf{T}_{\text{PS}} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{T}_{\text{PS}}^H$ in (B.7), (B.15), and (B.21).

We also note here that \mathbf{T}_{PS} is rank deficient. This can be seen from the fact that adding the matrix $(\mathbf{J} \otimes \mathbf{J}) \mathbf{L}_T$ to an identity matrix, as in (C.3), introduces off-diagonal unity elements; the columns of the resulting matrix are thus no longer linearly independent.

Appendix D

We prove Lemma 1 by exemplifying through the off-diagonal block (59) in (58); the remaining blocks of (58) follow from similar arguments. We thus want to show that

$$\Theta_{\text{PS}} = \Theta, \tag{D.1}$$

where, and below, we explicitly denote the persymmetric inheritance of a quantity through the subscript PS. The quantities of (D.1) are thus given by

$$\begin{aligned}
\Theta_{\text{PS}} &\triangleq \lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\boldsymbol{\theta}}_{\text{BPS}} \tilde{\boldsymbol{\theta}}_{\text{APS}}^H] \\
&= \left(\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \mathbf{P}_B \right)^{-1} \mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \\
&\quad \times \lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\mathbf{f}}_{\text{PS}(3)} \tilde{\mathbf{a}}_{\text{PS}(2)}^H] \\
&\quad \times \mathbf{W}_{\hat{\mathbf{a}}, \text{H}} \mathbf{P}_A \left(\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}}, \text{H}} \mathbf{P}_A \right)^{-1}, \tag{D.2}
\end{aligned}$$

and

$$\begin{aligned}
\Theta &\triangleq \lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\boldsymbol{\theta}}_{\text{Bh}} \tilde{\boldsymbol{\theta}}_{\text{Ah}}^H] \\
&= \left(\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \mathbf{P}_B \right)^{-1} \mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \\
&\quad \times \lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H] \\
&\quad \times \mathbf{W}_{\hat{\mathbf{a}}, \text{H}} \mathbf{P}_A \left(\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}}, \text{H}} \mathbf{P}_A \right)^{-1}, \tag{D.3}
\end{aligned}$$

where the subscript h denotes hermitian. From (40) and (42), one immediately finds

$$\begin{aligned}
\lim_{N \rightarrow \infty} N \mathbb{E}[\tilde{\mathbf{f}}_{\text{PS}(3)} \tilde{\mathbf{a}}_{\text{PS}(2)}^H] &= \frac{1}{2} \lim_{N \rightarrow \infty} N \mathbb{E} \left(\tilde{\mathbf{f}}_{(3)} \tilde{\mathbf{a}}_{(2)}^H \right) \\
&\quad + \frac{1}{4} \boldsymbol{\Xi}_{\hat{\mathbf{f}}} \mathbf{J} \mathbf{L}_{\text{T}} \left(\mathbf{R}^{\text{T}} \otimes \mathbf{R} \right) \boldsymbol{\Xi}_{\hat{\mathbf{a}}}^H \\
&\quad + \frac{1}{4} \boldsymbol{\Xi}_{\hat{\mathbf{f}}} \left(\mathbf{R}^{\text{T}} \otimes \mathbf{R} \right) \mathbf{L}_{\text{T}}^{\text{T}} \mathbf{J} \boldsymbol{\Xi}_{\hat{\mathbf{a}}}^H, \tag{D.4}
\end{aligned}$$

since $\mathbf{L}_{\text{T}} (\mathbf{X} \otimes \mathbf{Y}) \mathbf{L}_{\text{T}}^{\text{T}} = (\mathbf{Y} \otimes \mathbf{X})$ holds for arbitrary \mathbf{X}, \mathbf{Y} of appropriate dimensions and $\mathbf{J} \mathbf{X} \mathbf{J} = \mathbf{X}^{\text{T}}$ for \mathbf{X} PS. These relations, together with the (easily shown) equality $\mathbf{J} \mathbf{L}_{\text{T}} = \mathbf{L}_{\text{T}}^{\text{T}} \mathbf{J}$, immediately give

$$\left(\mathbf{R}^{\text{T}} \otimes \mathbf{R} \right) \mathbf{L}_{\text{T}}^{\text{T}} \mathbf{J} = \mathbf{J} \mathbf{L}_{\text{T}} \left(\mathbf{R}^{\text{T}} \otimes \mathbf{R} \right), \tag{D.5}$$

which, used in (D.4), leads to that it is sufficient to show that

$$\begin{aligned}
&\left(\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \mathbf{P}_B \right)^{-1} \mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \boldsymbol{\Xi}_{\hat{\mathbf{f}}} \mathbf{J} \mathbf{L}_{\text{T}} \left(\mathbf{R}^{\text{T}} \otimes \mathbf{R} \right) = \\
&\left(\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \mathbf{P}_B \right)^{-1} \mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, \text{H}} \boldsymbol{\Xi}_{\hat{\mathbf{f}}} \left(\mathbf{R}^{\text{T}} \otimes \mathbf{R} \right) \tag{D.6}
\end{aligned}$$

in order to conclude (D.1). From the facts that $\mathbf{J} \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}^c)$ for any hermitian and persymmetric matrix \mathbf{X} , and that $\mathbf{J} \text{vec}(\mathbf{R}_A) = \text{vec}(\mathbf{R}_A^c)$, $\mathbf{J} \text{vec}(\mathbf{R}_B) =$

$\text{vec}(\mathbf{R}_B^c)$, we have from (21) and (19) that $\mathbf{J}\mathbf{L}_A\mathbf{J} = \mathbf{L}_A$, and $\mathbf{J}\mathbf{L}_B\mathbf{J} = \mathbf{L}_B$. These observations used in (43) gives that $\Xi_{\hat{\mathbf{f}}}\mathbf{J} = \mathbf{J}\Xi_{\hat{\mathbf{f}}}^c$.

From the definition of \mathbf{L}_T , we have by a similar analysis that $\mathbf{L}_{T_1}\mathbf{L}_A = \mathbf{L}_B\mathbf{L}_{T_2}$, where the subscripts on T denote that the matrices are not identical. Using these, and the above, relations, we have that $\Xi_{\hat{\mathbf{f}}}^c\mathbf{L}_T = \mathbf{L}_{T_3}\Xi_{\hat{\mathbf{f}}}$.

Equivalently to (D.5), $\mathbf{W}_{\hat{\mathbf{f}},H}^c\mathbf{J}\mathbf{L}_{T_3} = \mathbf{J}\mathbf{L}_{T_3}\mathbf{W}_{\hat{\mathbf{f}},H}^c$, since $\mathbf{W}_{\hat{\mathbf{f}},H}^c$ is given by (53). Finally, we have that $\mathbf{P}_B^H\mathbf{J}\mathbf{L}_{T_3} = \mathbf{P}_B^H$, for any selection matrices \mathbf{P}_B mapping the (real) parameters of a PS matrix.

Thus, we have verified (D.6), and hence (D.1). The analysis is equivalent, mutatis mutandis, when operating on $\Xi_{\hat{\mathbf{a}}}$, etc.

Appendix E

The proof is made according to the following sequence:

1. Find closed-form expressions for \mathbf{C}_θ (58), (59), FIM (60), and HIM $\triangleq \mathbf{C}_\theta^{-1}$.
2. Show that $\text{HIM}^{-1} = \text{FIM}^\dagger + \mathbf{G}$.
3. Show that $\mathbf{K}\mathbf{G}\mathbf{K}^H = \mathbf{0}$.
4. Conclude that $\mathbf{K}\mathbf{C}_\theta\mathbf{K}^H = \mathbf{K}(\text{FIM})^\dagger\mathbf{K}^H$.

According to (58), \mathbf{C}_θ is partitioned as

$$\mathbf{C}_\theta = \begin{bmatrix} \mathbf{C}_{\theta 11} & \mathbf{C}_{\theta 12} \\ \mathbf{C}_{\theta 21} & \mathbf{C}_{\theta 22} \end{bmatrix}. \quad (\text{E.1})$$

Evaluate each of these blocks in turn. Use (35) in (59) to find:

$$\begin{aligned} N\mathbf{C}_{\theta 11} &= \left(n\mathbf{P}_A^H\mathbf{W}_{\hat{\mathbf{a}},H}\mathbf{P}_A\right)^{-1} + \left(n\mathbf{P}_A^H\mathbf{W}_{\hat{\mathbf{a}},H}\mathbf{P}_A\right)^{-1}\mathbf{P}_A^H \\ &\quad \times \frac{n}{m} \left[\frac{\text{Tr}(\mathbf{A}_I^2)}{m} \mathbf{a}_*^i(\mathbf{a}_*^i)^H - \mathbf{a}_*^i(\mathbf{a}_{\text{INIT}}^i)^H - \mathbf{a}_{\text{INIT}}^i(\mathbf{a}_*^i)^H \right] \\ &\quad \times \mathbf{P}_A \left(n\mathbf{P}_A^H\mathbf{W}_{\hat{\mathbf{a}},H}\mathbf{P}_A\right)^{-1}. \end{aligned} \quad (\text{E.2})$$

Looking at $\mathbf{C}_{\theta 21}$, we thus have that, using (B.26) in (59),

$$\begin{aligned} N\mathbf{C}_{\theta 21} &= \frac{1}{mn} \left(\mathbf{P}_B^H\mathbf{W}_{\hat{\mathbf{f}},H}\mathbf{P}_B\right)^{-1} \mathbf{P}_B^H\mathbf{W}_{\hat{\mathbf{f}},H}\bar{\mathbf{f}}_* \\ &\quad \times \left[\text{vec}^H(\mathbf{A}_I\mathbf{A}_*) - \frac{\text{Tr}(\mathbf{A}_I^2)}{m} \mathbf{a}_*^H \right] \\ &\quad \times \mathbf{W}_{\hat{\mathbf{a}},H}\mathbf{P}_A \left(\mathbf{P}_A^H\mathbf{W}_{\hat{\mathbf{a}},H}\mathbf{P}_A\right)^{-1}. \end{aligned} \quad (\text{E.3})$$

Obviously, $\mathbf{C}_{\theta 12} = \mathbf{C}_{\theta 21}^H$. Using (36) in (59),

$$N\mathbf{C}_{\theta 22} = \left(m\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, H} \mathbf{P}_B \right)^{-1} + \frac{\text{Tr}(\mathbf{A}_I^2) - m}{n} \left[\left(m\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, H} \mathbf{P}_B \right)^{-1} \mathbf{P}_B^H \hat{\mathbf{f}}_*^i \right] \left[\cdot \right]^H, \quad (\text{E.4})$$

where $[\cdot]$ denotes the immediately preceding factor.

Now we look at finding the closed-form expression for the FIM. We have from (60) that

$$\frac{1}{N}\text{FIM} = \mathbf{K}^H (\mathbf{R}_0^{-T} \otimes \mathbf{R}_0^{-1}) \mathbf{K} \quad (\text{E.5})$$

which, due to the block partition of the second factor in (47), can be written as a block matrix. Evaluating these blocks in turn gives

$$\begin{aligned} \frac{1}{N}\text{FIM} &\triangleq \begin{bmatrix} \text{FIM}_{11} & \text{FIM}_{12} \\ \text{FIM}_{21} & \text{FIM}_{22} \end{bmatrix} \\ &= \begin{bmatrix} n\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}}, H} \mathbf{P}_A & \mathbf{P}_A^H \mathbf{a}_*^i (\hat{\mathbf{f}}_*^i)^H \mathbf{P}_B \\ \mathbf{P}_B^H \hat{\mathbf{f}}_*^i (\mathbf{a}_*^i)^H \mathbf{P}_A & m\mathbf{P}_B^H \mathbf{W}_{\hat{\mathbf{f}}, H} \mathbf{P}_B \end{bmatrix}. \end{aligned} \quad (\text{E.6})$$

Note that (E.6), as given by [15], can be verified to be rank-deficient.

LEMMA 2

The inverse of \mathbf{C}_θ is given by

$$\begin{aligned} \text{HIM} &\triangleq \mathbf{C}_\theta^{-1} \\ &= \begin{bmatrix} \text{FIM}_{11} + k_1 \mathbf{v} \mathbf{v}^H & \text{FIM}_{12} \\ \text{FIM}_{21} & \text{FIM}_{22} \end{bmatrix}, \end{aligned} \quad (\text{E.7})$$

where

$$\begin{aligned} k_1 &\triangleq \frac{1}{\text{Tr}(\mathbf{A}_I^2) - k_2}, \\ k_2 &\triangleq \mathbf{v}^H (\mathbf{P}_A^H \mathbf{W}_{\hat{\mathbf{a}}, H} \mathbf{P}_A)^{-1} \mathbf{v}, \\ \mathbf{v} &\triangleq \mathbf{P}_A^H \text{vec}(\mathbf{A}_{\text{INIT}}^{-1}). \end{aligned}$$

Proof: Verified by direct insertion (using (E.1)). ■

Note that we can equivalently write (E.7) as

$$\text{HIM} = \text{FIM} + k_1 \bar{\mathbf{v}} \bar{\mathbf{v}}^H, \quad (\text{E.8})$$

where $\bar{\mathbf{v}} = [\mathbf{v}^T \quad \mathbf{0}_{n_B}^T]^T$, i.e. \mathbf{v} zero-padded to match the dimension of FIM.

To find the inverse of HIM (in terms of FIM, i.e. based on (E.8)), we cannot use the standard matrix inversion lemmas due to FIM being singular. We instead use the approach of [27]; accordingly (since $(\text{HIM})^\dagger = \text{HIM}^{-1}$, due to HIM being non-singular),

$$\text{HIM}^{-1} = (\text{FIM})^\dagger + \mathbf{G}, \quad (\text{E.9})$$

where

$$\begin{aligned} \mathbf{G} \triangleq & -k_3 \text{FIM}^\dagger \bar{\mathbf{v}} \bar{\mathbf{v}}^H \mathbf{\Pi}_{\text{FIM}}^\perp - k_3 \mathbf{\Pi}_{\text{FIM}}^\perp \bar{\mathbf{v}} \bar{\mathbf{v}}^H \text{FIM}^\dagger \\ & + k_3^2 \mathbf{\Pi}_{\text{FIM}}^\perp \bar{\mathbf{v}} \bar{\mathbf{v}}^H \mathbf{\Pi}_{\text{FIM}}^\perp (k_1^{-1} + \bar{\mathbf{v}}^H \text{FIM}^\dagger \bar{\mathbf{v}}), \end{aligned} \quad (\text{E.10})$$

in which $\mathbf{\Pi}_{\text{FIM}}^\perp = \mathbf{I} - \text{FIM}(\text{FIM})^\dagger$, $k_3^{-1} \triangleq \bar{\mathbf{v}}^H \mathbf{\Pi}_{\text{FIM}}^\perp \bar{\mathbf{v}}$ and we have exploited the fact that FIM (and hence FIM^\dagger) is hermitian, and that $\mathbf{\Pi}_{\text{FIM}}^\perp$ is idempotent.

From [15] and [28], we have that \mathbf{K} lies in the range space of FIM; since $\mathbf{\Pi}_{\text{FIM}}^\perp$ in (E.10) is the orthogonal projector onto the null-space of FIM, we have that $\mathbf{\Pi}_{\text{FIM}}^\perp \mathbf{K}^H = \mathbf{0}$ and $\mathbf{K} \mathbf{\Pi}_{\text{FIM}}^\perp = \mathbf{0}$; hence, $\mathbf{K} \mathbf{G} \mathbf{K}^H = \mathbf{0}$. Thus,

$$\begin{aligned} \mathbf{K} \mathbf{C}_\theta \mathbf{K}^H &= \mathbf{K} (\text{HIM})^{-1} \mathbf{K}^H = \mathbf{K} (\text{FIM}^\dagger + \mathbf{G}) \mathbf{K}^H \\ &= \mathbf{K} (\text{FIM})^\dagger \mathbf{K}^H, \end{aligned} \quad (\text{E.11})$$

and the proof is concluded.

References

- [1] P. Dutilleul, "The mle algorithm for the matrix normal distribution," *J. Stat. Comp. and Simulation*, vol. 64, no. 2, pp. 105–123, 1999.
- [2] A.P. Dawid, "Some matrix-variate distribution theory: Notational considerations and a bayesian application," *Biometrika*, vol. 68, no. 1, pp. 265–274, Apr. 1981.
- [3] N. Lu and D.L. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Stat. and Probability Lett.*, vol. 73, no. 4, pp. 449–457, 2005.
- [4] J.P. Kermoal, L. Schumacher, K.I. Pedersen, P.E. Mogensen, and F. Frederiksen, "A stochastic mimo radio channel model with experimental validation," *IEEE J. on Select. Areas in Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.
- [5] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.
- [6] J.C. de Munck, H.M. Huizenga, L.J. Waldorp, and R.A. Heethaar, "Estimating stationary dipoles from meg/leeg data contaminated with spatially and temporally correlated background noise," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1565–1572, Jul. 2002.
- [7] P. Strobach, "Low rank detection of multichannel gaussian signals using a constrained inverse," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, 1994, vol. 4, pp. 245–248.
- [8] N. Lu and D.L. Zimmerman, "On likelihood-based inference for a separable covariance matrix," Technical Report 337, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa, 2004.
- [9] G.I. Allen and R. Tibshirani, "Transposable regularized covariance models with an application to missing data imputation," *Ann. Appl. Stat.*, vol. 4, no. 2, pp. 764–790, 2010.
- [10] T. Tsiligkaridis, A.O. Hero, and S. Zhou, "On convergence of kronecker graphical lasso algorithms," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1743–1755, Apr. 2013.
- [11] P. Stoica and R.L. Moses, *Spectral analysis of signals*, Prentice Hall, 2005.
- [12] J.A. McEwen and G.B. Anderson, "Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity," *IEEE Trans. Biomed. Eng.*, vol. BME-22, no. 5, pp. 361–369, Sep. 1975.
- [13] A.F. Molisch, "A generic model for MIMO wireless propagation channels in macro- and microcells," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 61–71, Jan. 2004.
- [14] T.A. Barton and D.R. Fuhrmann, "Covariance structures for multidimensional data," *Multidimensional Syst. and Signal Process.*, vol. 4, no. 2, pp. 111–123, Apr. 1993.
- [15] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with Kronecker product structure," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 478–491, Feb. 2008.

- [16] M. Jansson, P. Wirfält, K. Werner, and B. Ottersten, "ML estimation of covariance matrices with Kronecker and persymmetric structure," in *IEEE Digital Signal Process. Workshop and 5th IEEE Signal Process. Educ. Workshop, (DSP/SPE)*, Jan. 2009, pp. 298–301.
- [17] K. Werner and M. Jansson, "Estimating MIMO channel covariances from training data under the Kronecker model," *Signal Process.*, vol. 89, no. 1, pp. 1–13, Jan. 2009.
- [18] G. H. Golub and C.F. Van Loan, *Matrix Computations*, Baltimore, MD: The John Hopkins University Press, 3rd edition, 1996.
- [19] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1998.
- [20] T.S. Ferguson, *A course in large sample theory*, Chapman and Hall/CRC Texts in Statistical Science Series. Chapman and Hall, 1996.
- [21] M. Jansson and P. Stoica, "Forward-only and forward-backward sample covariances – A comparative study," *Signal Process.*, vol. 77, no. 3, pp. 235–245, Sep. 1999.
- [22] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Signal Process.*, vol. 17, no. 4, pp. 383–387, Aug. 1989.
- [23] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International (UK) Ltd, 1989.
- [24] P. Wirfält and M. Jansson, "On toeplitz and kronecker structured covariance matrix estimation," in *2010 IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, Jerusalem, Israel, Oct. 2010, pp. 185–188.
- [25] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Process.*, vol. 8, no. 3, pp. 185–210, Jul. 1998.
- [26] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [27] C. Meyer, Jr., "Generalized inversion of modified matrices," *SIAM J. on Appl. Math.*, vol. 24, no. 3, pp. 315–323, 1973.
- [28] P. Stoica and T.L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.