

Robust Estimation and Model Order Selection for Signal Processing

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Michael Muma, Dipl.-Ing.
geboren am 28.04.1981 in Mannheim (Deutschland)

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof., Dr. Tech. Visa Koivunen
Tag der Einreichung:	15.11.2013
Tag der mündlichen Prüfung:	30.01.2014

D 17
Darmstadt, 2014

To my Family and Friends.

Acknowledgments

I would like to thank all the people who helped and inspired me in the last few years. Thanks to you, my doctoral study was a real pleasure.

I especially thank Prof. Dr.-Ing. Abdelhak Zoubir for his supervision. Prof. Zoubir provided me with the necessary freedom in research and guidance. His outstanding degree of enthusiasm and motivation helped me to accomplish things I would not have accomplished, otherwise.

I wholeheartedly thank Prof. Visa Koivunen for his co-supervision and his valuable scientific comments and Prof. Dr.-Ing Klaus Hofmann, Prof. Dr.-Ing. Ulrich Konigorski and Prof. Dr.-Ing Marius Pesavento for their hard work as Board of Examiners.

Special thanks go to Renate Koschella, Stefan Leier, Nevine Demitri, Dr.-Ing. Yacine Chakhchoukh, Dr.-Ing. Weaam Alkhalidi, Nils Bornhorst, Dr.-Ing. Uli Hammes, Dr.-Ing. Waqas Sharif, Tim Schck, Hauke Fath, Dr.-Ing. Raquel Fandos, Dr.-Ing. Christian Debes, Dr.-Ing. Philipp Heidenreich, Dr.-Ing. Marco Moebus, Jürgen Hahn, Sahar Khawatmi, Michael Leigsnering, Sara Al-Sayed, Dr.-Ing. Fiky Suratman, Zhihua Lu, Gökhan Gül, Feng Yin, Christian Weiss, Dr.-Ing. Ahmed Mostafa, Dr.-Ing. Gebremichael Teame, Mouhammad Alhumaidi, Michael Fauss, Adrian Sobic, Simon Rosenkranz, Wassim Suleiman, Dr.-Ing. Ramon Brcic, Roy Howard and ALL my colleagues and former colleagues at the Signal Processing Group at TU Darmstadt who provided me every day with a pleasant and lively environment.

I am grateful to Prof. em. Dr.-Ing. Eberhard Hänsler for sharing his valuable thoughts, to Prof. Dr.-Ing. Henning Puder and Dr.-Ing. Tobias Rosenkranz, not only for their input on speech signal processing. I wish to thank Prof. Dr. Augustin Kelava and Marlene Schmitt for creating a truly interdisciplinary research and introducing me to psychophysiology. Thank you Prof. Dr. D. Robert Iskander for your very special and personal guidance, much of what I do today stems from what I learned under your supervision. Many thanks go to Dr. Mengling Feng who introduced me to some interesting real-world biomedical tasks that require robustness. I am also grateful to Dr.-Ing Florian Roemer, Yao Cheng and Prof. Dr.-Ing. Martin Haardt for introducing me gently into the world of tensors. I also wish to thank Dr.-Ing. Holger Maune, a man of action who transforms a slowly progressing project into a power plant of research.

During my entire doctoral studies, I was extremely happy to supervise brilliant students whose efforts contributed to this thesis. My sincere thanks go to Stefan Vlaski,

Bin Han, Falco Strasser, Johannes Weise, Jack Dagdagan, Stefan Richter, Andrea Schnall, Thanh Minh Vu and Ivan Derwin.

I wish to express my gratitude to my parents Waltraud Wunsch-Muma and Craig Muma for their unconditional love and support throughout my life. You provided me with the ability to be the person I like to be.

Finally, I am most grateful to my wife Eli and my daughter Lillie for their understanding, love, encouragement, support and joy. Many of the smiles and happy moments I had while writing this document came from thinking about you.

Darmstadt, 20.03.2014

Kurzfassung

Die vorliegende Dissertation entwickelt und analysiert fortgeschrittene robuste Methoden für die Signalverarbeitung. Behandelte Probleme sind aus den Bereichen der Sensorgruppensignalverarbeitung, der robusten Modellordnungsschätzung, sowie der Robustheit für abhängige Daten. Die entwickelten Methoden wurden auf praktisch relevante Problemstellungen aus verschiedenen Gebieten der biomedizinischen Signalverarbeitung und der Sensorgruppensignalverarbeitung angewandt.

Insbesondere wurde für univariate unabhängige Daten ein robustes Modellordnungsschätzungskriterium entwickelt und für die Modellierung von Hornhautoberflächen Daten verwendet. Das vorgeschlagene Kriterium erweitert bestehende robuste Kriterien. Für echte gemessene Daten wählt das entwickelte Kriterium auch dann die Modellordnung in Übereinstimmung mit klinischen Erwartungen, wenn die Messbedingungen der Videokeratoskopie, welches die heute gängige Messmethode ist, schlecht sind.

Für die Sensorgruppensignalverarbeitung haben wir ein robustes Modellordnungsschätzungskriterium entwickelt und auf das Problem der Quellenschätzung angewandt. Das entwickelte Kriterium verwendet einen robusten und gleichzeitig effizienten Schätzer der Kovarianzmatrix der r -mode Auffaltungen des komplexwertigen Datentensors. Sowohl für Gaußverteiltes Sensorrauschen als auch für impulsives Rauschen, verursacht z. B. durch kurzzeitige Sensorfehler, ist die Leistung der vorgeschlagenen Tensor-basierten Modellordnungsschätzungskriterien besser als die der korrespondierenden Matrix-basierten Verfahren.

Im Kontext der robusten Verfahren der Sensorgruppensignalverarbeitung haben wir als nächstes das Problem der Schätzung der komplexwertigen Amplituden von Sinussignalen in einer komplett unbekanntem, impulsiven, räumlich und zeitlich unabhängig verteilten Rauschumgebung untersucht. Eine Auswahl aus nichtrobusten und robusten Schätzern wurde mit einem von uns vorgeschlagenen robusten semi-parametrischen Schätzer verglichen.

Eine dritte Forschungsfrage, die uns im Bereich der Signalverarbeitung für Antennengruppen beschäftigte, ist die Robustheitsanalyse von Raum-Zeit-Frequenz-Verteilungsschätzern. Zu diesem Zweck haben wir eine Robustheitsanalyse mittels der Einflussfunktion (influence function) durchgeführt. Die Einflussfunktion ist ein Robustheitsmaß, das den Einfluss einer infinitesimalen Kontamination auf den Bias eines Schätzers, normiert auf die Fraktion der Kontamination, beschreibt. Zusätzlich zur

analytischen asymptotischen Analyse haben wir eine praktisch implementierbare Einflussfunktion für begrenzte Datenlängen definiert. Simulationsergebnisse für endliche Datenlängen bestätigen die analytischen Ergebnisse und zeigen die Unempfindlichkeit kürzlich vorgeschlagener robuster Raum-Zeit-Frequenz-Verteilungsschätzer gegenüber geringen Verletzungen getroffener statistischer Annahmen.

Ein Schwerpunkt dieser Dissertation es, robuste Schätzer für abhängige Daten zu entwickeln und zu analysieren. Zuerst behandeln wir einige praktische Beispiele aus dem biomedizinischen Bereich, in denen Messartefakte mittels kombinierter robuster Schätzung und Datentransformationen erkannt und deren Effekte unterdrückt werden. Insbesondere schlagen wir einen Algorithmus zur Artefaktbereinigung von elektrokardiographischen (EKG) Messungen vor. Dies ist besonders wichtig für die Überwachung von Patienten mit mobilen EKG-Messgeräten, die aufgrund der Bewegung der Patienten von Artefakten besonders betroffen sind. Ein zweites biomedizinisches Problem, das wir in dieser Dissertation untersucht haben, ist die Vorhersage von Hirndrucksignalen für Patienten mit Schädel-Hirn-Trauma. Dies ermöglicht eine aktive und rechtzeitige Intervention zu einer effektiveren Kontrolle des Hirndrucks. Wir stellen hierzu eine Methode vor, die Artefaktbereinigung und eine Transformation in die Empirical-Mode-Domäne kombiniert.

Motiviert durch die Vielzahl an praktischen Anwendungen fokussieren wir uns dann auf die Herleitung und Analyse fortgeschrittener robuster Parameterschätzer und Modellordnungsschätzer für Autoregressive Moving-Average (ARMA) Modelle. Wir stellen für einen neu vorgeschlagenen Parameterschätzer einen schnellen Algorithmus vor und führen eine komplette statistische Robustheitsanalyse durch. Für diesen Schätzer, den wir den Bounded-Influence-Propagation (BIP) τ -Schätzer nennen, geben wir die statistischen Konvergenzbedingungen, sowie einen Beweis der quantitativen und qualitativen Robustheit. Die Robustheit wird gemessen durch die Einflussfunktion, die Maximal-Bias-Kurve (maximum bias curve) und den Ausfallpunkt (breakdown point) des Schätzers. Der schnelle Algorithmus des vorgeschlagenen Schätzers basiert auf einer Initialschätzung eines approximativen langen autoregressiven Modells, von dem die ARMA Parameter abgeleitet werden. Auf diese Weise kann ein weiteres Verwenden der ausreißerkontaminierten Daten vermieden werden. Der BIP τ -Schätzer ist sehr geeignet und von der Rechenzeit her attraktiv für die ARMA Modellordnungsschätzung, da die Rechenzeit für alle ARMA Kandidatenmodelle ungefähr der Rechenzeit entspricht, die das approximative autoregressive Modell benötigt. Im Bereich der robusten ARMA Modellordnungsschätzung schlagen wir verschiedene Kriterien basierend auf dem BIP τ -Schätzer vor und vergleichen diese Kriterien mit existierenden Ansätzen.

Abstract

In this thesis, advanced robust estimation methodologies for signal processing are developed and analyzed. The developed methodologies solve problems concerning multi-sensor data, robust model selection as well as robustness for dependent data. The work has been applied to solve practical signal processing problems in different areas of biomedical and array signal processing.

In particular, for univariate independent data, a robust criterion is presented to select the model order with an application to corneal-height data modeling. The proposed criterion overcomes some limitations of existing robust criteria. For real-world data, it selects the radial model order of the Zernike polynomial of the corneal topography map in accordance with clinical expectations, even if the measurement conditions for the videokeratoscopy, which is the state-of-the-art method to collect corneal-height data, are poor.

For multi-sensor data, robust model order selection criteria are proposed and applied to the problem of estimating the number of sources impinging onto a sensor array. The developed criteria are based on a robust and efficient estimator of the covariance of the r -mode unfoldings of a complex valued data tensor. Both in the case of Gaussian noise and for a brief sensor failure, the proposed robust multi-dimensional schemes outperform their matrix computation based counterparts.

In the context of robustness for multi-sensor data, we next investigate the problem of estimating the complex-valued amplitude of sinusoidal signals in a completely unknown heavy-tailed symmetric spatially and temporally independent and identically distributed (i.i.d.) sensor noise environment. A selection of non-robust and robust estimators are compared to a proposed semi-parametric robust estimator.

A third research focus in the area of multi-sensor data is that of analyzing the robustness of spatial time-frequency distribution (STFD) estimators. We provide a robustness analysis framework that is based on the influence function. The influence function is a robustness measure that describes the bias impact of an infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination. In addition to the asymptotic analysis, we also give a definition of the finite sample counterpart of the influence function. Simulation results for the finite sample influence function confirm the analytical results and show the insensitivity to small departures in the distributional assumptions for some recently proposed robust STFD estimators.

A large part of this thesis concerns the topic of obtaining and analyzing robust estimators in the dependent data setup. First, some practical issues concerning the detection, and robust estimation in presence of patient motion induced artifacts in biomedical measurements are addressed. In particular, we provide an artifact-cleaning algorithm for data collected with an electrocardiogram (ECG). This is especially important for the monitoring of patients with portable ECG recording devices, since these devices suffer severely from patient motion induced artifacts. A second real-world problem addressed in this doctoral project is that of forecasting the intracranial pressure (ICP) levels for patients who suffered a traumatic brain injury. This enables active and early interventions for more effective control of ICP levels. We propose a methodology which uses combined artifact detection and robust estimation after a data transformation into the empirical mode domain.

Motivated by plethora of practical applications, we then focus on deriving and analyzing sophisticated robust estimation and model selection techniques for autoregressive moving-average (ARMA) models. A fast algorithm as well as a detailed statistical and robustness analysis of a novel robust and efficient estimator is given. For the proposed estimator, which is termed the bounded influence propagation (BIP) τ -estimator, we compute a complete statistical robustness analysis, which includes conditions for the consistency, as well as a proof of qualitative and quantitative robustness. The robustness is measured by means of the influence function, the maximum bias curve and the breakdown point. The fast algorithm of the proposed estimator is based on first computing a robust initial estimate of an autoregressive (AR) approximation from which the ARMA model parameters are derived. In this way, the ARMA model parameters are derived from the long AR approximation without further use of the outlier-contaminated observations. The estimator is very suitable and attractive for ARMA model selection purposes, since the computational cost of estimating all the candidate ARMA models approximately reduces to that of computing one long AR model. In the area of model selection for ARMA models, we propose and compare different robust model order selection criteria that are based on the BIP τ -estimator.

Contents

1	Introduction to Robustness for Signal Processing	1
1.1	Introduction and Motivation	1
1.1.1	Aims of this Doctoral Project	2
1.1.2	Publications	2
1.2	What is Robustness?	4
1.2.1	Defining, Measuring and Ensuring Robustness	4
1.3	Important Open Robustness Issues for Signal Processing	9
1.3.1	Robust Model Selection Issues	9
1.3.2	Multi-Sensor Data Issues	10
1.3.3	Dependent Data Issues	10
1.3.4	Organization of this Thesis	10
2	Robustness in Regression Models with Independent Data	13
2.1	Introduction	13
2.1.1	Contributions in this Section	13
2.1.2	Robustifying the Linear Regression Model	14
2.1.3	Robustly Selecting the Model Order	20
2.1.4	Robust Model Selection Based on τ -Estimation	22
2.2	Robust Model Selection for Corneal-Height Data	25
2.2.1	Contributions in this Section	25
2.2.2	Corneal Topography Estimation	25
2.2.3	Model Order Selection for Corneal Topography Estimation	28
3	Robustness for Array Signal Processing (Multi-Sensor Data)	33
3.1	Introduction	33
3.2	Robust Source Number Enumeration in Case of Brief Sensor Failures (Model Order Selection for Tensor Data)	34
3.2.1	Contributions in this Section	34
3.2.2	Source Number Enumeration for R -D Arrays	34
3.2.3	Robust Covariance Matrix Estimation	37
3.2.4	Robust Source Number Enumeration for R -D Arrays	38
3.2.5	Simulations	41
3.3	Robust Semi-Parametric Estimation of Sinusoidal Signals	44
3.3.1	Contributions in this Section	44
3.3.2	Problem Statement	44
3.3.3	Simulations	50
3.3.4	Conclusions and Future Work	51

3.4	Measuring Robustness of Time-Frequency Distributions by Means of the Influence Function	53
3.4.1	Contributions in this Section	53
3.4.2	Robust Spatial Time-Frequency Distribution Estimation	53
3.4.3	Simulations	59
3.4.4	Conclusions and Future Work	62
3.5	Conclusion	62
4	Robustness for Dependent Data	65
4.1	Introduction	65
4.1.1	Contributions in this Chapter	65
4.2	Modeling Dependent Data	66
4.2.1	Autoregressive Moving-Average (ARMA) Models	66
4.2.2	Classical Parameter Estimation for ARMA Models	67
4.2.3	Outlier Models	70
4.3	Artifact Detection and Removal Using Data Transformations	71
4.3.1	Motion Artifact Removal in Electrocardiogram (ECG) Signals Using Multi-Resolution Thresholding	72
4.3.2	Artifact Detection in the Empirical Mode Domain for the Forecasting of Intracranial Pressure (ICP) Signals	74
4.3.3	Limitations of the Outlier Detection Approach	79
4.4	Robustness Theory for Dependent Data	79
4.4.1	Influence Function for Dependent Data	79
4.4.2	Maximum Bias Curve (MBC) for Dependent Data	80
4.4.3	Breakdown Point for Dependent Data	81
4.5	Robust Estimation of ARMA Models	81
4.5.1	Outlier Propagation	81
4.5.2	Cleaned Non-Robust Estimator	82
4.5.3	Median-of-Ratios-Estimator (MRE)	83
4.5.4	Robustly Filtered Estimators	84
4.5.5	Bounded Influence Propagation (BIP) Estimators	87
4.6	BIP τ -estimators	87
4.6.1	BIP τ -estimators: Asymptotic and Robustness Properties	88
4.6.2	BIP τ -estimators: Fast Algorithm to Obtain Stationary and Invertible ARMA Estimates	94
4.6.3	Simulations	104
4.7	Robust Model Order Selection for ARMA Models	113
4.7.1	Classical Model Order Selection Using Information Criteria	114
4.7.2	Robust ARMA Model Selection	114

4.7.3 Simulations	120
4.8 Summary and Conclusions	125
5 Summary, Conclusions and Future Work	127
5.1 Summary and Conclusions	127
5.2 Future Directions	129
List of Acronyms	133
List of Symbols	135
Bibliography	143
Curriculum Vitae	155

Chapter 1

Introduction to Robustness for Signal Processing

1.1 Introduction and Motivation

Increasing awareness about what we today call 'robustness' of an estimator has left its mark in scientific endeavours over the last two centuries. In an early contribution to the ongoing scientific discussion, the French mathematician Adrien-Marie Legendre commented on the necessity of rejecting outliers to provide stability to the least-squares method [1]. At the end of the 19th century, the idea of modeling heavy-tailed distributions as mixtures of Gaussian densities has been proposed by the Canadian-American astronomer and mathematician Simon Newcomb. The first formal theory of robustness was proposed in 1964 by Peter J. Huber in response to an article by John W. Tukey [2]. A second theory of robustness was later developed by Frank Hampel in 1968, who introduced important concepts, such as the influence function and the breakdown point [3].

In engineering, robust estimators and detectors have been of interest since the early days of digital signal processing, for a review, the interested reader is referred to the paper published by Salim Kassam and Vincent Poor in 1985 [4] and references therein. There are also excellent text books that consider robust estimation for signal processing [5, 6].

Measurement campaigns have reported the presence of impulsive noise in many of today's key signal processing applications, such as, e.g., radar, mobile communication channels, biomedical measurements, see [7] for a recent overview. Impulsive (heavy-tailed) noise can cause nominally optimal techniques, especially those derived using the Gaussian probability model to be biased and even to break down. These situations enforce the need for robust estimators, which are close-to optimal in nominal conditions and remain highly reliable for real-life data, even if the assumptions are only approximately valid. The increasing complexity of modern engineering problems poses new demands to robust methods, some of which are addressed in this doctoral project.

1.1.1 Aims of this Doctoral Project

The aim of this doctoral project is to analyze, introduce and improve robust estimation methodologies to solve current and future signal processing problems. The major open research questions which have been identified as highly relevant to the signal processing community and are therefore addressed in this project concern

- Robustness when dealing with multi-sensor data
- Robust model selection
- Robustness for dependent data.

1.1.2 Publications

The following publications have been produced during this doctoral project.

Internationally Refereed Journal Articles

- M. Muma and A. M. Zoubir, “A new robust estimation and model order selection method for ARMA processes”, to be submitted to *IEEE Trans. Signal Process.*, March 2014.
- T. Schäck, M. Muma and A. M. Zoubir, “Robust causality analysis of non-stationary multivariate time series”, to be submitted to *IEEE Trans. Biomed. Eng.*, March 2014.
- A. Kelava, M. Muma, M. Schmidt and A. M. Zoubir, “A new approach for quantifying the coherence of multivariate non-stationary data with an application to psychophysiological measures during emotion regulation”, under revision in *Psychometrica*.
- W. Sharif, M. Muma and A. M. Zoubir, “Robustness analysis of spatial time-frequency distributions based on the influence function”, *IEEE Trans. Signal Process.*, Vol 61, No 8, pp. 1958–1971, April 2013.
- A. M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, “Robust estimation in signal processing: a tutorial-style treatment of fundamental concepts”, *IEEE Signal Process. Magazine*, Vol 29, No 4, pp. 61–80, July 2012.

- M. Muma, D. R. Iskander and M. J. Collins, “The role of cardiopulmonary signals in the dynamics of the eye’s wavefront aberrations”, *IEEE Trans. Biomed. Eng.*, Vol 57, No 2, pp. 373–383, February 2010.

Internationally Refereed Conference Papers

- M. Muma, “Robust model order selection for ARMA models based on the bounded innovation propagation τ -estimator”, Submitted to the Proc. of the *IEEE Workshop on Statistical Signal Processing (SSP) 2014*, Gold Coast, Australia, June 2014.
- S. Vlaski, M. Muma and A. M. Zoubir, “Robust bootstrap methods with an application to geolocation in harsh LOS/NLOS environments ”, In the Proc. of the *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) 2014*, in Florence, Italy, May 2014, (accepted).
- J. Dagdagan, M. Muma and A. M. Zoubir, “Robust testing for stationarity in the presence of outliers”, In the Proc. of the *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) 2014*, in Florence, Italy, May 2014, (accepted).
- B. Han, M. Muma, M. Feng and A. M. Zoubir, “An online approach for ICP forecasting based on signal decomposition and robust statistics”, In the Proc. of the *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) 2013*, in Vancouver, Canada, pp. 6239–6243, May 2013.
- F. Strasser, M. Muma and A. M. Zoubir, “Motion artifact removal in ECG signals using multi-resolution thresholding”, In the Proc. of the *European Signal Processing Conference (EUSIPCO) 2012* in Bucharest, Romania , pp. 899–903. August 2012.
- M. Muma, Y. Cheng, F. Roemer, M. Haardt and A. M. Zoubir, “Robust source number enumeration for R -dimensional arrays in case of brief sensor failures”, In the Proc. of the *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) 2012* in Kyoto, Japan, pp. 3709–3712. March 2012
- M. Muma and A. M. Zoubir, “Robust model selection for corneal-height data based on τ -estimation”, In the Proc. of the *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) 2011* in Prague, Czech Republic, pp. 4096–4099, May 2011.
- M. Muma, U. Hammes and A. M. Zoubir, “Robust semi-parametric amplitude estimation of sinusoidal signals: the multi-sensor case”, In the Proc. of the

3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) 2009, Aruba, Dutch Antilles, December 2009.

1.2 What is Robustness?

The term “robust” has been used with different and sometimes contradicting annotations. The robustness we are concerned with in this doctoral project is that of analyzing the impact on statistical methods caused by a discrepancy between the (statistical) modeling assumptions and reality [2, 3, 7, 8]. It provides methods which trade-off some efficiency at the nominal model to minimize the effects of deviations.

1.2.1 Defining, Measuring and Ensuring Robustness

In the sequel, we highlight the aims of robust methods and then discuss how these aims can be fulfilled. For this, we first briefly revisit the most important measures of robustness that have been of great importance in this doctoral project.

1.2.1.1 Aims of Robust Methods

Clearly, depending on the application at hand, the aims of robust methods may vary somewhat. However, throughout the literature of robust statistics, three major aims regularly appear. Intuitively speaking, a robust method should fulfill:

Aim 1: Near Optimality

This aim states that a robust procedure should behave “reasonably good” (nearly optimal) at the assumed model [2]. When taking a closer look at Aim 1, it becomes clear that robust methods are approximate parametric methods. Robustness is relative to a nominal model, which is assumed to be approximately valid. Approximate validity can refer to (i) a majority of the data belonging to the nominal distribution in contrast to some outlying (contaminating) observations that do not follow the pattern of the majority or (ii) that the distribution of the data is within a specific class of neighboring distributions, as measured by some distance measure. This is referred to a model uncertainty. Clearly, near-optimality, in both cases is a necessary requirement, since the robust method must be competitive for the assumed model.

Aim 2: Qualitative Robustness

The second aim requires that the effect of an erroneous observation, even if it takes an arbitrary value, should not have a large impact on the method [3]. It is therefore related to the stability against small (infinitesimal) contamination. As will be seen in later Sections, qualitative robustness also ensures that a small change in the data only has a small effect on the estimates. Small changes can refer both to changing a small fraction of the data arbitrarily (e.g., outliers) or performing a minor change to a large fraction of the data (e.g., rounding). In analogy to the stability of a bridge, qualitative robustness proves the stability of a method against infinitesimal perturbations.

Aim 3: Quantitative Robustness

The third aim is that somewhat larger deviations from the model should not cause a catastrophe [2]. This means that, e.g., even if we increase the fraction of outlying data up to some point, the method should still provide us with reasonable information. A “catastrophe” for an estimator is, for example, when the bias of the estimate becomes infinite. Again speaking in the analogy of the bridge, the estimator should be stable enough not to break down in case of larger deviations.

1.2.1.2 Measures of Robustness

Analyzing the robustness of a method, and ensuring that it fulfills the above mentioned aims, requires tools (measures) derived from robust statistics. These are briefly introduced and defined in this Section.

The Relative Efficiency (Eff)

The relative efficiency the increase in the (asymptotic) variance σ^2 of the estimates *at the assumed model* compared to the optimal method σ_{opt}^2 :

$$\text{Eff} = \frac{\sigma_{\text{opt}}^2}{\sigma^2},$$

thus $0 \leq \text{Eff} \leq 1$. Figure 1.1 plots the distributions of three estimators of location for the standard Gaussian distribution $\mathcal{N}(0,1)$. For this distribution, the sample mean (average) is the optimal estimator. The variance of two robust estimators, i.e. the sample median $\text{Eff} = 0.64$ and Huber’s M-estimator $\text{Eff} = 0.95$ [2] is larger.

The Influence Function (IF)

The influence function (IF) describes the bias impact of an infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination.

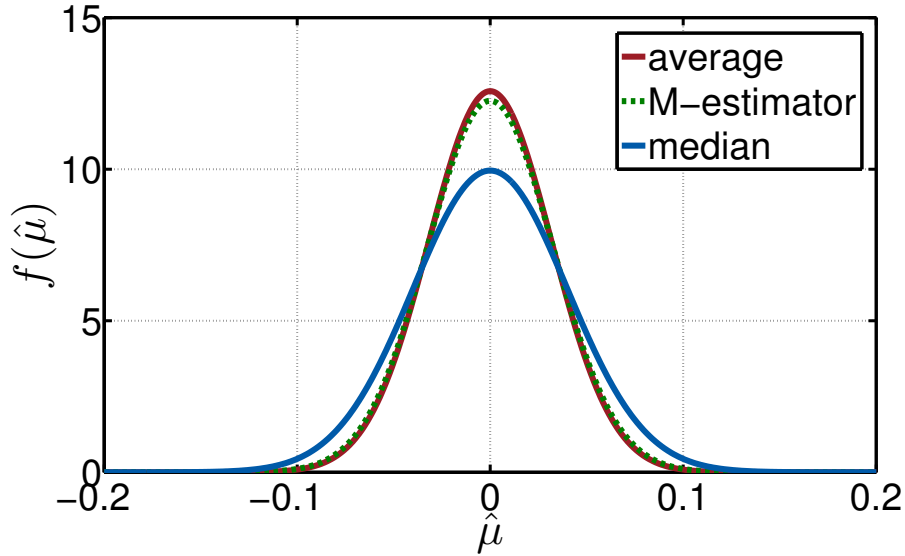


Figure 1.1: The distributions of three location estimators for the standard Gaussian distribution $\mathcal{N}(0, 1)$. Compared to the optimal estimator, which is the sample mean (average), the robust estimators have a larger variance, which corresponds to a relative efficiency $\text{Eff} < 1$. For the sample median $\text{Eff} = 0.64$ while Huber’s M-estimator is tuned to be highly efficient $\text{Eff} = 0.95$.

Boundedness and continuity of the influence function are required for the qualitative robustness (Aim 2) of an estimator. Furthermore, the relative efficiency of an estimator, i.e., the ratio of variances of the optimal method under perfect conditions and the (robust) method under consideration, can be derived from the influence function [3, 8].

When the limit exists, the influence function, which is basically the first derivative of the functional version of an estimator $\hat{\theta}$ at a nominal distribution F_θ , is defined by

$$\text{IF}(z; \hat{\theta}, F_\theta) = \lim_{\varepsilon \rightarrow 0} \frac{\hat{\theta}_\infty(F_\varepsilon) - \hat{\theta}_\infty(F_\theta)}{\varepsilon} = \left[\frac{\partial \hat{\theta}_\infty(F_\varepsilon)}{\partial \varepsilon} \right]_{\varepsilon=0}. \quad (1.1)$$

Here, $\hat{\theta}_\infty(F_\theta)$ and $\hat{\theta}_\infty(F_\varepsilon)$ are the asymptotic values of the estimator when the data is distributed following respectively F_θ and the contaminated distribution

$$F_\varepsilon = (1 - \varepsilon)F_\theta + \varepsilon\delta_z,$$

with δ_z being the point-mass probability on z and ε the fraction of contamination.

Figure 1.2 depicts the influence functions of three estimators of location for the standard Gaussian distribution $\mathcal{N}(0, 1)$. The influence function is plotted with respect to z , the position of the infinitesimal contamination. In Chapters 3 and 4, we derive the influence functions for some multi-channel and dependent data estimators.

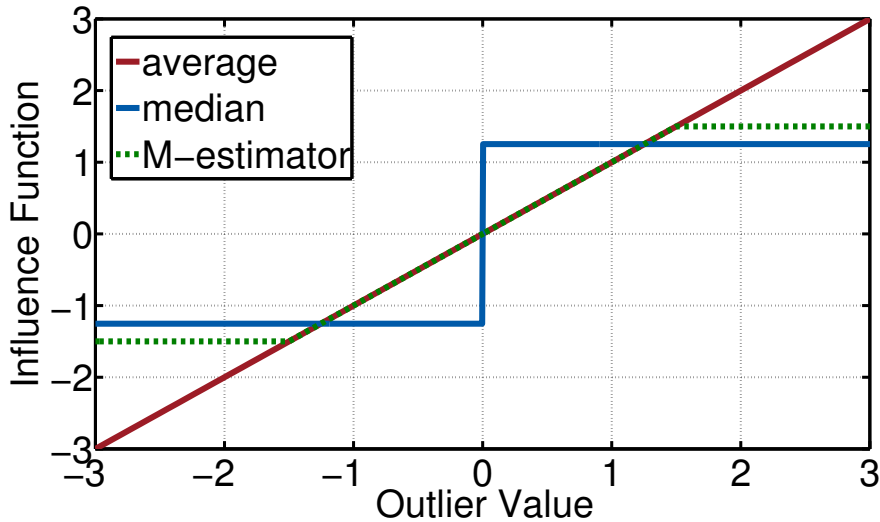


Figure 1.2: The influence functions of three location estimators for the standard Gaussian distribution $\mathcal{N}(0, 1)$. While the influence function of the sample mean (average) is unbounded, that of the sample median is bounded but not continuous at the origin. Only the influence function of the robust M-estimator [2], see Section 2.1.2.3, is bounded and continuous, which means that the estimator is qualitatively robust. For this estimator, the impact of infinitesimal large-valued outliers on the estimate is maximally suppressed, since the influence function re-descends to zero.

The Maximum Bias Curve (MBC)

The maximum bias curve (MBC) provides information on the bias introduced by a specific amount of contamination. It plots the absolute value of the maximum possible asymptotic bias

$$b_{\hat{\theta}}(F_{\theta}) = \hat{\theta}_{\infty}(F_{\theta}) - \theta$$

of an estimator $\hat{\theta}$ with respect to the fraction of contamination ε . This is particularly useful, when comparing the robustness of two estimators, since the more robust estimator minimizes the MBC.

Formally, the maximum bias curve is defined as

$$\text{MBC}(\varepsilon, \theta) = \max \{|b_{\hat{\theta}}(F_{\theta})| : F \in \mathcal{F}_{\varepsilon, \theta}\}. \quad (1.2)$$

Here, $\mathcal{F}_{\varepsilon, \theta} = \{(1 - \varepsilon)F_{\theta} + \varepsilon F_{\varepsilon}\}$ is an ε -neighborhood of distributions around the nominal distribution F_{θ} with F_{ε} being an arbitrary contaminating distribution. Figure 1.3 displays the maximum bias curves for the sample mean (average), the sample median and the α -trimmed mean at $\mathcal{N}(0, 1)$. In Chapter 4, we provide and discuss the maximum bias curves for some existing and a newly proposed robust autoregressive moving-average (ARMA) parameter estimator for dependent data.

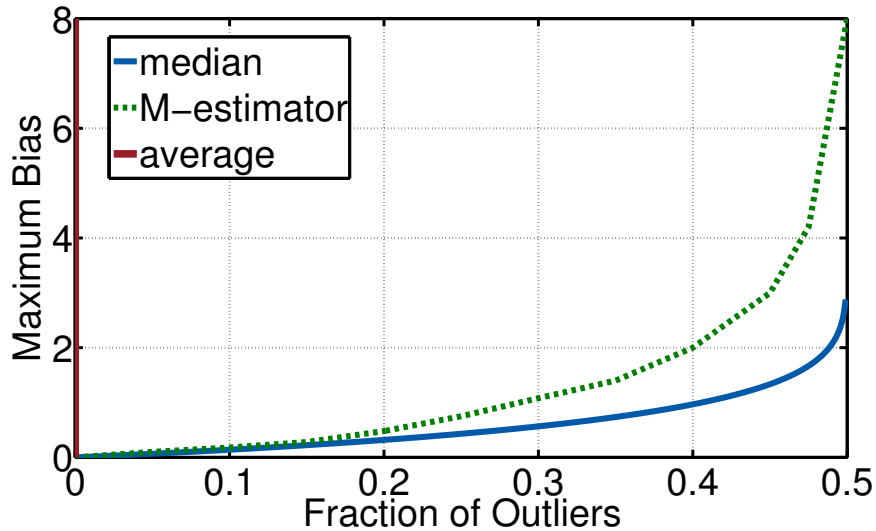


Figure 1.3: The maximum bias curves of three location estimators for the standard Gaussian distribution $\mathcal{N}(0, 1)$, i.e., the sample mean (average), the sample median and the α -trimmed mean, with $\alpha = 0.25$. The maximum bias is infinite beyond the BP, therefore, the maximum bias curve of a non-robust estimator, e.g., the sample mean is infinite at $\varepsilon > 0$.

The Breakdown Point (BP)

The breakdown point (BP) is perhaps the most intuitively comprehensible of the robustness measures. It is used to characterize the quantitative robustness (Aim 3) of an estimator. Loosely speaking, the breakdown point indicates the maximal fraction of outliers (highly deviating samples) in the observations, which an estimator can handle without breaking down. The breakdown point ranges from 0 to 50 %, where a higher BP value corresponds to larger quantitative robustness. As can be seen from Figure 1.3, for the location estimation problem, the breakdown point coincides with an infinite maximum bias.

Figure 1.4 illustrates the concept of the BP for the sample mean, the α -trimmed mean and the sample median. It can be seen that the breakdown point of the sample mean is zero, which means that a single outlier may throw the estimator completely off. The breakdown point of the sample median, on the other hand, is 50 %. Beyond 50 %, robust estimators are no longer applicable, since one cannot distinguish between the nominal and contaminating distributions.

In Section 2.1, we review some regression estimators that are highly efficient, while at the same time possess the maximal possible breakdown point. In Chapter 4, we discuss the breakdown point for the dependent data case, which is not yet established in the literature. We also propose a novel estimator for ARMA parameters and show that it

has the maximum possible breakdown point of 50 %, while, at the same time, being highly efficient.

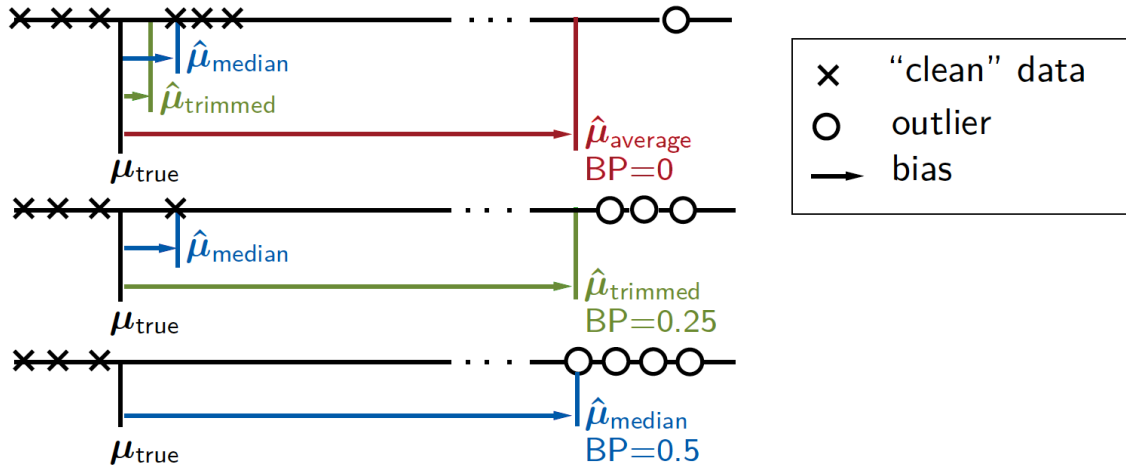


Figure 1.4: The bias and breakdown point of three location (μ) estimators, i.e., the sample mean (red), the α -trimmed mean (green) and the sample median (blue). “Clean” observations are depicted as crosses and outliers as circles. The breakdown point of the sample median is equal to 50 %, which means that its bias remains bounded even in situations when up to half of the observations, in this case three, are replaced by arbitrarily large values.

1.3 Important Open Robustness Issues for Signal Processing

Intuitively, robustness as defined above, is central to engineering applications, perhaps even more so than optimality. In this doctoral project, we identified and addressed the following open problems that are relevant to the signal processing community, today and will remain relevant in the future.

1.3.1 Robust Model Selection Issues

The key questions to be answered in this research area are: “How can I find a suitable statistical model to describe the majority of the data?” and “How can I prevent outliers or other contaminants from having overriding influence on the final conclusions?” In this thesis, we address these questions by means of robust model selection criteria. Applications arise in corneal topography estimation (Chapter 2), source number enumeration (Chapter 3) and biomedical signal modeling (Chapter 4).

1.3.2 Multi-Sensor Data Issues

The key questions to be answered in this research area are: “How can I robustly determine the number of sources impinging on an array, even when a random sensor briefly fails?”, “How can I use a sensor array to robustly estimate the complex-valued amplitudes of sinusoidal signals impinging on this array, even if the sensor noise distribution is heavy-tailed and/or completely unknown?”, “How can I analyze the robustness of spatial time-frequency distribution estimators?” and “Are recently proclaimed robust methods truly robust?” Answers to these questions are sought for in Chapter 3 of this thesis.

1.3.3 Dependent Data Issues

The key questions to be answered in this research area are: “How can I mitigate the effects of artifacts in biomedical measurements?” and “How can I robustly determine the parameters and the model orders of ARMA models in a way that is both computationally feasible and analytically justifiable?” Finding answers to such questions is the concern of Chapter 4.

1.3.4 Organization of this Thesis

Following this introduction, the dissertation is organized as follows:

Chapter 2 briefly reviews robust parameter estimation and robust model order selection for the linear regression model in case of independent data. A robust criterion to select the model order is proposed and applied to corneal-height data modeling. In particular, we derive an expression for the penalty term in robust information criteria when using the τ -estimator which is simultaneously robust and efficient. This is indispensable in the application of corneal topography estimation, where the fraction of outliers varies over the topographical map. Both simulated and real-world data examples gained from a videokeratoscope are given.

Chapter 3 deals with three challenges that demand robust multi-channel methods. The first problem that we investigate is that of robust source enumeration for R -dimensional data in presence of impulsive noise, as, e.g., caused by brief sensor failures. For this, we derive criteria that are based on robust and efficient estimation of the covariance matrix using the r -mode unfolding operation of the data tensor. The second

issue under investigation in this Chapter is the estimation of the complex-valued amplitudes of sinusoidal signals using multiple sensors. Here, our focus is to derive so-called robust semi-parametric estimators, which adapt to the underlying noise distribution by non-parametric transformation kernel density estimation followed by maximum likelihood estimation incorporating the estimated density. The third research focus in this Chapter is the robustness analysis of spatial time-frequency distributions. These are widely used, e.g., for direction-of-arrival and blind source separation in case of non-stationary signals. In particular, we provide the asymptotic and finite sample influence function expressions for a standard and a recently proposed robust estimators.

Chapter 4 is dedicated to robustness for dependent data. After briefly reviewing some dependent data models, we show some practical applications of robust dependent data methods to solve real-world problems. We briefly describe two artifact detection and mitigation strategies that are based on data transformations. We apply these to real-world electrocardiogram and intracranial pressure data. We also derive and analyze sophisticated robust estimation and model selection techniques for autoregressive moving-average models. A fast algorithm, as well as a detailed statistical and robustness analysis of a novel robust and efficient estimator, is given. For the proposed estimator, which is termed the bounded influence propagation (BIP) τ -estimator, we compute a complete statistical robustness analysis, which includes conditions for the consistency, as well as a proof of qualitative and quantitative robustness. The estimator is very suitable and attractive for ARMA model selection purposes and we propose and compare different robust model order selection criteria that are based on the BIP τ -estimator.

Chapter 5 concludes and summarizes the thesis. Furthermore, we give some future directions and describe preliminary steps we have undertaken. One future direction concerns robust bootstrap methods, which can be applied to confidence region estimation in geolocation position estimation. A further future research direction will be the integration of robustness concepts into the emerging and highly active research area of distributed detection and estimation.

Chapter 2

Robustness in Regression Models with Independent Data

2.1 Introduction

Linear regression models are amongst the most powerful and popular tools to tackle today's engineering problems. Depending on the data at hand, there exist a number of difficulties to deal with. In this Section, we discuss robust parameter estimation [2,7–10] and robust model order selection problems [11–14] for data that contains outliers. The term 'outliers' refers to a minority of observations which strongly deviate from the model which best fits the majority. The key research questions in robust model selection are: “How can I find a suitable statistical model to describe the majority of the data?” and “How can I prevent the outliers from having overriding influence on the final conclusions?”.

2.1.1 Contributions in this Section

In the sequel, we present a robust criterion to select the model order in a linear univariate regression setup and apply it to corneal-height data modeling. In particular, we derive an expression for the penalty term in robust information criteria [11–13], when using the τ -estimator [9, 10]. The τ -estimator is a simultaneously robust and efficient estimator, which is indispensable in corneal topography estimation. To derive the penalty term, we exploit the asymptotic equivalence of the τ -estimator to an M-estimator. Based on the proposed criterion, we introduce an algorithm to fit an appropriate set of Zernike polynomials to corneal-height data [15–17], even when measurements are contaminated by outliers, e.g., due to breakups in the pre-corneal tear film and reflections from the eyelashes.

The main contributions in this Section have been published in [14]. Some general illustrations of robustness issues in regression problems have been published in [7], more simulation results, some extensions and further real data examples can be found in [18].

2.1.2 Robustifying the Linear Regression Model

This Section contains a brief revisit of some key principles involved in robustly estimating the parameters of linear regression models.

2.1.2.1 The Linear Regression Model

The linear regression model is defined as

$$Y_n = \mathbf{X}_n^\top \boldsymbol{\theta} + V_n, \quad n = 1, \dots, N. \quad (2.1)$$

Here, we assume that the predictors $\mathbf{X}_n = (X_{1n}, \dots, X_{pn})^\top$, the errors V_n and data points (Y_n, \mathbf{X}_n^\top) for $n = 1, \dots, N$, are independent random variables. Furthermore, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ are the unknown parameters of interest and \mathbf{X}_n and V_n are mutually independent.

2.1.2.2 Outlier Types in Linear Regression Models

The occurrence of outliers in linear regression models has been reported in a plethora of engineering problems in areas as diverse as wireless communication [19–23], ultrasonic systems [24], computer vision [25, 26], electric power systems [27], automated detection of defects [28], biomedical signal analysis [29] and many more. For linear regression models, one can distinguish two different types of outliers:

1. vertical outliers, which are outliers in V_n ,
2. leverage points, which are outliers in \mathbf{X}_n^\top .

Vertical outliers are present, e.g., in case of an impulsive noise distribution of V_n and leverage points occur, e.g., when \mathbf{X}_n^\top is random and has a heavy-tailed distribution. The term leverage point comes from the fact that these points have a high leverage on the estimation of the slope defined by $\boldsymbol{\theta}$. Not all leverage points are harmful for the estimation of $\boldsymbol{\theta}$. In fact a large-valued point which lies on the true slope can even increase estimation performance. On the other hand, a 'bad leverage point' that tilts the slope into a different direction, is obviously harmful.

Figure 2.3 depicts a situation for a 2 vector-valued parameter $\boldsymbol{\theta}$, where we observe different kinds of outlying observations, which do not follow the linear pattern of the majority of the data. Vertical outliers are present, when their \mathbf{X}_n^\top is not outlying and leverage points are present, if it is. As depicted, the maximum likelihood (ML)-estimator, under the Gaussian data assumption, is not robust against any type of outliers which contribute in an unbounded fashion to its bias.

2.1.2.3 M-Estimation

The first systematic and perhaps the most frequently used robust estimator is the M-estimator. It was first defined by Huber in his seminal paper [30] on robustifying the location model. M-estimators generalize the ML-estimators by replacing the maximization of the likelihood function by a general function, which is usually denoted as the ρ -function. To obtain robust estimates, for symmetric distributions, $\rho(x)$ is an odd and bounded function with derivative

$$\psi(x) = \frac{d\rho(x)}{dx}.$$

Regression M-estimates solve

$$\sum_{n=1}^N \rho \left(\frac{R_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M} \right) = \min, \quad (2.2)$$

where $\hat{\sigma}_M$ is a robust M-estimate of scale, see Eq. (2.10), that is required for $\hat{\boldsymbol{\theta}}_M$ to be scale equivariant. M-estimates, as defined in Eq. (2.2), are also regression and affine equivariant. The residuals in Eq. (2.2) are given by

$$R_n(\hat{\boldsymbol{\theta}}_M) = Y_n - \mathbf{X}_n^\top \hat{\boldsymbol{\theta}}_M, \quad n = 1, \dots, N.$$

Differentiating Eq. (2.2) yields

$$\sum_{n=1}^N \psi \left(\frac{R_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M} \right) \mathbf{X}_n^\top = \mathbf{0}. \quad (2.3)$$

ML-estimators are included in the class of M-estimators by setting

$$\rho(x) = -\log(f_X(x)), \quad (2.4)$$

where $f_X(x)$ is the density function of the random variable X . For the Gaussian data assumption, the ML-estimator coincides with the least-squares (LS)-estimator and is thus given by $\rho(x) = \frac{x^2}{2}$. This estimator is non-robust, since its score function $\psi(x) = x$

is unbounded. Solutions to Eq. (2.3) with monotone and re-descending $\psi(x)$ are called monotone and re-descending regression M-estimates, respectively.

Huber's M-estimator [2] is also referred to as the soft limiter and uses

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & |x| \leq c_{\text{Hub}} \\ c_{\text{Hub}}|x| - \frac{1}{2}c_{\text{Hub}}^2 & |x| > c_{\text{Hub}} \end{cases} \quad (2.5)$$

or correspondingly

$$\psi(x) = \begin{cases} x & |x| \leq c_{\text{Hub}} \\ c_{\text{Hub}}\text{sign}(x) & |x| > c_{\text{Hub}}. \end{cases} \quad (2.6)$$

Tukey's biweight (or bisquare) M-estimator is obtained with

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c_{\text{Tuk}}^2} + \frac{x^6}{6c_{\text{Tuk}}^4} & |x| \leq c_{\text{Tuk}} \\ \frac{c_{\text{Tuk}}^2}{6} & |x| > c_{\text{Tuk}} \end{cases} \quad (2.7)$$

and

$$\psi(x) = \begin{cases} x - 2\frac{x^3}{c_{\text{Tuk}}^2} + \frac{x^5}{c_{\text{Tuk}}^4} & |x| \leq c_{\text{Tuk}} \\ 0 & |x| > c_{\text{Tuk}}. \end{cases} \quad (2.8)$$

Figures 2.1 and 2.2 illustrate the two classes of robust M-estimators by depicting the popular ρ - and ψ -functions proposed by Huber (monotone) and Tukey (re-descending).

The main advantages of the monotone estimators are, that all solutions of Eq. (2.2) yield solutions to Eq. (2.3) and that all solutions are unique. Furthermore, monotone M-estimates are important as starting point for re-descending M-estimates. This is necessary due to the non-convexity of the maximization problem for re-descending M-estimates. While M-estimators are robust against vertical outliers, they become non-robust with a BP equal to zero in the presence of leverage points. This can be seen from Eq. (2.6), where the influence of outliers in \mathbf{X}_n is not bounded by $\psi(x)$. M-estimates are consistent and asymptotically Gaussian distributed under some general conditions [8]. The tuning constants in Eqs. (2.5)-(2.8) trade-off robustness and efficiency. Achieving both simultaneously is not possible for regression M-estimators.

2.1.2.4 S-Estimation

A second fundamental class of robust estimators are the ones based on the minimization of a residual scale. This class includes estimators, such as the LS, least-trimmed-squares (LTS) [31] and the least-median-of-squares (LMS) [31] estimators.

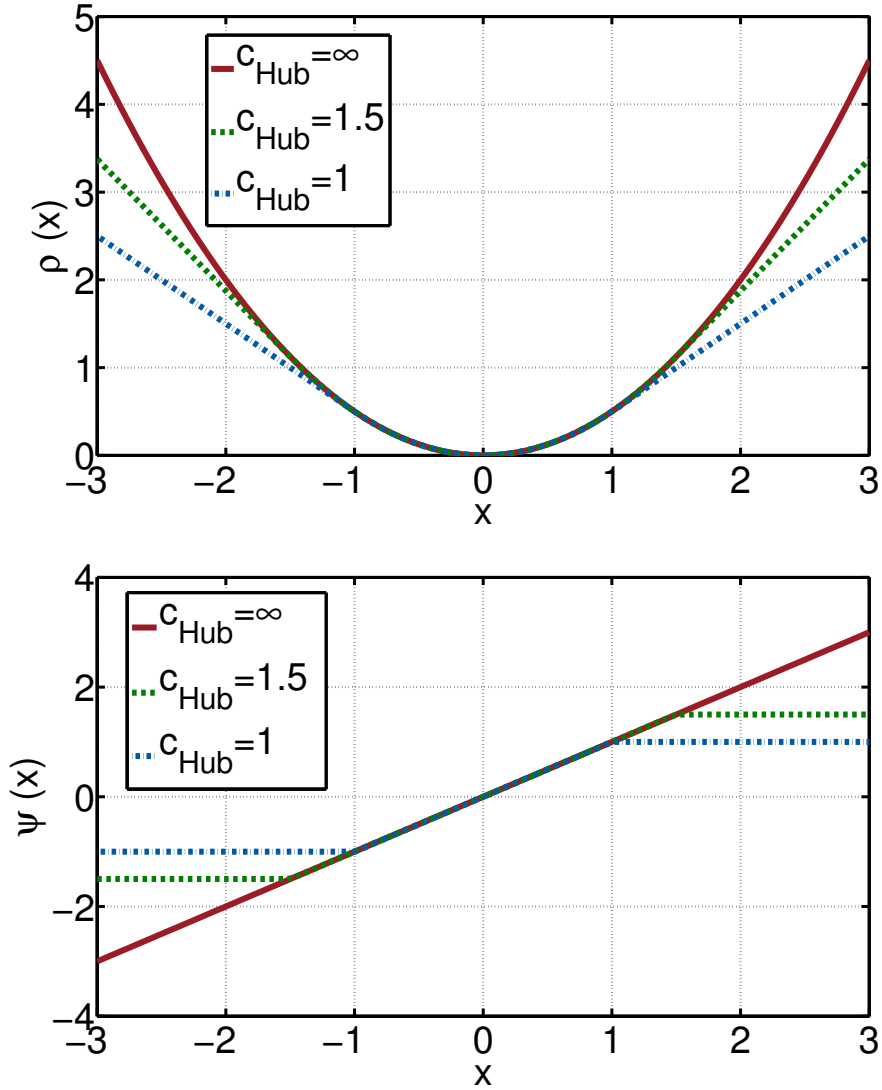


Figure 2.1: Examples of Huber's (monotone) ρ - and ψ -functions for different parameter choices for a standard Gaussian distribution.

S-estimators [32] minimize the M-estimate of scale of the residuals and are thus defined as

$$\hat{\boldsymbol{\theta}}_S = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \hat{\sigma}_M(\mathbf{R}(\boldsymbol{\theta})), \quad (2.9)$$

where $\hat{\sigma}_M(\mathbf{R}(\boldsymbol{\theta}))$ is the M-scale estimate defined by

$$\frac{1}{N} \sum_{n=1}^N \rho \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_M(\mathbf{R}(\boldsymbol{\theta}))} \right) = b, \quad (2.10)$$

with $\mathbf{R}(\boldsymbol{\theta}) = (R_1(\boldsymbol{\theta}), R_2(\boldsymbol{\theta}), \dots, R_N(\boldsymbol{\theta}))^\top$ and

$$b = \mathbb{E}_F [\rho(\mathbf{R}(\boldsymbol{\theta}))].$$

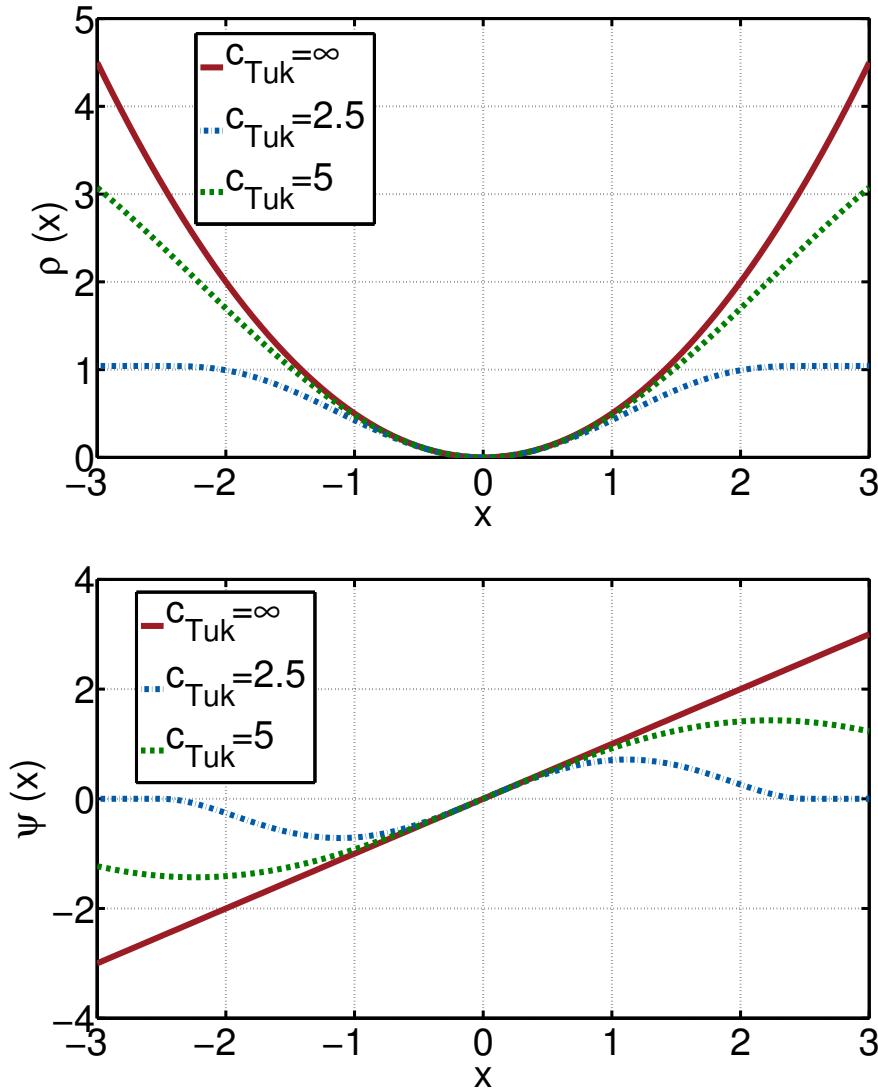


Figure 2.2: Example of Tukey's (re-descending) biweight ρ - and ψ -functions for different parameter choices for a standard Gaussian distribution.

Here, $E_F[\cdot]$ is the expectation w.r.t. the standard Gaussian distribution F .

S-estimates are regression, scale and affine equivariant and highly robust, if $\rho(x)$ is chosen accordingly, see [33] for details. However, as with M-estimates, BP=0.5 and high efficiency cannot be achieved simultaneously. Interestingly, S-Estimates are also M-estimates, as defined in Eq. (2.2) and Eq. (2.3), but with the condition that the scale $\hat{\sigma}_M(\mathbf{R}(\boldsymbol{\theta}))$ is estimated simultaneously [8].

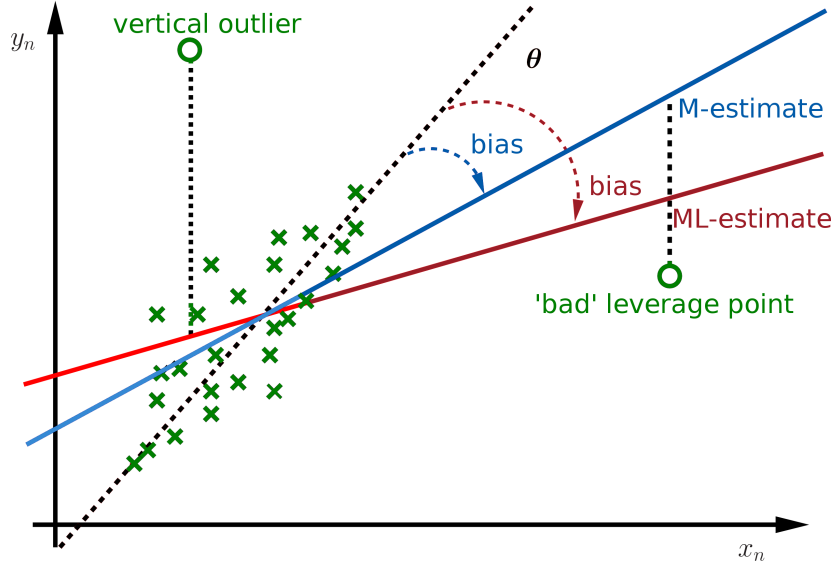


Figure 2.3: Different types of outliers in the regression context and their effect on the bias of the ML-estimator under the Gaussian assumption and on an M-estimator. The ML-estimator is not robust against any type of outliers. M-estimators are robust against vertical outliers, but they become non-robust with a BP equal to zero in the presence of leverage points.

2.1.2.5 τ -Estimation

The τ -estimate of scale [9] of the residuals $\hat{\sigma}_\tau(\mathbf{R}(\boldsymbol{\theta}))$ is defined as

$$\hat{\sigma}_\tau(\mathbf{R}(\boldsymbol{\theta})) = \hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta})) \sqrt{\frac{1}{N} \sum_{n=1}^N \rho_2 \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))} \right)}, \quad (2.11)$$

where $\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))$ is the robust M-scale estimate, as defined in Eq (2.10) with ρ -function given by ρ_1 . The τ -estimate of regression [9, 10] is defined as the $\boldsymbol{\theta}$ which minimizes Eq. (2.11)

$$\hat{\boldsymbol{\theta}}_\tau = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \hat{\sigma}_\tau(\mathbf{R}(\boldsymbol{\theta})). \quad (2.12)$$

Unlike the S-estimates, τ -estimates combine a high BP and a controllable efficiency. In fact, the choice of ρ_2 , controls the efficiency of the τ -estimate, e.g., full efficiency for the Gaussian distribution can be obtained by setting $\rho_2(r) = r^2$, which corresponds to the ML-estimate. The BP of the τ -estimate, on the other hand, is controlled by ρ_1 and is the same as that of an S-estimate that uses ρ_1 . In this way, one can obtain BP=0.5 simultaneously with high efficiency. This can be understood, by the fact that $\hat{\boldsymbol{\theta}}_\tau$ in Eq. (2.12) satisfies an M-estimating equation, as given in Eq. (2.3)

$$\sum_{n=1}^N \left(W_n(\boldsymbol{\theta}) \psi_1 \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))} \right) + \psi_2 \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))} \right) \right) \mathbf{X}_n^\top = \mathbf{0}. \quad (2.13)$$

Here, the weighting factor $W_n(\boldsymbol{\theta})$ is given as

$$W_n(\boldsymbol{\theta}) = \frac{2\mathbb{E}_F \left[\rho_2 \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))} \right) \right] - \mathbb{E}_F \left[R_n(\boldsymbol{\theta}) \hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta})) \psi_2 \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))} \right) \right]}{\mathbb{E}_F \left[R_n(\boldsymbol{\theta}) \hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta})) \psi_1 \left(\frac{R_n(\boldsymbol{\theta})}{\hat{\sigma}_{M_1}(\mathbf{R}(\boldsymbol{\theta}))} \right) \right]}. \quad (2.14)$$

If

$$\rho_2(x) \geq x\psi_2(x)/2$$

is satisfied, $W_n(R_n(\boldsymbol{\theta})) \geq 0$ and we can think of the τ -estimate as an M-estimate with the adaptive ψ_τ -function

$$\psi_\tau = W_n(\boldsymbol{\theta})\psi_1 + \psi_2. \quad (2.15)$$

Clearly, if $\psi_1 = \psi_2$, the τ -estimate becomes an S-estimate.

2.1.3 Robustly Selecting the Model Order

When fitting a model to the data, questions that arises are: "Which of the candidate models under consideration is the best?" or given a specific type of models "Which variables to include?" To find answers, a fundamental approach is to trade-off data fit and model complexity. Robustness furthermore requires that a minority of the data does not override the decision that is made for the majority.

2.1.3.1 Classical Criteria

Much model selection research has been conducted since the late 1960's, overviews are given, e.g., by [34–36].

2.1.3.2 Information Criteria

An important class of approaches to model selection are the information criteria, e.g., [13, 37–44]. These can be categorized into two paradigms: efficient criteria, e.g., [13, 37–40] and consistent criteria, e.g., [41, 42, 44].

When assuming that the data generating model, henceforth denoted as the 'true model', is of infinite dimension or not contained in the set of candidates, a reasonable approach

is to choose the model with minimum mean-squared-error (MSE) distribution [45]. Based on the estimation of the Kullback-Leibler discrepancy and ML-estimation, the general expression of Akaike's information criterion (AIC) [39] is:

$$\text{AIC} = \underbrace{-2 \log(\text{likelihood})}_{\text{data fit}} + \underbrace{2 \times \text{number of model parameters}}_{\text{model complexity penalty}}$$

For the univariate linear regression model, assuming Gaussian distribution of the data, AIC becomes

$$\text{AIC}(p) = \log(\hat{\sigma}_{ML}(\mathbf{R}(\hat{\boldsymbol{\theta}}_{ML}))) + \frac{2p}{N}, \quad (2.16)$$

where $\hat{\sigma}_{ML}(\mathbf{R}(\hat{\boldsymbol{\theta}}_{ML}))$ is the sample standard deviation of the residuals of the ML-estimate $\hat{\boldsymbol{\theta}}_{ML}$ which is of dimensions $p \times 1$ and N is the sample length.

A small sample bias corrected version is the AICc [13]

$$\text{AICc}(p) = \log(\hat{\sigma}_{ML}(\mathbf{R}(\hat{\boldsymbol{\theta}}_{ML}))) + \frac{N + p}{N(N - p - 2)}. \quad (2.17)$$

A second paradigm in model selection are the consistent criteria. When the true model is of finite dimension and is included in the set of the candidate models, consistency requires asymptotically identifying the correct model with probability one. This leads to a stronger penalty term for over-fitting, compared to the AIC. E.g., the Schwarz information criterion [42] for univariate regression assuming Gaussian data becomes

$$\text{SIC}(p) = \log(\hat{\sigma}_{ML}(\mathbf{R}(\hat{\boldsymbol{\theta}}_{ML}))) + \frac{\log(N)p}{N}. \quad (2.18)$$

2.1.3.3 Robust Criteria

Since the 1980's, robust versions of classical criteria of model selection for univariate regression have been proposed [8, 11, 12, 46–53]. The most general robustification of the AIC and AICc has been proposed by [12]. By generalizing the Kullback-Leibler information so that it measures the discrepancy between a robust function, evaluated both under the true and the fitted model, the authors derived criteria that have two important properties: (i) they include the classical criteria as special case and (ii) they can be combined with M-estimators.

For the univariate linear regression model, assuming a Gaussian distribution of the majority of the data the robust criteria that are suggested by [12] are

$$\text{AICR}_M^*(p) = \sum_{n=1}^N \rho \left(\frac{R_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M(\mathbf{R}(\hat{\boldsymbol{\theta}}_M))} \right) + p \frac{\mathbb{E}_F[\psi(x)^2]}{\mathbb{E}_F[\psi(x)']} \quad (2.19)$$

with the corresponding small sample bias corrected

$$\text{AICCR}_M^*(p) = \sum_{n=1}^N \rho \left(\frac{R_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M(\mathbf{R}(\hat{\boldsymbol{\theta}}_M))} \right) + \frac{pN}{N-p-2} \frac{\mathbb{E}_F[\psi(x)^2]}{\mathbb{E}_F[\psi(x)']}. \quad (2.20)$$

2.1.4 Robust Model Selection Based on τ -Estimation

In order to utilize the advantageous properties of the τ -estimator in model selection, we propose an extension to the criteria defined in [12] to incorporate this type of estimator. By exploiting the equivalence of the τ -estimator to an M-estimator with adaptive ψ -function [9], we define

$$\text{AICR}_\tau^*(p) = \sum_{n=1}^N \rho_\tau \left(\frac{R_n(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau(\mathbf{R}(\hat{\boldsymbol{\theta}}_\tau))} \right) + p \frac{\mathbb{E}_F[\psi_\tau(x)^2]}{\mathbb{E}_F[\psi_\tau(x)']}, \quad (2.21)$$

where

$$\rho_\tau = W_n(\boldsymbol{\theta})\rho_1 + \rho_2.$$

and ψ_τ is given in Eq. (2.13). The small sample bias corrected AICCR_τ^* becomes

$$\text{AICCR}_\tau^*(p) = \sum_{n=1}^N \rho_\tau \left(\frac{R_n(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau(\mathbf{R}(\hat{\boldsymbol{\theta}}_\tau))} \right) + \frac{pN}{N-p-2} \frac{\mathbb{E}_F[\psi_\tau(x)^2]}{\mathbb{E}_F[\psi_\tau(x)']}. \quad (2.22)$$

2.1.4.1 Simulations

To evaluate the performance of the proposed criterion, simulated data is generated for the polynomial trend from [54] in additive noise

$$Y_n = 0 + 0.4X_n + 0.1X_n^2 + V_n \quad n = 1, \dots, N, \quad (2.23)$$

where $\mathbf{X} = (0, 1, \dots, N-1)^\top$ and $N = 30$. The candidate orders range from $0 \leq p \leq 6$, and the true model order is 2. The performance is evaluated both for Gaussian and additive outlier-contaminated noise. The outlier-contaminated noise is generated from an ϵ -contaminated Gaussian mixture model [2]

$$(1 - \epsilon)\mathcal{N}(0, \sigma^2) + \epsilon\mathcal{N}(0, \kappa\sigma^2) \quad (2.24)$$

with probability of contamination $\epsilon = 0.3$, impulsiveness factor $\kappa = 100$, and nominal variance $\sigma^2 = 25$. These parameters were chosen in order to gain an outlier-contaminated noise scenario with an SNR comparable to the Gaussian noise case.

For the AICCR*, when using Huber's M-estimator, $c_{\text{Hub}} = 1.345$, which ensures an efficiency of 0.95 for Gaussian errors, yields

$$\frac{\mathbb{E}_F [\psi(x)^2]}{\mathbb{E}_F [\psi(x)']^2} = 1.729.$$

To compute $\hat{\theta}_\tau$ and $\hat{\sigma}_\tau^2$, we use the fast- τ algorithm [10], which combines re-sampling with local iteratively re-weighted least-squares improvements and for which the statistical properties are very close to that of the τ -estimator. The parameter settings for the fast- τ -algorithm are: 15 random sub-samples, 2 initial iteratively re-weighted least-squares steps on each candidate and 5 candidates to fully improve. The use of Tukey's ρ -functions (see Eq. (2.7) with $c_{\text{Tuk},1} = 1.56$ and $c_{\text{Tuk},2} = 6.08$ ensure that the resulting τ -estimate has BP=0.5 and efficiency of 0.95 under Gaussian errors [9]. The penalty terms in Eqs. (2.21) and (2.22) are approximated by numerical integration and we estimate the expectations in Eq. (2.14) with their sample means.

The results of 1000 Monte Carlo runs for Gaussian noise with $\sigma^2 = 100$ are displayed in Figure 2.4. Only results of the bias corrected versions of the AIC are given here, since they were in general slightly better with similar differences between the model selection procedures, for an extensive overview, see [18]. It can be seen, that for Gaussian noise, both the robust and classical model order selection criteria choose the correct model order with a high empirical probability. Excellent results are obtained for the AICC : 86.5% and AICCR*_M : 88.2%, and slightly poorer results for the AICCR*_ τ : 78.1%. Furthermore, it is noticeable, that all criteria choose candidate models within a bounded range of model orders.

The results for the impulsive noise scenario are displayed in Figure 2.5. It can be seen, that the AICC drastically loses in performance and only chooses the correct model order with an empirical probability of 31.0%, while the AICCR*_M only degrades slightly 81.2% and the AICCR*_ τ even chooses the correct model order with a higher empirical probability of 92.9%, compared to the Gaussian noise case. Furthermore, the AICCR*_ τ is the only criterion, which remains bounded within the same range of model orders ($p = 1$ to 4), in both noise scenarios.

Further extensive simulation studies and a comparison of robust extensions for different information criteria [42, 44] have been performed in the Bachelor project by Andrea Schnall [18].

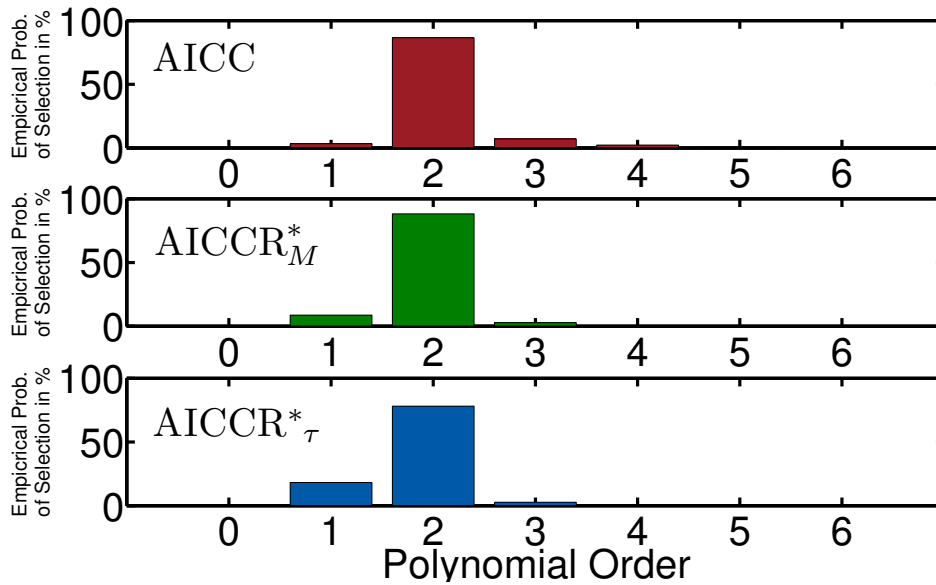


Figure 2.4: The results of model order selection for a polynomial trend in additive Gaussian noise. All criteria choose the correct model order with a high empirical probability. Excellent results are obtained with the AICC : 86.5% and AICCR_M^{*} : 88.2%, and slightly poorer results for the AICCR_τ^{*} : 78.1 %.

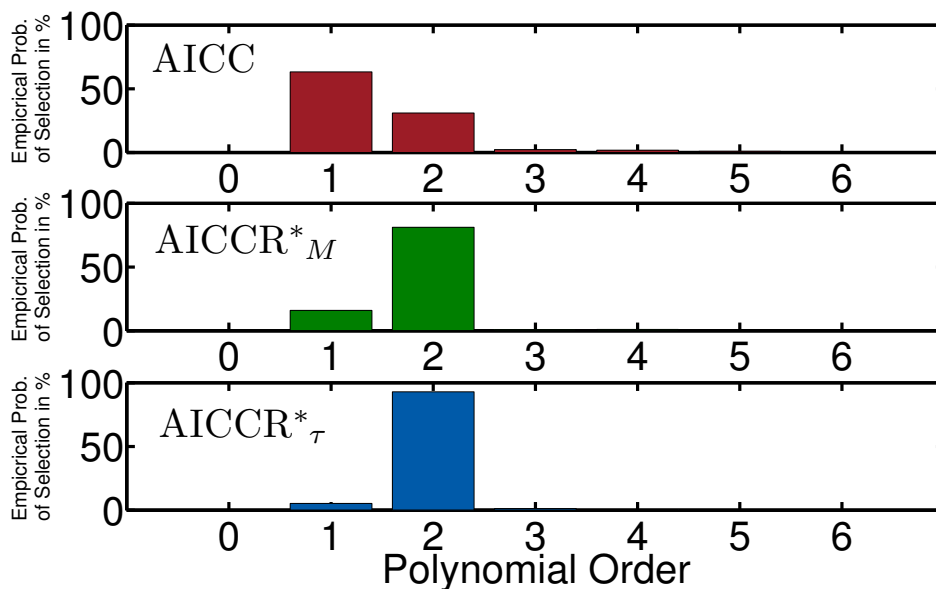


Figure 2.5: The results of model order selection for a polynomial trend in additive outlier-contaminated noise. It can be seen, that the AICC drastically loses in performance AICC : 31.0%, while the robust criteria still exhibit a high empirical probability (AICCR_M^{*} : 81.2%, AICCR_τ^{*} : 92.9%) of selecting the correct model order of $p = 2$.

2.1.4.2 Conclusions

In this Section, we derived an expression for the penalty term in robust information criteria [11–13] when using τ -estimators [9, 10]. Some simulation results were given. In [18], it is shown that the proposed criteria outperform robust information criteria based on M-estimation in a wide range of impulsive noise settings.

2.2 Robust Model Selection for Corneal-Height Data

Corneal-height data, typically measured with a videokeratoscope, is modeled as a set of Zernike polynomials [15–17]. Videokeratoscopy requires a good quality of the pre-corneal tear film and sufficiently wide eyelid aperture, which is not always fulfilled in practice. This results in missing values or outliers in the corneal topography map.

2.2.1 Contributions in this Section

In the following Sections, we show how to reliably estimate corneal topography models with a new two-step model selection procedure based on τ -estimation and the $\text{AICCR}_\tau^*(p)$ criterion.

2.2.2 Corneal Topography Estimation

The cornea is the transparent front most part of the human eye. Together with the crystalline lens, it constitutes the eye's optical system, which focuses incident light onto the retina. The cornea contributes to about two thirds of the total optical power of the eye and subtle changes in the corneal shape can strongly influence the vision process [16].

2.2.2.1 Measuring Corneal-Height Data

The state-of-the-art technique for capturing corneal-height information is videokeratoscopy, where a set of illuminated concentric rings of pre-defined geometry is projected onto the corneal surface, as illustrated in Figure 2.6. The rings are reflected off

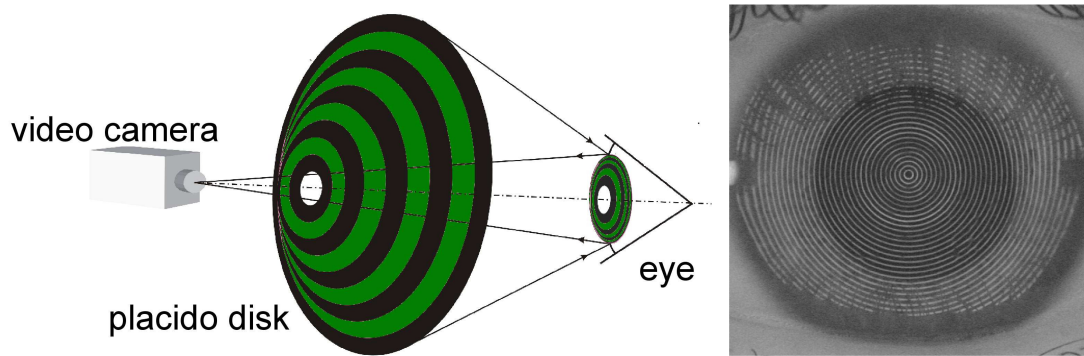


Figure 2.6: (left) Videokeratometry, illuminated concentric rings of pre-defined geometry projected onto cornea; (right) videokeratoscopic image in good measurement conditions.

the pre-corneal tear film and captured by a video camera. From these reflections, a topographical map of the cornea is generated, which reveals distortions or irregularities in the corneal shape.

2.2.2.2 Outlier Sources

In practice, corneal topography data includes measurements with break-ups in pre-corneal tear film (dry eyes), reflections from eyelashes (narrow eyelid aperture) and mucus. See Figure 2.7 for an example of such severe measurement conditions. These lead to outliers in the topographical map and therewith an inaccurate modeling of the corneal-height data.

2.2.2.3 Modeling Corneal Topography

Corneal modeling can be used in a variety of applications [16], [15], [17], e.g., corneal refractive surgery requires accurate modeling of the corneal shape prior to surgery and evaluation of changes resulting from the surgery. Also, contact lens design and fitting requires corneal topography characterization and corneal modeling can be used as a tool for screening corneal diseases, such as Keratoconus, which distorts the corneal shape and results in a significant vision loss [17].

Commonly, a Zernike polynomial expansion [15] is applied as a model to decompose the corneal-height or wavefront aberration data [55] into surfaces, which have optical interpretations. Zernike polynomials are a set of functions, which consist of a polynomial variation in the radial and a sinusoidal variation in the azimuthal direction.

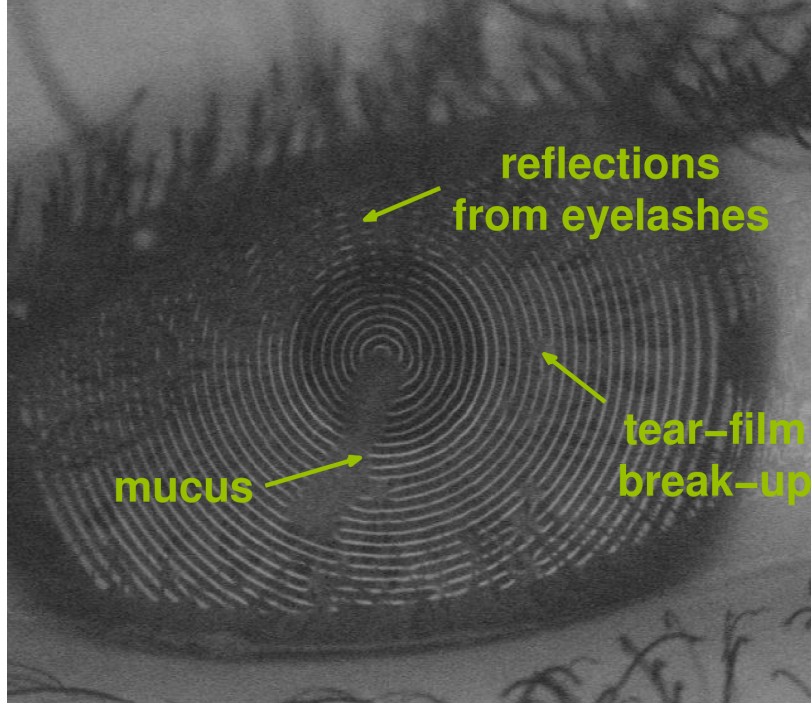


Figure 2.7: Videokeratoscopic image in severe measurement conditions.

Zernike polynomials are especially useful for representing a circular sphere, such as a section of the pupil. Because they are orthogonal in a unit circle, they may also be broken into a sum of independent components, which represent specific types of aberrations. Let $0 \leq r \leq 1$ be the normalized radial distance, and $0 \leq \phi \leq 2\pi$ the angle, as known from the polar coordinate system. A corneal surface $S(r, \phi)$ is then modeled as

$$S(r, \phi) = \sum_{m,n} a_{m,n} Z_m^n(r, \phi) + v(r, \phi),$$

where $a_{m,n}$ are the Zernike coefficients with corresponding Zernike polynomials

$$Z_m^n(r, \phi) = \begin{cases} N_n^m R_n^{|m|}(r) \cos(m\phi) & \text{for } m \geq 0 \\ -N_n^m R_n^{|m|}(r) \sin(m\phi) & \text{for } m < 0. \end{cases}$$

In the double indexing scheme, n describes the order of the radial polynomial and m describes the azimuthal frequency of the sinusoidal component. Here, m and n are always integer and satisfy $m \leq n$ and $n - |m| = \text{even}$. The radial polynomial $R_n^{|m|}(r)$ is described by the following equation

$$R_n^{|m|}(r) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s! [(n+|m|)/2 - s]! [(n-|m|)/2 - s]!} r^{n-2s},$$

where N_n^m is a normalization factor which is given by

$$N_n^m = \sqrt{\frac{2(n+1)}{1+\delta_{m0}}},$$

with δ_{m0} being the Kronecker delta function

$$\delta_{m0} = \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Today, there is an ANSI standard [56], which derives the single indexed Zernike polynomials from the double indexed scheme:

$$i = \frac{n(n+2)+m}{2}$$

The first six Zernike terms Z_1, Z_2, \dots, Z_6 are referred to as the lower order aberrations, while the remaining coefficients constitute the higher order aberrations, which cannot be corrected by sphero-cylindrical lens corrections.

An important issue encountered in eye research, which has been a subject of research for many years [14, 16, 17, 57], is to choose the appropriate set of Zernike terms. Since, in most practical cases, it is sufficient to choose only the model order of the Zernike polynomial expansion, rather than determining the particular subset [16], we suggest, in the sequel, to estimate only the radial order and choose the complete set of azimuthal and radial terms up to the determined order. This is formulated as a model order selection problem in a linear regression.

2.2.3 Model Order Selection for Corneal Topography Estimation

In this Section, we describe an approach how robust model selection can be used to model corneal-height data. Figure 2.8 shows the corneal topography map (top) of a healthy cornea, gained from a Medmont E300 videokeratoscope (Medmont Pty. Ltd., Melbourne, Australia), which has a radial resolution of 33 samples and an azimuthal resolution of 300 samples. Figure 2.8 (bottom) represents an example of a radial profile, in which missing data points are treated as outliers with a height value of $z = 0$.

2.2.3.1 Proposed Algorithm

For the determination of the radial order of the corneal-height data, we suggest the following two-step procedure:

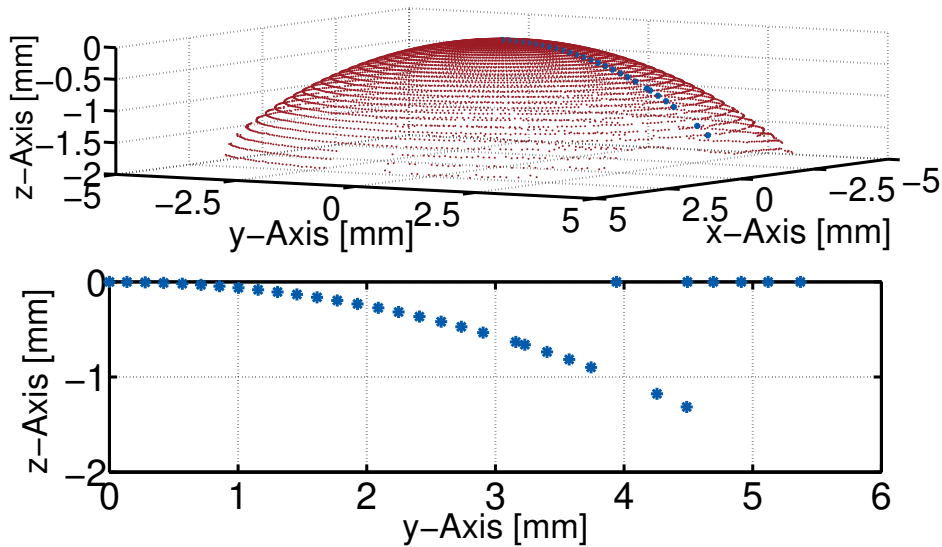


Figure 2.8: The upper plot depicts the corneal topography map of a healthy cornea, which consists of maximally 9900 data points with radial resolution of 33 samples and an azimuthal resolution of 300 samples. Breakups in the pre-cornial tear film and reflections from the eyelashes, result in missing regions. The lower plot exemplifies a radial profile, for which missing data points are treated as outliers with a height value of $z = 0$.

1. Compute the polynomial order for each radial profile individually, which gives us an empirical probability of selection for each model order.
2. Compute a statistical measure, e.g., the mode or the median of this empirical distribution, resulting in the overall radial order estimate.

2.2.3.2 Real Data Results

The following example illustrates the applicability of the suggested robust model selection procedure to determine the radial order of the Zernike polynomial expansion to fit corneal data. As described above and depicted in Figure 2.8, 300 radial profiles may be used to estimate the radial model order of the Zernike polynomial expansion. For a normally shaped, healthy cornea the clinically expected radial order is 4 to 5 [16]. Figure 2.9 displays the empirical probability with which each candidate model order chosen by the criteria. To determine the overall radial model order, a statistical measure, such as the mode or the median of this empirical distribution, can be taken. While the AICC, and even the robust AICCR_M^* overestimate the complexity of the corneal surface due to outliers (mode and median is 10), the suggested AICCR_T^* (mode and median is 5) gives results, which are in close accordance with clinical expectations for a normally shaped cornea.

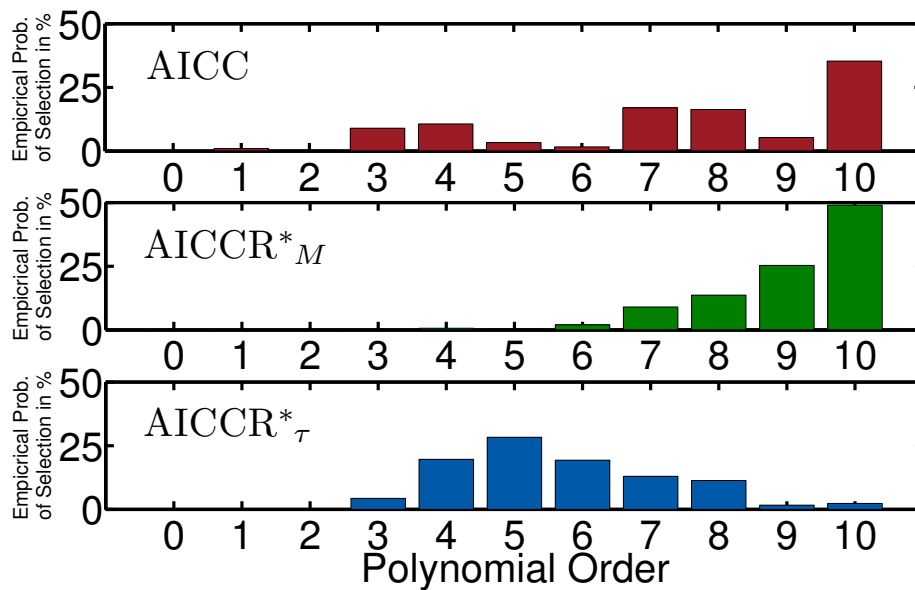


Figure 2.9: The percentage with which each candidate model order is chosen by the criteria is displayed above. The AICC, and the AICCR*_M overestimate (overall estimated order is $\hat{p} = 10$) the complexity of the corneal surface due to the outliers, while the AICCR*_τ ($\hat{p} = 5$) gives results, which are in accordance with clinical expectations.

Extensions to other robust criteria based on M-estimation are straightforward and have been investigated in the Bachelor project of Andrea Schnall, see [18] for details. Her results are based on further videokeratographic data and show that the τ -estimator based criteria are a good choice, and outperform the ones based on M-estimation. It was also shown that the τ -estimator based criteria outperform ones based on semi-parametric robust estimation [58], which can be contributed to the small sample size ($N=30$). The performance evaluation showed that a robustification of the SIC, called the SICR*_τ performs slightly better than AICCR*_τ.

2.2.3.3 Conclusions and Future Work

An algorithm was proposed and tested to reliably estimate the radial order of corneal-height data in severe measurement conditions. Based on robust information criteria and τ -estimation, a two-step model order selection was performed, which, even for moderately contaminated data, was able to reliably estimate the radial order of the Zernike polynomials in accordance with clinical expectations.

Future research will be to investigate whether it is possible to surpass the intrinsic outlier detection of the Medmont E300 videokeratoscope which causes large regions of missing data that we treat as outliers by assigning them zero height information. In

fact, these regions may contain useful information and an approach based completely on robust estimation will be promising, especially when the data has a high level of outliers and heavy mucus, which still causes trouble to the existing approach. A second future direction will be to investigate a robustification of re-sampling based techniques which have been proposed in [57]. Furthermore, an evaluation using corneal topography data of more complex structure (higher radial order), e.g., subjects with Keratoconus, will be necessary to evaluate whether the proposed algorithm is able to distinguish between irregularities in the surface and outliers due to measurement conditions.

Acknowledgments: The authors wish to thank W. Alkhaldi for helpful discussions and D.R. Iskander and the Contact Lens and Visual Optics Laboratory, QUT, Brisbane, Australia for providing data.

Chapter 3

Robustness for Array Signal Processing (Multi-Sensor Data)

3.1 Introduction

The use of multiple sensors which form an array is essential in many of today's key signal processing applications, such as, e.g., multiple-input multiple-output (MIMO) communication systems, phased array and MIMO radar systems and biomedical measurement systems, such as used in magnetoencephalography (MEG) or electroencephalography (EEG). The Gaussian assumption is often justified by the central limit theorem and is also convenient in terms of mathematical tractability. There are, however, scenarios, where the Gaussian assumption is violated, or even cases where the noise distribution is completely unknown or strongly varying in time. See, e.g., [7] for a recent review.

In this Chapter, we identified and addressed three challenges that demand robust multi-channel methods. The first research problem that we investigate is that of robust source enumeration for R -dimensional data in presence of impulsive noise, as, e.g., caused by brief sensor failures. For this, we derive criteria that are based on robust and efficient estimation of the covariance matrix using the r -mode unfolding operation of the data tensor. The second issue under investigation in this Chapter is the estimation of the complex-valued amplitudes of sinusoidal signals using multiple sensors. Here, our focus is to derive so-called robust semi-parametric estimators, which adapt to the underlying noise distribution by non-parametric transformation kernel density estimation followed by ML-estimation incorporating the estimated density. The third research focus in this Chapter is the robustness analysis of spatial time-frequency distributions (STFD). These are widely used, e.g., for direction-of-arrival (DOA) and blind source separation in case of non-stationary signals. In particular, we derive the asymptotic and finite sample influence functions for some recently proposed robust estimators.

3.2 Robust Source Number Enumeration in Case of Brief Sensor Failures (Model Order Selection for Tensor Data)

3.2.1 Contributions in this Section

In this Section, we define robust model order selection schemes for multi-dimensional data. The results presented here, are based on the MM-estimator of the covariance of the r -mode unfoldings of the complex valued data tensor. We treat the R -dimensional extensions of the AIC [40] and the minimum description length (MDL) [43], which are denoted as R -D AIC and R -D MDL, respectively.

In the context of source enumeration, we provide simulation examples for 2-D and 3-D uniform rectangular arrays. Both in the case of Gaussian noise and for a brief sensor failure, the proposed robust multi-dimensional schemes outperform their matrix computation based counterparts significantly.

Some of the contributions of this Section have been published in [59]. Some general illustrations of robustness concepts for array signal processing problems have been published in [7].

3.2.2 Source Number Enumeration for R -D Arrays

Estimating the number of signal components impinging on a sensor array is an elementary step in various signal processing tasks, such as, e.g., source separation, DOA estimation and Doppler frequency estimation. There has been much research on matrix-based array signal processing techniques during the last decades which consider various signal and noise constellations, depending on the application at hand. A number of model order selection criteria have been developed, see, e.g., [36, 60–62] and references therein, and shown to be optimal in some sense, e.g., efficient or consistent under given assumptions, such as the Gaussian distribution of the noise [63, 64].

In the last few years, there has been an increased interest in multi-dimensional array signal processing, which is advantageous, since problems are seen from multiple perspectives. Multiple dimensionality can refer to spatial dimensions, e.g., 2-dimensional or 3-dimensional arrays, but also refer to combinations of several dimensions like space, time,

frequency, and polarization. A further important advantage of using multi-dimensional data lies in the identifiability due to the higher rank of the multi-dimensional data.

A first step has been undertaken in [65] to extend classic model order selection criteria [60] to the multi-dimensional case by using tensor notation. It has been shown that taking into account the multi-dimensional structure of the data improves the estimation of the model order. Not much attention has been paid, however, how these criteria perform for small departures from the assumptions. This is crucial, since the occurrence of impulsive noise has been frequently reported in array processing applications.

Impulsive noise occurs, for example, in outdoor mobile communication channels because of switching transients in power lines or due to automobile ignitions [66], in radar and sonar systems as a result of natural or man-made electromagnetic and acoustic interference [67, 68] and in indoor wireless communication channels, owing, e.g., to microwave ovens and devices with electromechanical switches, such as electric motors in elevators, printers, and copying machines [69, 70]. It was also shown that biomedical sensor array measurements of the brain activity, such as in magnetic resonance imaging (MRI) were found to have non-Gaussian noise and interference in various regions of the human brain, where the complex tissue structure is known to exist [71]. Furthermore, when arrays are used in wireless sensor networks, the low cost and low quality sensors in harsh and unattended environments make the generated data unreliable and inaccurate [72].

A practical case which we investigate in this Section is what happens when a sensor fails for a short period of time. One could assume that the above techniques can deal with this problem, especially when the number of dimensions and correspondingly sensors, increases. We show in the following that this is not the case. In fact, even a single sensor failure during one snapshot can cause classical matrix and tensor-based model order selection criteria to break down, i.e., the ability to estimate the correct number of sources impinging onto the array decreases drastically.

3.2.2.1 Tensor Notation

For tensor notation, we follow [59, 65, 73]: scalars are represented by italic letters, column vectors by lower-case bold-face letters, matrices by bold-face capital letters and tensors are written as bold-face calligraphic letters. The (i, j) -th element of the matrix \mathbf{A} is denoted as $a_{i,j}$ and the (i, j, k) -th element of a third order tensor \mathcal{A} as $a_{i,j,k}$. The superscripts \top , $\#$, $^{-1}$ and $*$ denote transposition, Hermitian transposition,

matrix inversion, and complex conjugation, respectively. Furthermore, r -mode vectors of a tensor are obtained by keeping all indices fixed, except for the r -th index which is varied within its range. The r -mode unfolding of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_R}$ is denoted by $[\mathcal{A}]_{(r)} \in \mathbb{C}^{I_r \times (I_1 \dots I_{r-1} I_{r+1} \dots I_R)}$. The r -mode unfolding is therefore nothing else than a matrix containing the r -mode vectors of the tensor. The r -mode product of a tensor \mathcal{A} and matrix $\mathbf{U} \in \mathbb{C}^{J_r \times I_r}$ is denoted as $\mathcal{A} \times_r \mathbf{U} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times J_r \times \dots \times I_R}$. It is obtained by multiplying all r -mode vectors of \mathcal{A} from the left-hand side by the matrix \mathbf{U} .

3.2.2.2 Source Number Enumeration for R -D Arrays

The criteria that we robustify are based on the structure of the signal and noise subspace, i.e., the eigenvalues. The R -D extension consists in replacing the eigenvalues in the classical criteria by the global eigenvalues, i.e., the R -D subspace and adjusting the free number of parameters in the penalty terms.

The R -D AIC [65] chooses the model order \hat{d} as the value $k \in \{1, \dots, K\}$, which minimizes

$$R\text{-D AIC}(k) = -N(\alpha^{(G)} - p) \log \left(\frac{g^{(G)}(p)}{a^{(G)}(p)} \right) + p(2\alpha^{(G)} - p). \quad (3.1)$$

Here, $\alpha^{(G)}$ is the total number of sequentially defined eigenvalues, $a^{(G)}(p)$ and $g^{(G)}(p)$ are the arithmetic and geometric means of the smallest $p = K - k$ global eigenvalues, which are given by the product of the eigenvalues computed for every r -mode of the tensor. Similarly, the R -D MDL criterion [65] is given by

$$R\text{-D MDL}(k) = -N(\alpha^{(G)} - p) \log \left(\frac{g^{(G)}(p)}{a^{(G)}(p)} \right) + \frac{1}{2}p(2\alpha^{(G)} - p) \log(N) \quad (3.2)$$

and only differs in the penalty term. The r -mode eigenvalues are estimated by use of the sample covariance matrix of the r -mode unfolding of the data tensor \mathcal{X} , i.e.,

$$\hat{\mathbf{R}}_{XX}^{(r)} = \frac{M_r}{\prod_{i=1}^R M_i} [\mathcal{X}]_{(r)} [\mathcal{X}]_{(r)}^H \in \mathbb{C}^{M_r \times M_r}. \quad (3.3)$$

Here, M_r and M_i are the array dimensions and the total number of sensors equals $M_r \cdot M_i$.

3.2.3 Robust Covariance Matrix Estimation

3.2.3.1 Gaussian ML-Estimator of Covariance

It is well known in the array signal processing community that the sample covariance matrix is optimal in the ML-sense under the Gaussian noise assumption. However for slight deviations from the Gaussian assumption it loses drastically in performance [7, 74]. This can be explained by means of the influence function which describes the bias impact of infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination [7].

Figure 3.1 depicts the $\|\text{IF}(\mathbf{z}, \mathbf{g}_1, F)\|$ at the bi-variate complex valued standard Gaussian distribution F for a complex-valued outlier $\mathbf{z} = (z_1, z_2)^\top$ such that $z_2 = \|1\|$ is fixed and $z_1 = x_1 + jy_1$ varies. The expression of the $\text{IF}(\mathbf{z}, \mathbf{g}_1, F)$ is given in terms of the eigenvector functional \mathbf{g}_1 corresponding to the eigenvalue λ_1 , for details, see [7, 74]. In the case of the sample covariance matrix, the influence of an outlier z_1 , e.g., generated by a sensor failure is unbounded and has linear influence on the bias of the eigenvector estimate. Hence, none of the classical subspace based source enumeration methods are robust against a sensor failure and the optimality of these methods is quickly lost.

3.2.3.2 MM-Estimator of Covariance

As described for the regression case in Section 2.1.2, also for robust estimates of covariance, it is desirable to attain both a given asymptotic efficiency and BP simultaneously [75, 76]. One such estimator is the MM-estimator [77] that has been extended to the multi-variate data case by Lopuhaã [76].

An MM-estimator of covariance is basically an M-estimator based upon the observations obtained after scaling with an affine equivariant S-estimator. The resulting MM-estimator is affine equivariant and only the M-estimator determines the asymptotic efficiency independently of the initial S-estimator of covariance, as long as it is consistent. The influence function of the MM-estimator is the same as that of the M-estimator, as long as it is continuous. The BP of the MM-estimator is inherited from the S-estimator, i.e., combining a high BP S-estimator with a highly efficient M-estimator yields an MM-estimator with very favorable robustness properties. For details, see [76, 78, 79].

Figure 3.2 plots $\|\text{IF}(\mathbf{z}, \mathbf{g}_1, F)\|$ of the MM-covariance matrix estimator at the bi-variate complex valued standard Gaussian distribution F for a complex-valued outlier $\mathbf{z} =$

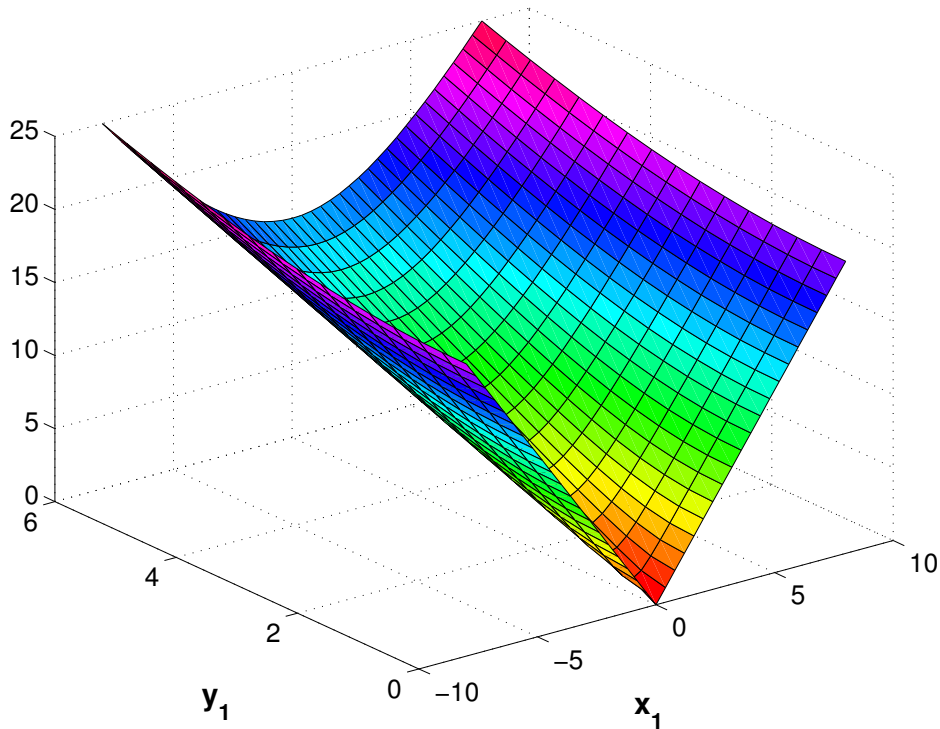


Figure 3.1: $\|\text{IF}(\mathbf{z}, \mathbf{g}_1, F)\|$ of the Gaussian ML-estimator at the bi-variate complex valued standard Gaussian distribution F for a complex-valued outlier $\mathbf{z} = (z_1, z_2)^\top$ such that $z_2 = \|1\|$ is fixed and $z_1 = x_1 + jy_1$ varies. Since the influence function is unbounded, it is non-robust against outliers, which in turn leads to unreliable source enumeration in case of sensor failures.

$(z_1, z_2)^\top$ such that $z_2 = \|1\|$ is fixed and $z_1 = x_1 + jy_1$ varies. The bounded influence function is inherited from the M-estimator. In this Figure, we depict the $\|\text{IF}(\mathbf{z}, \mathbf{g}_1, F)\|$ of an M-estimator that coincides with the t_1 -ML-estimator, for details, see [7, 74].

3.2.4 Robust Source Number Enumeration for R -D Arrays

By applying robust MM-estimation of covariance to all r -mode unfoldings of the data tensor, we obtain a robust R -D AIC that chooses the model order \hat{d} as the value $k \in \{1, \dots, K\}$, which minimizes

$$R\text{-D AIC}_{\text{rob}}(k) = -N(\alpha^{(G)} - p) \log \left(\frac{g_{\text{rob}}^{(G)}(p)}{a_{\text{rob}}^{(G)}(p)} \right) + p(2\alpha^{(G)} - p). \quad (3.4)$$

Here, $a_{\text{rob}}^{(G)}(p)$ and $g_{\text{rob}}^{(G)}(p)$ are computed analogously to $a^{(G)}(p)$ and $g^{(G)}(p)$ with the difference that the eigenvalues are estimated using the robust MM-covariance matrix estimator. Similarly, the robust R -D MDL is given by

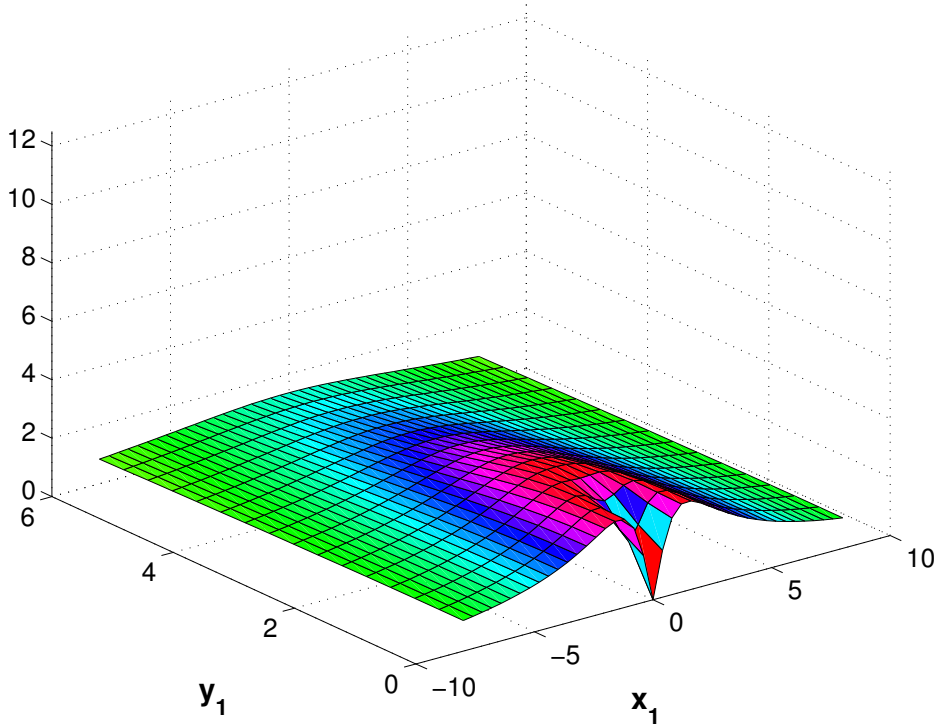


Figure 3.2: $\|\text{IF}(\mathbf{z}, \mathbf{g}_1, F)\|$ of the MM-covariance matrix estimator at the bi-variate complex valued standard Gaussian distribution F for a complex-valued outlier $\mathbf{z} = (z_1, z_2)^\top$ such that $z_2 = \|1\|$ is fixed and $z_1 = x_1 + jy_1$ varies. The $\text{IF}(\mathbf{z}, \mathbf{g}_1, F)$ is bounded and continuous, i.e., the MM-estimator is qualitatively robust, which yields reliable source enumeration in case of sensor failures.

$$R\text{-D MDL}_{\text{rob}}(k) = -N(\alpha^{(G)} - p) \log \left(\frac{g_{\text{rob}}^{(G)}(p)}{a_{\text{rob}}^{(G)}(p)} \right) + \frac{1}{2}p(2\alpha^{(G)} - p) \log(N). \quad (3.5)$$

Since the MM-estimator was designed for real-valued data, for complex valued data, stacking of the real- and imaginary-parts of the r -mode unfoldings must be performed, i.e.,

$$[\tilde{\mathcal{X}}]_{(r)} = \begin{bmatrix} \Re \left\{ [\mathcal{X}]_{(r)} \right\} \\ \Im \left\{ [\mathcal{X}]_{(r)} \right\} \end{bmatrix}. \quad (3.6)$$

By employing the MM-estimator on $[\tilde{\mathcal{X}}]_{(r)}$, as defined in Eq. (3.6), we obtain the robust estimate of its covariance matrix $\hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}, \text{robust}}^{(r)}$ that can be expressed as

$$\hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}, \text{robust}}^{(r)} = \begin{bmatrix} \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_A}^{(r)} & \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_B}^{(r)} \\ \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_C}^{(r)} & \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_D}^{(r)} \end{bmatrix}. \quad (3.7)$$

Therewith, the robust estimate of the covariance matrix of the r -mode unfolding of the data tensor \mathcal{X} can be identified as

$$\hat{\mathbf{R}}_{\mathcal{X}\mathcal{X},\text{robust}}^{(r)} = \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_A}^{(r)} + \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_D}^{(r)} + j(\hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_C}^{(r)} - \hat{\mathbf{R}}_{\tilde{\mathcal{X}}\tilde{\mathcal{X}}_B}^{(r)}). \quad (3.8)$$

3.2.4.1 Robust Source Number Enumeration R -D Uniform Rectangular Arrays

In this Section, we give two examples which illustrate the applicability of the proposed methods. We consider the relatively simple setup of 2-D and 3-D rectangular arrays. Extensions into higher dimensions are straight-forward and other applications such as EEG, where the dimensions are time, frequency, and channels can be formulated analogously.

3.2.4.2 Example 1: Robust Source Number Enumeration for a 2-D Uniform Rectangular Array

Consider a two-dimensional uniform rectangular array (URA) of dimensions $M_1 \times M_2$ with d sources impinging onto the array. The spatial frequencies for the i -th source for the two dimensions are represented by $\boldsymbol{\mu}_i = [\mu_i^{(1)}, \mu_i^{(2)}]^T$, $i = 1, \dots, d$. The vector $\mathbf{a}^{(r)}(\boldsymbol{\mu}_i^{(r)})$ denotes the array response in the r -th dimension for the i -th source, where $r = 1, 2$ in this example. Let N denote the number of available snapshots and $M = M_1 \cdot M_2$ be the total number of sensors. Let \mathbf{S} be the complex valued source symbol matrix of dimensions $d \times N$. In the classical matrix-based approach, all the spatial dimensions are stacked into column vectors. Here, we construct a measurement tensor $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times N}$ as

$$\mathcal{X} = \mathcal{A} \times_3 \mathbf{S}^T + \mathcal{W}, \quad (3.9)$$

where \mathcal{W} is i.i.d. complex circular stationary noise tensor and \mathcal{A} is the array steering tensor constructed as

$$\mathcal{A} = [\mathbf{A}_1 \sqcup_3 \mathbf{A}_2 \dots \sqcup_3 \mathbf{A}_d], \quad (3.10)$$

where \sqcup_r represents the concatenation operation along mode r , and matrix \mathbf{A}_i is obtained from the outer product of the array response vectors $\mathbf{a}^{(1)}(\mu_i^{(1)})$ and $\mathbf{a}^{(2)}(\mu_i^{(2)})$, $i = 1, \dots, d$. For this case, the global (robust) eigenvalue is given by the product of the three (robust) eigenvalues of the r -mode unfoldings of the measurement data tensor \mathcal{X} [65].

3.2.4.3 Example 2: Robust Source Number Enumeration for a 3-D Uniform Rectangular Array

For the 3-D case, with M_3 denoting the number of sensors on the third dimension of the array, we model the measurement tensor $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times M_3 \times N}$ as

$$\mathcal{X} = \mathcal{A} \times_4 \mathbf{S}^T + \mathcal{W}, \quad (3.11)$$

where \mathcal{W} is i.i.d. complex circular stationary noise tensor. Note that in this 3-D example, the spatial frequencies for the i -th source for the three dimensions are represented by $\boldsymbol{\mu}_i = [\mu_i^{(1)}, \mu_i^{(2)}, \mu_i^{(3)}]^T$, $i = 1, \dots, d$. The vector $\mathbf{a}^{(r)}(\mu_i^{(r)})$ represents the array response in the r -th dimension for the i -th source, where $r = 1, 2, 3$. The array steering tensor \mathcal{A} is then constructed as

$$\mathcal{A} = [\mathcal{A}_1 \sqcup_4 \mathcal{A}_2 \dots \sqcup_4 \mathcal{A}_d], \quad (3.12)$$

where tensor \mathcal{A}_i is obtained from the outer product of the array response vectors $\mathbf{a}^{(1)}(\mu_i^{(1)})$, $\mathbf{a}^{(2)}(\mu_i^{(2)})$ and $\mathbf{a}^{(3)}(\mu_i^{(3)})$, $i = 1, \dots, d$.

3.2.5 Simulations

3.2.5.1 Simulation Setup

The simulation setup for Example 1 is a 2-D URA array setup with parameters as follows: $M_1 = 8$, $M_2 = 8$, $d = 3$, $N = 6$, $\boldsymbol{\mu}_1 = [-\pi/2, -\pi/4]^T$, $\boldsymbol{\mu}_2 = [-\pi/4, -\pi/2]^T$, $\boldsymbol{\mu}_3 = [0, \pi]^T$, \mathbf{S} contains complex valued Gaussian source symbols. Our results are given for varying SNR are based on an average over 100 Monte Carlo runs.

The second simulation setup is a 3-D scenario with the following parameters: $M_1 = 5$, $M_2 = 7$, $M_3 = 9$, $d = 3$, $N = 10$, $\boldsymbol{\mu}_1 = [-\pi/4, 0, \pi/4]^T$, $\boldsymbol{\mu}_2 = [0, \pi/4, \pi/2]^T$, $\boldsymbol{\mu}_3 = [\pi/4, \pi/2, 3\pi/4]^T$. The other parameters were chosen as in the previous example.

To simulate a scenario where we have a very short sensor failure, we randomly replaced a single observation at a random sensor position with a complex i.i.d. impulsive noise, i.e., the contaminating density is complex, zero mean Gaussian with variance equal to $\kappa\sigma^2$, where $\kappa > 1$ determines the impulsiveness of the outliers. In the example presented here, $\kappa = 50$, other impulsive noise types produced similar results.

3.2.5.2 Simulation Results

Figure 3.3 illustrates the performance of robust and classical methods for source number enumeration. It is clearly visible here that 2-D based methods perform better than matrix-based methods. Furthermore, the robust methods based on the MM-estimator perform similarly to the non-robust ML-estimator based methods which are optimal for this setup.

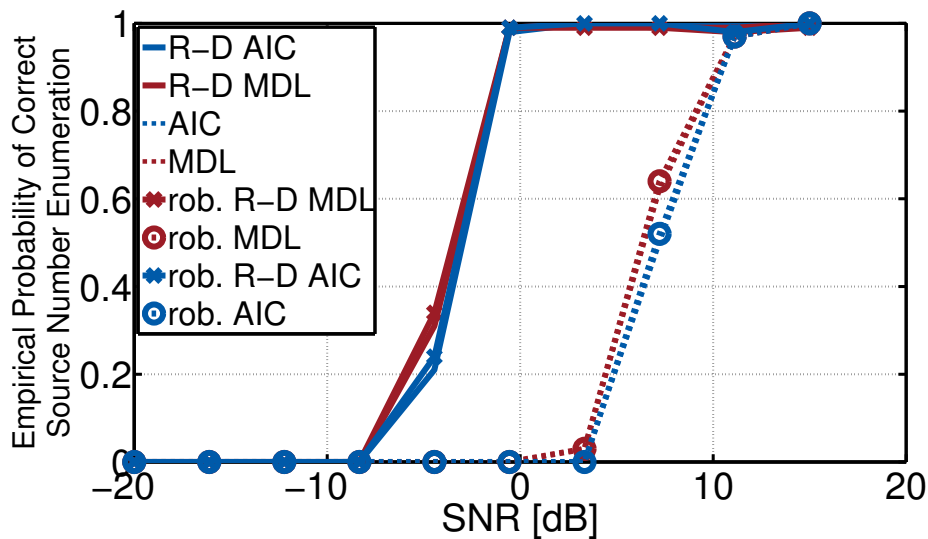


Figure 3.3: Average probability of detecting the correct number of sources for different SNR using source enumeration with a 2-D URA in case of Gaussian noise.

Figure 3.4 depicts results for a scenario where we have a very short sensor failure. The non-robustness of classical methods is apparent. In fact, the classical methods fail, while the robust ones nearly maintain the performance of the Gaussian noise case. This can be traced back to the high efficiency of the MM-estimator of covariance. Again, the 2-D based methods, in general, perform better than matrix-based methods.

Figure 3.5 shows the results for the 3-D URA source enumeration study for the Gaussian noise case. Again, the *R-D* methods provide a considerable gain in detecting the correct number of sources compared to matrix-based methods. Differences between non-robust and robust methods, as well as between AIC and MDL based methods are small.

Figure 3.6 displays the effects of a brief single sensor failure. It is evident that even for the increased number of sensors that are used in the 3-D case, a single failure causes the non-robust methods to break down.

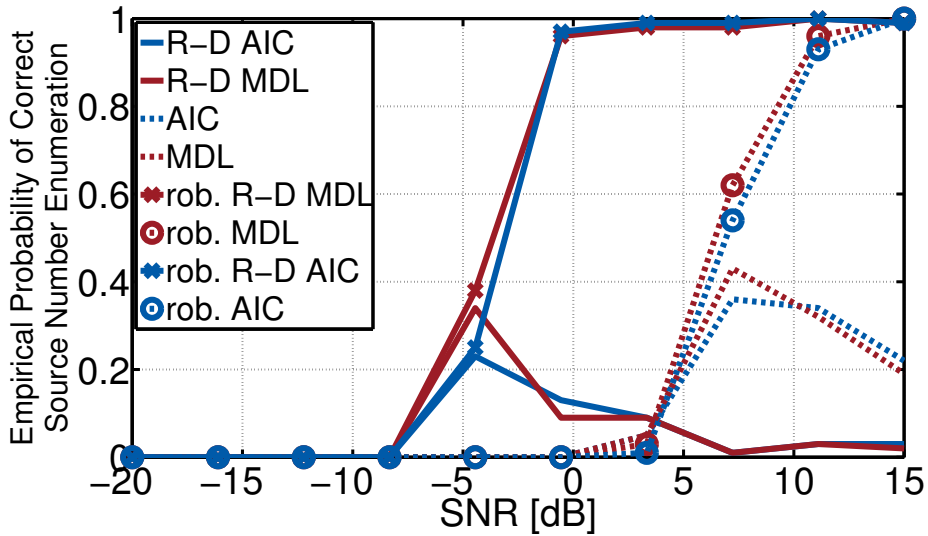


Figure 3.4: Average probability of detecting the correct number of sources for different SNR using source enumeration with a 2-D URA in case of brief sensor failure.

3.2.5.3 Conclusions and Future Work

In this Section, we investigated the problem of source number enumeration for array signal processing in the presence of brief sensor failures. For this, we introduced robust model selection criteria for multi-dimensional array data based on the robust MM-estimator of the covariance matrix using the r -mode unfolding operation of the data tensor. The proposed method is applicable to complex as well as real-valued data. In this way, we obtained robust MM-estimates of the r -mode eigenvalues which are multiplied to get global eigenvalues. The global eigenvalues were used to robustify R -D model order selection criteria.

The proposed criteria showed nearly optimal performance with 2-D and 3-D URA settings for the Gaussian noise case. For a brief sensor failure, simulated by an impulsive noise at a single random sensor position, they provided a similar performance while non-robust methods broke down. It could be noted that, in general, the R -D criteria outperformed the matrix-based ones. The difference between applying all variants of AIC and MDL was not significant for our simulations.

Future work will consider different non-Gaussian noise scenarios, dropping of the independence assumption and applying the proposed criteria to real EEG data.

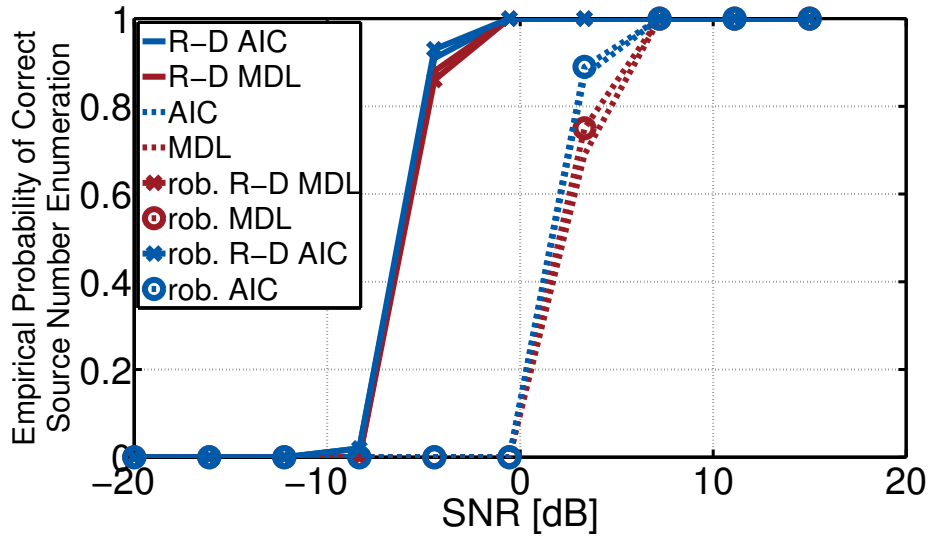


Figure 3.5: Average probability of detecting the correct number of sources for different SNR using source enumeration with a 3-D URA in case of Gaussian noise.

3.3 Robust Semi-Parametric Estimation of Sinusoidal Signals

3.3.1 Contributions in this Section

In this Section, we present a semi-parametric robust estimation approach for the estimation of the complex-valued amplitudes of sinusoidal signals using multiple sensors. This approach adapts to the sensor measurements using a compact, and conceptually simple non-parametric transformation density estimation followed by ML-estimation incorporating the estimated density. It is conceptually favorable to robust methods, when the noise distribution is not even approximately known. Simulation results are provided, which compare the performance of the proposed method with robust and non-robust estimation procedures, i.e., Huber's M-estimator, the τ -estimator and the ML-estimator under the Gaussian noise assumption.

Some of the contributions of this Section have been published in [80].

3.3.2 Problem Statement

Estimation of a parameter using multiple sensor observations is a task, which occurs in different applications, such as radar, radio and underwater applications. A widespread

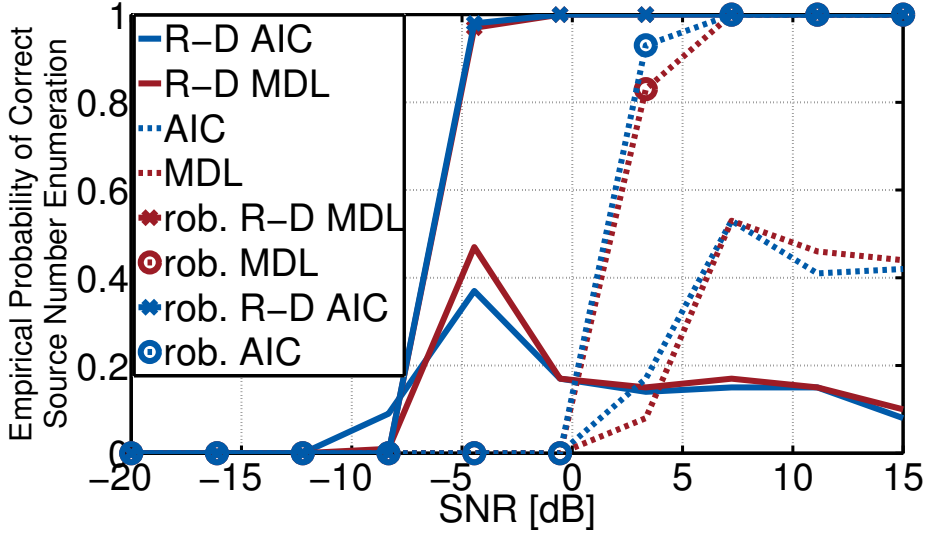


Figure 3.6: Average probability of detecting the correct number of sources for different SNR using source enumeration with a 3-D URA in case of a brief sensor failure.

noise model, for which various optimal procedures exist, is the Gaussian model. In practical array signal processing applications, however, measurement campaigns revealed the presence of impulsive noise [4, 66, 67, 69, 70], due to natural and man made electromagnetic interferences, see [7] for a recent overview.

3.3.2.1 Signal Model

In the sequel, the signal model for the estimation of the complex amplitudes of sinusoidal signals of known frequency, based on multiple sensor observations is given. Let $\mathbf{y}_m \in \mathbb{C}^N$ denote the observation vector at sensor m .

$$\mathbf{y}_m = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{v}_m \quad m = 1, 2, \dots, M \quad (3.13)$$

Here, $\mathbf{y}_m = [y_{0,m}, y_{1,m}, \dots, y_{N-1,m}]^T$, where $y_{n,m}$ is the observation at time instant n at sensor m , $\mathbf{s}(\cdot)$ is a known functional, $\boldsymbol{\theta} \in \mathbb{C}^K$ is an unknown K -dimensional parameter vector containing the complex amplitudes of the sinusoidal signals, and \mathbf{v}_m is spatially and temporally i.i.d. circular complex noise, whose elements have the p.d.f. $f_V(v)$, which is modeled by the complex-valued ε -contaminated mixture model

$$(1 - \varepsilon) \mathcal{N}_c(v; 0, \nu^2) + \varepsilon \mathcal{H}_c. \quad (3.14)$$

The ε -contaminated model, as given in Eq. (3.14), is used to describe the effect of deviations from the nominal p.d.f. $\mathcal{N}_c(v; 0, \nu^2)$. Here, $0 \leq \varepsilon \leq 0.5$ is the contamination

parameter and \mathcal{H}_c is the contamination density, which is only assumed to be symmetric, and centered around zero. The task that we consider in the sequel, is to estimate $\boldsymbol{\theta} \in \mathbb{C}^K$, given time series measurements of the $m = 1, 2, \dots, M$ sensors, in a heavy-tailed noise environment.

For $k = 1, 2, \dots, K$ sinusoidal signals of length N , $\mathbf{s}(\boldsymbol{\theta})$ in (3.13) becomes a vector, whose elements are given by

$$\sum_{k=1}^K \theta_k e^{j\omega_k n} \quad n = 0, 1, \dots, N-1, \quad (3.15)$$

where θ_k is the unknown, complex amplitude of the k -th sinusoid of known frequency ω_k . Combining (3.13) and (3.15), yields the following matrix notation [81]

$$\mathbf{y}_m = \mathbf{S}\boldsymbol{\theta} + \mathbf{v}_m \quad m = 1, 2, \dots, M, \quad (3.16)$$

where

$$\mathbf{S} = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\omega_1} & \dots & e^{j\omega_K} \\ \vdots & & \vdots \\ e^{j(N-1)\omega_1} & \dots & e^{j(N-1)\omega_K} \end{bmatrix}. \quad (3.17)$$

By separating the real and imaginary parts of (3.16), we obtain

$$\tilde{\mathbf{y}}_m = \tilde{\mathbf{S}}\tilde{\boldsymbol{\theta}} + \tilde{\mathbf{v}}_m \quad m = 1, 2, \dots, M, \quad (3.18)$$

with

$$\begin{aligned} \tilde{\mathbf{y}}_m &= \begin{bmatrix} \Re\{\mathbf{y}_m\} \\ \Im\{\mathbf{y}_m\} \end{bmatrix} & \tilde{\mathbf{S}} &= \begin{bmatrix} \Re\{\mathbf{S}\} & -\Im\{\mathbf{S}\} \\ \Im\{\mathbf{S}\} & \Re\{\mathbf{S}\} \end{bmatrix} \\ \tilde{\boldsymbol{\theta}} &= \begin{bmatrix} \Re\{\boldsymbol{\theta}\} \\ \Im\{\boldsymbol{\theta}\} \end{bmatrix} & \tilde{\mathbf{v}}_m &= \begin{bmatrix} \Re\{\mathbf{v}_m\} \\ \Im\{\mathbf{v}_m\} \end{bmatrix}. \end{aligned}$$

From this, the extended signal model can be formulated as a linear regression, cf. Eq. (2.1).

$$\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\theta}} + \mathbf{v} \quad (3.19)$$

with

$$\mathbf{y} = \begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}}_2 \\ \vdots \\ \tilde{\mathbf{y}}_M \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \tilde{\mathbf{S}} \\ \tilde{\mathbf{S}} \\ \vdots \\ \tilde{\mathbf{S}} \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} \tilde{\mathbf{v}}_1 \\ \tilde{\mathbf{v}}_2 \\ \vdots \\ \tilde{\mathbf{v}}_M \end{bmatrix}.$$

For reasons of visual clarity, in the following, estimates of $\tilde{\boldsymbol{\theta}}$ are denoted by $\hat{\boldsymbol{\theta}}$.

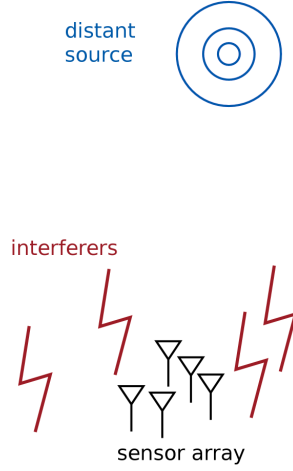


Figure 3.7: The schematic illustration of a sensor array that is used to estimate the complex amplitudes of multiple sinusoidal signals of known frequencies from distant sources is depicted. A nearby electromagnetic interferer causes impulsive noise.

3.3.2.2 Impulsive and Unknown Distribution

While in many situations, the distribution of the sensor observations can at least approximately be modeled as a Gaussian distribution, in this setup, we consider the case, where the noise distribution is not even approximately known. Figure 3.7 depicts a distant source whose sinusoidal signal impinges on a sensor array that is disturbed by a nearby impulsive interferer.

3.3.2.3 Maximum Likelihood (ML) Estimation

When the noise distribution $f_V(v)$ is known, the ML-estimator of $\tilde{\boldsymbol{\theta}}$ is obtained by maximizing the log-likelihood function $\log f_Y(y|\tilde{\boldsymbol{\theta}})$. Based on the linear regression model as given in Eq. (3.19), ML-estimation requires solving

$$\sum_{j=1}^{2MN} x_{j,k} \varphi \left(y_j - \sum_{k'=1}^{2K} x_{j,k'} \hat{\theta}_{k'} \right) = \mathbf{0}, \quad k = 1, 2, \dots, 2K, \quad (3.20)$$

where $\varphi(v) = -d \log f_V(v)/dv$ is the location score function, $x_{j,k}$ denotes the (j, k) -th element of \mathbf{X} , and $2MN$ is the length of the observation vector \mathbf{y} . If $f_V(v)$ is the density function of a $\mathcal{N}_c(v; 0, \nu^2)$ distribution, the ML-estimator of $\tilde{\boldsymbol{\theta}}$ coincides with the LS-estimator.

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (3.21)$$

When outliers occur in the observations, the ML-estimator fails, as their impact on the estimator is unbounded.

3.3.2.4 Robust Estimation

Approaches based on robust statistics [2, 3, 8] can be applied to the estimation of $\tilde{\boldsymbol{\theta}}$ in order to be less sensitive to small deviations from the assumed parametric model. M-estimation replaces $\varphi(\cdot)$ by a bounded location score function $\psi(\cdot)$, as detailed in Section 2.1.2.3. Applying Huber's $\psi(\cdot)$, as given in Eq. (2.6) yields the minimax-optimal estimator of $\tilde{\boldsymbol{\theta}}$.

The minimax estimator is the ML-solution to the least favorable distribution \mathcal{H}_c (3.14), assuming a nominal distribution, e.g., $\mathcal{N}_c(v; 0, \nu^2)$. The least favorable distribution has the interpretation of being the density for which the observations are least informative about the parameter one wishes to estimate [4]. However, the possibility of Huber's estimator in particular and M-estimators in general, to adapt to the data, is limited and therefore they may be far from optimal for completely unknown noise distributions $f_V(v)$.

In principle, any type of robust regression estimator can be applied, to estimate $\tilde{\boldsymbol{\theta}}$. In this study, we consider τ -estimators, which have some ability to adapt their ψ -function to the distribution of the data (see Section 2.1.2.5).

3.3.2.5 Semi-Parametric Robust Estimation

While the aim of robust estimation is to robustify the estimator against small deviations from the assumed model, further adaptability to the data can be achieved by semi-parametric estimation, which compensates for model misspecifications by estimating a data-dependent $\varphi(\cdot)$ non-parametrically.

Given a preliminary consistent (e.g., the ML) initial estimate $\hat{\boldsymbol{\theta}}_0$, the residuals of (3.19) can be estimated by

$$\hat{\mathbf{v}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}, \quad (3.22)$$

and based on these estimates, the score function $\varphi(v)$ can be estimated non-parametrically. Different approaches exist for the estimation of $\varphi(v)$ [82]. Here we

consider a common approach, which is to use kernel density estimation (KDE) [20], [58], where

$$\hat{\varphi}(v) = \frac{-\hat{f}'_V(v)}{\hat{f}_V(v)}. \quad (3.23)$$

However, for a heavy-tailed $f_V(v)$, KDE with a global bandwidth fails, since outliers produce peaks at the tails, causing $\hat{f}'_V(v)$ to become multi-modal, and thus a unique solution of (3.20) is no longer guaranteed. In contrast, KDE with local bandwidth selection requires the estimation of nuisance parameters, for which no optimal solutions exist [83].

A less complex, and conceptually simple approach, which is based on transformation density estimation [84], has been presented by Hammes *et al.* [58]. Here, the heavy-tailed residuals $\hat{\mathbf{v}}$ are transformed [84], so that the transformed sample distribution is approximately Gaussian [85]. For the transformed residual, KDE with a global bandwidth is performed, and $\hat{f}_V(v)$ is gained via back-transform.

One such transformation is the modulus transformation which is a monotonic and point symmetric function, that is defined as

$$w = m(v, \lambda) = \begin{cases} \text{sign}(v) \frac{(|v|+1)^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \text{sign}(v) \log(|v| + 1), & \lambda = 0. \end{cases} \quad (3.24)$$

In the case of a symmetric and heavy-tailed $f_V(v)$, the modulus transformation tends to approximately normalize the data [85], [86]. By constraining $\lambda < 1$, so that $w = m(v, \lambda)$ is concave for $v > 0$ and convex for $v < 0$, the residuals become concentrated around zero, which allows for KDE using a global bandwidth. The parameter λ can be estimated e.g., by ML-estimation [58], when we assume that $f_W(w)$, the p.d.f. of the transformed residuals (3.24), is a Gaussian p.d.f. with mean μ_W and variance σ_W^2 .

The probability distribution $f_W(w)$ of the transformed residuals is estimated by KDE

$$\hat{f}_W(w, h) = \frac{1}{2MNh} \sum_{j=1}^{2MN} \mathcal{K} \left(\frac{w - \hat{w}_j}{h} \right), \quad (3.25)$$

where $\mathcal{K}(\cdot)$ is the standard Gaussian kernel density function, \hat{w}_j is the j -th transformed residual, and h is the global bandwidth. It was found [58], that the choice of h is not critical, and the plug-in rule [83] $\hat{h} = 1.06\hat{\sigma}_W^2 N^{-1/5}$, where $\hat{\sigma}_W = \text{med}(|\hat{\mathbf{w}} - \text{med}(\hat{\mathbf{w}})|) \cdot 1.483$ is the standardized median absolute deviation (MAD) estimate, is suitable.

The density of the residuals is obtained via back-transformation as

$$\hat{f}_V(v) = \hat{f}_W(m(v, \lambda)) \cdot |m'(v, \lambda)|, \quad (3.26)$$

where differentiation of (3.26) w.r.t. v yields $\hat{f}'_V(v)$, and the final estimate $\hat{\boldsymbol{\theta}}$ is obtained by performing an iterative Newton-Raphson procedure [2], until convergence is achieved. Different algorithms for the adaptation of the step size ν , yield similar results, a possible choice is given in the algorithm of Table 3.1.

Table 3.1: The algorithm of the proposed semi-parametric robust estimation of the complex valued amplitudes of sinusoids.

1. *Initialization*
Set $i=0$. Obtain an initial estimate $\hat{\boldsymbol{\theta}}_0$
2. *Determine the residuals*
 $\hat{\mathbf{v}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}_i$
3. *Estimate the score function*
Determine $\hat{\lambda}$, perform the modulus transformation $\mathbf{w} = m(\hat{\mathbf{v}}, \hat{\lambda})$ of the residuals, estimate $\hat{f}_W(w)$ and $\hat{f}'_W(w)$, obtain $\hat{f}_V(v)$ and $\hat{f}'_V(v)$ by back-transform and estimate the score function $\hat{\varphi}(v) = -\hat{f}_V(v)/\hat{f}'_V(v)$.
4. *Update the parameter estimates*
 $\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i + \nu(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\hat{\varphi}(\hat{\mathbf{v}})$, where $\nu = 1/(1.25\max\{|\hat{\varphi}'(\hat{\mathbf{v}})|\})$
5. *Check for convergence*
For a small number $\xi \in \mathbb{R}$, stop if $\frac{\|\hat{\boldsymbol{\theta}}_{i+1} - \hat{\boldsymbol{\theta}}_i\|}{\|\hat{\boldsymbol{\theta}}_{i+1}\|} < \xi$ otherwise $i \leftarrow i + 1$ and go to step 2

3.3.3 Simulations

In this Section, the performance of the discussed estimators for the estimation of a complex amplitude of a sinusoidal signal of known frequency is compared for Gaussian and impulsive noise environments.

3.3.3.1 Simulation Setup

The frequencies and complex amplitudes of the $k = 3$ sinusoidal signals are chosen as in [81] to be $\omega_1 = 0.2\pi, \theta_1 = e^{j\pi/4}, \omega_2 = 0.22\pi, \theta_2 = e^{j\pi/3}$ and $\omega_3 = 0.6\pi, \theta_3 = e^{j\pi/4}$. The clipping point for Huber's M-estimator is set to be $c_{\text{Hub}} = 1.5 \cdot \hat{\sigma}_V$, where $\hat{\sigma}_V$ is the standardized MAD estimate. All results that are presented here, are based on $M = 4$ sensors, $N = 32$ samples, and averaging over 1000 Monte Carlo simulations. For smaller sample sizes down to $N \geq 4$, the MSE increases similarly for all estimators. The results for all $\hat{\theta}_k$ are also similar, hence only the results for $\hat{\theta}_1$ are given.

In Eq. (3.14), \mathcal{H}_c is set as $\mathcal{N}_c(v; 0, \kappa\nu^2)$, where κ controls the impulsiveness of the noise. The parameter ν^2 is chosen such, that for given ε and κ , the fixed overall variance becomes $\sigma_V^2 = (1 - \varepsilon)\nu^2 + \varepsilon\kappa\nu^2$.

3.3.3.2 Simulation Results

Figure 3.8 shows the MSE for the estimation of $\hat{\theta}_1$ for different noise environments for the ML-, M-, τ - and semi-parametric robust estimators. By varying σ_V^2 the effect of the signal-to-noise-ratio (SNR) on the MSE of the estimators is evaluated for $-5 \text{ dB} < \text{SNR} < 30 \text{ dB}$. The Gaussian noise case is included by setting $\varepsilon = 0$, while different impulsive noise cases are modeled by $\varepsilon = 0.25, 0.4$ and $\kappa = 100$.

It can be seen from the top plot of Figure 3.8 that the performance of all robust estimators is similar to the optimal ML-estimator in the Gaussian noise case. For increasing levels of impulsiveness and contamination Huber's M-estimator is increasingly outperformed by the τ -estimator and the proposed semi-parametric robust estimator, which yields best results in all impulsive noise scenarios.

3.3.4 Conclusions and Future Work

In this Section, we investigated the task of estimating the complex-valued amplitude of sinusoidal signals, based on multiple sensor observations in an unknown heavy-tailed symmetric spatially and temporally i.i.d. noise environment. A selection of non-robust and robust estimators were compared to a semi-parametric robust estimator and simulation results were provided. The performance of all robust estimators was similar to the optimal ML-estimator in the Gaussian noise case. For increasing levels

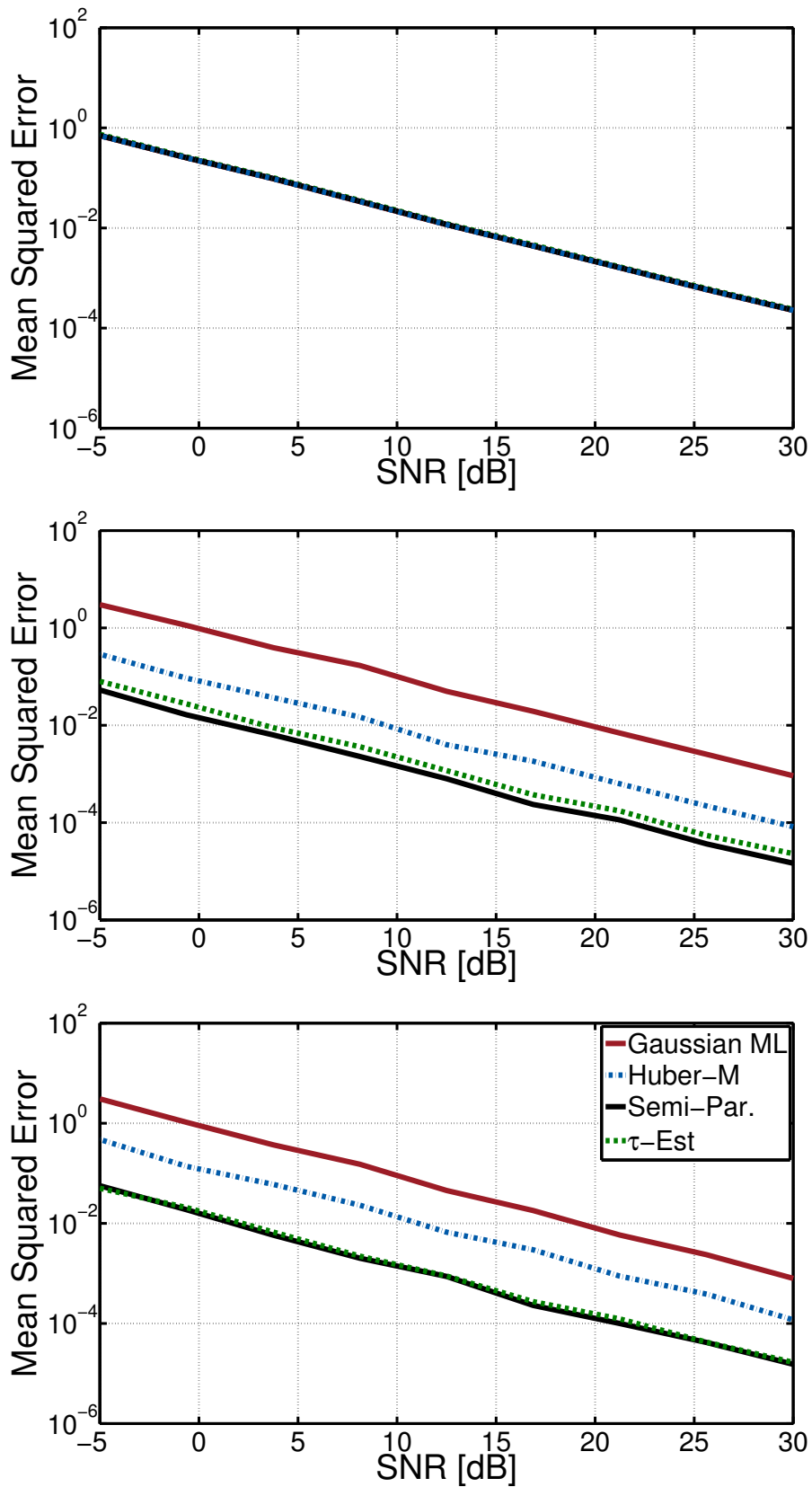


Figure 3.8: The mean squared error (MSE) for the estimation of $\hat{\theta}_1$ for different noise environments (from top to bottom $\varepsilon = 0, 0.25, 0.4$), $\kappa = 100$, $N = 32$ samples, $M = 4$ sensors, based on averaging 1000 Monte Carlo simulations.

of impulsiveness and contamination Huber's M-estimator was outperformed by the τ -estimator and the proposed semi-parametric robust estimator, which performed best in all impulsive noise scenarios.

Further studies will investigate the sensitivity of the estimator towards other deviations from the model assumptions, such as multi-modality of the noise distribution, and violations of the i.i.d. assumption.

3.4 Measuring Robustness of Time-Frequency Distributions by Means of the Influence Function

3.4.1 Contributions in this Section

In this Section, we define a generally applicable framework of robustness analysis of spatial time-frequency distribution (STFD) estimators. For this, we apply the definition of the influence function, see Section 1.2.1.2, and also that of its finite sample version, which is called sensitivity curve or empirical influence function, to STFD matrix estimators.

The results presented here, are based on the M-estimator of the STFD, for which we prove qualitative robustness, i.e., continuity and boundedness of the influence function, in case of Huber's estimator. For detailed derivations and the analysis of further estimators, the reader is referred to [87]. Array processing examples for the influence functions and empirical influence functions in the case of a uniform linear array observing an FM source are given.

It should be noted, that major derivations as well as all simulations have been performed by W. Sharif to whom we give full credit. The main contribution of the author of this thesis, in his own view, was to rigorously ensure technical correctness and to introduce uniformity in the notation for the derivations of the influence functions, as well as to contribute with some minor ideas of how to define robust statistics for non-stationary signals.

3.4.2 Robust Spatial Time-Frequency Distribution Estimation

Non-stationary signals, such as, e.g., frequency modulated (FM) signals occur in many practical applications [88, 89]. For sensor array applications, e.g., in radar, sonar and

mobile communication applications, STFD matrices have been developed and widely used over the last decade for DOA estimation [90–101] and blind source separation of FM sources [102–106].

Standard STFD estimators [90–93, 96, 103] were derived under strict distributional assumptions, which, in practice, often are not valid [7, 66, 67, 70, 107] and a serious degradation of performance is experienced in nominally optimal methods [108–112].

While nominal optimality for the assumed distribution is clearly desirable for a DOA estimator of an FM source, in practice, consideration of the performance in case of minor deviations from the assumptions made, is at least equally important. Practical situations enforce the need for robust STFD estimators [108–112], which are close-to-optimal in nominal conditions and highly reliable for real-life data, even if the assumptions are only approximately valid. Up to now, the statements on the robustness of DOA estimators are only based on simulations. In the following, we perform a formal robustness analysis of STFD estimators by means of the influence function.

3.4.2.1 Array Signal Model

In narrow-band array processing, the baseband signal model for K signals, impinging on an m -element sensor array is given by

$$\mathbf{x}(t) = \sum_{k=1}^K \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t) \quad t \in \{1, 2, \dots, N\}. \quad (3.27)$$

Here, $\mathbf{x}(t)$ is the $M \times 1$ array output vector, $\mathbf{a}(\theta_k)$ is the $M \times 1$ array steering vector for a source impinging from direction θ_k and $s_k(t)$ denotes the source waveform of the k -th signal at time t . The source's waveforms are considered to be constant amplitude deterministic FM signals with unknown parameters

$$s_k(t) = e^{j2\pi\Phi_k(t)} \quad k \in \{1, 2, \dots, K\}$$

with corresponding instantaneous frequencies $f_k(t) = d\Phi_k(t)/dt$. Here, $\Phi_k(t)$ is a continuously differentiable function, and e.g., $\Phi_k(t) = t^2$ corresponds to a linear FM signal with $f_k(t) = 2t$.

Polynomial phase signals are encountered e.g., in radar, oceanography, and ultrasound imaging [89, 113]. The noise at the sensor array $\mathbf{n}(t)$, $t = 1, \dots, N$ is from the $\mathcal{N}_c(\mathbf{0}, \sigma^2 \mathbf{I})$

distribution, hence $\mathbf{x}(t) \sim \mathcal{N}_c(\boldsymbol{\mu}(t), \sigma^2 \mathbf{I})$ and the means at the p -th and the q -th sensors are given by:

$$\mu_p(t+l) = \sum_{k=1}^K a_p(\theta_k) s_k(t+l), \quad \mu_q(t-l) = \sum_{k=1}^K a_q(\theta_k) s_k(t-l) \quad (3.28)$$

Since our focus is on the robustness analysis, we assume that the instantaneous frequencies and the DOA's of the sources are known. Tools to estimate instantaneous frequencies and DOA's exist in the literature, see, for example, [96], [114].

3.4.2.2 Classical Spatial Time-Frequency Distribution Estimation

The STFD matrix consists of auto and cross-TFDs of sensor signals as its diagonal and off-diagonal elements, respectively. Let $\mathbf{D}_{\mathbf{xx}}(t, f)$ denote the STFD matrix of the signal $\mathbf{x}(t)$ and

$$d_{p,q}(t, f) = D_{x_p x_q}(t, f), \quad t, f \in \mathbb{R}, \quad p, q \in \{1, \dots, m\}, \quad (3.29)$$

where (t, f) represents a point in the time-frequency plane, $d_{p,q}(t, f)$ stands for the (p, q) -th element of the STFD matrix $\mathbf{D}_{\mathbf{xx}}(t, f)$, and x_p and x_q are the signals corresponding to the p -th and the q -th sensor, respectively. Given observations for $t \in \{1, \dots, N\}$, $\hat{D}_{x_p x_q}(t, f)$ is assumed to be a bi-linear TFD estimate of Cohen's class [115], e.g., the pseudo Wigner-Ville distribution (PWVD) in discrete form, given by

$$\hat{D}_{x_p x_q}(t, f) = \sum_{l=-L/2}^{L/2} x_p(t+l) x_q^*(t-l) e^{-j4\pi f l}, \quad (3.30)$$

where L denotes the window length. Since the PWVD is the most commonly used STFD matrix estimator, in the sequel, we refer to it as the standard STFD estimator. It is evident from Eq. (3.30) that any outlier in either x_p or x_q has an unbounded (linear) influence on the estimate $\hat{D}_{x_p x_q}(t, f)$.

3.4.2.3 Spatial Time-Frequency Distribution M-Estimation

In this Section, we consider the M-estimator for which the (p, q) -th element of the STFD matrix is defined as a solution to the minimization of the following cost function:

$$\hat{D}_{x_p x_q}(t, f) = \arg \min_D \sum_{l=-L/2}^{L/2} \rho(d(t, l)) \quad (3.31)$$

Here, $d(t, l)$ is the distance, which is given by

$$d(t, l) = \frac{x_p(t+l)x_q^*(t-l)e^{-j4\pi fl} - D}{\hat{\sigma}_M(t)}, \quad (3.32)$$

where $\hat{\sigma}_M(t)$ is the M-estimate of scale of $|x_p(t+l)x_q^*(t-l)e^{-j4\pi fl} - D|$ for $l = -L/2, \dots, L/2$, as given in [2] p. 107. ρ is a function which ensures robustness if it is chosen in such a way that it leaves the ‘good’ data untouched and bounds the influence of an outlier. As discussed in Section 2.1.2.3, a monotone ρ -function, yields convexity of the estimation problem, e.g., Huber’s ρ -function [2] becomes

$$\rho(d(t, l)) = \begin{cases} |d(t, l)|^2/2 & \text{if } |d(t, l)| < c_{\text{Hub}} \\ c_{\text{Hub}}(|d(t, l)| - c_{\text{Hub}}/2) & \text{if } |d(t, l)| \geq c_{\text{Hub}}, \end{cases} \quad (3.33)$$

where c_{Hub} is the parameter which determines the threshold to weigh the normal and the outlying observations differently. For details on how to compute $\hat{\sigma}_M(t)$ and how to choose c_{Hub} , the reader is referred to [95]. The solution to the minimization of the cost function in Eq. (3.31) can be computed by a gradient descent based approach:

$$\frac{\partial \hat{D}_{x_p x_q}(t, f)}{\partial D} = \sum_{l=-L/2}^{L/2} \psi(d(t, l)) = 0 \quad (3.34)$$

For Huber’s ψ -function, we obtain

$$\psi(d(t, l)) = \begin{cases} d(t, l) & \text{if } |d(t, l)| < c_{\text{Hub}}, \\ c_{\text{Hub}} \text{sign}(d(t, l)) & \text{if } |d(t, l)| \geq c_{\text{Hub}}. \end{cases} \quad (3.35)$$

For the complex-valued distance $d(t, l) \in \mathbb{C}$

$$\text{sign}(d(t, l)) = \frac{d(t, l)}{|d(t, l)|}. \quad (3.36)$$

The solution to Eq. (3.31) when using a monotone ρ , can be computed e.g., by finding the root of Eq. (3.34) with iteratively re-weighted least-squares.

3.4.2.4 The Influence Function of STFD Matrix Estimators

For the (p, q) -th element of the STFD matrix, the influence function of STFD matrix estimators is defined as:

$$\begin{aligned} \text{IF}(z_p, z_q; \hat{D}_{x_p x_q}(t, f), F) &= \left. \frac{\mathbf{E}_{F_\epsilon} [\hat{D}_{x_p x_q}(t, f)] - \mathbf{E}_F [\hat{D}_{x_p x_q}(t, f)]}{\epsilon} \right|_{\epsilon \downarrow 0} \\ &= \left. \frac{\partial \mathbf{E}_{F_\epsilon} [\hat{D}_{x_p x_q}(t, f)]}{\partial \epsilon} \right|_{\epsilon \downarrow 0} \end{aligned} \quad (3.37)$$

where z_p and z_q represent the contamination in x_p and x_q , respectively, and $E_F[\cdot]$ is the expectation operator under the distribution F . For the signal model defined in Eq. (3.27), F is complex circular Gaussian $\mathcal{N}_c(\mu(t), \sigma^2)$ with variance σ^2 and time-varying mean $\mu(t)$ and F_ε is an unknown contaminating distribution.

The empirical influence function for STFD matrices, which is the finite sample counterpart of the influence function in Eq. (3.37), is obtained by replacing the statistical expectations in Eq. (3.37) by their Monte Carlo estimates. Let $\bar{D}_{x_px_q}(t, l)$ be an estimate of $E_F[\hat{D}_{x_px_q}(t, f)]$ obtained by averaging Monte Carlo realizations of $\hat{D}_{x_px_q}(t, f)$ based on the signal model defined in Eq. (3.27). Further, let $\bar{D}_{x_px_q}(t, l, z_p, z_q)$ be the Monte Carlo estimate of $E_{F_\varepsilon}[\hat{D}_{x_px_q}(t, f)]$ which is obtained by adding single point outliers z_p and z_q to $x_p(t+l)$ and $x_q(t-l)$ with $l \in \{-L/2, \dots, L/2\}$, respectively. The corresponding empirical influence function is then given by:

$$\text{EIF}(z_p, z_q; \hat{D}_{x_px_q}(t, f), F) = \frac{\bar{D}_{x_px_q}(t, l, z_p, z_q) - \bar{D}_{x_px_q}(t, l)}{1/(L+1)}, \quad (3.38)$$

where L is the window length.

3.4.2.5 Robustness Analysis of the Classical STFD Estimator

Considering the sensor signal model of Eq. (3.27) and the definition of the influence function in Eq. (3.37), we find

$$\begin{aligned} \text{IF}(z_p, z_q; \hat{D}_{x_px_q}(t, f), F) = & -2 \sum_{l=-L/2}^{L/2} \sum_{k=1}^K a_p(\theta_k) s_k(t+l) s_k^*(t-l) a_q^*(\theta_k) e^{-j4\pi fl} \\ & + z_q^* \sum_{l=-L/2}^{L/2} \sum_{k=1}^K a_p(\theta_k) s_k(t+l) e^{-j4\pi fl} \\ & + z_p \sum_{l=-L/2}^{L/2} \sum_{k=1}^K a_q^*(\theta_k) s_k^*(t-l) e^{-j4\pi fl}. \end{aligned} \quad (3.39)$$

The derivation is given in [87]. The influence function for the auto-sensor TFD is obtained by following the same steps, and differs only by an additional term for $p \neq$

q given by $-2\sigma^2$ that is equal to zero in Eq. (3.39) because of the independence assumption. The first term in Eq. (3.39) is dependent only on the signal parameters while the second and the third terms lead to non-robustness of $\hat{D}_{x_p x_q}(t, f)$, since they are linearly proportional to the magnitude of the contamination z_q and z_p , respectively. The influence function is unbounded which means that a small fraction ε of outliers approximately cause a large bias of $\varepsilon \text{IF}(z_p, z_q; \hat{D}_{x_p x_q}(t, f), F)$, for details, see [8]. This analytically confirms the non-robustness of standard STFD matrix estimators.

3.4.2.6 Robustness Analysis of the STFD M-Estimator

The derivation of the influence function is given in [87]. The resulting expression for a general ψ is:

$$\begin{aligned} \text{IF}(z_p, z_q; \hat{D}_{x_p x_q}(t, f), F) = & \left(2 \sum_{l=-L/2}^{L/2} \mathbf{E}_F [\psi(d(t, l))] - \sum_{l=-L/2}^{L/2} \mathbf{E}_{F(x_p)} [\psi(d(t, l, z_q))] \right. \\ & - \sum_{l=-L/2}^{L/2} \left(\mathbf{E}_{F(x_q)} [\psi(d(t, l, z_p))] \right. \\ & \left. \left. - \mathbf{E}_F \left[\psi'(d(t, l)) \frac{\partial d(t, l)}{\partial \hat{\sigma}_M(t)} \right] \text{IF}(z_p, z_q; \hat{\sigma}_M(t), F) \right) \right) \\ & \times \left(\sum_{l=-L/2}^{L/2} \mathbf{E}_F \left[\psi'(d(t, l)) \frac{\partial d(t, l)}{\partial D} \right] \right)^{-1} \end{aligned} \quad (3.40)$$

Here, $\mathbf{E}_{F(x_p)}[\cdot]$ and $\mathbf{E}_{F(x_q)}[\cdot]$ denote the expectations w.r.t. the distribution of x_p and x_q . $d(t, l, z_q)$ and $d(t, l, z_p)$ are the distances obtained by setting $x_q = z_q$ and $x_p = z_p$, respectively, in Eq. (3.32).

The computation of statistical expectations in Eq. (3.40) for the example of Huber's ρ -function is given in [87]. Here, it was shown that in case of Huber's M-estimator, for all terms that include z_p or z_q , down-weighting of the outliers is performed for $|d(t, l, z)| > c_{\text{Hub}}$, i.e., even for $z_{p,q} \rightarrow \infty$ the influence function remains bounded. Boundedness and continuity, which is given as long as $\psi(\cdot)$ is continuous ensure qualitative robustness of the STFD M-estimator.

3.4.3 Simulations

3.4.3.1 Simulation Setup

In our simulations, we use the array signal model defined in Eq. (3.27) for a sensor array of $m = 2$ sensors in a uniform linear array (ULA) geometry observing a single linear FM source ($K = 1$) from the broadside angle of $\theta_1 = -5^\circ$. The instantaneous frequency f_1 of the linear FM source varies from the normalized frequency value of 0.1 to 0.3. An SNR of -5 dB is used for plotting the influence function for different STFD estimators. In the case of two sensors, the single point outlier is $\mathbf{z} = [z_1, z_2]^T$, where $z_1, z_2 \in \mathbb{C}$ are complex-valued. In the following, the influence function for the $p = 1, q = 1$ auto-sensor TFD is considered and the range of z_1 is $\Re\{z_1\}, \Im\{z_1\} \in [-50, 50]$. The window length $L = 25$ is used for the computation of the PWVD based STFD matrix. For the computation of the influence function of $\hat{\sigma}_M(t)$, we approximate σ and $\mu(t)$ to be constant on an interval of length L . We divide our results into two parts. The statistical expectations for the finite sample counterpart of the influence function, i.e., the empirical influence function (EIF) as defined in Eq. (3.38) are estimated by the sample means of the STFDs which are obtained by averaging the STFD estimates from 1000 Monte Carlo runs.

3.4.3.2 Simulation Results

Figure 3.9 (left) depicts the influence function of the standard STFD matrix estimator. The influence function increases linearly with increasing magnitude of contamination, meaning that the influence of outliers is unbounded. Figure 3.9 (right) shows that the empirical influence for the standard STFD matrix estimator increases with an increase in the magnitude of the contamination and which confirms the findings based on the analytical influence function, see Figure 3.9.

The influence function of the M-estimation based STFD matrix estimator is plotted in Figure 3.10 (left). The influence of the contamination decreases with an increasing magnitude of the outlier. The largest peak of the influence function for the M-estimator of the STFD matrix is obtained for the case when the contamination $\Re\{z_1\} = \Im\{z_1\} = 0$. This means that the uncontaminated data in the observations is given the largest influence on the estimate. For visual clarity, this peak is not plotted in Figure 3.10 (left). The EIF of the M-estimation based STFD matrix estimator is plotted in Figure 3.10 (right). Again, the influence function is bounded.

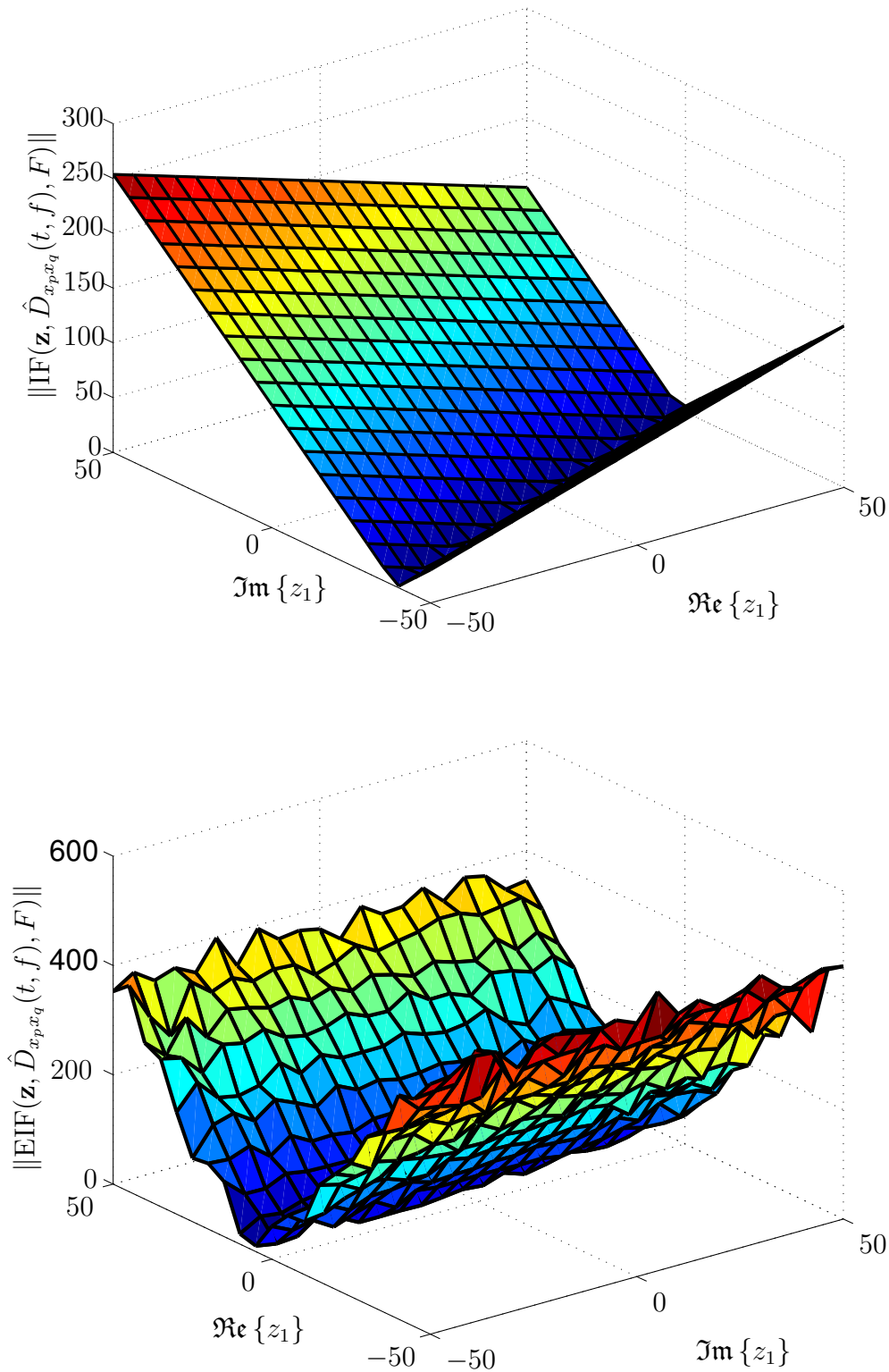


Figure 3.9: (top) Influence function of the PWVD based STF matrix estimator. (bottom) Empirical influence function of the standard STF estimator. In both cases, $p = q = 1$ and the parameters are set as described at the beginning of this Section.

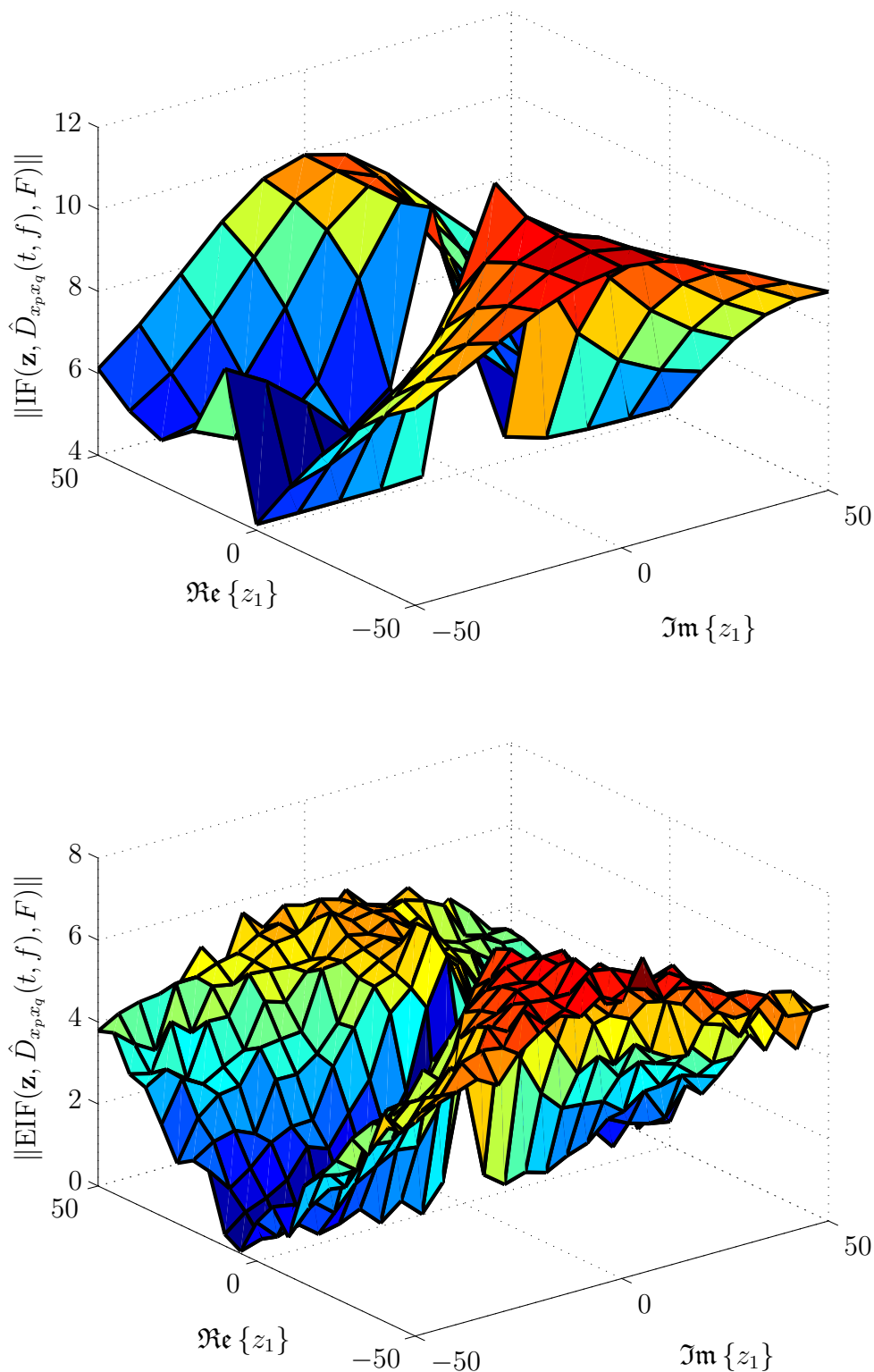


Figure 3.10: (top) Influence function of the M-estimation based STFD matrix estimator. (bottom) Empirical influence function of the M-estimation based STFD matrix estimator. In both cases, $c_{\text{Hub}} = 5$ for $p = q = 1$ and the parameters are set as described at the beginning of this Section.

3.4.4 Conclusions and Future Work

We introduced a framework for the robustness analysis of STFDs. A definition for the influence function of the STFD matrix estimator was given and the analytical expressions for the influence functions of different types of STFD matrix estimators were provided. Although herein, we treated mainly the robust STFDs, this analysis can be applied to any type of quadratic TFDs. The influence functions for the robust estimators are bounded and continuous which confirms their qualitative robustness. In addition to the asymptotic analysis, we also gave a definition for the finite sample counterpart of the influence function also called the EIF or the sensitivity curve. The simulation results for the EIF confirmed the analytical results and showed the insensitivity to small departures in the distributional assumptions for the robust techniques.

Future research on STFD estimation will allow for the design of robust and efficient estimators based on the influence function. Estimators, such as the τ - and the MM-, can be derived for STFD estimation, along with their influence functions. Furthermore, it would be of great interest to study robustness when using real-data. To do so, one could create a controlled setup where a real-data measurements which follow the nominal noise distribution as well as contaminated measurements which contain interference from an impulsive noise sources are taken. From these measurements, STFD estimates must be computed. Based on repeated measurements, one would estimate the distributions of the estimates and obtain an (empirical) influence function for the real-data case.

3.5 Conclusion

In this Chapter, we contributed to three research foci within the multi-channel setting. First, we proposed robust R -dimensional information criteria that are based on the estimator of the (complex-valued) covariance matrix of the r -mode unfoldings of the data tensor. Simulated scenarios of brief sensor failures at random sensor positions showed the superior performance of the proposed criteria compared to matrix-based criteria. Furthermore, the vulnerability of both non-robust R -dimensional and matrix-based criteria to such events was demonstrated and remedied by the proposed criteria. We then presented a robust semi-parametric estimation approach for the estimation of the complex-valued amplitudes of sinusoidal signals impinging onto a sensor array. This approach, in a simulation study, outperformed even advanced robust methods, such as the τ -estimator for all levels of contamination. Finally, in this Chapter, we

performed influence function based robustness analysis of STFD estimators. Simulation examples were given for a single linear FM source impinging onto a uniform linear sensor array. Both the Monte Carlo averaged and asymptotically derived influence functions confirmed the non-robustness of the standard PWVD-based estimator and the qualitative robustness of the M-estimation based STFD estimator.

Further research directions concerning the above mentioned issues have been detailed in Sections 3.3.4, 3.2.5.3 and 3.4.4. A promising new line of research, which goes beyond the scope of this project, but is of very high future interest, is the robustness in case of distributed signal processing, see Section 5.2.

Chapter 4

Robustness for Dependent Data

4.1 Introduction

Although there exists a considerable body of literature on robust estimation methods for engineering practice under the assumption of i.i.d. observations, for a long time, estimators for dependent data were limited. While the first contributions were made in the 1980's [4, 116, 117], there was not much progress on robust estimation for dependent data for some time. This is mainly due to the fact that existing robust estimators and concepts for i.i.d. data are not easily extendible to dependent data, and thus, new robust approaches are sought for. Recently, research increased significantly in this area [8, 27, 118–131] and some novel estimators have been proposed. In this Section, we consider the robust estimation, both of the parameters and the model orders for autoregressive moving-average (ARMA) models.

4.1.1 Contributions in this Chapter

The contributions in this Chapter include some practical applications of robust dependent data methods to solve real-world problems. In particular, we develop artifact detection and mitigation strategies that are based on data transformations. These are applied to real-world electrocardiogram and intracranial pressure data. We also derive and analyze sophisticated robust estimation and model order selection techniques for ARMA models. A fast algorithm as well as a detailed statistical and robustness analysis of a novel robust and efficient estimator, are given. For the proposed estimator, which is termed the bounded influence propagation (BIP) τ -estimator, we compute a complete statistical robustness analysis, which includes conditions for the consistency, as well as a proof of qualitative and quantitative robustness. The estimator is very suitable and attractive for ARMA model order selection purposes and we propose and compare different robust model order selection criteria that are based on the BIP τ -estimator.

Some contributions in this Chapter have been published in [7, 130, 131]. A paper concentrating on the ARMA parameter estimation and model order selection with the BIP τ -estimator [132] is to be submitted in December 2013.

4.2 Modeling Dependent Data

4.2.1 Autoregressive Moving-Average (ARMA) Models

One of the most popular approaches the signal processing practitioner uses to model dependent data is the ARMA model [133]. This is not much different in robust statistics.

An ARMA(p, q) model is defined as

$$X_n + \sum_{k=1}^p a_k X_{n-k} = Z_n + \sum_{k=1}^q b_k Z_{n-k}, \quad n \in \mathbb{Z}, \quad (4.1)$$

where Z_n is a sequence of zero-mean i.i.d. random variables with finite variance σ_Z^2 (i.e., white-noise). Important subclasses of ARMA models are the autoregressive AR(p) models

$$X_n + \sum_{k=1}^p a_k X_{n-k} = Z_n, \quad n \in \mathbb{Z}, \quad (4.2)$$

which are obtained from Eq. (4.1) when $q = 1$ and $b_1 = 0$, as well as the moving-average MA(q) models:

$$X_n = Z_n + \sum_{k=1}^q b_k Z_{n-k}, \quad n \in \mathbb{Z}, \quad (4.3)$$

which are obtained by setting $p = 0$ in Eq. (4.1).

A convenient representation of Eq. (4.1) is

$$A_p(z)X_n = B_q(z)Z_n, \quad n \in \mathbb{Z}. \quad (4.4)$$

Here, $A_p(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p}$ and $B_q(z) = 1 + b_1 z^{-1} + \dots + b_q z^{-q}$, where z^{-k} is the lag operator defined by $z^{-k} X_n = X_{n-k}$, $k \in \mathbb{Z}$. To ensure stationarity of the process X_n , the polynomial $A_p(z)$ must have roots inside the unit circle. Invertibility of the system is obtained when the same condition is satisfied for $B_q(z)$.

By defining

$$B^\infty(z) = A^{-1}(z)B(z) = 1 + \sum_{k=1}^{\infty} b_k^\infty z^{-k}, \quad (4.5)$$

we obtain the MA(∞) representation of Eq. (4.1)

$$X_n = Z_n + \sum_{k=1}^{\infty} b_k^\infty Z_{n-k}, \quad n \in \mathbb{Z}.$$

4.2.2 Classical Parameter Estimation for ARMA Models

According to Stoica and Moses [134], there is no well-established algorithm for estimating the parameters of ARMA models from both theoretical and practical standpoints. As in other estimation problems, ML-estimators are known to have desirable asymptotic properties, e.g., they are consistent and achieve the Cramér-Rao lower bound asymptotically. However, for the estimation of ARMA models, ML-estimation requires a numerical search procedure to find the maximum of the likelihood function given constraints on the parameter space, such that one will end up with a stationary and invertible model. Because of finite sample performance, as well as for computational complexity reasons, ML-estimation has never become very popular for estimating ARMA models in practice [134–136]. Instead, different statistically sub-optimal, but computationally cheaper methods, e.g., [134, 136–140], have been proposed and compared to the Cramér-Rao lower bound.

4.2.2.1 Durbin’s Method and Reduced Statistics Estimators for ARMA Models

In the sequel, we describe a set of methods [136, 141] that improve Durbin’s two stage method [138]. The key idea is to perform an initial splitting of the dynamics of an ARMA model into the AR and MA parts, based on a long autoregressive model. Given the initial estimate of the AR parameters, the second stage consists of first computing the final MA parameters using the initial estimates and then improving the first stage AR estimate. The attractiveness of this method in terms of robustness is, that the second stage only uses the parameter estimates from the first stage, not the data itself. This means that if we are able to estimate a long AR model robustly, we can derive the corresponding ARMA model without further use of the outlier-contaminated observations. Also, the computational cost of computing the ARMA model robustly is basically that of computing the long AR model, since the reduced statistics estimator’s contribution to the computational cost is negligible compared to that of the robust estimator. There exist different possibilities to perform the initial separation, none of which are superior to the others in all given settings.

We describe four possibilities to compute the MA parameters in the first stage, i.e.,

- a) The “long AR” estimator [139], which employs the relation

$$\frac{\hat{B}(z)}{\hat{A}(z)} \approx \frac{1}{\hat{A}^{p_0}(z)}, \quad (4.6)$$

where $\hat{A}(z)$ and $\hat{B}(z)$ are estimates of the polynomials, as given in Eq. (4.4), and $\hat{A}^{p_0}(z)$ is the AR(p_0) approximation of the ARMA(p, q) model obtained by using a high order denoted as p_0 . Since $p_0 > p$ holds, $a_k = 0$ for $k > p$ and therewith the first stage MA(q) parameters b_1, b_2, \dots, b_q can be estimated without knowing the first AR parameters by finding the LS solution of

$$\sum_{k=0}^q \hat{b}_k \hat{a}_{m-k}^{p_0} = \zeta_m, \quad m = p+1, \dots, p_0. \quad (4.7)$$

Here, ζ_m represents the inaccuracies of the approximation made in Eq. (4.6) and a solution exists, if $p_0 \geq p+q$. The initial AR(p) parameters can then be computed by substituting the estimated initial MA(q) parameters into

$$\sum_{k=0}^q \hat{b}_k \hat{a}_{m-k}^{p_0} = \hat{a}_m, \quad m = 0, \dots, p, \quad (4.8)$$

or by using the second stage AR method which is described later.

- b) The “long MA” estimator [136], which uses

$$\frac{1}{\hat{A}^{p_0}(z)} \approx \hat{B}^{q_0}(z), \quad (4.9)$$

where $q_0 \gg p_0$ is chosen such that the impulse response computed from $\hat{A}^{p_0}(z)\hat{B}^{q_0}(z) = 1$ dies out at q_0 . Knowing that the MA(k) parameters are zero for $k > q$, the initial AR parameters are computed as the LS solution to

$$\sum_{k=0}^p \hat{a}_k \hat{b}_{m-k}^{q_0} = \zeta_m, \quad m = q+1, \dots, q_0. \quad (4.10)$$

- c) The “long COV” estimator [136], where, first, the estimate of the covariance function $\hat{c}_{XX}(\kappa)$ is obtained from the Yule-Walker equations:

$$\sum_{k=1}^{p_0} \hat{a}_k^{p_0} \hat{c}_{XX}(m-k) = 0, \quad m = 1, 2, \dots, p_0 \quad (4.11)$$

with $\hat{c}_{XX}(-|\kappa|) = \hat{c}_{XX}(\kappa)$ for $\kappa > 0$. Using $\hat{c}_{XX}(\kappa)$, the initial AR parameters are obtained from the LS solution to

$$\sum_{k=0}^p \hat{a}_k \hat{c}_{XX}(m-k) = \zeta_m, \quad m = q+1, \dots, p_0. \quad (4.12)$$

Again, a solution exists, if $p_0 \geq p+q$. $\hat{B}(z)$ can then be computed by substituting the estimates into Eq. (4.6).

- d) The “long RINV” estimator [136] uses the inverse correlations $r_{\text{inv}}(\kappa)$, which are defined by interchanging the AR and MA polynomials [140]. Given $\hat{A}^{p_0}(z)$, $r_{\text{inv}}(\kappa)$ can be estimated with

$$\hat{r}_{\text{inv}}(\kappa) = \sum_{k=0}^{p_0-\kappa} \hat{a}_k^{p_0} \hat{a}_{k+\kappa}^{p_0}, \quad \kappa = 0, 1, \dots, p_0. \quad (4.13)$$

Additionally, $\hat{r}_{\text{inv}}(\kappa > p_0) = 0$ and $\hat{r}_{\text{inv}}(-|\kappa|) = \hat{r}_{\text{inv}}(\kappa)$. The initial MA parameters are thus found as the LS solution to

$$\sum_{k=0}^q \hat{b}_k \hat{r}_{\text{inv}}(m-k) = \zeta_m, \quad m = p+1, \dots, p_0 \quad (4.14)$$

and the initial AR parameters are found in the same way as described for method a).

Clearly, none of these estimators are statistically optimal in any sense. It is hoped that at least one of the four yields an adequate initialization $\hat{A}^0(z)$ for the second stage of Durbin’s method [138] that consists of two steps: (i) estimate the final MA parameters $\hat{B}(z)$ using approximate relations between the polynomials $\hat{A}^{p_0}(z)$ and $\hat{A}^0(z)$. (ii) estimate the final AR parameters $\hat{A}(z)$ with the help of $\hat{B}(z)$ and $\hat{A}^{p_0}(z)$. For details, the interested reader is referred to [137, 138]. An interesting frequency domain formulation of the second stage of Durbin’s method has been provided by [136] as follows:

- (i) Estimate the final MA parameters $\hat{B}(z)$ via

$$\hat{B}(e^{j\omega}) \approx \underset{B(e^{j\omega})}{\text{argmin}} \frac{N}{2\pi} \int_{-\pi}^{\pi} \left| \frac{B(e^{j\omega}) \hat{A}^{p_0}(e^{j\omega})}{\hat{A}^0(e^{j\omega})} \right|^2 d\omega. \quad (4.15)$$

This is intuitively understandable by defining

$$\text{ME} \left(\frac{1}{B(e^{j\omega})}, \frac{\hat{A}^{p_0}(e^{j\omega})}{\hat{A}^0(e^{j\omega})} \right) := \frac{N}{2\pi} \int_{-\pi}^{\pi} \left| \frac{B(e^{j\omega}) \hat{A}^{p_0}(e^{j\omega})}{\hat{A}^0(e^{j\omega})} \right|^2 d\omega,$$

where $\text{ME} \left(\frac{\hat{A}(z)}{\hat{B}(z)}, \frac{A(z)}{B(z)} \right)$ is a measure of the model error for ARMA models estimated from N observations, see [136] for details.

- (ii) Estimate the final AR parameters $\hat{A}(z)$ via

$$\hat{A}(e^{j\omega}) \approx \underset{A(e^{j\omega})}{\text{argmin}} \frac{N}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A(e^{j\omega})}{\hat{B}(e^{j\omega}) \hat{A}^{p_0}(e^{j\omega})} \right|^2 d\omega, \quad (4.16)$$

where again the intuition is to minimize the ME

$$\text{ME} \left(\frac{1}{A(e^{j\omega})}, \frac{1}{\hat{A}^{p_0}(e^{j\omega}) \hat{B}(e^{j\omega})} \right) = \frac{N}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A(e^{j\omega})}{\hat{B}(e^{j\omega}) \hat{A}^{p_0}(e^{j\omega})} \right|^2 d\omega.$$

Models estimated with Durbin's second method have the important property that they are always stable and invertible.

4.2.3 Outlier Models

There exist several probability models for outliers in dependent data. Outliers can, e.g., replace an observation or be an additive contaminant. Furthermore, they can occur in X_n or in Z_n . Outliers may also appear as isolated events or in patches. In the sequel, we introduce some frequently used outlier models.

4.2.3.1 Additive Outlier (AO) Model

For the so-called additive outlier (AO) model, a contaminated process Y_n is given by

$$Y_n = X_n + V_n^\varepsilon, \quad n \in \mathbb{Z}, \quad (4.17)$$

where

$$V_n^\varepsilon = \xi_n^\varepsilon W_n, \quad n \in \mathbb{Z}. \quad (4.18)$$

Here, W_n denotes the contaminating process and ξ_n represents a stationary random process for which

$$\xi_n^\varepsilon = \begin{cases} 1 & \text{with probability } \varepsilon \\ 0 & \text{with probability } (1 - \varepsilon). \end{cases} \quad (4.19)$$

4.2.3.2 Replacement Outlier Model

For the so-called replacement outlier model, a contaminated process Y_n is given by

$$Y_n = X_n + V_n^\varepsilon, \quad n \in \mathbb{Z},$$

where

$$V_n^\varepsilon = \xi_n^\varepsilon (W_n - X_n), \quad n \in \mathbb{Z},$$

with W_n denoting the contaminating process which is independent of X_n and ξ_n as defined in Eq. (4.19).

4.2.3.3 Innovation Outlier (IO) Model

For the so-called innovation outlier (IO) model, the innovations Z_n in Eq. (4.1) contain outliers. This can happen either when the distribution of Z_n is heavy-tailed, see e.g., [142], or when Z_n is contaminated by outliers, e.g.,

$$Z_n^\varepsilon = Z_n + V_n^\varepsilon, \quad (4.20)$$

with V_n^ε as given in Eq. (4.18) for the additive innovation outliers and where for replacement innovation outliers

$$V_n^\varepsilon = \xi_n^\varepsilon(W_n - Z_n), \quad n \in \mathbb{Z}.$$

Here, W_n is independent of Z_n .

4.2.3.4 Patchy and Isolated Outlier Models

For dependent data, outliers may occur isolated or they may be grouped in patches. In case of isolated outliers, ξ_n^ε takes the value 1, such that at least one non-outlying observation is between two outliers (e.g., follows an independent Bernoulli distribution). For patchy outliers on the other hand $\xi_n^\varepsilon, n \in \mathbb{Z}$ takes value 1, for $n_p \leq N/2$ subsequent values. Patchy outliers occur in many practical situations, e.g., when switching between line-of-sight/non-line-of-sight environments in geolocation position estimation [22], as a result of a loose connection in a sensor [59], which creates patchy impulsive noise [72] or in case of motion artifacts in patient movement for biomedical measurements [130,131].

4.3 Artifact Detection and Removal Using Data Transformations

In some cases, it is possible to mitigate the effects of outliers by means of outlier detection. Outlier detection should be based on robust estimation procedures. Finding outliers in dependent data is not always straight forward. In the following Section, we describe two practical examples, where we apply data transformations which enable outlier detection in the transformed domain. After detecting the outlier positions, one can either apply a form of data cleaning followed by classical or robust estimation, or use estimators that can handle missing data.

4.3.1 Motion Artifact Removal in Electrocardiogram (ECG) Signals Using Multi-Resolution Thresholding

The electrocardiogram (ECG) is a powerful non-invasive tool which contains information that helps in the diagnosis of a wide range of heart conditions. Figure 4.1 depicts an ECG of a healthy subject. The P-,Q-,R-,S-,T- waves, as indicated in the plot, are features which are useful in a broad range of applications. Today's health applications include monitoring patients with portable ECG recording devices that are equipped with a transmitter in order to communicate health related information and to trigger alarms in case of life threatening situations. However, these devices suffer severely from patient motion-induced artifacts. While much research has been conducted to remove time-invariant noise from ECG signals, the removal of motion-induced artifacts remains an unsolved problem [130].

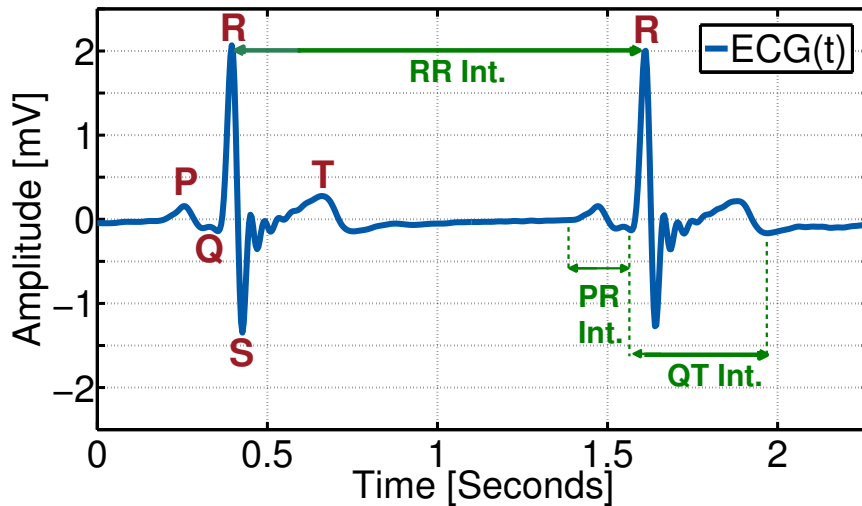


Figure 4.1: The ECG is a powerful non-invasive tool which contains information that helps in the diagnosis of a wide range of heart conditions. The P-,Q-,R-,S-,T- waves, and corresponding intervals, are features which are useful in a broad range of applications.

In this Section, we therefore briefly describe a method that we introduced in [130], which removes these artifacts by obtaining an estimate of the artifacts in the stationary wavelet domain (SWD) [143]. The idea of our method is to estimate a signal $\hat{s}_{\text{outliers}}$, that represents the motion-induced artifacts and impulsive noise. This estimate is then subtracted from the recorded signal to obtain a cleaned ECG signal, i.e.,

$$\hat{s}_{\text{cleaned}} = s_{\text{recorded}} - \hat{s}_{\text{outliers}}. \quad (4.21)$$

The additive outlier model is well motivated in this case, since motion-induced artifacts are additive disturbances. We use the diversity of the coefficient sequences obtained by

the stationary wavelet transform (SWT) and a multi-resolution thresholding methodology based on robust estimation to find $\hat{s}_{\text{outliers}}$. The scheme of the algorithm is shown in Figure 4.2. A real-data illustration of the signals given in Eq. (4.21), is given in Figure 4.3. In this case, the motion artifacts follow a patchy additive outlier model. Figure 4.4 plots an example of the fifth detail coefficient \mathbf{d}_5 of the stationary wavelet transformed data from Figure 4.3 (top). Much of $\hat{s}_{\text{outliers}}$ is concentrated in this coefficient, while signal content is reduced to a simple structure, which allows for a good separation of signal and artifacts. For an extensive performance analysis and an application to R-peak detection based on simulated and real measured data, the interested reader is referred to [130].

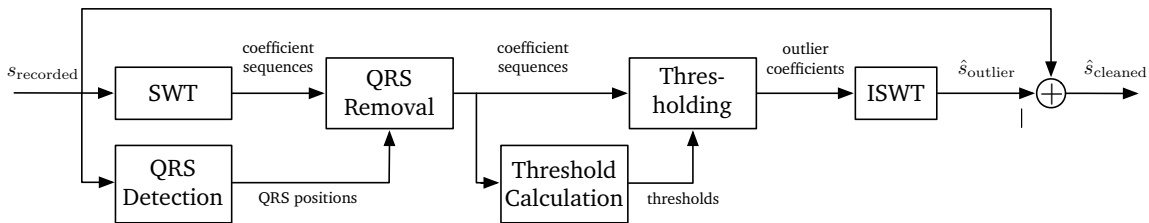


Figure 4.2: The ECG motion artifact removal algorithm performs an estimation of the artifact signal via outlier detection in the stationary wavelet domain (SWD). The data is transformed to allow for better separability of the signal and artifact components. Robust estimation is used to detect the QRS complexes and to set the thresholds.

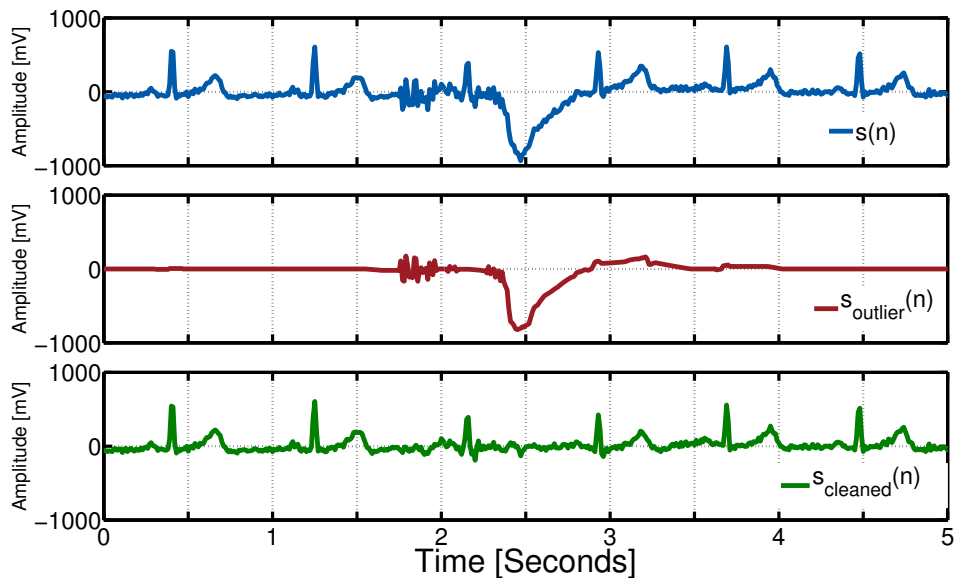


Figure 4.3: (top) An example of a recorded ECG signal s_{recorded} . (middle) An estimate \hat{s}_{outlier} of the motion artifact signal obtained by the multi-resolution thresholding algorithm. (bottom) The artifact cleaned signal \hat{s}_{cleaned} , see Eq. (4.21).

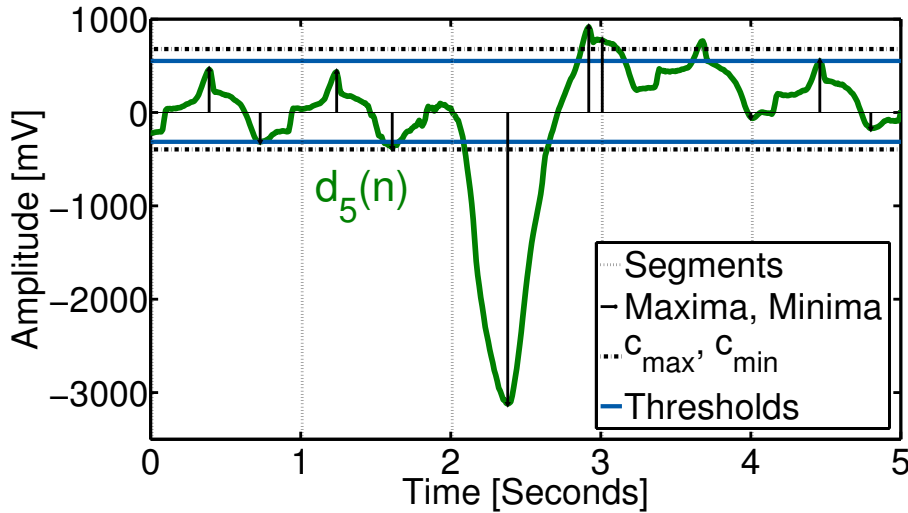


Figure 4.4: An example of the 5th detail coefficient d_5 of the stationary wavelet transformed ECG measurement. The thresholds are based on robustly estimated averages of maxima and minima calculated within segments of one second length. Each coefficient series has independent thresholds, which allows for exploiting the diversity of the signal and artifact components in the stationary wavelet domain.

4.3.2 Artifact Detection in the Empirical Mode Domain for the Forecasting of Intracranial Pressure (ICP) Signals

The monitoring of intracranial pressure (ICP) signals is common practice for patients who suffered a traumatic brain injury. A danger for these patients is that the primary brain damage, caused by the accident, can lead to a secondary pathophysiological damage, which usually occurs together with a significantly high or low ICP value. Currently, manual observation and judgment of the ICP signals by the nurses is used to predict whether ICP levels are likely to raise or drop significantly and to call a doctor who gives medication. This method risks human errors and suffers from ineffectiveness [144]. Accurate ICP forecasting enables active and early interventions for more effective control of ICP levels. The major difficulties that arise in this application are (i) the non-stationarity of ICP signals, which is too high to be canceled by methods like time-differentiation or time-segmentation and (ii) the inevitable artifacts, which are caused by motion of the patients or by equipment errors [145]. Figure 4.5 plots an example of an ICP measurement, where some artifacts are highlighted with red crosses.

In this Section, we briefly present a method that we introduced in [131] which uses combined artifact detection and robust estimation after a data transformation into the empirical mode domain. The empirical mode decomposition (EMD) [146] decomposes a non-stationary ICP signal \mathbf{x} into a set of second order stationary components via

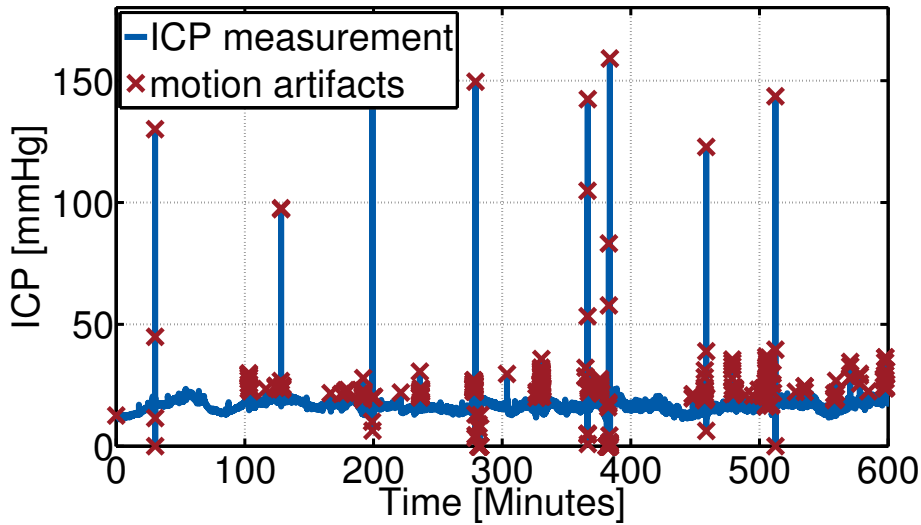


Figure 4.5: A ten hour excerpt of a typical ICP measurement (blue); artifacts highlighted by red crosses.

$$\mathbf{x} = \sum_{m=1}^M \mathbf{c}_m + \mathbf{r}_M.$$

Here, $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M)^\top$ must satisfy two conditions: (i) the number of extrema and the number of zero crossings must either equal or differ at most by one; and (ii) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero [146]. Figure 4.6 gives an example of a the decomposition of a one hour excerpt of an ICP signal into $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M, \mathbf{r}_M)^\top$. Only the first two IMFs and residual from the EMD are plotted.

Figure 4.7 shows an overview of the algorithm we proposed in [131]. The steps are briefly explained in the sequel, for full details, see [131]. The artifact detection and signal reconstruction consists of two artifact detectors, whose decisions ξ_1 and ξ_2 are 'or'-fused to yield the overall outlier label ξ_{out} .

In Artifact Detector 1, \mathbf{x} is filtered with an i -th order median filter with output $\tilde{\mathbf{x}}$. The residuals given by $\mathbf{e}_1 = \mathbf{x} - \tilde{\mathbf{x}}$ and a $3\text{-}\sigma_{e_1}$ rejection rule yields the outlier label ξ_1 .

Artifact Detector 2 works in the EMD, i.e., \mathbf{x} is decomposed into $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M, \mathbf{r}_M)^\top$ and then, the following steps are performed:

For every row in $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M, \mathbf{r}_M)^\top$:

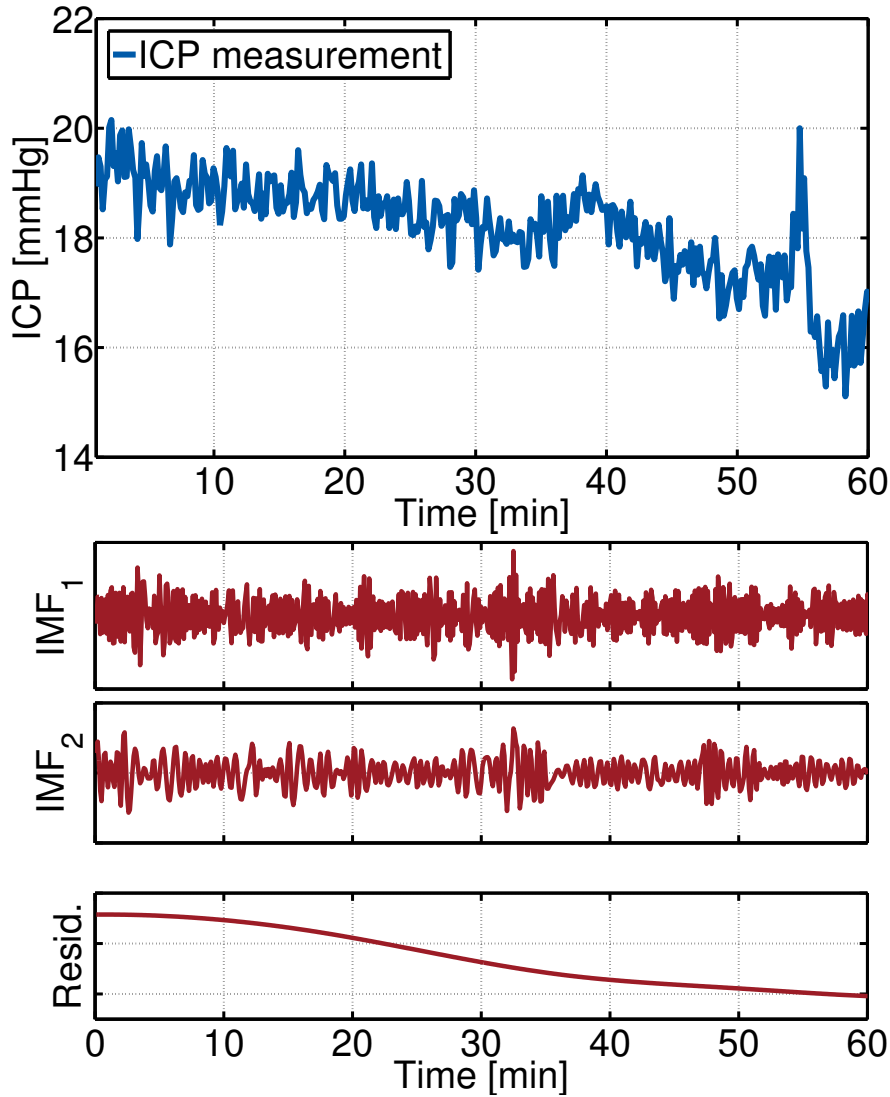


Figure 4.6: (top) one hour excerpt of an ICP signal. (bottom) First two intrinsic mode functions (IMF) and residual from empirical mode decomposition (EMD).

1. Robustly estimate its integrated autoregressive moving-average (ARIMA) model order (p, d, q) .
2. Estimate $\text{ARMA}(\hat{p}, \hat{q})$ parameters with median-of-ratios-estimator (MRE), see Section 4.5.3.
3. Apply an $\text{ARMA}(\hat{p}, \hat{q})$ filter-cleaner with output \mathbf{x}^f , see Section 4.5.4.
4. Obtain overall residuals $\boldsymbol{\epsilon}_i \rightarrow \mathbf{x} - \mathbf{x}^f$, $i = 1, \dots, k$.

The back-transform is simply the summation $\mathbf{e}_2 = \sum_{i=1}^k \boldsymbol{\epsilon}_i$, with corresponding overall residual scale $\hat{\sigma}_{e_2} = \sum_{i=1}^k \hat{\sigma}_i$. Finally, a $3\text{-}\sigma_{e_2}$ rejection rule yields the outlier label ξ_2 .

The signal reconstruction is performed via

$$\hat{s}_R(n) = \begin{cases} x(n) & \xi_{\text{out}}(n) = 0 \\ \tilde{x}(n) & \xi_{\text{out}}(n) = 1. \end{cases}$$

For the robust forecasting, the following steps are performed:

1. Decompose \hat{s}_R with the EMD.
2. For every component, robustly estimate its ARIMA model.
3. Compute a robust ARIMA forecast with the MRE, as described in [124].
4. Sum all forecast components up to obtain the overall forecast signal $\hat{s}_{F,\text{curr}}$.

In some cases, long patches of outliers occur. For this reason, forecasts based on some data blocks are much worse than others, even when using robust methods. Our suggested solution is to:

1. Save the current forecast in a buffer:
 - forecasts $\hat{s}_{F,\text{curr}}$ and $\hat{s}_{F,\text{curr-D}}$, which overlap on an interval of length D .
2. Assess forecasting quality: compute (non-robust) sample standard deviations of $\hat{s}_{F,\text{curr}}$ and $\hat{s}_{F,\text{curr-D}}$ on D and compare them to a robustly estimated scale of measurements $x(n)$ for which $\xi_{\text{out}}(n) = 0$.
3. Final forecast \hat{s}_F : prediction whose standard deviation is closer to a robust scale estimate.

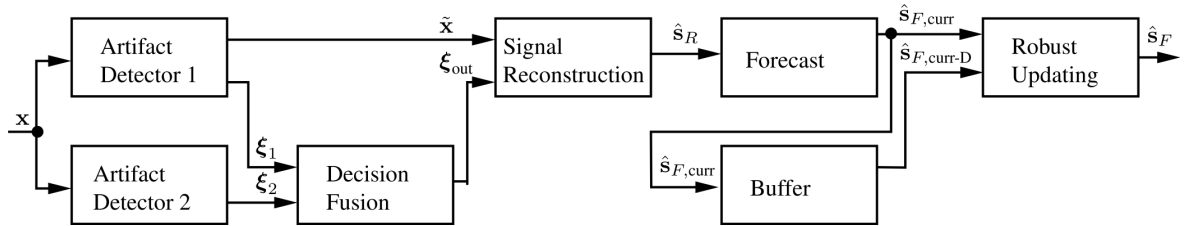


Figure 4.7: A schematic overview of the intracranial pressure signal forecasting algorithm that is based on outlier detection, robust ARIMA estimation and a data transformation into the empirical mode domain.

Figure 4.8 plots three real-data forecasts for (top) a measurement with a small amount of artifacts, (middle) a highly contaminated measurement and (bottom) a highly contaminated measurement with quickly varying ICP level. Our method is compared to a recently proposed algorithm by [145], which becomes unreliable in presence of outliers. For an extensive performance evaluation based on simulated data, the interested reader is referred to [131].

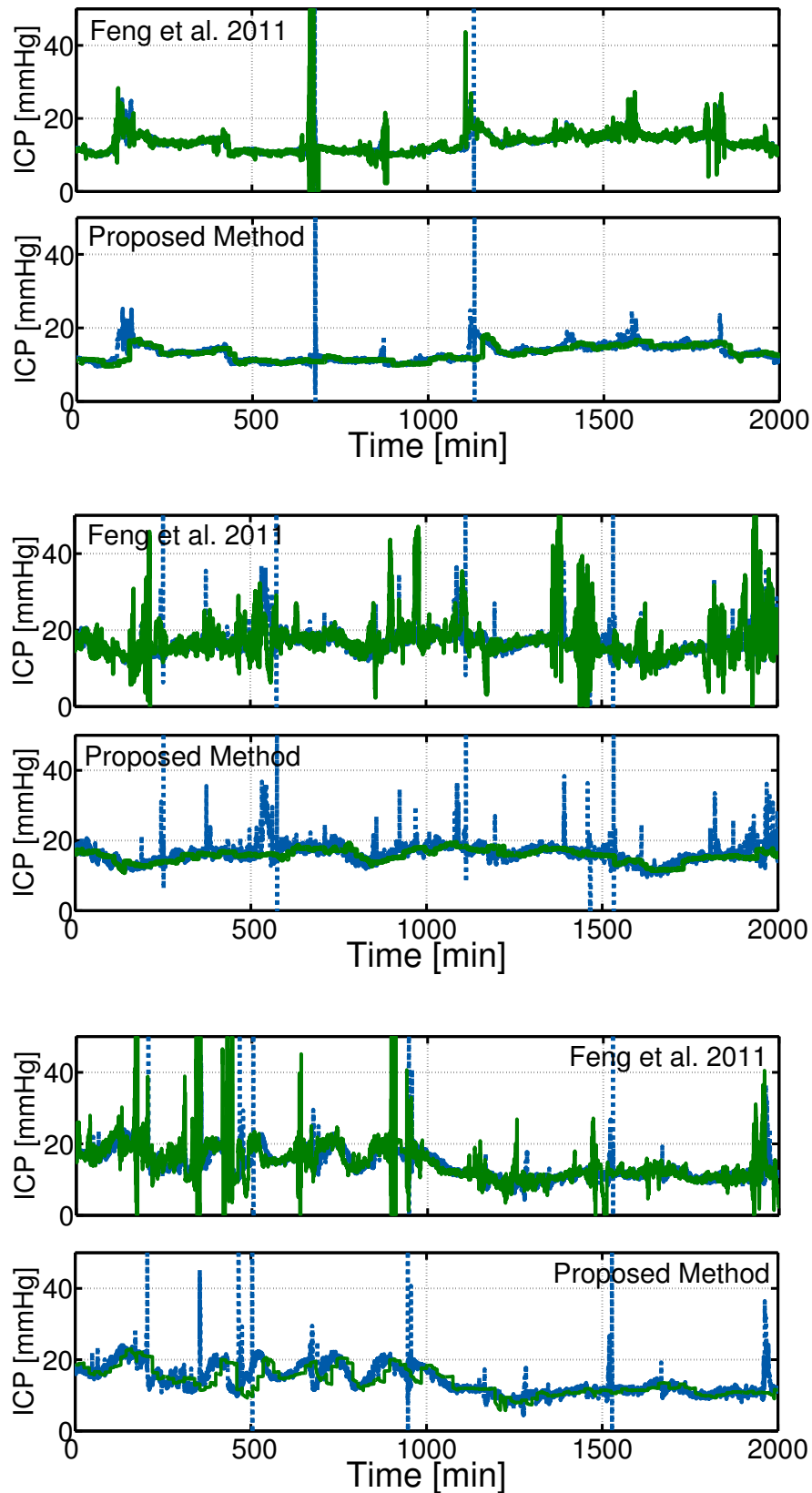


Figure 4.8: Three real-data examples of ICP measurements (blue) and on-line forecasts of the proposed and a neural networks based existing method [145] (green).

Acknowledgment

We would like to thank clinicians and nurses from National Neuroscience Institute, Singapore, for their efforts in data collection and clinical advice. The work of M. Feng was partially supported by the SERC grant 092-148-0067 of Agency for Science, Technology and Research, Singapore.

4.3.3 Limitations of the Outlier Detection Approach

While the outlier detection and data cleaning approaches are highly useful in practical situations, as shown in the examples above, performance can be usually only measured via extensive simulations. No formal statements on robustness or optimality can be made. Furthermore, we will show for the example of ARMA parameter estimation that methods based on outlier detection can be outperformed by sophisticated robust methods. Nevertheless, we believe that the research areas of outlier detection and robust statistics are highly related and can be used together in many practical situations.

4.4 Robustness Theory for Dependent Data

There are some significant difficulties that appear, when analyzing the robustness of estimators for dependent data.

4.4.1 Influence Function for Dependent Data

The IF describes the bias impact of an infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination. There exist two definitions of the influence function for the dependent data case [147], [148], which differ, but are mathematically related, see [8] for a comparative discussion. The influence functions of some robust estimators in the case of an AR(1) have been evaluated theoretically in [124, 147], however, there is no evaluation for an autoregressive process of order superior than $p = 1$ or for an ARMA(p, q) process with $p \neq 0$ and $q \neq 0$. This is explained by the complexity introduced by the correlation, where for an AR(p), for example, one must consider the joint distribution of $(Y_n, Y_{n-1}, \dots, Y_{n-p})$.

For dependent data, the influence functions also change depending on the outlier model, which makes the definition more general than in the i.i.d. case, since a contamination process does not have to be represented by a Dirac distribution. It can, for example, be a Gaussian process with a different variance and correlation [124, 147].

In this Section, we use the definition of the influence function by [147] which is given by the functional derivative at F_X

$$\text{IF}(F_{X,\xi,W}^\varepsilon; \hat{\boldsymbol{\theta}}) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left(\hat{\boldsymbol{\theta}}(F_Y^\varepsilon) - \hat{\boldsymbol{\theta}}(F_X) \right). \quad (4.22)$$

Here, $F_{X,\xi,W}^\varepsilon$ is the joint distribution of the processes X_n , ξ_n and W_n . According to this definition, the influence function is a functional on a distribution space. Influence functions for LS and some robust estimators for the AR(1) and MA(1) models are computed in [147]. We apply this definition to compute the influence functions of some existing and a newly proposed estimator in Section 4.6.1.2.

4.4.2 Maximum Bias Curve (MBC) for Dependent Data

The MBC gives an idea on the maximum asymptotic bias of an estimator w.r.t. a fraction of contamination ε . For dependent data, the MBC is defined as for the i.i.d. case, but also depends on the outlier model. Its definition is analogous to the i.i.d. data case, however, it is more complex and harder to compute for the same reasons as for the influence function. The MBC is generally obtained using Monte Carlo simulations.

As described in [8], the maximum bias curve can be computed by

$$\text{MBC}(\varepsilon) = \sup_c \left| \hat{\boldsymbol{\theta}}_N(\varepsilon, c) - \boldsymbol{\theta} \right|, \quad (4.23)$$

where $\hat{\boldsymbol{\theta}}_N(\varepsilon, c)$ is the worst-case estimate of $\boldsymbol{\theta}$ for a given sample size N and contamination probability ε . c is a deterministic value that is varied on a grid such that for each value of c , the distribution of W_n is given by $P(w_n = -c) = P(w_n = c) = 0.5$. We perform a maximum bias analysis of some existing and a newly proposed estimator in Section 4.6.1.3.

4.4.3 Breakdown Point for Dependent Data

The definition of the BP in the correlated data case has not been established in the literature, yet [149]. This is partly due to the fact that outliers may not drive the bias of the estimate to infinity, but to a border of the parameter space. Furthermore, the point that the bias converges to, depends on the type of outlier. E.g., when we consider an AR(1) process, additive independent outliers will drive the estimate of a to zero.

An intuitive definition of the breakdown point has been given by [149]. It is that the breakdown occurs for a certain percentage of outliers, beyond which the estimator gets stuck at some value (fixed maximum bias) even when adding more outliers. This definition is useful in many practical situations to estimate the breakdown point. BP analysis for some existing and a proposed estimator of ARMA models is given in Section 4.6.1.4.

4.5 Robust Estimation of ARMA Models

In this Section, we discuss the estimation of ARMA(p, q) models in the presence of outliers. This consists in estimating the parameter vector $\boldsymbol{\theta} = (a_1, \dots, a_p, b_1, \dots, b_q)^\top$ and the innovations scale σ_Z based on the contaminated observations y_1, \dots, y_N . In the sequel, we describe the propagation of outliers that makes direct application of robust estimators designed for the i.i.d. case impossible. Then, we briefly review some existing robust estimators and introduce and analyze a novel estimator which we call the BIP τ -estimator.

4.5.1 Outlier Propagation

ARMA parameter estimation is often based on fitting a model, such that some measure of the estimated innovations sequence \hat{Z}_n is minimized. While these approaches work fine for cases when the observed process Y_n equals the true process X_n , in presence of additive or replacement outliers, as given in Eqs. (4.17) and (4.2.3.2), a single outlier in the observations can spoil the estimates of multiple innovations. In the worst case, a single outlier can even contaminate an infinite number of subsequent innovations estimates, making the straight forward adaptation of robustness concepts, known from the independent data case impossible. E.g., the BP of M-, GM-, τ , S- or MM-estimators

quickly decreases for increasing autoregressive orders p and even equals zero for moving-average orders $q > 0$ [8].

Figure 4.9 illustrates the propagation of outliers for an ARMA(2,1) model with parameters $\mathbf{a} = (-0.39, -0.3)^T$, $b = 0.9$. The top plot shows that the estimates (red-dashed) of the innovations series obtained by inverse filtering a realization of the clean ARMA(2,1) are indistinguishable from the true innovations process (blue). The lower plot, on the other hand depicts the innovations series estimates derived from an ARMA(2,1) that is contaminated by additive outliers. The red crosses mark the additive outlier positions, where, for this example, the additive outliers are equally spaced with amplitude 10. The “smearing” of a single outlier onto multiple innovations estimates can be seen when comparing the true innovations process Z_n with the innovations estimates that are obtained by inversely filtering the observations Y_n with the true parameters of the ARMA(2,1) model. It can be clearly seen that robust estimators, as known from the independent data case, are only applicable, if they include a mechanism that prevents the “smearing” of a single outlier onto multiple innovations estimates. Sections 4.5.4 and 4.5.5.1 describe two possible mechanisms to suppress the propagation of outliers. Figure 4.11 displays the innovations estimates using the bounded influence propagation model that is described in Section 4.5.5.1.

In the sequel, we briefly revisit some existing robust estimators for ARMA models.

4.5.2 Cleaned Non-Robust Estimator

The cleaned non-robust estimator is a simple robust method that is frequently used among practicing engineers [7]. It uses classical estimation after a ‘3- σ ’ rejection. The ‘3- σ ’ rejection is a simple rule of thumb that rejects observations beyond three times the standard deviation estimated from the underlying signal. Its justification is explained by the fact that for $X_n \sim \mathcal{N}(\mu, \sigma^2)$, the probability of X_n taking a value above 3σ is unlikely, i.e., $P(|X_n - \mu| > 3\sigma) = 0.003$. Robust estimates of the mean μ and the standard deviation σ should be used to avoid the masking effect, which means that some outliers will be masked or will not be detected because of other outliers that inflate the estimate of σ and cause a bias to $\hat{\mu}$. After rejecting outliers, a classical estimator that handles missing observations, e.g., [150], is used.

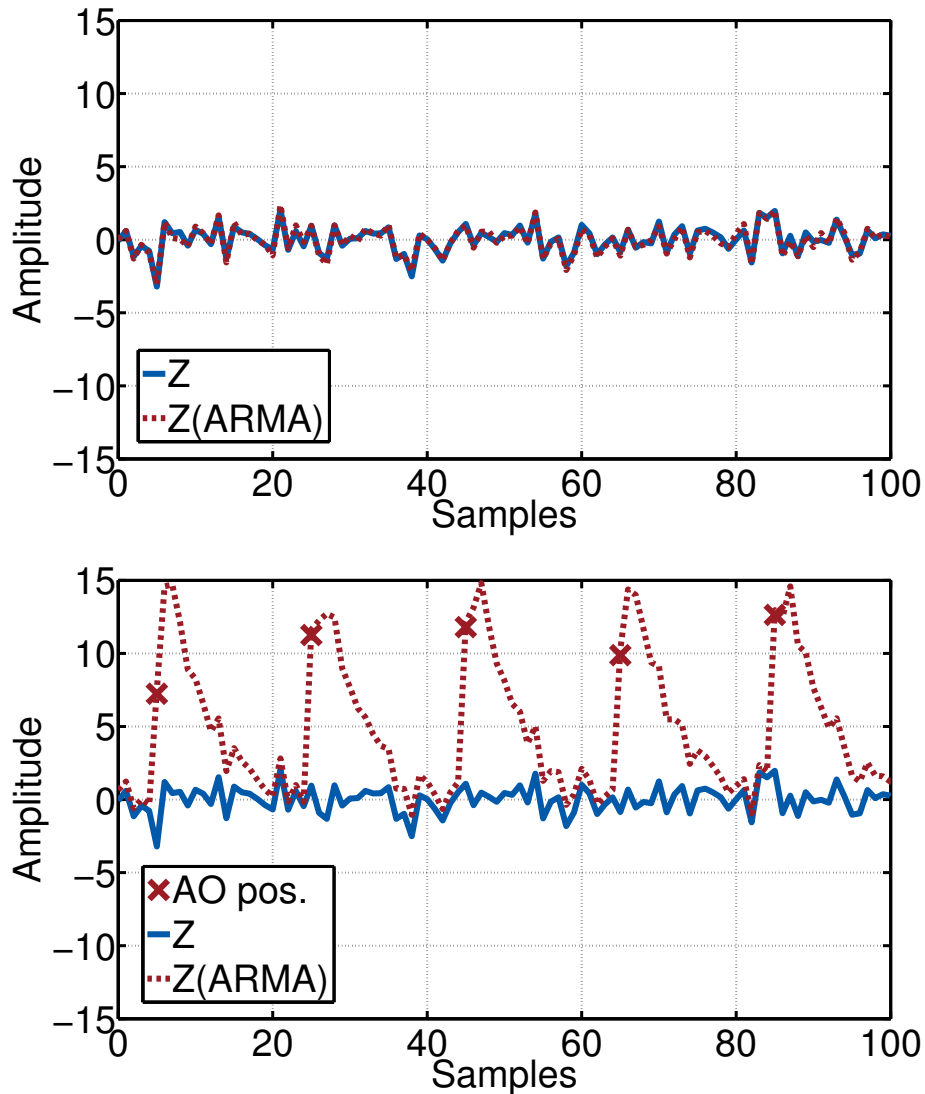


Figure 4.9: True innovations process Z_n (blue); innovations estimates using ARMA(2,1) model fit with true parameters $\mathbf{a} = (-0.39, -0.3)^T$, $b = 0.9$ (red-dashed); additive outlier positions (red). (top) innovations derived from a clean ARMA(2,1) (bottom) illustration of the “smearing effect” of additive outliers.

4.5.3 Median-of-Ratios-Estimator (MRE)

The median-of-ratios-estimator (MRE) [7, 124] uses robust autocorrelation estimates based on sample medians coupled with a robust filter-cleaner, which rejects outlying observations. This improves the efficiency and prevents outlier propagation. An ARMA(p, q) model is estimated by the MRE as follows:

1. Fit a high order AR(p^*) using the median of Y_n/Y_{n-k} to estimate the correlation. Here, the order $p^* > p$ is obtained by a robust order selection criterion.

2. Discard the outliers by filtering the signal using a robust filter-cleaner [116] with the estimated parameters of the high order AR(p^*) and apply a classical estimation method of ARMA models that handles missing data [150].

The method offers good performance in practice and is easy to implement. However, its BP is limited to 0.25 [7, 124].

4.5.4 Robustly Filtered Estimators

Robust filters [7, 116, 120, 151] work in the state-space domain and bound the propagation of outliers by filtering the innovation residuals.

4.5.4.1 State-Space Model for Autoregressive Processes

Given the definition of an AR(p) process from Eq. (4.2), the so-called state equation relates the unobservable p -dimensional state vector $\mathbf{X}_n = (X_n, X_{n-1}, \dots, X_{n-p+1})^\top$ to the previous state \mathbf{X}_{n-1} via the state transition matrix \mathbf{A} :

$$\mathbf{X}_n = \mathbf{A}\mathbf{X}_{n-1} + \mathbf{Z}_n. \quad (4.24)$$

Here, $\mathbf{Z}_n = (Z_n, 0, \dots, 0)^\top$, with $\mathbf{R} = \sigma_Z^2$ and

$$\mathbf{A} = \begin{bmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

The measurement equation represents the contaminated observations, as given in Eq. (4.17).

4.5.4.2 Approximate Conditional Mean Filter for Autoregressive Models

The approximate conditional mean (ACM) type filter, see [116, 120, 151] has as its output a filter-cleaned version of Y_n , henceforth denoted as Y_n^f , which is an estimate of X_n , given contaminated observations $\mathbf{Y}_n = (Y_n, Y_{n-1}, \dots, Y_{n-p+1})^\top$. The state recursion of the ACM filter is given by

$$\hat{\mathbf{X}}_{n|n} = \mathbf{A}\hat{\mathbf{X}}_{n-1|n-1} + \frac{\Sigma_n^1}{\sigma_n^2} \sigma_n \psi \left(\frac{Y_n - \hat{Y}_{n|n-1}}{\sigma_n} \right),$$

where Σ_n^1 is the first column of the prediction error covariance matrix Σ_n , which is computed recursively via

$$\Sigma_{n+1} = \mathbf{A}\mathbf{P}_n\mathbf{A}^T + \mathbf{Q}. \quad (4.25)$$

Here,

$$\mathbf{Q} = \begin{bmatrix} \hat{\sigma}_Z^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}$$

and the filtering error covariance matrix \mathbf{P}_n is given by

$$\mathbf{P}_n = \mathbf{M}_n - W \left(\frac{Y_n - \hat{Y}_{n|n-1}}{\sigma_n} \right) \frac{\Sigma_n^1 \Sigma_n^{1\top}}{\sigma_n^2}.$$

The weighting function $W(x) = \frac{\psi(x)}{x}$ and $\psi(\cdot)$ is usually chosen as Hampel's three-part re-descending [116, 120]. The one-step prediction error variance σ_n^2 is the first element of Σ_n^1 . Furthermore, $\hat{Y}_{n|n-1}$ denotes the robust one step ahead prediction of Y_n and is given by the first element of $\mathbf{A}\hat{\mathbf{X}}_{n-1|n-1}$. Finally, the robustly filter-cleaned process Y_n^f at time n is given by the first element of $\hat{\mathbf{X}}_{n|n}$.

An example of a contaminated AR(2) process Y_n , $n = 1, \dots, N$, along with the unobservable process X_n and its estimate Y_n^f obtained by the filter-cleaner is given in Figure 4.10.

4.5.4.3 Approximate Conditional Mean Filter for ARMA Models

For ARMA(p, q) models, it has been shown [8] that the state-space model can be extended such that for a state vector \mathbf{X}_n of dimension $r = \max(p, q + 1) \times 1$, the following state-space representation holds

$$\mathbf{X}_n = \mathbf{A}\mathbf{X}_{n-1} + \mathbf{B}\mathbf{Z}_n. \quad (4.26)$$

Here, $\mathbf{B} = (1, -b_1, -b_2, \dots, -b_{r-1})^\top$ with $b_k = 0$ for $k > q$ in case $p > q$. The state transition matrix becomes

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_{d-1} & \mathbf{I}_{d-1} \\ \mathbf{a}_d & \mathbf{0}_{d-1} \end{bmatrix}$$

with $\mathbf{a}_{d-1} = (a_1, \dots, a_{d-1})$ and $a_k = 0$ for $k > p$. The filter recursions are computed analogously to the AR case. However, implementing this filter for ARMA parameter estimation requires a good robust initial value and obtaining stable and invertible

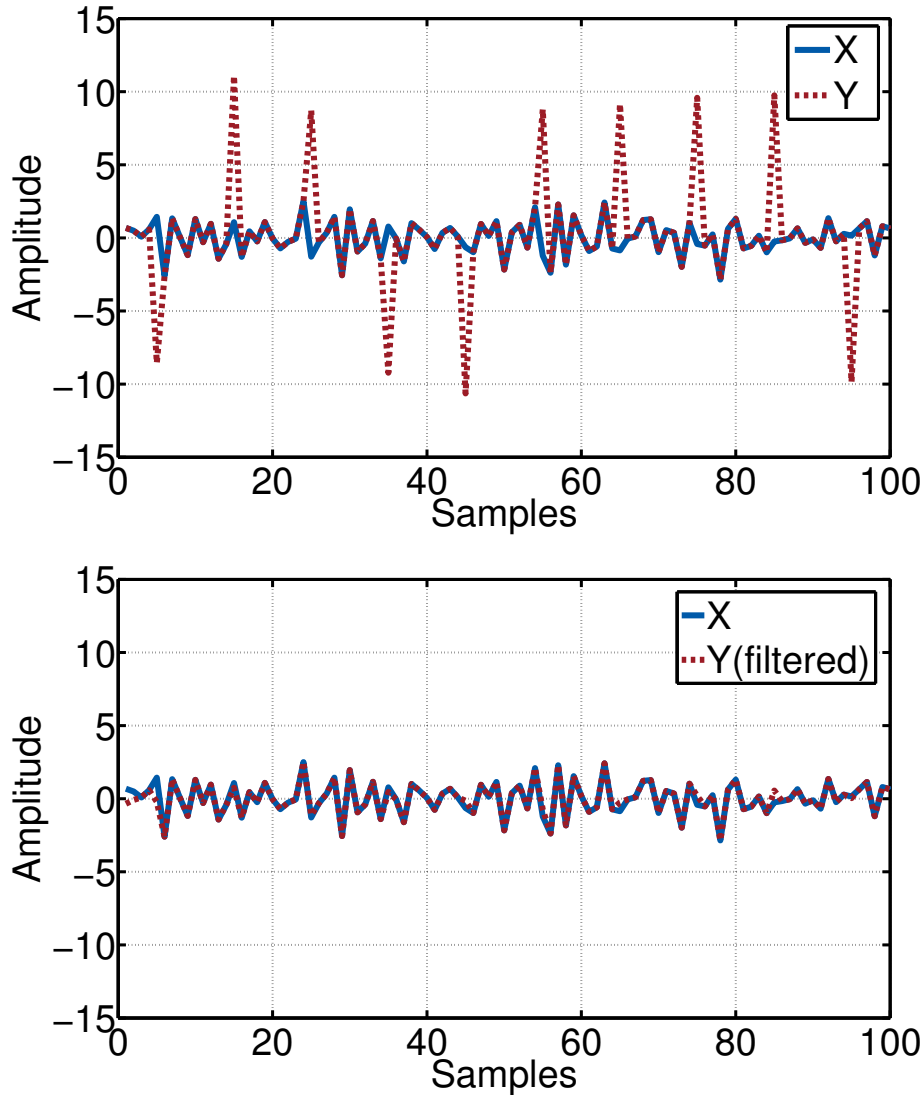


Figure 4.10: (top) example of a contaminated AR(2) process $\mathbf{Y} = \mathbf{X} + \mathbf{V}$ along with the unobservable process \mathbf{X} . (bottom) the estimate \mathbf{Y}^f obtained by the filter-cleaner that uses the true parameters $\mathbf{a} = (-0.5, 0.1)^\top$ as filter coefficients. In this example, \mathbf{V} is an isolated additive outlier process of constant amplitude $\xi = 10$, random sign and contamination probability $\varepsilon = 0.1$.

models is somewhat difficult. A suggestion has been given in [8], an alternative suggestion is to adapt reduced statistics estimators to robust estimation, analogously to the methodology described in Section 4.6.2.2.

The filtered innovation estimate $\mathbf{Z}^f(\boldsymbol{\theta})$ from the one-step robust forecast obtained from robust filters for ARMA models

$$Z_n^f(\boldsymbol{\theta}) = Y_n + \sum_{k=1}^p a_k Y_{n-k}^f(\boldsymbol{\theta}) - \sum_{k=1}^q b_k \sigma_n \psi \left(\frac{Z_{n-k}^f(\boldsymbol{\theta})}{\sigma_n} \right). \quad (4.27)$$

Here, Y_{n-k}^f is the filter-cleaned value for Y_{n-k} and is given by

$$Y_{n-k}^f = Y_{n-k} - Z_{n-k}^f(\boldsymbol{\theta}) + \sigma_{n-k} \psi \left(\frac{Z_{n-k}^f(\boldsymbol{\theta})}{\sigma_{n-k}} \right). \quad (4.28)$$

4.5.5 Bounded Influence Propagation (BIP) Estimators

An auxiliary model to robustly estimate ARMA models is the bounded innovation propagation autoregressive moving-average (BIP-ARMA) model that has been recently introduced by Muler *et al.* [123].

4.5.5.1 Bounded Influence Propagation (BIP) Models

$$X_n + \sum_{k=1}^p a_k X_{n-k} = Z_n + \sum_{k=1}^r \left((b_k - a_k) \sigma_Z \psi \left(\frac{Z_{n-k}}{\sigma_Z} \right) + a_k Z_{n-k} \right) \quad (4.29)$$

Here, $r = \max(p, q)$; if $r > p$, $a_{p+1} = \dots = a_r = 0$, while if $r > q$, $b_{q+1} = \dots = b_r = 0$; σ_Z is a robust M-scale of Z_n which coincides with the standard deviation when the distribution of Z_n is Gaussian. $\psi(x)$ is an odd, bounded and continuous function. The models described by Eq. (4.1) are included in Eq. (4.29) by setting $\psi(x) = x$. For statistical properties of the BIP-ARMA model, cf. [123].

By applying the definition in Eq. (4.5), the MA(∞) representation of Eq. (4.29) becomes

$$X_n = Z_n + \sum_{k=1}^{\infty} b_k^{\infty} \sigma_Z \psi \left(\frac{Z_{n-k}}{\sigma_Z} \right).$$

Figure 4.11 displays the innovations estimates for the ARMA(2,1) model with parameters $\mathbf{a} = (-0.39, -0.3)^T$, $b = 0.9$ discussed in Section 4.5.1 fit with a BIP-ARMA(2,1) model with the same parameters. The smearing effect of the outliers is significantly suppressed compared to using the ARMA(2,1) model, cf. Figure 4.9.

4.6 BIP τ -estimators

In this Section, we propose a novel estimator, the BIP τ -estimator, which is based on the BIP model. We analyze the robustness and asymptotic properties and provide a

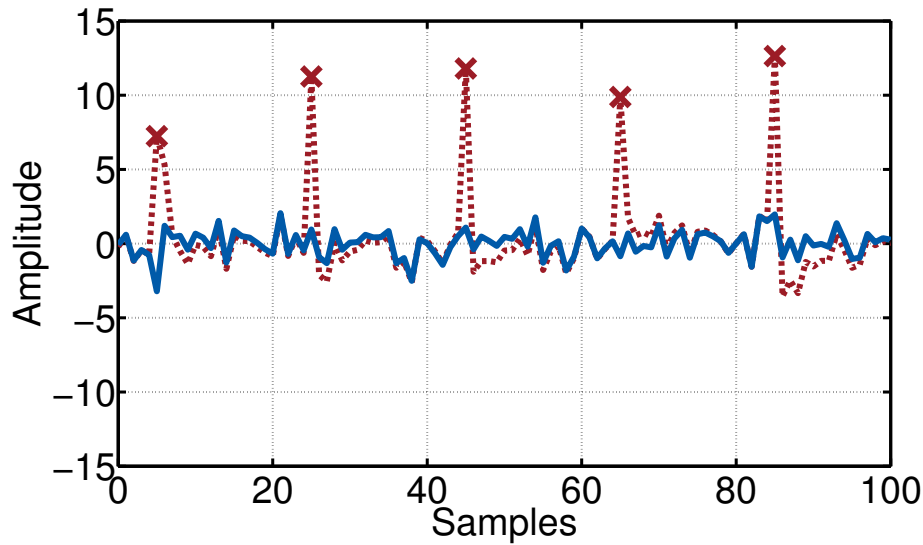


Figure 4.11: True innovations process Z_n (blue); additive outlier positions (red); innovations estimates using BIP-ARMA(2,1) model fit with true parameters $\mathbf{a} = (-0.39, -0.3)^T$, $b = 0.9$ (red-dashed).

fast algorithm to compute stable and invertible ARMA models. We also compare the proposed estimator to some existing estimators and provide real-data examples that illustrate its applicability to real-world problems. The proposed BIP τ -estimator is similar in its statistical properties to the BIP MM-estimator by [123]. An advantage over the MM-estimator is that the τ -estimator, inherently with the estimate of θ , provides us with a highly efficient and robust τ -scale of the innovations sequence, which is useful, e.g., in model order selection, see Section 4.7.

4.6.1 BIP τ -estimators: Asymptotic and Robustness Properties

4.6.1.1 Consistency

It has been shown by [123] that under the following assumptions an S-estimate is strongly consistent:

A1^S Y_n , $n \in \mathbb{Z}$, is a stationary and invertible ARMA process, where polynomials $A(z)$ and $B(z)$ do not have common roots.

A2^S The innovations Z_n , $n \in \mathbb{Z}$, have an absolutely continuous distribution with a symmetric and strictly uni-modal density and $E_F[\max\{\log(|Z_n|, 0)\}] < \infty$.

A3^S $\rho(x)$ is continuous, even and non-constant and non-decreasing in $|x|$ with $\sup(\rho) > b$ and $\psi(x)$ is bounded and continuous.

In the sequel, we prove the consistency of the τ -estimate under the Gaussian assumption. The τ -estimate has been shown to be a linear combination of two S-estimates where the weight $W_n(\boldsymbol{\theta})$, Eq. (2.14), depends on the data [9]. When the innovations Z_n follow a Gaussian distribution, asymptotically, $W_n(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{b})$ equals 0. Therefore, since the scale estimate $\hat{\sigma}_{M_1}$ in Eq. (2.11) is consistent for the ARMA model, see [123], the τ -estimate becomes an S-estimate which satisfies **A3^S**, as long as $\rho_2(x)$ is chosen accordingly. This proves that the τ -estimate is strongly consistent under the following assumptions:

A1 $Y_n, n \in \mathbb{Z}$, is a stationary and invertible ARMA process, where polynomials $A(z)$ and $B(z)$ do not have common roots.

A2 The innovations $Z_n, n \in \mathbb{Z}$ follow a Gaussian distribution with $E_F[\max\{\log(|Z_n|, 0)\}] < \infty$.

A3 $\rho_2(x)$ is continuous, even and non-constant and non-decreasing in $|x|$ with $\sup(\rho_2) > b$ and $\psi_2(x)$ is bounded and continuous.

The assumption of $\rho_2(x)$ being non-constant is only required to ensure that for all σ

$$E_F \left[\rho \left(\frac{Z_n - \mu}{\sigma} \right) \right] \quad (4.30)$$

has a unique minimum at $\mu = 0$. In practice, this condition on $\rho_2(x)$ may be relaxed, as long it is ensured that the minimum at $\mu = 0$ is found. Otherwise, the estimates will converge with probability 1 to $\mu_1 \neq 0$.

Asymptotic equivalence of the S-estimates under an ARMA model and the S-estimates of a BIP-ARMA has been established by [123]. Following the same steps, the asymptotic equivalence of the τ -estimates under an ARMA model and the τ -estimates of a BIP-ARMA is given as long as **A1**, **A2**, **A3** and

A4 $P(Z_n \in C)$ for any compact C ,

A5 $\psi(x)$ in Eq. (4.29) is odd, bounded and continuous,

hold.

4.6.1.2 Influence Function

As described in Section 4.4, the influence function in the dependent data case depends on the type of outlier model. Furthermore, due to the dependence of the variables involved in the joint probability distribution function, computing the influence functions for orders $p > 1$ or $q > 1$ quickly becomes infeasible. However, to obtain an impression of the robustness of the BIP τ -estimator, we compute the influence function for the AR(1) model in the case of additive isolated outliers, see Section 4.2.3.

With the definition of the influence function by [147], and the equivalence given in Eq. (2.15), following the same steps as taken for M-estimators in [147], under the assumption that

A6 $\psi_\tau(x)$ is bounded, odd and continuous and $\frac{d}{dx}\psi_\tau(x)$ is bounded and continuous,

the influence function of the τ -estimator for AR(1) model is given by

$$\text{IF}(\xi; \hat{a}_\tau, a, F_{X_1, X_0, V_0}) = \frac{\sqrt{1-a^2}}{\mathbb{E}_{F_U}[U\psi_\tau(U)]} \mathbb{E}_{F_{X_1, X_0, V_0}} \left[(X_1 + aX_0 - aV_0)\psi_\tau((X_0 + V_0)\sqrt{1-a^2}) \right],$$

where U is a standard Gaussian random variable. The computation of the influence function requires evaluation of the following integrals:

$$\mathcal{E}_0 = \int_{-\infty}^{\infty} u\psi_\tau(u) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du,$$

$$\mathcal{E}_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - ax_0)\psi_\tau((x_0 + v_0)\sqrt{1-a^2}) f_{X_1, X_0}(x_1, x_0; a) f_{V_0}(v_0) dx_1 dx_0 dv_0,$$

$$\mathcal{E}_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_0\psi_\tau((x_0 + v_0)\sqrt{1-a^2}) f_{X_1, X_0}(x_1, x_0; a) f_{V_0}(v_0) dx_1 dx_0 dv_0,$$

with

$$f_{X_1, X_0}(x_1, x_0; a) = f_{X_1|X_0}(x_1|x_0; a) f_{X_0}(x_0; a),$$

where

$$f_{X_1|X_0}(x_1|x_0; a) = \frac{1}{\sqrt{2\pi}\sigma_Z} e^{-\frac{1}{2} \frac{(x_1 - ax_0)^2}{\sigma_Z^2}},$$

$$f_{X_0}(x_0; a) = \frac{\sqrt{1-a^2}}{\sqrt{2\pi}\sigma_Z} e^{-\frac{1}{2} \frac{x_0^2(1-a^2)}{\sigma_Z^2}},$$

and to compute $\text{IF}(\xi; \hat{a}_\tau, a, F_{X_1, X_0, V_0})$ as a function of the outlier magnitude ξ , [147] suggests

$$f_{V_0}(v_0) = \delta(\xi),$$

which yields additive outliers of constant value ξ . The influence function of the τ -estimator for the AR(1) with isolated additive outliers has the following form:

$$\text{IF}(\xi; \hat{a}_\tau, a, F_{X_1, X_0, V_0}) = \frac{\sqrt{1-a^2}}{\mathcal{E}_0} (\mathcal{E}_1 - a\mathcal{E}_2).$$

From Eqs. (4.6.1.2), (4.6.1.2) and (4.6.1.2), it can be seen that as $\xi \rightarrow \infty$, $\text{IF}(\xi; \hat{a}_\tau, a, F_{X_1, X_0, V_0})$ remains bounded, as long as $\psi_\tau(\xi \rightarrow \infty) = 0$, which holds, e.g., when $\psi_1(x)$ and $\psi_2(x)$ follow Eq. (2.8). Figure 4.12 plots $\text{IF}(\xi; \hat{a}_\tau, a, F_{X_1, X_0, V_0})$ and $\text{IF}(\xi; \hat{a}_{ML}, a, F_{X_1, X_0, V_0})$ for $a = -0.5$ in dependence of the outlier magnitude ξ . Clearly, $\text{IF}(\xi; \hat{a}_\tau, a, F_{X_1, X_0, V_0})$ is bounded, continuous and re-descending to zero as $\xi \rightarrow \infty$, while $\text{IF}(\xi; \hat{a}_{ML}, a, F_{X_1, X_0, V_0})$ is unbounded.

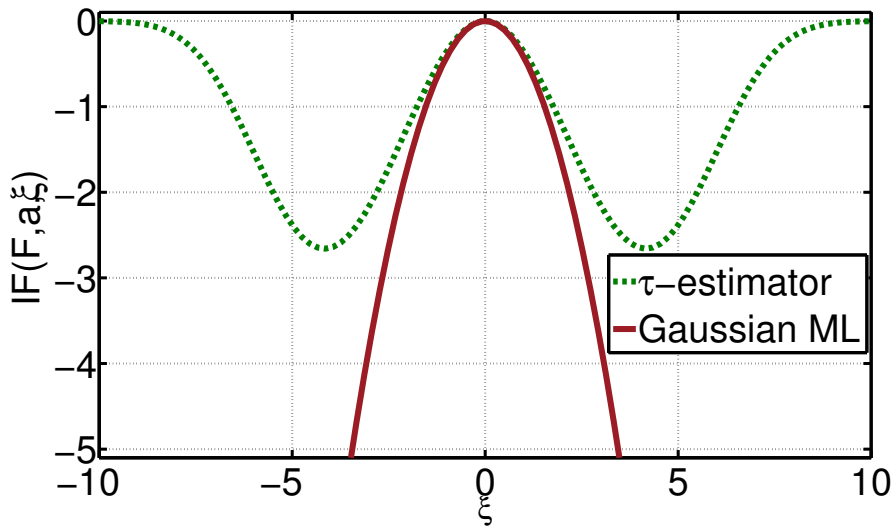


Figure 4.12: Example of the influence functions of the τ -estimator and the Gaussian ML-estimator for the AR(1) model with $a = -0.5$ and independent additive outliers of magnitude ξ . The parameter a was chosen to be comparable to [147].

It can be proved [123] that the influence functions of robust estimates under the ARMA model and under the BIP-ARMA model asymptotically coincide. Therefore, the influence function computed in Eq. (4.6.1.2) is the influence function of the BIP-AR(1) τ -estimate. However, because the influence function only measures robustness against infinitesimal contamination, one should consider further measures of robustness, which is done in the sequel. These measures show that the robust BIP estimates outperform the ARMA estimates for a positive fraction of outlier-contamination.

4.6.1.3 Maximum Bias Curve (MBC)

In practice, the MBC is computed by a Monte Carlo procedure [8, 123]. The MBC of the BIP τ -estimator for the AR(1) model and independent additive outliers is plotted in Figure 4.15. Figures 4.13 and 4.14 plot the MBC of the Gaussian ML-estimator and the Median-of-Ratios-estimator for comparison. To be comparable to [8], p. 306, we chose $a = 0.9$ and set $N = 2000$. c was varied from 0 to 600 and ε from 0 to 0.5.

The MBC was estimated by

$$\text{MBC}(\varepsilon) = \sup_c \{|\hat{a}(N, \varepsilon, c) - a|\}. \quad (4.31)$$

We also define the Median BC

$$\text{Median BC}(\varepsilon) = \text{median} \{|\hat{a}(N, \varepsilon, c) - a|\} \quad (4.32)$$

and the Quantile BC

$$Q_x\text{BC}(\varepsilon) = Q_x \{|\hat{a}(N, \varepsilon, c) - a|\}. \quad (4.33)$$

Here, Q_x means that x percent of the sorted data is to the left of Q_x . Therefore, e.g., $Q_{75}\text{BC}(\varepsilon)$ represents the MBC obtained in 75 % of the cases for varying c and fixed ε . $Q_{50}\text{BC}(\varepsilon)$ corresponds to Eq. (4.32) and $Q_{100}\text{BC}(\varepsilon)$ is the $\text{MBC}(\varepsilon)$ given in Eq. (4.31).

4.6.1.4 Breakdown Point

As discussed in Section 4.4.3, the concept of the BP is not fully established in dependent data. For the AR(1) model and additive outliers, one can estimate the BP by evaluating the bias curves. This yields the following results:

- The BP is the value of $\varepsilon < \varepsilon_0$ with $\text{MBC}(\varepsilon_0) - a = 0$.
- The definition by [149] is also fulfilled, i.e., $\text{MBC}(\varepsilon \geq \varepsilon_0)$ does not change.

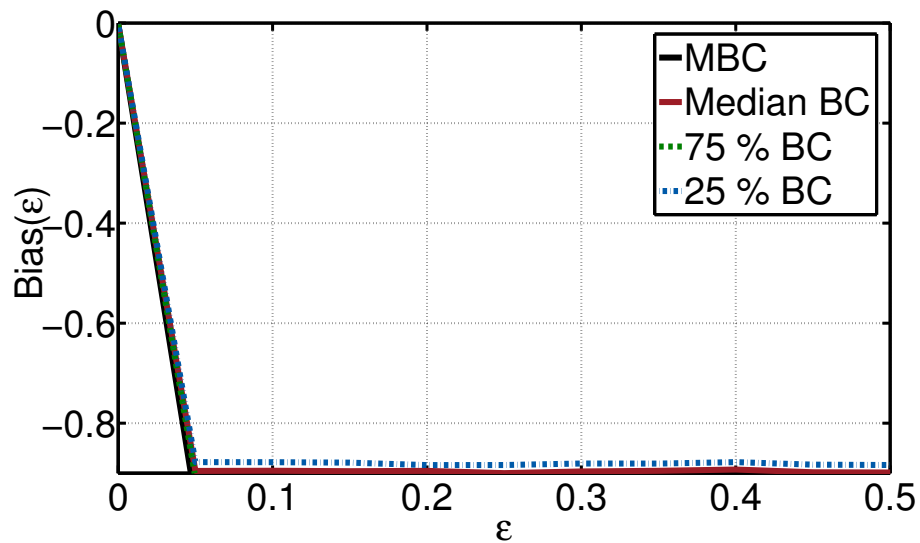


Figure 4.13: Example of the bias curves of the Gaussian ML-estimator for the AR(1) model with $a = 0.9$ w.r.t. the fraction of contamination ε . The MBC; the Median BC, $Q_{75}BC(\varepsilon)$ and $Q_{25}BC(\varepsilon)$, i.e., the BC maintained by 75 % and 25 % of the data, respectively. The parameter a was chosen to be comparable to [8], p. 306.

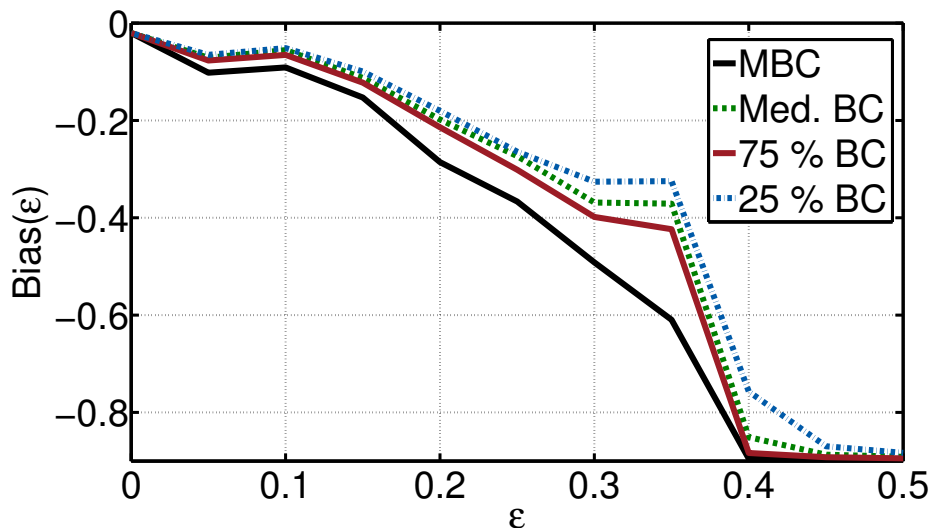


Figure 4.14: Example of the bias curves of the MRE for the AR(1) model with $a = 0.9$ w.r.t. the fraction of contamination ε . The MBC; the Median BC, $Q_{75}BC(\varepsilon)$ and $Q_{25}BC(\varepsilon)$, i.e., the BC maintained by 75 % and 25 % of the data, respectively. The parameter a was chosen to be comparable to [8], p. 306.

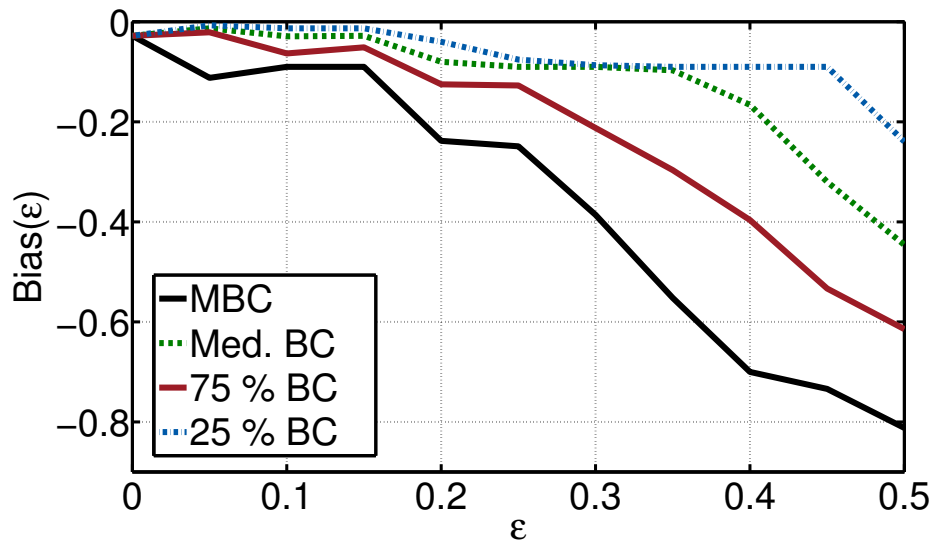


Figure 4.15: Example of the bias curves of the BIP τ -estimator for the AR(1) model with $a = 0.9$ w.r.t. the fraction of contamination ϵ . The MBC; the Median BC, $Q_{75}BC(\epsilon)$ and $Q_{25}BC(\epsilon)$, i.e., the BC maintained by 75 % and 25 % of the data, respectively. The parameter a was chosen to be comparable to [8], p. 306.

- A further interesting observation is that for $\epsilon \geq \epsilon_0$, $Q_xBC(\epsilon)$ coincides for all values of x .

Figure 4.13 clearly reveals that the BP of the Gaussian ML-estimator equals zero, since all three conditions are fulfilled. For the MRE, evaluating Figure 4.14 shows, that in this case, the BP ≈ 0.35 . For the proposed BIP τ -estimator, evaluating the bias curves in Figure 4.14 yields BP ≈ 0.5 , which is the maximally possible BP.

4.6.2 BIP τ -estimators: Fast Algorithm to Obtain Stationary and Invertible ARMA Estimates

In this Section, we describe an algorithm to robustly compute ARMA model parameter estimates which yield stationary and invertible ARMA models. In the first step, the parameters of a long AR model are computed. Robust AR model order selection for this step is discussed in Section 4.6.2.1. In the second step, from the long AR model, we propose an ARMA parameter estimator based on reduced statistics. This estimator compares favorably to computing an ARMA estimate with missing data after outlier detection based on the long AR model. We also illustrate with some simulated data examples that the proposed estimator outperforms a set of existing robust estimators.

4.6.2.1 Estimating the AR Model Parameters

In the sequel, we show how to compute BIP τ -estimates for the AR(p) model. Our algorithm applies a robust Durbin-Levinson type procedure, similar to the one described in [8], where the non-robustness sources of the standard Durbin-Levinson method are eliminated. We begin by computing the AR(1) as follows:

1. For $-1 < \zeta < 1$, compute:

- (i) The AR(1) innovations estimate $\mathbf{Z}^{(1)}(\zeta) = (Z_2^{(1)}(\zeta), \dots, Z_N^{(1)}(\zeta))^\top$, see Eq. (4.2) with $p = 1$, which is recursively defined for $n = 2, \dots, N$ by

$$Z_n^{(1)}(\zeta) = Y_n + \zeta Y_{n-1}.$$

- (ii) The BIP-AR(1) innovations estimate $\mathbf{Z}^{\text{BIP}(1)}(\zeta) = (Z_2^{\text{BIP}(1)}(\zeta), \dots, Z_N^{\text{BIP}(1)}(\zeta))^\top$, see Eq. (4.29) with $p = 1, q = 0$, which is recursively defined for $n = 2, \dots, N$ by

$$Z_n^{\text{BIP}(1)}(\zeta) = Y_n + \zeta Y_{n-1} - \zeta Z_{n-1}^{\text{BIP}(1)}(\zeta) + \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(1)}(\zeta)) \psi \left(\frac{Z_{n-1}^{\text{BIP}(1)}(\zeta)}{\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(1)}(\zeta))} \right),$$

where

$$\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(1)}(\zeta)) = \frac{\hat{\sigma}_\tau(\mathbf{Y})}{1 + \gamma^2 \sum_{k=1}^{\infty} b_k^\infty(\zeta)}. \quad (4.34)$$

Here, $\hat{\sigma}_\tau(\mathbf{Y})$ is the τ -estimate of the scale (see Eq. (2.11)) of $\mathbf{Y} = (Y_1, \dots, Y_N)^\top$, $\mathbf{b}^\infty(\zeta) = (b_1^\infty(\zeta), \dots, b_\infty^\infty(\zeta))$ is the MA(∞) representation of the AR(1) with $a_1 = \zeta$ and $\gamma^2 = E[\psi(X)]^2$, when the distribution of X is standard Gaussian. In practice, one approximates $\mathbf{b}^\infty(\zeta) \approx (b_1^\infty(\zeta), \dots, b_{q_0}^\infty(\zeta))$, where q_0 is a sufficiently large number.

2. Compute $\hat{\sigma}_\tau(\mathbf{Z}^{(1)}(\zeta))$, see Eq. (2.11), and $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(1)}(\zeta))$, see Eq. (4.34). In our algorithm, for the estimation of $\hat{\sigma}_{M_1}$, we use a biweight M-estimate of scale, which is computed by the iterative algorithm described in [8] pages 40-41.
3. Find $-1 < \zeta < 1$ which minimizes $\hat{\sigma}_\tau(\mathbf{Z}^{(1)}(\zeta))$ and $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(1)}(\zeta))$:

$$\hat{a}_1 = \underset{\zeta}{\operatorname{argmin}} \left\{ \hat{\sigma}_\tau(\mathbf{Z}^{(1)}(\zeta)), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(1)}(\zeta)) \right\}. \quad (4.35)$$

Figure 4.16 illustrates the estimation of $a_1 = -0.5$ for $\sigma_Z = 1$ using simulated data of length $N = 1000$. The top plot depicts the clean data case, i.e., $Y_n = X_n$. The bottom plot shows an example, where $Y_n = X_n + V_n$ and V_n contains 10 % equally spaced additive outliers of amplitude 10. In this case, the τ -estimator

based on the AR(1) model yields a biased estimate of a_1 due to the propagation of outliers, see Section 4.5.1. The τ -estimator based on the BIP-AR(1) model, on the other hand, maintains a reliable estimate of a_1 . The overall estimate of the proposed estimator is, in the case of additive outliers, given by the estimate based on the BIP-AR(1) model. Figure 4.17 shows a detailed zoom into the minimization of $\hat{\sigma}_\tau(\mathbf{Z}^{(1)}(\zeta))$ and $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}^{(1)}}(\zeta))$ for the clean data case. It can be noticed that the $\hat{\sigma}_\tau(\mathbf{Z}^{(1)}(\zeta))$ curve lies below the $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}^{(1)}}(\zeta))$ curve, meaning that the true AR(1) model is used. It has been shown by [123] for the BIP MM-estimator, that in general, for the clean data case, the minimum is provided by the ARMA model based estimator. The same holds for the τ -estimator.

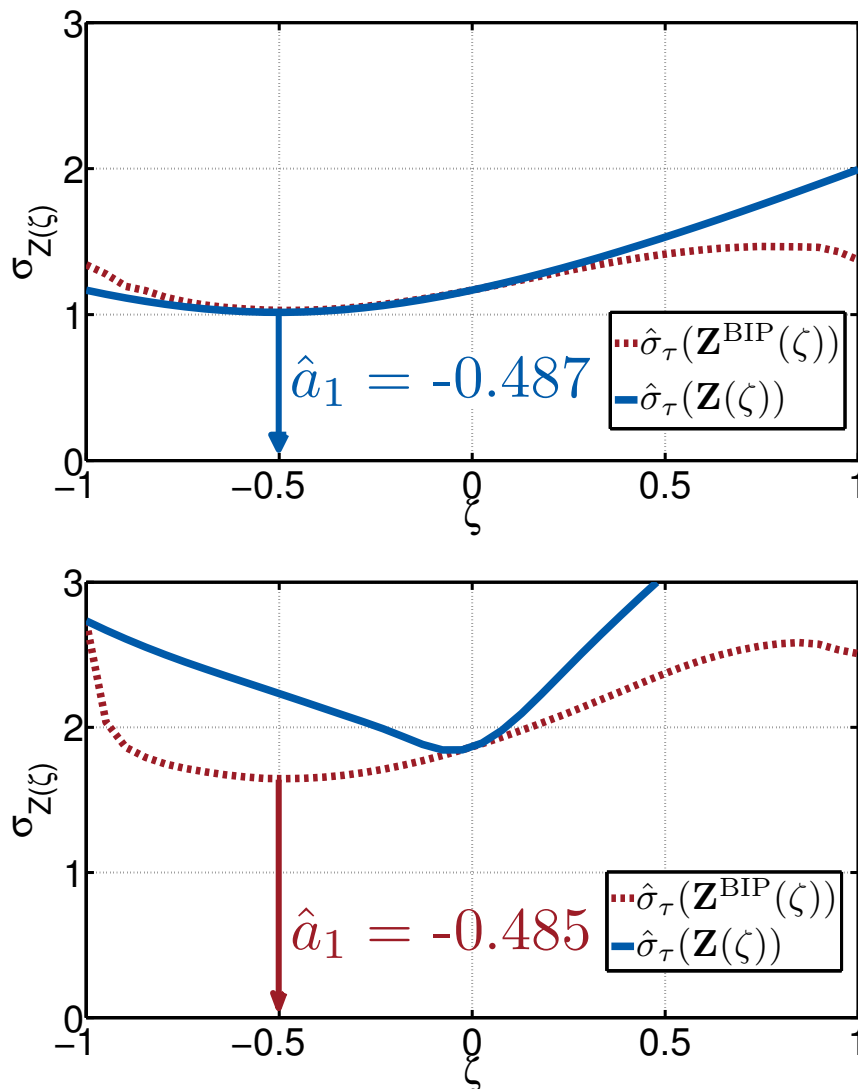


Figure 4.16: Example of finding $-1 < \zeta < 1$ which minimizes $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\zeta))$ and $\hat{\sigma}_\tau(\mathbf{Z}(\zeta))$ for an AR(1) process with $a_1 = -0.5$ and $\sigma_Z = 1$. (top) $Y_n = X_n$ clean data example; (bottom) $Y_n = X_n + V_n$, where V_n contains 10 % equally spaced additive outliers of amplitude 10.

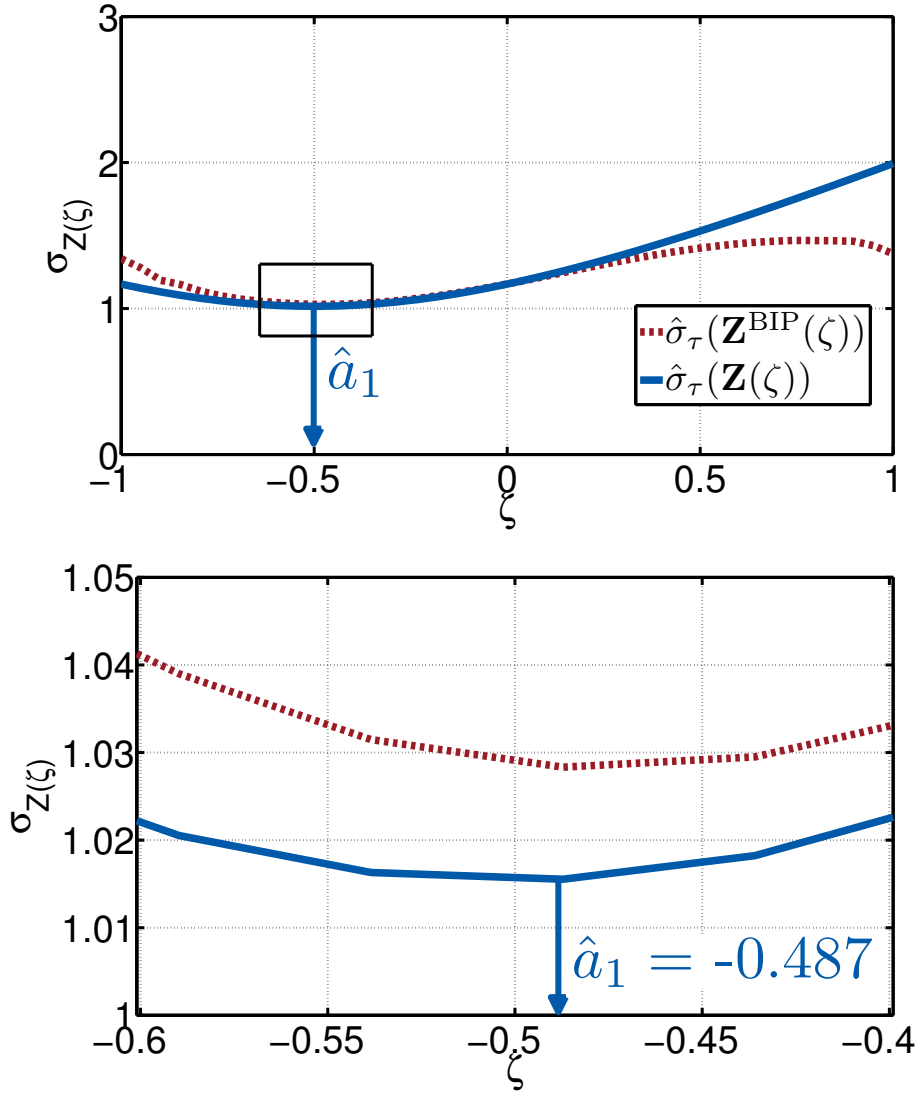


Figure 4.17: Example of finding $-1 < \zeta < 1$ which minimizes $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\zeta))$ and $\hat{\sigma}_\tau(\mathbf{Z}(\zeta))$ for an AR(1) process with $a_1 = -0.5$ and $\sigma_Z = 1$. (top) $Y_n = X_n$ clean data example; (bottom) A zoom into the region of interest shows that for the clean data the AR(1) model is used.

Since, in practice, finding $-1 < \zeta < 1$ which minimizes Eq. (4.35) requires a grid search, in the sequel, we present a suggestion to significantly speed up the algorithm. We suggest to evaluate Eq. (4.35) on a coarse grid, e.g., using a step size of $\Delta\zeta_0 = 0.05$ requires computing 41 τ -estimates of scale. In the next step, we model the true curves $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\zeta))$ and $\hat{\sigma}_\tau(\mathbf{Z}(\zeta))$ by a polynomial model estimated from $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\zeta_0))$ and $\hat{\sigma}_\tau(\mathbf{Z}(\zeta_0))$, with $\zeta_0 = (-0.99, 0.99 + \Delta\zeta_0, \dots, 0.99 - \Delta\zeta_0, 0.99)^\top$. In practice, least-squares estimation of the polynomials of degree five produced results which practically did not differ to an evaluation performed on a finer grid. In this way, we are able to reduce the computational costs roughly by the factor $2 \times \text{length}(\zeta_{\text{fine}}) / \text{length}(\zeta_0)$. E.g., for $\Delta\zeta_0 = 0.05$ and $\Delta\zeta_{\text{fine}} = 10^{-3}$, the computational costs are approximately reduced

by the factor 50.

To obtain a BIP τ -estimate for an AR(2), given $\hat{a}_{1,1} = \hat{a}_1$ from the previous step, again requires minimizing the τ -estimate of scale of two innovations estimates. These are:

- (i) The AR(2) innovations estimate $\mathbf{Z}^{(2)}(\zeta) = (Z_3^{(2)}(\zeta), \dots, Z_N^{(2)}(\zeta))^\top$, see Eq. (4.2) with $p = 2$, is recursively defined for $n = 3, \dots, N$ by

$$Z_n^{(2)}(\zeta) = Y_n + \hat{a}_{2,1}(\zeta)Y_{n-1} + \hat{a}_{2,2}(\zeta)Y_{n-2}.$$

- (ii) The memory two BIP innovations τ -estimate $\mathbf{Z}^{\text{BIP}(2)}(\zeta) = (Z_2^{\text{BIP}(1)}(\zeta), \dots, Z_N^{\text{BIP}(1)}(\zeta))^\top$, which is recursively defined for $n = 3, \dots, N$ by

$$Z_n^{\text{BIP}(2)}(\zeta) = Y_n + \hat{a}_{2,1}(\zeta)\hat{X}_{n-1}^{\text{BIP}(2)}(\zeta) + \hat{a}_{2,2}(\zeta)\hat{X}_{n-2}^{\text{BIP}(2)}(\zeta), \quad n = 3, \dots, N,$$

where $\hat{a}_{2,1}(\zeta) = (1 - \zeta)\hat{a}_{1,1}$ and $\hat{a}_{2,2}(\zeta) = \zeta$ come from the Durbin-Levinson recursion, see Eq. (4.38). Here, $\hat{X}_{n-1}^{\text{BIP}(2)}(\zeta)$, is the BIP τ -prediction of $X_{n-1}(\zeta)$:

$$\hat{X}_{n-1}^{\text{BIP}(2)}(\zeta) = Y_n - Z_{n-1}^{\text{BIP}(2)}(\zeta) + \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(2)}(\zeta))\psi_\tau \left(\frac{Z_{n-1}^{\text{BIP}(2)}(\zeta)}{\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(2)}(\zeta))} \right), \quad n = 3, \dots, N,$$

where $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(2)}(\zeta))$ is defined analogously to Eq. (4.34), but letting $\mathbf{b}^\infty(\zeta) = (b_1^\infty(\zeta), \dots, b_\infty^\infty(\zeta))^\top$ stand for the MA(∞) representation of the AR(2) process with parameter vector $\mathbf{a} = (a_1 = \hat{a}_{2,1}(\zeta), a_2 = \hat{a}_{2,2}(\zeta))^\top$.

The BIP τ -estimates for an AR(2) are then obtained by

$$\hat{a}_2 = \underset{\zeta}{\operatorname{argmin}} \{ \hat{\sigma}_\tau(\mathbf{Z}^{(2)}(\zeta)), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(2)}(\zeta)) \}. \quad (4.36)$$

Figure (4.18) plots an exemplary contaminated AR(2) process $\mathbf{Y} = \mathbf{X} + \mathbf{V}$ along with $\hat{\mathbf{X}}^{\text{BIP}(2)}(\hat{\mathbf{a}}_\tau)$, which can be interpreted as an outlier-cleaned version of \mathbf{Y} . \mathbf{V} is an isolated additive outlier process of constant amplitude $\xi = 10$ and random sign. The contamination probability $\varepsilon = 0.1$. The true parameter vector is given by $\mathbf{a} = (-0.5, 0.1)^\top$, $\sigma_Z = 1$ and using simulated data of length $N = 1000$ for this example, yields $\hat{\mathbf{a}}_\tau(\mathbf{Y}) = (-0.472, 0.085)^\top$. For comparison, $\hat{\mathbf{a}}_{ML}(\mathbf{Y}) = (-0.066, -0.013)^\top$ and $\hat{\mathbf{a}}_{ML}(\mathbf{X}) = (-0.48, -0.105)^\top$. For visual clarity, we plot an excerpt of 100 samples.

Figure (4.19) illustrates the estimation a_2 for a contaminated AR(2) process. In this case, the τ -estimator based on the AR(2) model yields a biased estimate of a_2 due to

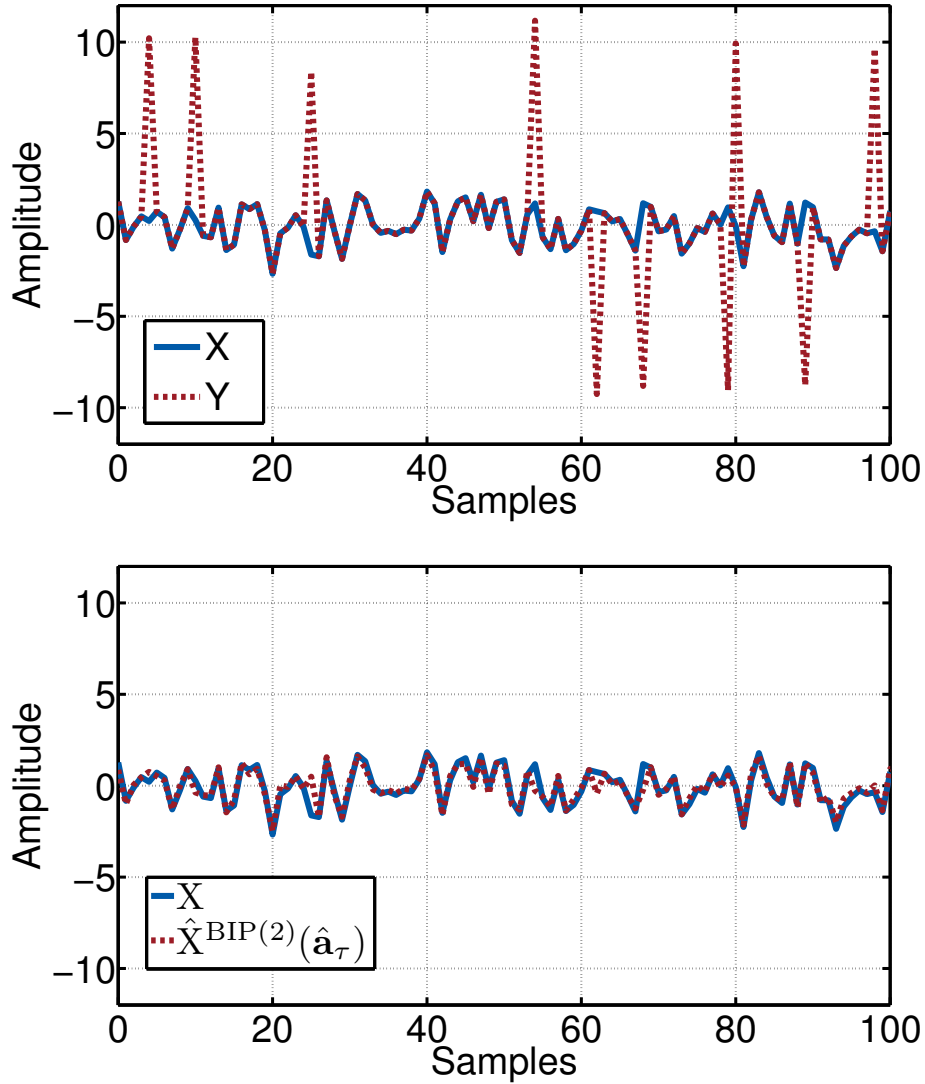


Figure 4.18: (top) Example of a contaminated AR(2) process $\mathbf{Y} = \mathbf{X} + \mathbf{V}$ along with the unobservable true AR(2) process \mathbf{X} . \mathbf{V} is an isolated additive outlier process of constant amplitude $\xi = 10$, random sign and contamination probability $\varepsilon = 0.1$. (bottom) $\hat{\mathbf{X}}^{\text{BIP}(2)}(\hat{\mathbf{a}}_\tau)$, which can be interpreted as an outlier-cleaned version of \mathbf{Y} .

the propagation of outliers, see Section 4.5.1. The τ -estimator based on the BIP-AR(2) model, on the other hand, maintains a reliable estimate of $a_2 = 0.085$.

For a general AR(p) process, after estimating the AR(1) as described above, the same steps as described for the AR(2) are performed for $m = 2, \dots, p$ to find

$$\hat{a}_{m,m} = \operatorname{argmin}_{\zeta} \{ \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(m)}(\zeta)), \hat{\sigma}_\tau(\mathbf{Z}^{(m)}(\zeta)) \} \quad (4.37)$$

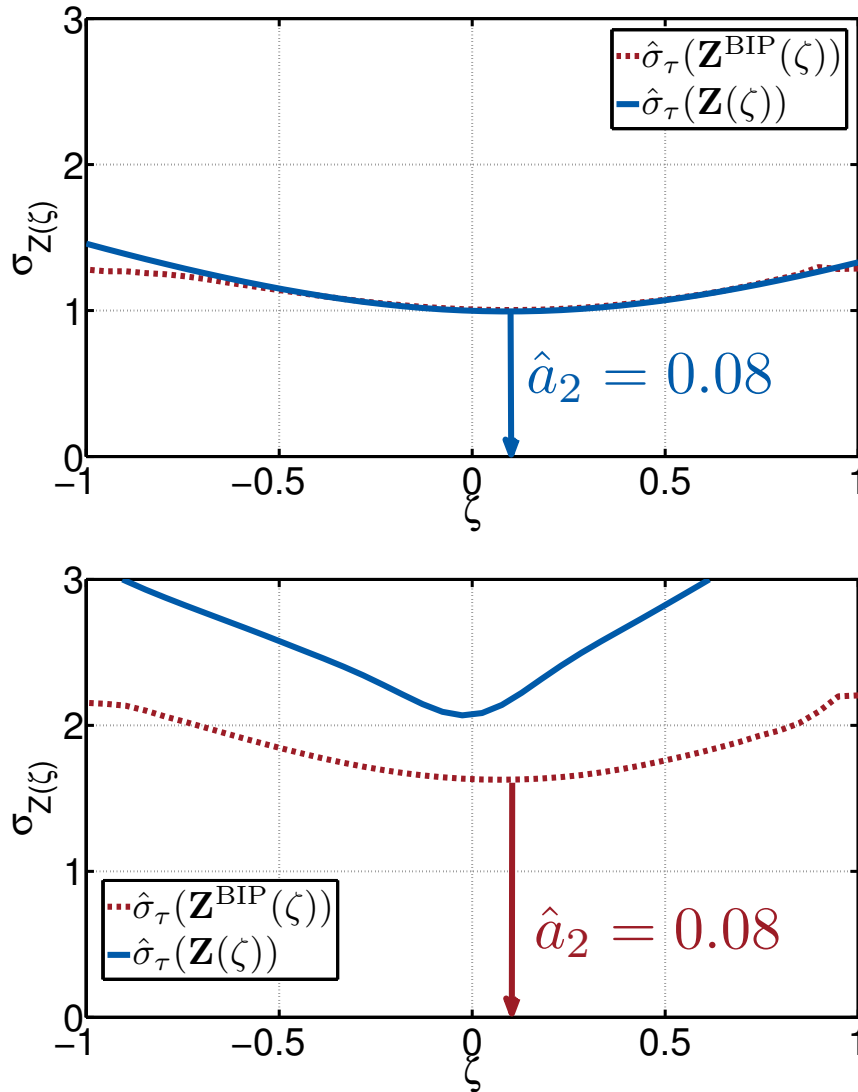


Figure 4.19: Example of finding $-1 < \zeta < 1$ which minimizes $\hat{\sigma}_{\tau}(\mathbf{Z}^{\text{BIP}(2)}(\zeta))$ and $\hat{\sigma}_{\tau}(\mathbf{Z}^{(2)}(\zeta))$ for an AR(2) process with $\mathbf{a} = (-0.5, 0.1)^{\top}$ and $\sigma_{\mathbf{Z}} = 1$. (top) $Y_n = X_n$ clean data example (bottom) $Y_n = X_n + V_n$, where V_n contains 10 % isolated additive outliers of amplitude 10 and random sign.

with the help of the Durbin-Levinson recursion:

$$\hat{a}_{m,m} = \begin{cases} \zeta & \text{if } n = m \\ \hat{a}_{m-1,n} - \zeta \hat{a}_{m-1,m-n} & \text{if } 1 \leq n \leq m-1, \end{cases} \quad (4.38)$$

where in Eq. (4.37),

$$\mathbf{Z}_n^{\text{BIP}(m)}(\zeta) = Y_n + \sum_{k=1}^m \hat{a}_{m,k}(\zeta) \hat{X}_{n-k}^{\text{BIP}(m)}(\zeta), \quad n = m+1, \dots, N, \quad (4.39)$$

with

$$\hat{X}_{n-k}^{\text{BIP}(m)}(\zeta) = Y_{n-k} - Z_{n-k}^{\text{BIP}(m)}(\zeta) + \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(m)}(\zeta)) \psi_\tau \left(\frac{Z_{n-k}^{\text{BIP}(m)}(\zeta)}{\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}(m)}(\zeta))} \right), \quad n = m+1, \dots, N,$$

and

$$Z_n^{(m)}(\zeta) = Y_n + \sum_{k=1}^m \hat{a}_{m,k}(\zeta) Y_{n-k}, \quad n = m+1, \dots, N. \quad (4.40)$$

The final estimates are then given by $\hat{\mathbf{a}}_\tau = (\hat{a}_{m,1}, \hat{a}_{m,2}, \dots, \hat{a}_{m,m})^\top$ from Eq. (4.37) and Eq. (4.38) with $m = p$.

4.6.2.2 Estimating the ARMA Model Parameters

In this Section, we describe a computationally attractive algorithm to robustly estimate ARMA(p, q) models, such that stationary and invertible models are obtained. The general idea is to derive the ARMA(p, q) parameters from a long AR(p_0) model. We estimate this model robustly with the BIP τ -estimator, as described in the previous Section. The final ARMA(p, q) parameter estimates are obtained by the reduced statistics estimator, see Section 4.2.2.1, which minimizes a τ -estimate of scale.

Many simulation studies have been done to find a suitable choice for p_0 , i.e., the intermediate AR order for the ARMA estimation. From a theoretical viewpoint, [136, 152], none of the existing criteria are able to select an intermediate AR order for which the parameters are most accurate (rather than the model best suited for prediction). To determine an appropriate value for p_0 , we propose to use robust model order selection criteria, as will be described in Section 4.7. Adapting the results by [136, 152], we divide the penalty factor by two, which can be made plausible by considering that a factor of one gives the same weight to bias and variance contributions [136].

Depending on the procedure applied in the first stage to obtain an initial separation into the AR and MA parts, we define the reduced statistics estimates of $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{b})$ in Eq. (4.1) to be:

$$\text{a) } \hat{\boldsymbol{\theta}}_\tau^{(\text{AR})} := (\hat{\mathbf{a}}_\tau^{(\text{AR})}, \hat{\mathbf{b}}_\tau^{(\text{AR})})^\top$$

with $\hat{\mathbf{a}}_\tau^{(\text{AR})} := (\hat{a}_1^{(\text{AR})}, \dots, \hat{a}_p^{(\text{AR})})^\top$ and $\hat{\mathbf{b}}_\tau^{(\text{AR})} := (\hat{b}_1^{(\text{AR})}, \dots, \hat{b}_q^{(\text{AR})})^\top$, where the parameters are estimated from $\hat{\mathbf{a}}_\tau^0 = (a_1, \dots, a_{p_0})^\top$ using the ‘‘long AR’’ estimator defined by Eq. (4.7) and Eq. (4.8).

$$\text{b) } \hat{\boldsymbol{\theta}}_{\tau}^{(\text{MA})} := (\hat{\mathbf{a}}_{\tau}^{(\text{MA})}, \hat{\mathbf{b}}_{\tau}^{(\text{MA})})^{\top}$$

with $\hat{\mathbf{a}}_{\tau}^{(\text{MA})} := (\hat{a}_1^{(\text{MA})}, \dots, \hat{a}_p^{(\text{MA})})^{\top}$ and $\hat{\mathbf{b}}_{\tau}^{(\text{MA})} := (\hat{b}_1^{(\text{MA})}, \dots, \hat{b}_q^{(\text{MA})})^{\top}$, where the parameters are estimated from $\hat{\mathbf{a}}_{\tau}^0 = (a_1, \dots, a_{p_0})^{\top}$ using the “long MA” estimator defined by Eq. (4.9) and Eq. (4.10).

$$\text{c) } \hat{\boldsymbol{\theta}}_{\tau}^{(\text{CO})} := (\hat{\mathbf{a}}_{\tau}^{(\text{CO})}, \hat{\mathbf{b}}_{\tau}^{(\text{CO})})^{\top}$$

with $\hat{\mathbf{a}}_{\tau}^{(\text{CO})} := (\hat{a}_1^{(\text{CO})}, \dots, \hat{a}_p^{(\text{CO})})^{\top}$ and $\hat{\mathbf{b}}_{\tau}^{(\text{CO})} := (\hat{b}_1^{(\text{CO})}, \dots, \hat{b}_q^{(\text{CO})})^{\top}$, where the parameters are estimated from $\hat{\mathbf{a}}_{\tau}^0 = (a_1, \dots, a_{p_0})^{\top}$ using the “long COV” estimator defined by Eq. (4.11) and Eq. (4.12).

$$\text{d) } \hat{\boldsymbol{\theta}}_{\tau}^{(\text{RI})} := (\hat{\mathbf{a}}_{\tau}^{(\text{RI})}, \hat{\mathbf{b}}_{\tau}^{(\text{RI})})^{\top}$$

with $\hat{\mathbf{a}}_{\tau}^{(\text{RI})} := (\hat{a}_1^{(\text{RI})}, \dots, \hat{a}_p^{(\text{RI})})^{\top}$ and $\hat{\mathbf{b}}_{\tau}^{(\text{RI})} := (\hat{b}_1^{(\text{RI})}, \dots, \hat{b}_q^{(\text{RI})})^{\top}$, where the parameters are estimated from $\hat{\mathbf{a}}_{\tau}^0 = (a_1, \dots, a_{p_0})^{\top}$ using the “long RINV” estimator defined by Eq. (4.13) and Eq. (4.14).

From this, we obtain eight different estimates of $\mathbf{Z}(\boldsymbol{\theta})$:

- (i) The ARMA innovations $\mathbf{Z}(\hat{\boldsymbol{\theta}}_{\tau}) = (Z_{p+1}(\hat{\boldsymbol{\theta}}_{\tau}), \dots, Z_N(\hat{\boldsymbol{\theta}}_{\tau}))^{\top}$ by plugging $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{AR})}$, $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{MA})}$, $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{CO})}$ and $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{RI})}$ into

$$Z_n(\hat{\boldsymbol{\theta}}_{\tau}) = Y_n + \sum_{k=1}^p \hat{a}_k Y_{n-k}(\hat{\boldsymbol{\theta}}_{\tau}) - \sum_{k=1}^q \hat{b}_k Z_{n-k}(\hat{\boldsymbol{\theta}}_{\tau}) \quad (4.41)$$

for $n = p + 1, \dots, N$.

- (ii) The BIP-ARMA innovations $\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}) = (Z_{p+1}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}), \dots, Z_N^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}))^{\top}$ by plugging $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{AR})}$, $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{MA})}$, $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{CO})}$ and $\hat{\boldsymbol{\theta}}_{\tau}^{(\text{RI})}$ into

$$\begin{aligned} Z_n^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}) &= Y_n + \sum_{k=1}^p \hat{a}_k \hat{X}_{n-k}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}) \\ &+ \sum_{k=1}^r \left((\hat{b}_k - \hat{a}_k) \hat{\sigma}_{\tau}(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau})) \psi_{\tau} \left(\frac{Z_{n-k}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau})}{\hat{\sigma}_{\tau}(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}))} \right) + \hat{a}_k Z_{n-k}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}) \right), \end{aligned} \quad (4.42)$$

where $n = p + 1, \dots, N$ and $r = \max\{p, q\}$, if $r > p$, $a_{p+1} = \dots = a_r = 0$, while if $r > q$, $b_{q+1} = \dots = b_r = 0$. Here,

$$\hat{X}_{n-k}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau) = Y_{n-k} - Z_{n-k}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau) + \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))\psi_\tau \left(\frac{Z_{n-k}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))} \right) \quad (4.43)$$

and

$$\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)) = \frac{\hat{\sigma}_\tau(\mathbf{Y})}{1 + \gamma^2 \sum_{k=1}^{\infty} b_n^\infty(\hat{\boldsymbol{\theta}}_\tau)}, \quad (4.44)$$

where $\hat{\sigma}_\tau(\mathbf{Y})$ is the τ -estimate of the scale of $\mathbf{Y} = (Y_1, \dots, Y_N)^\top$, $\mathbf{b}^\infty(\hat{\boldsymbol{\theta}}_\tau) = (b_1^\infty(\hat{\boldsymbol{\theta}}_\tau), \dots, b_\infty^\infty(\hat{\boldsymbol{\theta}}_\tau))$ is the MA(∞) representation of the ARMA(p, q) with parameter estimates $\hat{\boldsymbol{\theta}}_\tau$ and $\gamma^2 = \text{Var}[\psi_\tau(X)]$, when the distribution of X is standard Gaussian.

Having computed these estimates, which, in the sequel, are denoted depending on the procedure applied to obtain the initial separation into the AR and MA parts in the first stage, the final BIP-ARMA(p, q) τ -estimate $\hat{\boldsymbol{\theta}}_\tau = (\hat{\mathbf{a}}_\tau, \hat{\mathbf{b}}_\tau)^\top$ is given by

$$\hat{\boldsymbol{\theta}}_\tau = \underset{\hat{\boldsymbol{\theta}}_\tau}{\text{argmin}} \left\{ \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{AR})})), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{MA})})), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{CO})})), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{RI})})), \right. \\ \left. \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{AR})})), \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{MA})})), \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{CO})})), \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{RI})})) \right\}. \quad (4.45)$$

This means that the final parameter estimate is the one which minimizes the τ -estimate of scale of the innovations of all candidates, evaluated both under the BIP-ARMA(p, q) and the ARMA(p, q) model.

Additionally, the BIP τ -estimator intrinsically supplies us with a robust innovations scale estimate

$$\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)) = \min \left\{ \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{AR})})), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{MA})})), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{CO})})), \hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau^{(\text{RI})})), \right. \\ \left. \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{AR})})), \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{MA})})), \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{CO})})), \hat{\sigma}_\tau(\mathbf{Z}(\hat{\boldsymbol{\theta}}_\tau^{(\text{RI})})) \right\}. \quad (4.46)$$

The final innovations sequence estimate $\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)$ is obtained by using the results from Eq. (4.45) and Eq. (4.46) in Eq. (4.42).

4.6.3 Simulations

In this Section, we give some simulation results for the parameter estimation of different $\text{ARMA}(p,q)$ that are contaminated by various outlier models. We compare the proposed BIP τ -estimator to some existing robust and classical estimators.

4.6.3.1 $\text{AR}(p)$ Models With Different Outlier Models

Figure 4.20 plots a realization of an isolated AO process V_n^ε (see Eq. (4.17)), as used in the simulations in this Section. In this setup, the probability of contamination is $\varepsilon = 0.1$, the positions ξ_n^ε are Bernoulli distributed and the p.d.f. of $W_n, n = 1, \dots, 500$ is zero-mean Gaussian with $\sigma_W = \kappa\sigma_X$, where $\kappa = 10$. In case of an IO process, V_n^ε contaminates the innovations process Z_n , see (see Eq. (4.20)). For a clean process, the observations $Y_n = X_n$. In this setup, X_n follows a Gaussian $\text{AR}(1)$ model with parameters $a_1=0.5, \sigma_Z = 1, \mu_Z = 0, N = 500$.

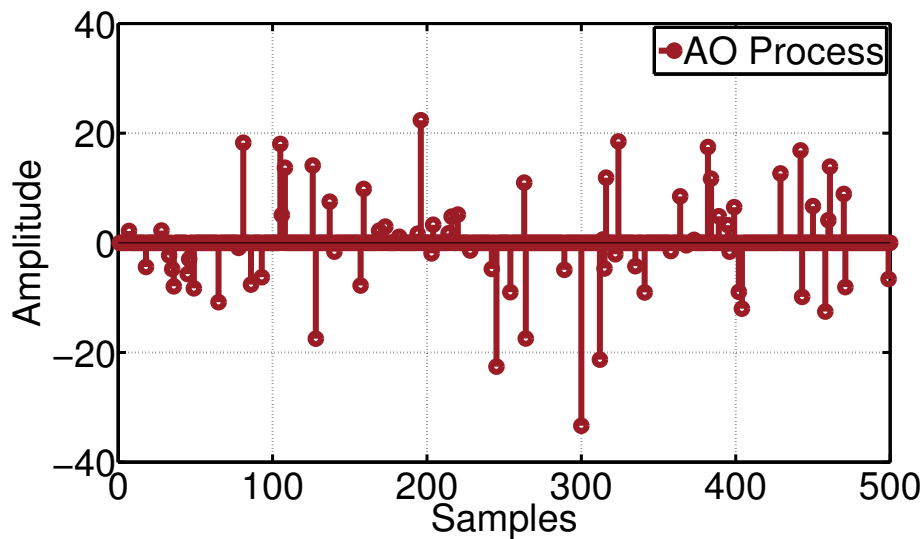


Figure 4.20: An example of a realization of an isolated AO process V_n^ε (see Eq. (4.17)), as used in the simulations in this Section.

Figure 4.21 plots the parameter estimates of six robust estimators and the classical ML-estimator for three cases: clean data, innovations outliers and additive outliers for 100 Monte Carlo runs.

The estimators considered are:

1. Gaussian ML: the maximum likelihood-estimator under the Gaussian assumption.
2. 3- σ -Cleaned Gaussian ML: all samples with amplitude greater than $3\hat{\sigma}_{\text{MAD}}$ are rejected. A missing data ML-approach is performed with the remaining samples, see Section 4.5.2.
3. MRE: median-of-ratios-estimator (MRE) [7, 124], as described in Section 4.5.3.
4. Ranks: the estimator proposed in [121]. It is obtained by minimizing a rank-based residual dispersion function. According to [121], the estimator can have the same asymptotic efficiency as ML-estimators and are claimed to be robust.
5. Filtered τ : the estimator proposed in [8]. It consists of an approximate conditional mean robust filter combined with the τ -estimator, see Section 4.5.4. The estimate minimizes the τ -estimate of the scale of $\mathbf{Z}^f(\hat{\boldsymbol{\theta}})$, as given in Eq. (4.27) with $p = 1$ and $q = 0$.
6. BIP τ : the proposed bounded influence propagation τ -estimator, as described in full detail in Section 4.6.2.

Table 4.1 gives the Monte Carlo averaged MSE for parameter estimation. Breakdown is highlighted with red bold font. Best results are highlighted by black bold font. From this example, which is the simplest possible ARMA model, we conclude that for:

- The clean data case: highly efficient estimators (ML, BIP τ) do best, some loss in optimality for non-efficient estimators. All estimators, for this setup are biased $a_1=0.5$, while $\hat{a}_1 \approx 0.4$, even for $N = 500$.
- Innovations outliers: the ranks-based estimator yields best results, classical estimators are next followed by BIP τ . The overall conclusion is that innovations outliers are not harmful to classical or robust estimators.
- Additive outliers: BIP τ performs best and sophisticated robust estimators outperform the 3- σ -cleaned Gaussian ML. Clearly, additive outliers cause the Gaussian ML-estimator to break down. We also see that the robustness of the ranks-based estimator [121] is only against heavy-tailed innovations and the estimator breaks down for a contamination level of $\varepsilon = 0.1$ (10 %) additive outliers.

Overall, it can be seen that the BIP τ -estimator is the only estimator that yields similar results for all three scenarios. In two cases, it performs best, and for innovations

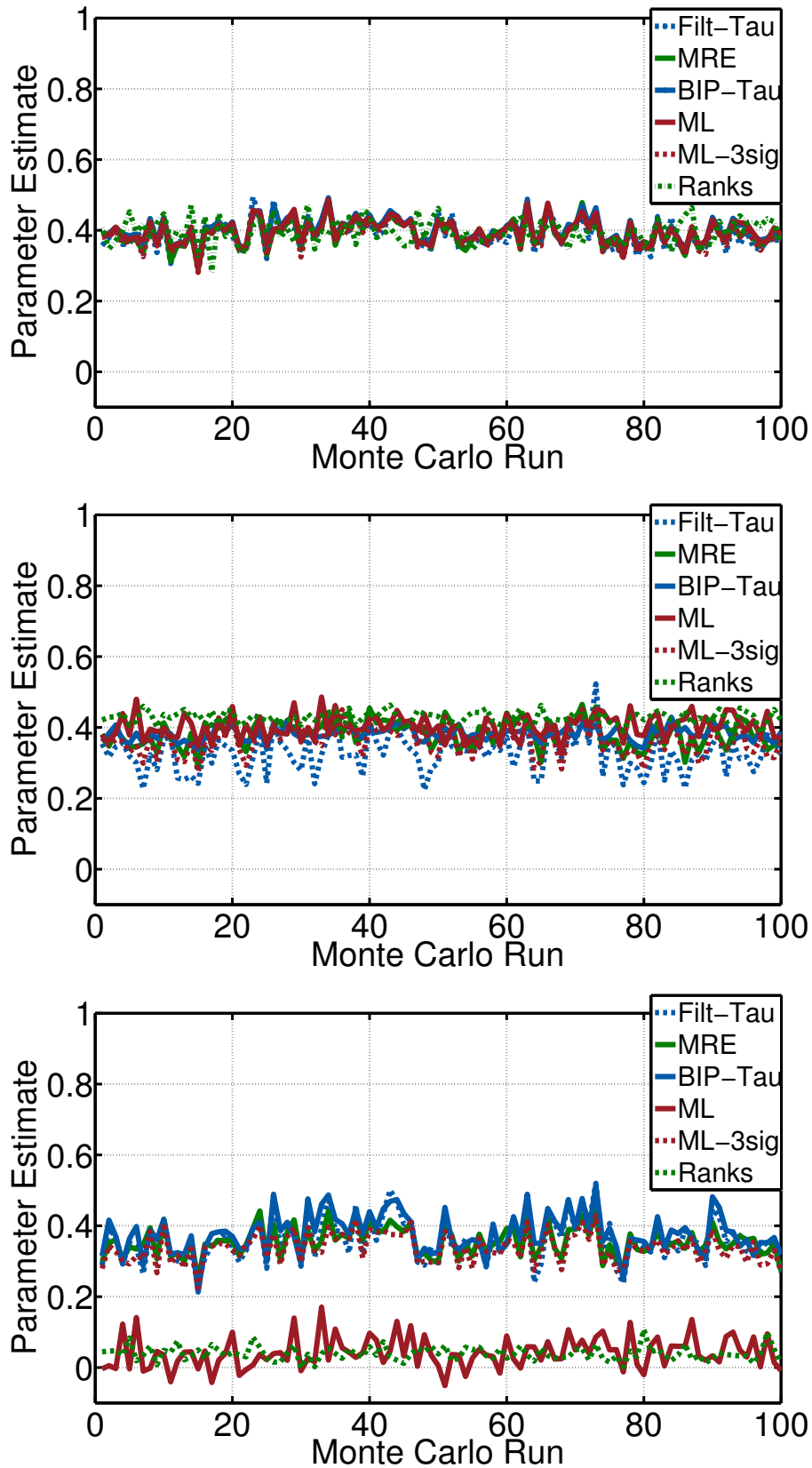


Figure 4.21: An example of 100 MC runs of parameter estimation of an AR(1) with $a_1=0.5$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 500$. The three cases considered are: (top) Clean data; (middle) innovations outliers (IO); and (bottom) additive outliers (AO). In both outlier models $\varepsilon = 0.1$, $\kappa = 10$. All estimators, for this setup are biased, even for $N = 500$.

outliers, it is among the top three. Furthermore, estimators that robustify against heavy-tailed innovations, e.g., [121] do not ensure any form of robustness against additive outliers, which are most harmful to classical estimators.

	MSE for Clean Data	MSE for IO	MSE for AO
Gaussian ML	0.0119	0.0109	0.2119
3- σ -cleaned Gaussian ML	0.0142	0.0200	0.0280
MRE	0.0132	0.0152	0.0229
Ranks	0.0118	0.0056	0.2125
Filtered τ	0.0145	0.0331	0.0237
BIP τ	0.0112	0.0144	0.0182

Table 4.1: Monte Carlo averaged (100 MC runs) MSE for parameter estimation of an AR(1) with $a_1=0.5$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 500$. The three cases considered are: clean data, IO and AO. In both outlier models $\varepsilon = 0.1$, $\kappa = 10$. Breakdown is highlighted with red bold font. Best results highlighted by black bold font.

4.6.3.2 ARMA(p,q) Models with Additive Outliers

In this Section, we give two examples of estimating ARMA(p,q) models in presence of additive outliers.

Example 1: ARMA(2,1)

For the first example, we consider the ARMA(2,1) process that, for the clean data case, was investigated in [139]. X_n has the following parameters: $\mathbf{a} = (1.5, 0.625)^\top$, $b = -0.5$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 1000$. The outliers follow the isolated additive outlier model with V_n^ε (see Eq. (4.17)), as used in the simulations in this Section. In this setup, the probability of contamination is $\varepsilon = 0.05$, the positions ξ_n^ε are Bernoulli distributed and the p.d.f. of W_n , $n = 1, \dots, 1000$ is zero-mean Gaussian with $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. For the clean data case, $Y_n = X_n$.

Figure 4.22 plots the parameter estimates of the proposed BIP τ -estimator for the ARMA(2,1) process given in [139], evaluated over 100 Monte Carlo runs. It can be seen that the estimator shows excellent performance for the clean data case. The variability of the parameter estimates is increased in presence of additive outliers, however, in none of the Monte Carlo runs, the estimator broke down.

Example 2: ARMA(2,2)

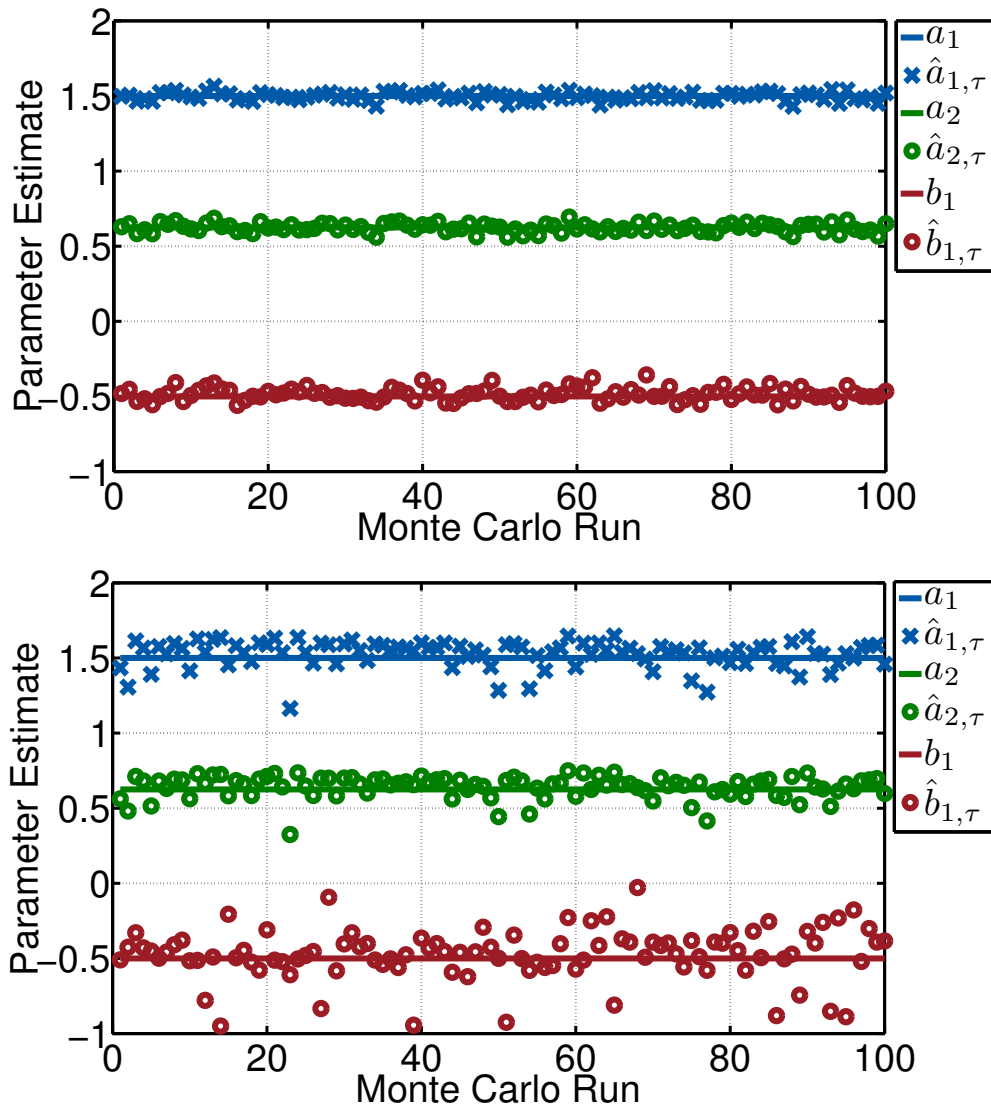


Figure 4.22: Example of parameter estimation of an ARMA(2,1) [139] with $\mathbf{a} = (1.5, 0.625)^\top$, $b = -0.5$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 1000$. (top) clean data case $Y_n = X_n$ (bottom) isolated additive outliers with probability of contamination $\varepsilon = 0.05$ and amplitude $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$.

As a second example, we consider the ARMA(2,2) process that, for the clean data, case was investigated in [153]. Here, X_n has the following parameters: $\mathbf{a} = (-1.65, 0.845)^\top$, $\mathbf{b} = (-1.112, 0.596)^\top$, $\sigma_Z = 1.112$, $\mu_Z = 0$, $N = 1000$. The outliers again follow the isolated additive outlier model with V_n^ε having the same parameters, as for the ARMA(2,1) study.

Figure 4.23 plots the parameter estimates of the proposed BIP τ -estimator for the ARMA(2,2) process given in [153] evaluated over for 100 Monte Carlo runs. Figure 4.24 shows the reduced statistics with missing data refinement, i.e., computing an ML-

estimate after a $3\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))$ rejection. It can be seen that, with this refinement, the estimator is nearly able to maintain the performance obtained in the clean data case in presence of outliers.

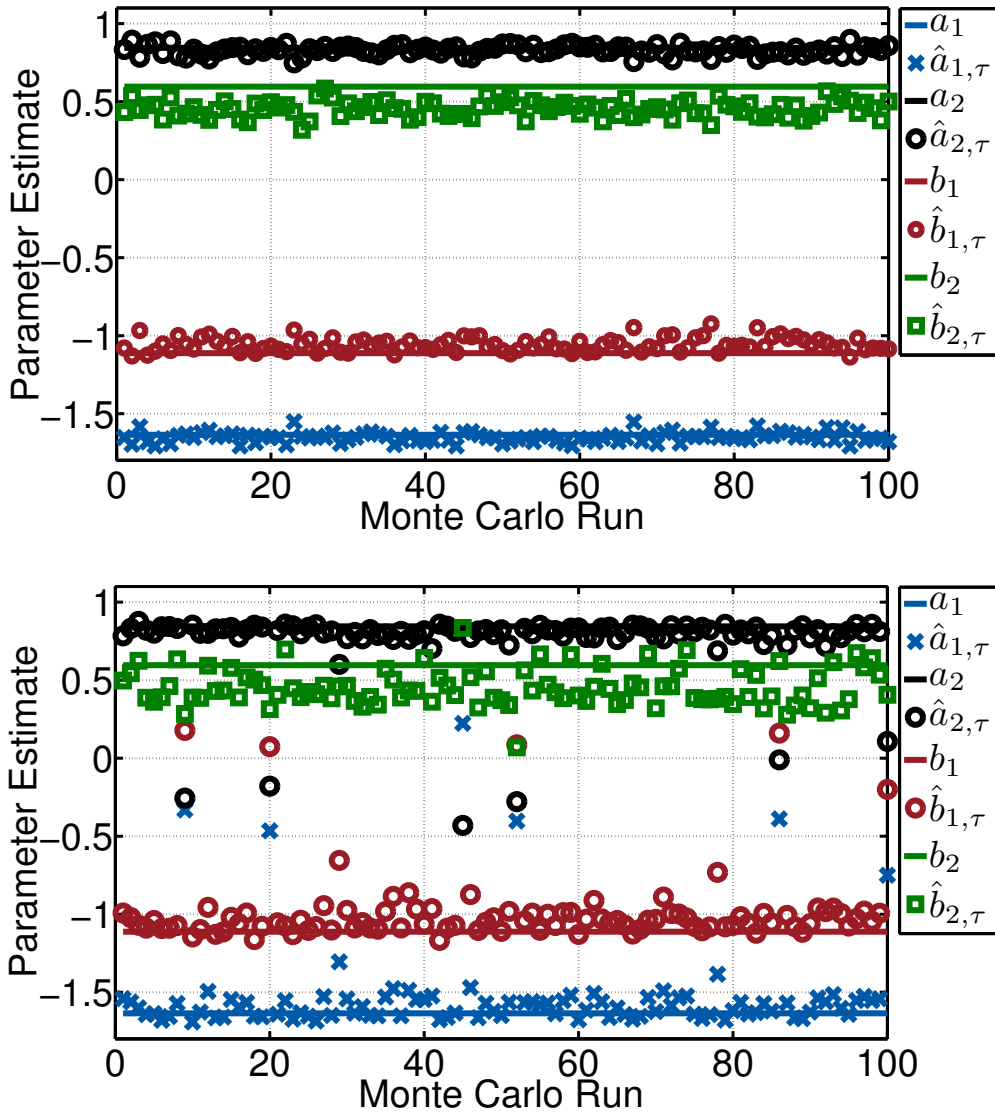


Figure 4.23: Example of parameter estimation of an ARMA(2,2) [153] with $\mathbf{a} = (-1.65, 0.845)^\top$, $\mathbf{b} = (-1.112, 0.596)^\top$, $\sigma_Z = 1.112$, $\mu_Z = 0$, $N = 1000$. (top) clean data case $Y_n = X_n$ (bottom) isolated additive outliers with probability of contamination $\varepsilon = 0.05$ and amplitude $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. The plots show the procedure using reduced statistics only.

4.6.3.3 Fast BIP τ Robust Spectrum Estimation

In this Section, we give an example of how the fast BIP τ -estimator can be used to obtain robust spectrum estimates. We compare the obtained results for the ARMA(2,2)

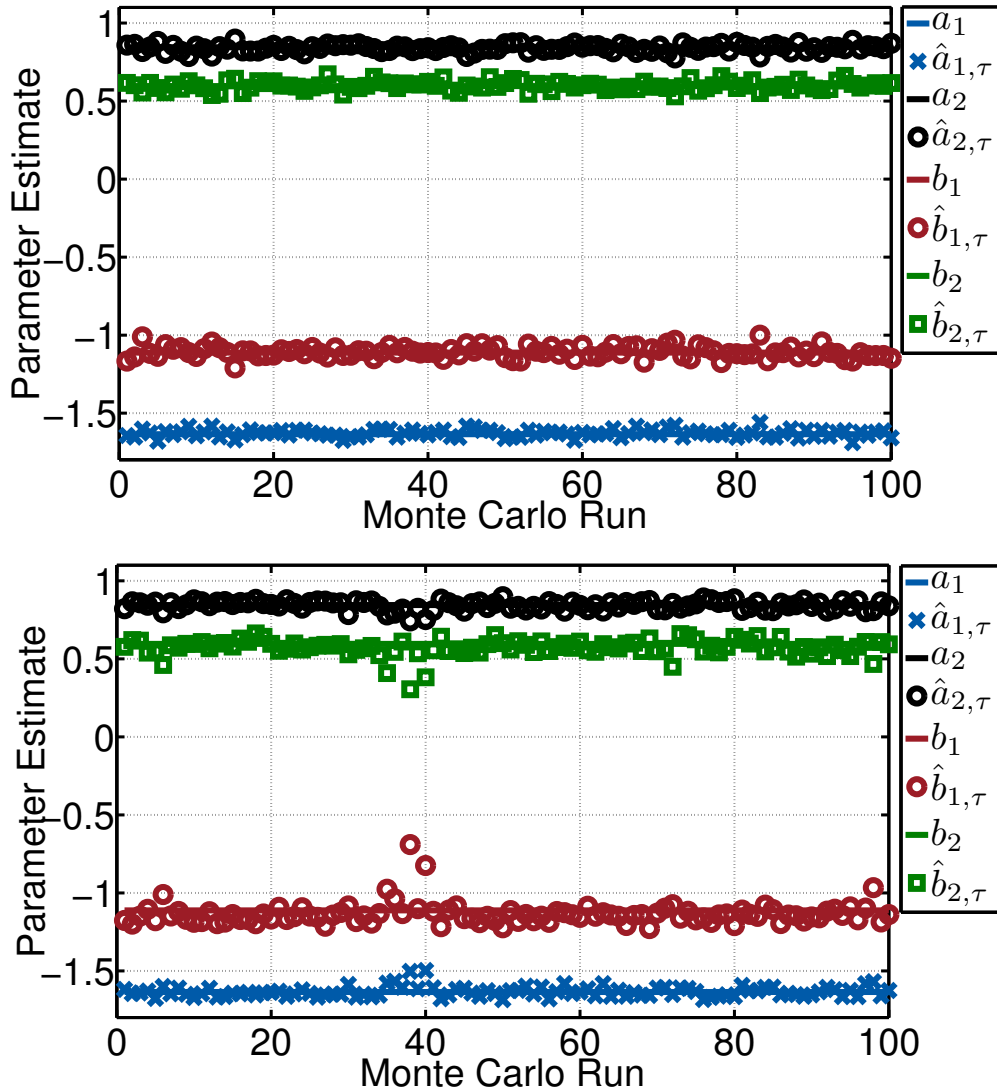


Figure 4.24: Example of parameter estimation of an ARMA(2,2) [153] with $\mathbf{a} = (-1.65, 0.845)^\top$, $\mathbf{b} = (-1.112, 0.596)^\top$, $\sigma_Z = 1.112$, $\mu_Z = 0$, $N = 1000$. (top) clean data case $Y_n = X_n$ (bottom) isolated additive outliers with probability of contamination $\varepsilon = 0.05$ and amplitude $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. The plots depict the reduced statistics with missing data refinement.

process [153] with simulation parameters, as defined above. The proposed estimator is compared to (i) the classical Bartlett estimator, (ii) a simple nonparametric robust estimator described in [8] that replaces the sample averages in the Bartlett estimator by robust location estimates and (iii) the ℓ_1 -Bartlett estimator that is based on the ℓ_1 -periodogram estimator proposed by [126] which interprets the periodogram as the least-squares solution of a (nonlinear) harmonic regression and replaces the ℓ_2 -norm minimization by an ℓ_1 -norm minimization. This estimator is a special case of the M-periodogram proposed by [154], where the ℓ_2 -norm cost function is replaced by an

M-estimation cost function.

The BIP τ spectrum estimate is based on the robustly estimated ARMA parameters, i.e., for a wide sense stationary process X_n , the ARMA(p, q) BIP τ spectrum estimate is given by

$$\hat{C}_{XX}^{\text{BIP-ARMA}}(e^{j\omega}, \hat{\boldsymbol{\theta}}_\tau) = \frac{\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))}{2\pi} \frac{\left|1 + \sum_{k=1}^q \hat{b}_k e^{-j\omega k}\right|^2}{\left|1 + \sum_{k=1}^p \hat{a}_k e^{-j\omega k}\right|^2},$$

with the parameter estimates $\hat{\boldsymbol{\theta}}_\tau$ as given in Eq. (4.45) and $\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))$ computed as in Eq. (4.46). Figure 4.25 plots the power spectral density estimates averaged over 100 Monte Carlo runs along with the true spectral density for the clean data and additive outlier cases. It can be noticed that only the fast BIP τ -estimator and the ℓ_1 -periodogram estimator [126] provide similar results in both noise conditions. Furthermore, in case of additive outliers, these two estimators provide results that are quite close to the true spectrum. In all cases, the fast BIP τ -estimator is closest to the true spectrum.

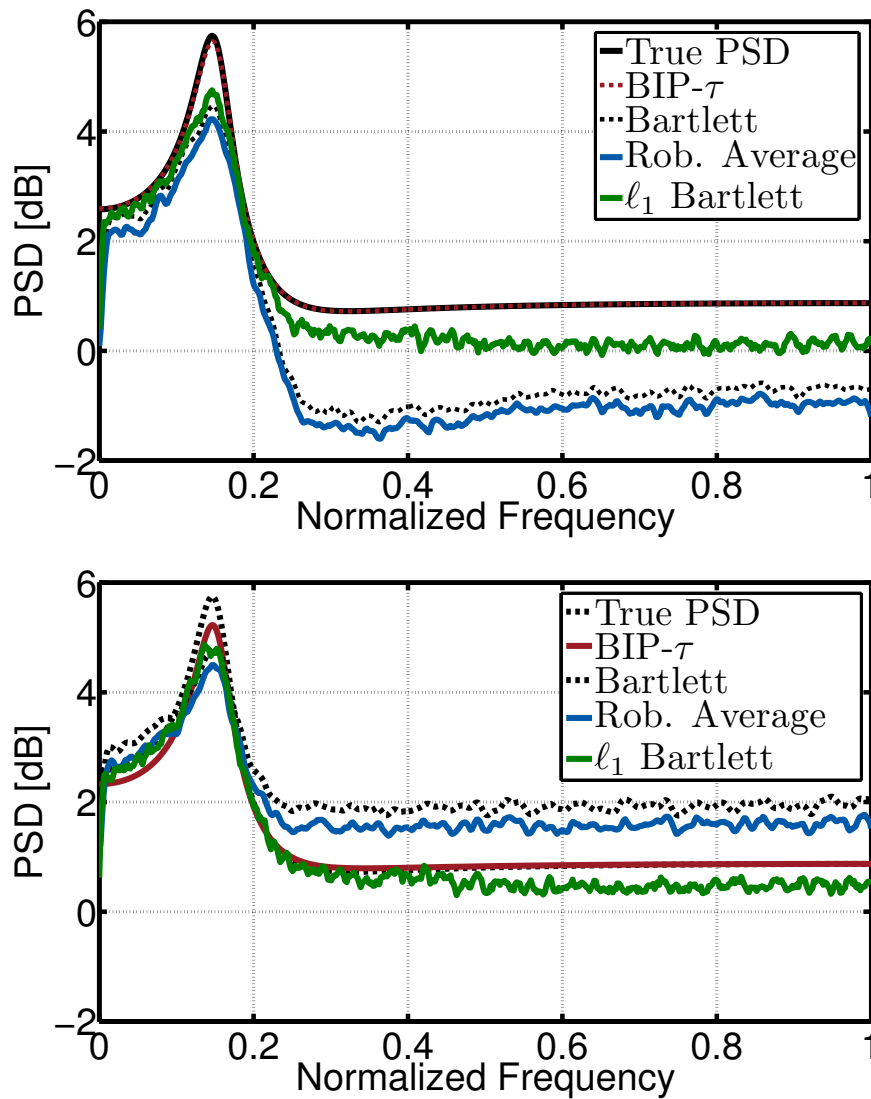


Figure 4.25: Example of spectrum estimation of an ARMA(2,2) [153] with $\mathbf{a} = (-1.65, 0.845)^\top$, $\mathbf{b} = (-1.112, 0.596)^\top$, $\sigma_Z = 1.112$, $\mu_Z = 0$, $N = 1000$. The power spectral density estimates are averaged over 100 Monte Carlo runs. (top) clean data case $Y_n = X_n$ and (bottom) isolated additive outliers with probability of contamination $\varepsilon = 0.05$ and amplitude $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$.

4.7 Robust Model Order Selection for ARMA Models

A fundamental task when fitting an ARMA model to real-world data is the selection of the autoregressive and moving-average orders. As real-world data frequently contains outliers and artifacts, robust model order selection becomes crucial. Robust model order selection aims at finding a suitable statistical model that describes the majority of the data while preventing atypical observations (outliers) from having overriding influence on the final conclusions.

In the classical non-robust setting, order determination for ARMA models is a diverse field of research that is scattered over several scientific disciplines. Up to date, no single general method has been established that is superior for all purposes. As with general order selection problems, methods have been classified into statistical hypothesis testing, methods based on the analysis of (extended) autocorrelation functions, methods based on prediction errors and methods based on information criteria, amongst others.

Literature on robust order determination for ARMA models is much more sparse, especially in presence of additive outliers. The demand for robust order selection methods, however, has been repeatedly stated as an important issue that needs to be addressed, e.g.:

1. *“Although, some convincing examples of the use of a proposed robust order-selection rule based on BIFAR estimates may be found ..., careful studies of the problem are needed.”* [116]
2. *“Many other theoretical as well as empirical problems should be investigated. For example, how sensitive are some of the order-determination methods to deviations from the familiar but restrictive assumption that the white noise process is Gaussian distributed? How robust are the various methods in the presence of outliers and bad data? It is our hope and expectation that these problems will be solved in the very near future.”* [155]
3. *“One important problem that will be only briefly mentioned here is that of the robust model selection.”* [123]
4. *“... un autre travail de recherche qui est de dériver un critère de selection d’ordre robuste...”* [128]

In this thesis, we focus our attention on robust model order selection based on information criteria.

4.7.1 Classical Model Order Selection Using Information Criteria

The classical information criteria for Gaussian ARMA models discussed in this Section all share the general form

$$\underbrace{\log(\hat{\sigma}_{ML}^2(\mathbf{Z}(\hat{\boldsymbol{\theta}}_{ML})))}_{\text{data fit}} + \underbrace{\alpha(p, q, N)}_{\text{model complexity penalty}} \quad (4.47)$$

Here, α is a penalty term that depends on p, q and N . The following table sums up the penalty terms of some important classical information criteria.

	$\alpha(p, q, N)$
AIC [39]	$\frac{2(p+q)}{N}$
AICc [13]	$\frac{2(p+q)}{N-(p+q)-2}$
SIC [42]	$\frac{p+q}{N} \log(N)$
HQ [44]	$2(p+q) \frac{\log(\log(N))}{N}$

Table 4.2: Penalty terms $\alpha(p, q, N)$ of classical information criteria to select ARMA(p, q) models.

4.7.2 Robust ARMA Model Selection

Figure 4.26 illustrates the model order selection for an AR(2) process with $\mathbf{a} = (0.7, -0.4)$, given $N = 200$ samples, in the presence of isolated additive outliers with probability of contamination $\varepsilon = 0.05$ and amplitude $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. The parameters of the robust data fit are estimated with the BIP τ -estimator. The penalty is of the SIC type. The robust criterion consists of the superposition of a robust data fit term and a model complexity penalty and has its minimal value at the true AR order of $p = 2$.

4.7.2.1 Existing Approaches for Robust ARMA Model Selection

To the best of our knowledge, robust model order selection for dependent data in presence of additive outliers has not been treated much in the robust statistics community

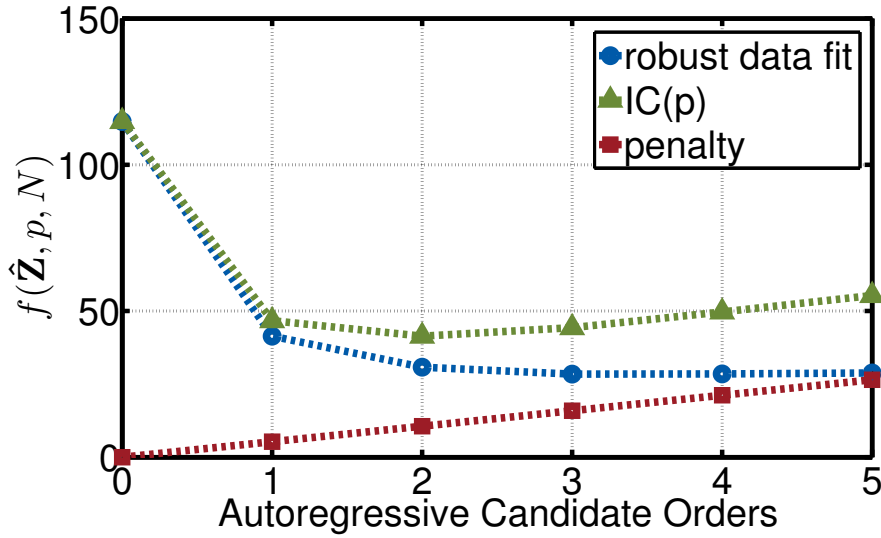


Figure 4.26: Example of model order selection for an AR(2) process. The robust criterion consists of the superposition of a robust data fit term and a model complexity penalty and has its minimal value at the true AR order of $p = 2$.

with the notable exceptions of [8, 12, 53, 156–159] in case of AR processes and [160] for ARMA processes. These are briefly reviewed in the sequel.

Existing Approach 1: “Criterion based on a robust likelihood obtained with robustly filtered estimates”

The approach by [158] is based on robust filters and yields criteria via a robust likelihood

$$\log(f(\mathbf{y}|p, \hat{\boldsymbol{\theta}}^f)) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{n=p+1}^N \log(\sigma_n^2) - \frac{1}{2} \sum_{n=p+1}^N \rho \left(\frac{Z_n^f(\hat{\boldsymbol{\theta}}^f)}{\sigma_n} \right).$$

In this equation, $\hat{\boldsymbol{\theta}}^f$ is obtained by sequentially maximizing the robust likelihood as a function of the partial autocorrelations with a Newton-Raphson method. The prediction error innovations scale is obtained from the filter recursions, see Eq. (4.25), and $\rho(\cdot)$ is chosen to be Hampel’s two part re-descending.

This, for autoregressive processes, gives a Schwarz-type criterion [158]

$$\text{SIC}^f(p) = -2 \log(f(\mathbf{y}|p, \hat{\boldsymbol{\theta}}^f)) + p \log(N)$$

and an Akaike-type criterion [158]

$$\text{AIC}^f(p) = -2 \log(f(\mathbf{y}|p, \hat{\boldsymbol{\theta}}^f)) + 2p.$$

Further criteria of this type, e.g., [13, 44], can be found by varying the penalty term.

The approach by [158] is applicable to AR(p) processes and can be extended to ARMA(p, q) processes, by using a robust ARMA filter, see Eq. (4.27), however, the asymptotic analysis and robustness properties are not tractable in this case.

Existing Approach 2: “Criterion based on M-estimation”

The approach by [156, 157] is based on M-estimators. The order of an AR process is selected by the candidate order p that minimizes

$$\text{RAIC}_M(p) = \sum_{n=p+1}^N \rho \left(\frac{Z_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M} \right) + \alpha(p).$$

The penalty term used by [156] is

$$\alpha(p) = 2p$$

and [157] suggests

$$\alpha(p) = \text{tr}(\mathcal{E}_1^{-1} \mathcal{E}_2),$$

where

$$\mathcal{E}_1 := \text{E}_F \left[\psi' \left(\frac{Z_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M} \right) (Y_{n-1}, \dots, Y_{n-p}) (Y_{n-1}, \dots, Y_{n-p})^\top \right]$$

and

$$\mathcal{E}_2 := \text{E}_F \left[\psi^2 \left(\frac{Z_n(\hat{\boldsymbol{\theta}}_M)}{\hat{\sigma}_M} \right) (Y_{n-1}, \dots, Y_{n-p}) (Y_{n-1}, \dots, Y_{n-p})^\top \right].$$

This is the extension of Eq. (2.19) to AR models.

The above approach is limited in its applicability to robust ARMA model order selection, since the BP of the M-estimator for ARMA(p, q) processes is zero, and even for AR(p) processes, it is limited to $0.5/(p+1)$.

Existing Approach 3: “Criterion based on a robustly estimated innovations scale obtained with robustly filtered estimates”

A further preliminary suggestion has been made by [8], who suggest the following AIC type criterion

$$\text{AIC}^{f\tau}(p) = \log(\hat{\sigma}_\tau^2(\mathbf{Z}_\tau^f(\hat{\boldsymbol{\theta}}_\tau^f))) + \frac{2p}{N-p}.$$

Concerning the penalty term, the authors state that *“it would be better to multiply the penalty term by a factor depending on the choice of the scale and on the distribution of the residuals. This area requires further research.”* As the distribution of the residuals obtained by the filtered- τ -estimator $\hat{\boldsymbol{\theta}}_\tau^f$ (see [8], p.271) is not known, this is, up to now, an impossible task. Similar to the approach by [158], the approach by [8] can be extended to ARMA(p, q) models by combining a robust ARMA filter (see Eq. (4.27)) with a τ -estimator of scale, as described in [8], p.290.

Existing Approach 4: “Criterion based on a weighted likelihood”

A different approach has been proposed by [53], where a weighted likelihood is found via

$$\text{WAIC}(p, q) = -2 \sum_{n=1}^N w_n \log(f_n(y|p, q, \hat{\boldsymbol{\theta}}^w)) + 2(p + q).$$

Here, the weighted likelihood estimate $\hat{\boldsymbol{\theta}}^w$ is computed with weights obtained by adjusting the Pearson residuals, based on an outlier classification scheme and on kernel density estimation, for details the reader is referred to [53].

4.7.2.2 Robust BIP τ ARMA Model Selection

In the sequel, we describe and compare a set of possible model order selection criteria that are based on the proposed robust BIP τ -estimator.

Proposed Approach 1: “Criterion based on a robust likelihood obtained with the BIP τ -estimate”

This approach extends the ideas of Existing Approach 1 by estimating a robust likelihood $f(\mathbf{y}|p, q, \hat{\boldsymbol{\theta}}_\tau)$ with the BIP τ -estimator. The criterion trades off model fit, as measured by this robust likelihood, with complexity as penalized by $\alpha(p, q)$. An advantage, compared to Existing Approach 1, is that the distribution of the residuals for the BIP τ -estimator can be derived.

The form of this criterion is given by

$$\text{ICR}_{\text{LR}}^{\text{BIP}\tau}(p, q) = \underbrace{-2 \log(f(\mathbf{y}|p, q, \hat{\boldsymbol{\theta}}_\tau))}_{\text{rob. data fit}} + \underbrace{\alpha(p, q)}_{\text{model complexity penalty}} \quad (4.48)$$

and includes, e.g., criteria of the Schwarz-type

$$\text{SICR}_{\text{LR}}^{\text{BIP}\tau}(p, q) = -2 \log(f(\mathbf{y}|p, q, \hat{\boldsymbol{\theta}}_\tau)) + (p + q) \log(N) \quad (4.49)$$

and Akaike-type

$$\text{AICR}_{\text{LR}}^{\text{BIP}\tau}(p, q) = -2 \log(f(\mathbf{y}|p, q, \hat{\boldsymbol{\theta}}_\tau)) + 2(p + q) \quad (4.50)$$

by varying $\alpha(p, q)$. Here the robust likelihood is given by

$$\log(f(\mathbf{y}|p, q, \hat{\boldsymbol{\theta}}_\tau)) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))) \quad (4.51)$$

$$-\frac{1}{2} \sum_{n=p+1}^N \rho_\tau \left(\frac{Z_n^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))} \right). \quad (4.52)$$

An approximation of the posterior probability of the ARMA(p, q) model can be obtained via

$$P(\text{ARMA}(p, q)|\mathbf{y}) = \frac{P(\mathbf{y}|\text{ARMA}(p, q))P(\text{ARMA}(p, q))}{\sum_{k=0}^K \sum_{m=0}^M P(\mathbf{y}|\text{ARMA}(k, m))P(\text{ARMA}(k, m))}, \quad (4.53)$$

where $P(\mathbf{y}|\text{ARMA}(p, q)) \approx \text{SIC}_\tau(p, q)$ and $P(\text{ARMA}(p, q))$ is the prior probability of the ARMA(p, q) model.

Proposed Approach 2: “Extension of the criterion based on M-estimation to τ -estimation”

Existing Approach 2 can be extended to τ -estimation by using the equivalence given in Eq. (2.15). In this way, an RAIC-type criterion can be combined with an estimator that is more robust for ARMA parameter estimation than the M-estimator used in RAIC_M . This gives

$$\text{RAIC}_{\text{BIP}\tau}(p, q) = \sum_{n=p+1}^N \rho_\tau \left(\frac{Z_n^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))} \right) + \alpha(p, q),$$

where $\alpha(p, q)$ is given by

$$\alpha(p, q) = 2(p + q)$$

or, in accordance with Eq. (4.7.2.1)

$$\alpha(p, q) = \text{tr}(\mathcal{E}_{1,\tau}^{-1} \mathcal{E}_{2,\tau}),$$

where

$$\mathcal{E}_{1,\tau} := E_F[\psi'_\tau \left(\frac{Z_n^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))} \right) (Y_{n-1}, \dots, Y_{n-p})(Y_{n-1}, \dots, Y_{n-p})^\top]$$

and

$$\mathcal{E}_{2,\tau} := E_F[\psi_\tau^2 \left(\frac{Z_n^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau)}{\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))} \right) (Y_{n-1}, \dots, Y_{n-p})(Y_{n-1}, \dots, Y_{n-p})^\top].$$

This is the extension of the criterion defined in Eq. (2.19) to BIP τ -estimators for ARMA models. In our view, a conceptual weakness exists with this criterion: large values in the innovations estimates due to underestimation of the model order are down-weighted by the re-descending ρ_τ -function, instead of being penalized. This leads to a poor performance, even for clean data, as seen, e.g., in Table 4.3. Future research will be necessary to make this criterion practically useful.

Proposed Approach 3: “Criterion based on the τ -estimate of the BIP τ innovations scale”

Similarly to the Existing Approach 3, a robust criterion can be obtained via a robustly estimated innovations scale estimate, which yields

$$\text{IC}_{\hat{\sigma}_\tau}^{\text{BIP}\tau}(p, q) = \log(\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))) + \alpha(p, q),$$

where choosing $\alpha(p, q)$ as in Table 4.2 yields the robust extensions to existing model order selection criteria. The performance of this criterion is expected to be similar to that of Existing Approach 3. An advantage, in terms of computational cost, compared to computing the filtered- τ estimates is obtained, through deriving all ARMA candidate models from the long AR approximation.

Proposed Approach 4: “Criterion based on a weighted likelihood obtained with the BIP τ -estimate”

In this approach, we define a weighted likelihood type criterion based on the residuals obtained with the BIP τ -estimate. This adapts the ideas of Existing Approach 4, in a way that the BIP τ -estimator can be applied.

$$\text{WAIC}_{\text{BIP}\tau}(p, q) = -2 \sum_{n=1}^N w_n \log(f_n(y|p, q, \hat{\boldsymbol{\theta}}_\tau)) + 2(p + q),$$

where

$$\log(f_n(y|p, q, \hat{\boldsymbol{\theta}}_\tau)) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))) - \frac{\sum_{n=p+1}^N (Z_n^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))^2}{\hat{\sigma}_\tau^2(\mathbf{Z}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_\tau))}.$$

Details on possible choices and calculation of the weights w_n are given in [160]. The key idea is to compute weights that down-weight the influence of outlying observations on the $\text{WAIC}_\tau(p, q)$. As described in [160], weights can be obtained via Pearson residuals and a residual adjustment function that operates on Pearson residuals and allows to create different measures of data fit, such as, e.g., the squared Hellinger or the Kullback Leibler.

4.7.3 Simulations

First simulated examples for ARMA model order selection are given in this Section. Further extensive simulations, as well as theoretical analysis will be necessary to evaluate the performance of the proposed criteria. This is part of current and future work.

Example 1: Model Order Selection for AR(2)

As a first example, we evaluate the proposed and some existing model order selection criteria for the autoregressive model described in [158]. X_n has the following parameters: $\mathbf{a} = (-0.7, 0.4)^\top$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 200$. The outliers follow the isolated additive outlier model with V_n^ε (see Eq. (4.17)), as used in the simulations in this Section. In this setup, the probability of contamination ε is varied from 0 to 0.25, the outlier positions ξ_n^ε are Bernoulli distributed and the p.d.f. of $W_n, n = 1, \dots, 200$, is zero-mean Gaussian with $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. For the clean data case, $Y_n = X_n$.

Table 4.3 displays the empirical probabilities of selecting the true model order, obtained by averaging the result of 100 Monte Carlo runs. Candidate orders ranged from $p = 0, 1, \dots, 5$. The proposed criterion that is based on BIP τ -estimation and a robust likelihood $\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$ outperforms all competitors by a margin. The classical SIC, which uses ML-estimation, as well as the $\text{SIC}_{\hat{\sigma}_\tau}$ and WSIC_τ only provide high empirical probabilities of selecting the true model order for the clean data case.

MOS Criterion	$\varepsilon = 0$	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$	$\varepsilon = 0.25$
SIC	1.00	0.28	0.03	0.01	0.00	0.00
$\text{SIC}_{\hat{\sigma}_{RA}}$	0.45	0.32	0.16	0.21	0.18	0.17
$\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$	1.00	0.91	0.74	0.45	0.26	0.13
$\text{RSIC}_{\text{BIP}\tau}$	0.01	0.00	0.00	0.00	0.00	0.00
$\text{SIC}_{\hat{\sigma}_\tau}^{\text{BIP}\tau}$	0.93	0.34	0.02	0.01	0.01	0.00
$\text{WSIC}_{\text{BIP}\tau}$	0.96	0.32	0.01	0.01	0.01	0.00

Table 4.3: Empirical probabilities of selecting the true model order for the AR(2) model [158] with $\mathbf{a} = (-0.7, 0.4)^\top$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 200$ in dependence of the probability of contamination ε . The proposed criterion based on BIP τ -estimation and a robust likelihood $\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$ outperforms all competitors by a margin. Best results are highlighted by black bold font.

Example 2: Model Order Selection for ARMA(1,1)

As a second example, we evaluate the proposed and some existing model order selection criteria for the ARMA described in [123]. X_n has the following parameters: $a = 0.5$, $b = -0.5$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 1000$. The outliers follow the isolated additive outlier model with V_n^ε , see Eq. (4.17), as used in the simulations in this Section. In this setup, the probability of contamination ε is varied from 0 to 0.05, the outlier positions ξ_n^ε are Bernoulli distributed and the p.d.f. of W_n , $n = 1, \dots, 1000$ is zero-mean Gaussian with $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. For the clean data case, $Y_n = X_n$. Candidate orders ranged from $p = 0, 1, \dots, 7$ and $q = 0, 1, \dots, 7$, resulting in 64 candidate models, in total.

Table 4.4 displays the empirical probabilities of selecting the true model order, obtained by averaging the result of 100 Monte Carlo runs. Again, in case of outliers, the proposed criterion based on BIP τ -estimation and a robust likelihood $\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$ outperforms all competitors by a margin. An interesting observation is the poor performance of the criterion based on a robust τ -estimate of scale ($\text{SIC}_{\hat{\sigma}_\tau}$). In accordance with [8], we agree that the penalty term should be made dependent on the choice of the scale and the distribution of the residuals, which requires further research.

Figure 4.27 displays the empirical probabilities of selecting the candidate models with $p = 0, 1, \dots, 7$ and $q = 0, 1, \dots, 7$ for the proposed criterion based on BIP τ -estimation and a robust likelihood $\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$. Both in the clean data case (top plot) and for additive outliers that occur with a contamination probability $\varepsilon = 0.05$, the criterion performs nearly optimal and selects the true model with empirical probabilities of 0.96 and 0.92, respectively. In both cases, the probability of under-estimating the model order is equal to zero.

	Clean Data	AO($\varepsilon = 0.05$)
SIC	0.97	0.13
$\text{SIC}_{\hat{\sigma}_{RA}}$	0.15	0.04
$\text{SICR}_{LR}^{\text{BIP}\tau}$	0.96	0.92
$\text{SIC}_{\hat{\sigma}_\tau}^{\text{BIP}\tau}$	0.5	0.79
$\text{WSIC}_{\text{BIP}\tau}$	0.58	0.70

Table 4.4: Empirical probabilities of selecting the true model order for the ARMA(1,1) model [123] with $a = 0.5$, $b = -0.5$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 1000$. For the clean data case, $\varepsilon = 0$, and in the additive outlier scenario, the contamination probability $\varepsilon = 0.05$.

Figure 4.28 plots the empirical probabilities of selecting the candidate models for the classical SIC criterion, which performs nearly optimal for the clean data case (top plot), however, under-estimates the model order in presence of additive outliers. The empirical probabilities of selecting the true model are 0.97 and 0.13 respectively. The probability of under-estimating the model order in the case of additive outliers is 0.83.

Example 3: Model Order Selection for ARMA(2,1)

As a third example, we evaluate the proposed and some existing model order selection criteria for the ARMA model described in [136]. X_n has the following parameters: $\mathbf{a} = (0.39, 0.30)^\top$, $b = -0.9$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 1000$. The outliers follow the isolated additive outlier model with V_n^ε (see Eq. (4.17)), as used in the simulations in this Section. In this setup, the probability of contamination ε is varied from 0 to 0.05, the outlier positions ξ_n^ε are Bernoulli distributed and the p.d.f. of W_n , $n = 1, \dots, 1000$ is zero-mean Gaussian with $\sigma_W = \kappa\sigma_X$, where $\kappa = 5$. For the clean data case, $Y_n = X_n$. Candidate orders ranged from $p = 0, 1, \dots, 7$ and $q = 0, 1, \dots, 7$, resulting in 64 candidate models, in total.

Table 4.5 displays the empirical probabilities of selecting the true model order, obtained by averaging the results of 100 Monte Carlo runs. Again, in case of outliers, the proposed criterion based on BIP τ -estimation and a robust likelihood $\text{SICR}_{LR}^{\text{BIP}\tau}$ outperforms all competitors by a margin. For the classical SIC, in case of additive outliers, the probability of estimating the true model is zero and the probability of under-estimating the model order is 0.93.

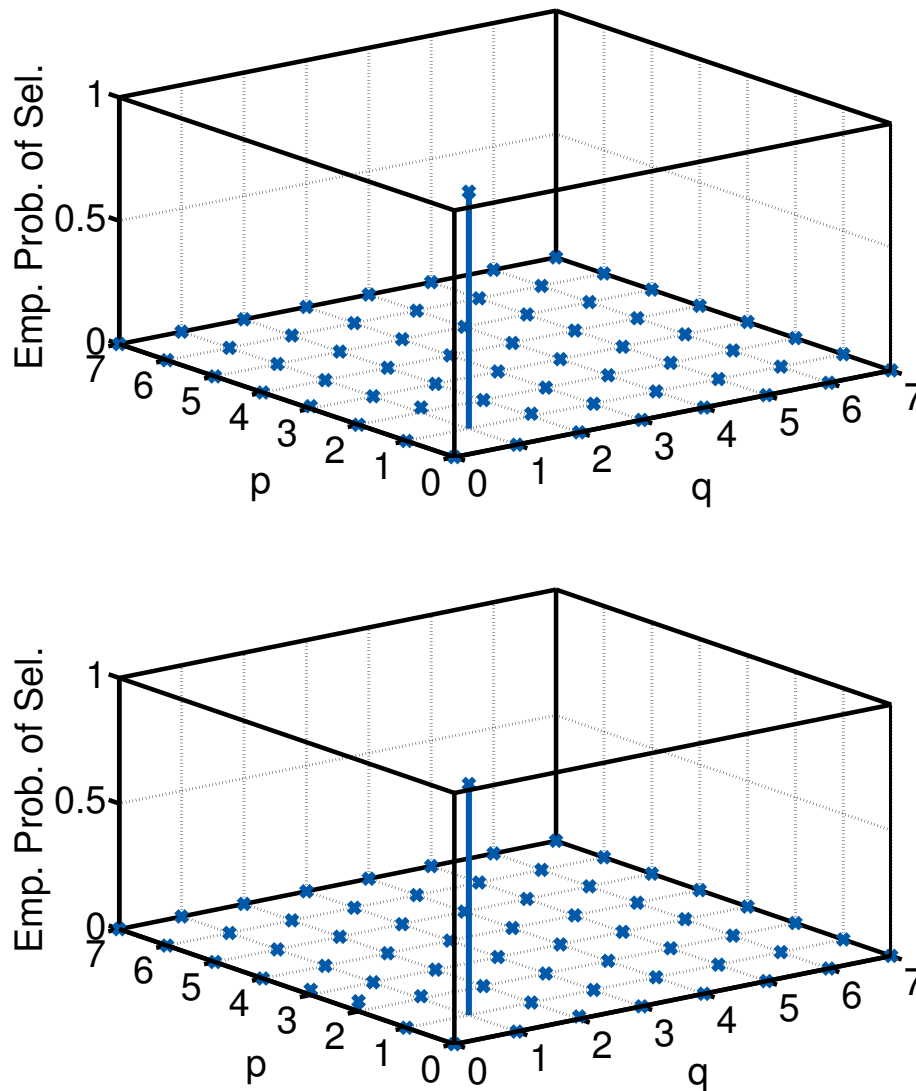


Figure 4.27: The empirical probabilities of selecting the candidate models with $p = 0, 1, \dots, 7$ and $q = 0, 1, \dots, 7$ for the proposed criterion based on BIP τ -estimation and a robust likelihood $\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$. The true model is the ARMA(1,1) model by [123]. (top) clean data case; (bottom) results in presence of additive outliers that occur with a contamination probability $\varepsilon = 0.05$.

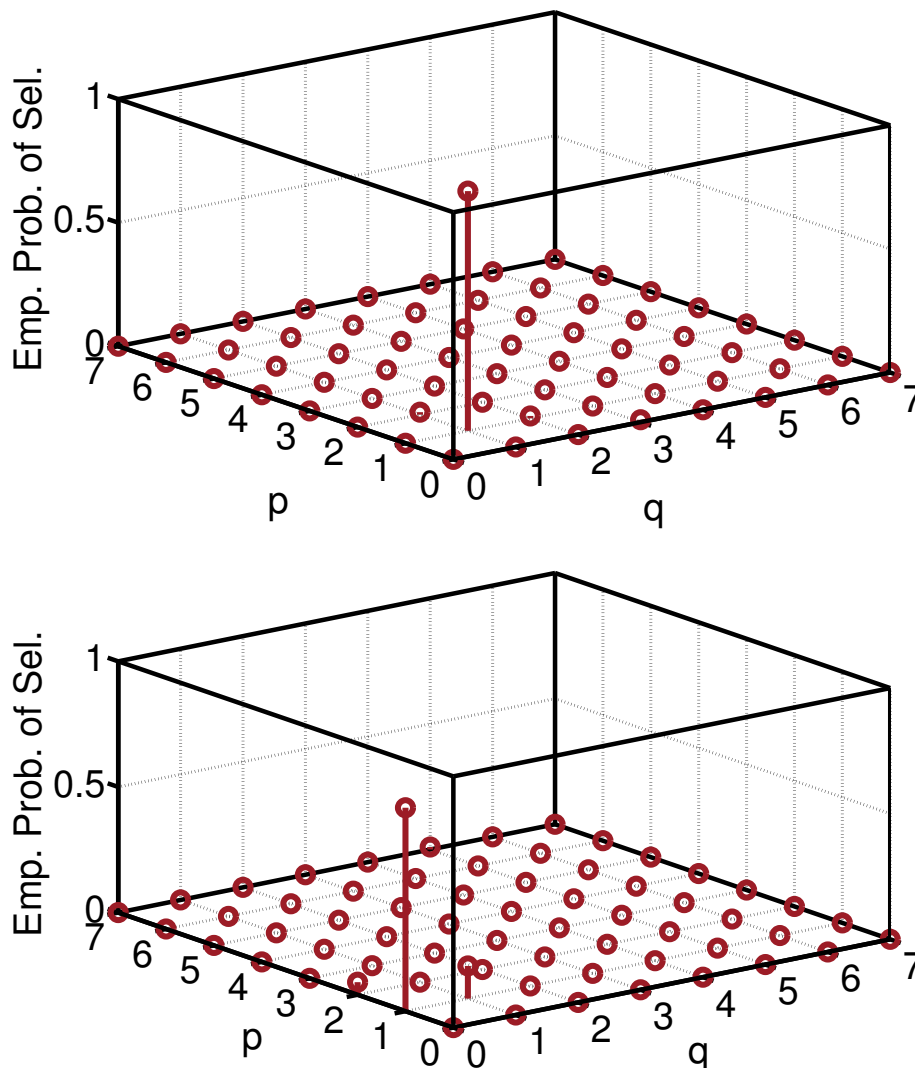


Figure 4.28: The empirical probabilities of selecting the candidate models with $p = 0, 1, \dots, 7$ and $q = 0, 1, \dots, 7$ for the classical SIC criterion. The true model is the ARMA(1,1) model by [123]. (top) clean data case; (bottom) results in presence of additive outliers that occur with a contamination probability $\varepsilon = 0.05$.

	Clean Data	AO($\varepsilon = 0.05$)
SIC	0.97	0.00
$\text{SIC}_{\hat{\sigma}_{RA}}$	0.85	0.03
$\text{SIC}_{LR}^{\text{BIP}\tau}$	0.95	0.86
$\text{SIC}_{\hat{\sigma}_\tau}^{\text{BIP}\tau}$	0.94	0.77
$\text{WSIC}_{\text{BIP}\tau}$	0.95	0.65

Table 4.5: Empirical probabilities of selecting the true model order for the ARMA(2,1) model [136] with $\mathbf{a} = (0.39, 0.30)^\top$, $b = -0.9$, $\sigma_Z = 1$, $\mu_Z = 0$, $N = 1000$. For the clean data case, $\varepsilon = 0$, and in the additive outlier scenario, the contamination probability $\varepsilon = 0.05$.

4.8 Summary and Conclusions

In this Chapter, we briefly reviewed robustness concepts and some existing estimators for dependent data. We then proposed a novel robust and efficient estimator called the BIP τ -estimator. We performed statistical robustness analysis of the proposed estimator and also established the conditions required for the estimator to be consistent. We next derived the influence function of the BIP τ -estimator for an AR(1) model contaminated by additive outliers. The boundedness and continuity of the influence function proved the qualitative robustness of the proposed estimator. We then estimated the maximum bias curve of the BIP τ -estimator and compared it to some existing robust and non-robust estimators. We also defined the quantile bias curve, which displays additional information on the robustness of the estimator for a given fraction of outlier-contamination at a finite sample size. Finally, we evaluated the breakdown point of the proposed BIP τ -estimator by means of the maximum bias curve. It was shown to possess the maximum possible breakdown point of 0.5 for the AR(1) model, which proves the quantitative robustness of the proposed estimator.

After analyzing the statistical and robustness properties, we presented a fast algorithm to robustly obtain stationary and invertible ARMA parameter estimates with the BIP τ -estimator. The estimator minimizes the robust and efficient τ -estimate of scale of the innovations estimates from the BIP τ -estimator. The key idea of this algorithm is to first compute a robust initial estimate of an autoregressive approximation from which the ARMA model parameters are derived. In this way, the ARMA model parameters are derived from the long AR approximation without further use of the outlier-contaminated observations. This is computationally very attractive, since the computational cost of the ARMA parameter estimation approximately reduces to that of computing the long AR model.

This advantage multiplies, when performing robust model order selection for ARMA models, since all the candidate models can be derived from the long AR approximation, while other robust estimators require robustly computing the estimates for all ARMA models, individually. Different robust model order selection criteria that are based on the BIP τ -estimator were derived and we have shown via extensive simulations that the most robust criterion is that based on a robust likelihood.

Future work will consider the adaptation of the penalty term in model order selection criteria based on the distribution of the residuals, which is still an open question. Furthermore, re-sampling based robust model order selection methodologies, which were not treated in this thesis, will be adapted to the BIP τ -estimator. On this topic,

we recently submitted our first contribution in robust bootstrap techniques [161]. In the area of robust stationarity testing, we have recently undertaken some preliminary steps [162]. This interesting research area is closely related to problems, such as robust spectrum estimation and outlier-cleaning, which can be addressed with the BIP τ -estimator.

The proposed BIP τ -estimator will also be applied to a real-world problem in a biomedical application. The necessary data for this evaluation has already been collected and includes, amongst others, the intracranial pressure and mean arterial pressure collected in a large clinical study. Further improvement compared to the methodology described in Section 4.3.2, as well as novel problem evaluations along the line of [163, 164] to be expected.

First research is currently being conducted on robust vector-valued AR parameter estimation [165, 166] and a major and difficult step will be to extend the BIP τ -estimator to the multi-variate setting. Recently, we have conducted a first promising step in the multi-variate analysis of non-stationary data [163]. The proposed approach is based on a vector valued robust Kalman filter. An extension of the BIP τ -estimator or other computationally feasible robust estimators to such problems seems to be possible and is currently under investigation.

Chapter 5

Summary, Conclusions and Future Work

5.1 Summary and Conclusions

This doctoral project analyzed, introduced and improved robust estimation methodologies to solve current and future signal processing problems in the areas of robustness when dealing with multi-sensor data, robust model selection and robustness for dependent data. The work was applied to solve practical signal processing problems in different areas of biomedical and array signal processing.

In particular, for uni-variate independent data, a robust criterion was presented to select the model order in a linear uni-variate regression setup with an application to corneal-height data modeling. With the proposed robust criteria, the correct radial model order of the Zernike polynomial of the corneal topography map could be selected, even if the measurement conditions for the videokeratoscopy were poor, i.e., the quality of the pre-corneal tear film was bad and the eyelid aperture was not sufficiently wide.

We then defined robust model order selection schemes for multi-dimensional data. The results presented in this thesis were based on the MM-estimator of the covariance of the r -mode unfoldings of the complex-valued data tensor. In detail, we proposed and evaluated the performance of robust R -dimensional extensions of some well-known information criteria. In the context of source enumeration, we provided simulation examples for 2-D and 3-D uniform rectangular arrays. Both in the case of Gaussian noise and for a brief sensor failure, the proposed robust multi-dimensional schemes outperformed their matrix computation based counterparts significantly.

In the context of robustness for multi-sensor data, we next investigated the task of estimating the complex-valued amplitude of sinusoidal signals in a completely unknown heavy-tailed symmetric spatially and temporally i.i.d. noise environment. A selection of non-robust and robust estimators were compared to a semi-parametric robust estimator. The performance of all robust estimators was similar to the optimal ML-estimator in the Gaussian noise case. For increasing levels of impulsiveness and contamination Huber's M-estimator was outperformed by the τ -estimator and the proposed semi-parametric robust estimator, which performed best in all impulsive noise scenarios.

A third research focus concerning multi-sensor data was that of analyzing the robustness of spatial time-frequency distribution (STFD) estimators. Here, we provided a

robustness analysis framework that is based on the influence function. The influence function of different robust and non-robust STFD matrix estimators was given and the analytical expressions for the influence functions of different types of STFD matrix estimators were provided. Although in this doctoral project, we treated mainly the robust STFDs, this analysis can be applied to any type of quadratic time-frequency distribution. The influence functions for the robust estimators were shown to be bounded and continuous, which confirms their qualitative robustness. In addition to the asymptotic analysis, we also gave a definition for the finite sample counterpart of the influence function. The simulation results for the finite sample influence function confirmed the analytical results and show the insensitivity to small departures in the distributional assumptions for the robust techniques.

A large portion of this doctoral work concerned the topic of obtaining and analyzing robust estimators in the dependent data setup. First, some practical issues concerning the detection, and robust estimation in presence of patient motion induced artifacts in biomedical measurements, were addressed. In particular, we provided an artifact-cleaning algorithm for the electrocardiogram (ECG), which is a powerful non-invasive tool containing information that helps in the diagnosis of a wide range of heart conditions. This algorithm, which works in the stationary wavelet domain, enables the monitoring of patients with portable ECG recording devices that suffer severely from patient motion induced artifacts. Since the portable devices are equipped with a transmitter in order to communicate health related information and are able to trigger alarms in case of life threatening situations, robustness is a fundamental requirement.

A second real-world problem addressed in this doctoral project was that of forecasting the intracranial pressure (ICP) levels for patients who suffered a traumatic brain injury. This enables active and early interventions for more effective control of ICP levels. We proposed a methodology which uses combined artifact detection and robust estimation after a data transformation into the empirical mode domain.

Motivated by plethora of practical applications, we then focused on deriving and analyzing sophisticated robust estimation and model selection techniques for autoregressive moving-average (ARMA) models. A fast algorithm as well as a detailed statistical and robustness analysis of a novel robust and efficient estimator were given. For the proposed estimator, which is termed the bounded influence propagation (BIP) τ -estimator, we computed a complete analysis, which included conditions for the consistency, as well as qualitative and quantitative robustness analysis by means of robustness measures, such as the influence function, the maximum bias curve and the breakdown point. The fast algorithm of the proposed estimator is based on first computing a robust initial

estimate of an autoregressive approximation from which the ARMA model parameters are derived. In this way, the ARMA model parameters are derived from the long AR approximation without further use of the outlier contaminated observations. For this reason, the estimator is very suitable and attractive for ARMA model selection purposes, since the computational cost of estimating all the candidate ARMA models, approximately reduces to that of computing one long autoregressive model. In the area of model selection for ARMA models, we proposed and compared different robust model order selection criteria that are based on the BIP τ -estimator.

5.2 Future Directions

In addition to the open research questions detailed in the end of each Chapter, we have identified future research directions and, in some cases, undertaken first steps to address them.

One major future research concerns robust bootstrap methods. The bootstrap is a powerful computational tool for statistical inference that allows for the estimation of the distribution of an estimate without distributional assumptions on the underlying data, reliance on asymptotic results or theoretical derivations. However, robustness properties of the bootstrap in the presence of outliers are very poor, irrespective of the robustness of the bootstrap estimator. This motivates the need to robustify the bootstrap procedure itself. In [161], we present first improvements of existing robust bootstrap methods, as well as a new methodology, which we call the robust starting point bootstrap (RSPB). The first application of the RSPB we investigated, is the problem of geolocation in harsh mixed line-of-sight (LOS) / non-line-of-sight (NLOS) environments. Here, we extended point estimates to distribution estimates, hence extracting additional information from a given sample, without additional data or assumptions.

Figure 5.1 shows an example of RSPB distribution estimates of two location estimates. The left plot of Figure 5.1 displays the RSPB distribution estimate (contour plot) of the MM-estimate, the MM-point estimate and the true position of the MT for an LOS scenario, while the right plot depicts an NLOS scenario with a NLOS probability of 0.4. Unlike for the classical bootstrap, the RSPB distribution estimate inherits the robustness properties of the underlying MM-estimate and is similar in terms of area to the LOS case.

The proposed improvements and the new method compare favorably to existing techniques, as measured by the maximum bias curve and empirical coverage probabilities of

the confidence intervals derived from the bootstrap, for details, see [161]. Further improvements, applications, analysis and development of new advanced robust bootstrap methods are future research topics. This is also highly relevant for bootstrap-based robust model order selection. Initially, we plan to investigate extensions of existing approaches for i.i.d. data, and then to propose novel and robust bootstrap model order selection schemes for dependent data.

A second significant research area will concern robustness for distributed signal processing. Distributed processing without a central processing unit (fusion center) [167–170] is an emerging and highly active research field. A future research direction will be the integration of robustness concepts into the framework of distributed detection and estimation. Robustness, while highly demanded by the application, is only now beginning to be considered in this young research area [171, 172]. In distributed signal processing applications, robustness is not limited to deviations from distributional statistical assumptions on the sensor noise, but also includes robustness against sensor failures which lead to erroneous, noisy or incomplete data [7, 72]. Furthermore, in a wireless system, network state information may be imprecise leading to a severe performance degradation if robustness is not properly taken into account in the algorithm design. Only integrating robustness into distributed signal processing will truly enable its application in critical real-world scenarios.

This doctoral project and resulting discussions have provided evidence, that robustness, today, is more important than ever before, and will form the basis of countless research projects to be conducted in the future.

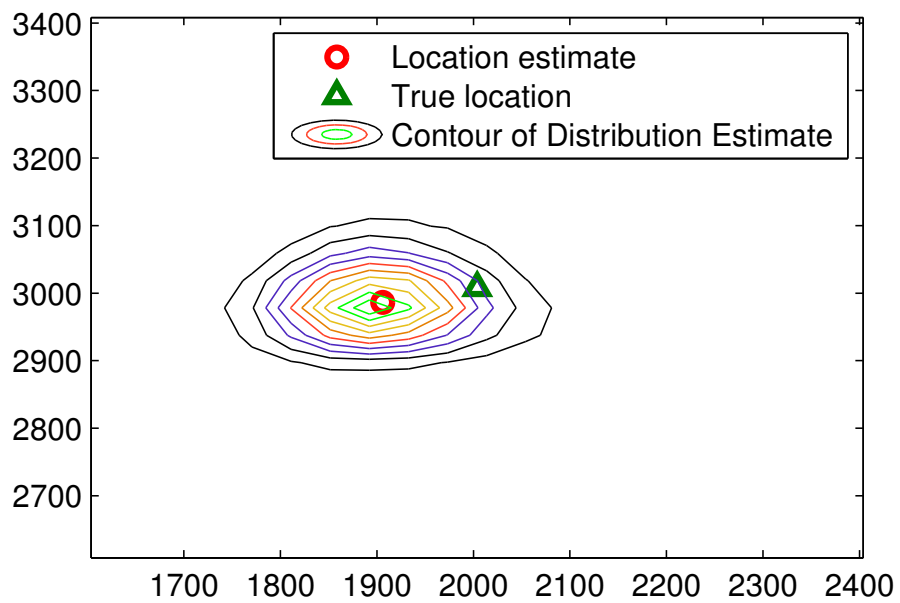
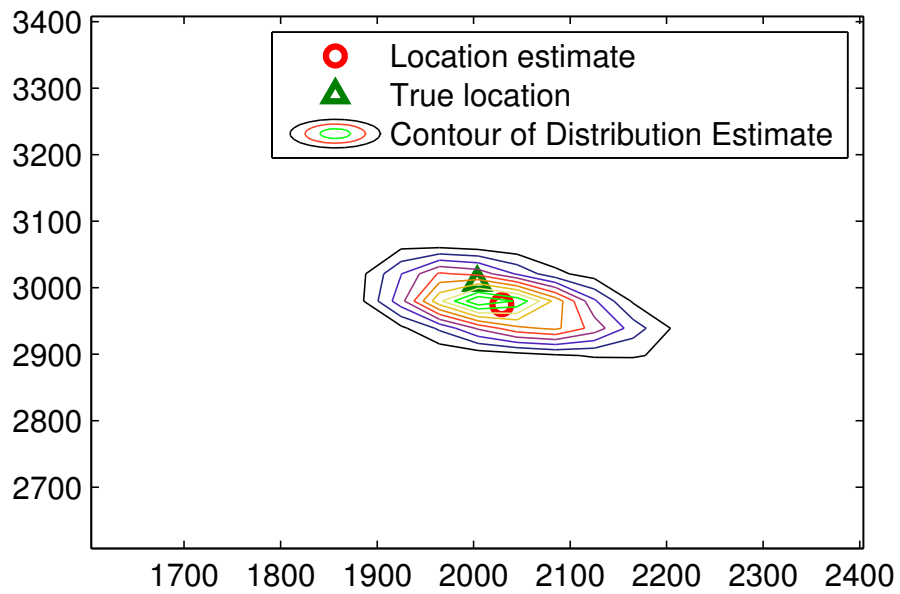


Figure 5.1: RSPB distribution estimate of a geolocation estimate for a LOS scenario and an NLOS scenario with NLOS probability of 0.5. The contour plot is plotted in 10% steps.

List of Acronyms

AIC	Akaike's information criterion
AICc	small sample bias corrected version of Akaike's information criterion
ACM	approximate conditional mean
AO	additive outlier
ARIMA	integrated autoregressive moving-average
ARMA	autoregressive moving-average
BC	bias curve
BIP	bounded influence propagation
BP	breakdown point
DOA	direction-of-arrival
ECG	electrocardiogram
EEG	electroencephalography
EIF	empirical influence function
EMD	empirical mode decomposition
FM	frequency modulated
i.i.d.	independent and identically distributed
ICP	intracranial pressure
IF	influence function
KDE	kernel density estimation
LMS	least-median-of-squares
LOS	line-of-sight
LS	least-squares
LTS	least-trimmed-squares
MAD	median absolute deviation

MBC	maximum bias curve
MC	Monte Carlo
MDL	minimum description length
MEG	magnetoencephalography
MIMO	multiple-input multiple-output
ML	maximum likelihood
MRE	median-of-ratios-estimator
MRI	magnetic resonance imaging
MSE	mean-squared-error
NLOS	non-line-of-sight
p.d.f.	probability density function
PWVD	pseudo Wigner-Ville distribution
RSPB	robust starting point bootstrap
SIC	Schwarz information criterion
SNR	signal-to-noise-ratio
STFD	spatial time-frequency distribution
SWD	stationary wavelet domain
SWT	stationary wavelet transform
ULA	uniform linear array
URA	uniform rectangular array

List of Symbols

\top	transposition
H	Hermitian transposition
-1	matrix inversion
$*$	complex conjugation
$P(\cdot)$	probability of
\mathbf{I}	identity matrix
$\mathbf{0}$	zero-vector
$a_{i,j}$	(i, j) -th element of the matrix \mathbf{A}
$\Re\{\mathbf{A}\}$	real part of a matrix \mathbf{A}
$\Im\{\mathbf{A}\}$	imaginary part of a matrix \mathbf{A}
$\text{IF}(z; \hat{\theta}, F)$	influence function of an estimator $\hat{\theta}$ at distribution F (Chapter 1)
$\hat{\theta}$	estimator/estimate of the parameter θ (Chapter 1)
F_θ	nominal distribution parametrized by θ (Chapter 1)
F_ε	contaminated distribution (Chapter 1)
ε	fraction of contamination (Chapter 1)
δ_z	point-mass probability on z (Chapter 1)
$\hat{\theta}_\infty(F)$	asymptotic values of an estimator when the data is distributed following F (Chapter 1)
$b_{\hat{\theta}}(F_\theta)$	asymptotic bias of an estimator $\hat{\theta}$ evaluated at the nominal distribution F_θ (Chapter 1)
$\mathcal{F}_{\varepsilon,\theta}$	ε -neighbourhood of distributions around the nominal distribution (Chapter 1)
\mathbf{X}_n	predictors in the linear regression model (Chapter 2)
$\boldsymbol{\theta}$	unknown (vector-valued) parameter of interest (Chapter 2)
V_n	error term in the linear regression model (Chapter 2)
(Y_n, \mathbf{X}_n^\top)	data point in the linear regression model (Chapter 2)
R_n	residual from the linear regression model (Chapter 2)
$\rho(x)$	ρ -function of an M-estimator (Chapter 2)
$\psi(x)$	ψ -function of an M-estimator (Chapter 2)
$\hat{\boldsymbol{\theta}}_M$	M-estimate of $\boldsymbol{\theta}$ (Chapter 2)
$\hat{\sigma}_M$	M-estimate of scale (Chapter 2)

$f_X(x)$	density function of the random variable X with realization x (Chapter 2)
c_{Hub}	parameter of Huber's M-estimator (Chapter 2)
c_{Tuk}	parameter of Tukey's M-estimator (Chapter 2)
$\hat{\boldsymbol{\theta}}_S$	S-estimate of $\boldsymbol{\theta}$ (Chapter 2)
$E_F[\cdot]$	expectation operation taken w.r.t. the distribution F (Chapter 2)
$\hat{\sigma}_\tau$	τ -estimate of scale (Chapter 2)
$\hat{\boldsymbol{\theta}}_\tau$	τ -estimate of $\boldsymbol{\theta}$ (Chapter 2)
$W_n(\boldsymbol{\theta})$	weighting factor of the M-estimator equivalent to a τ -estimator (Chapter 2)
$\psi_\tau(x)$	ψ -function of the M-estimator equivalent to a τ -estimator (Chapter 2)
$\rho_\tau(x)$	ρ -function of the M-estimator equivalent to a τ -estimator (Chapter 2)
$\text{AIC}(p)$	Akaike's criterion for order p (Chapter 2)
$\text{AICc}(p)$	small sample bias corrected version of Akaike's criterion for order p (Chapter 2)
$\text{SIC}(p)$	Schwarz's criterion for order p (Chapter 2)
$\text{AICR}_M^*(p)$	Shi's robust Akaike criterion for M-estimators for order p (Chapter 2)
$\text{AICCR}_M^*(p)$	small sample bias corrected version of Shi's robust Akaike criterion for M-estimators for order p (Chapter 2)
$\text{AICR}_\tau^*(p)$	extension of Shi's robust Akaike criterion for τ -estimators for order p (Chapter 2)
$\text{AICCR}_\tau^*(p)$	small sample bias corrected version of the extension of Shi's robust Akaike criterion for τ -estimators for order p (Chapter 2)
$S(r, \phi)$	corneal surface with normalized radial distance r and angle ϕ (Chapter 2)
$a_{m,n}$	Zernike coefficients with corresponding Zernike polynomials $Z_m^n(r, \phi)$ in double indexing scheme (Chapter 2)
$R_n^{ m }(r)$	radial component of Zernike polynomial (Chapter 2)
N_n^m	normalization factor of radial component of Zernike polynomial (Chapter 2)
δ_{m0}	Kronecker delta function (Chapter 2)
$R\text{-D AIC}$	R -dimensional extension of the AIC (Chapter 2)
$R\text{-D MDL}$	R -dimensional extension of the MDL (Chapter 2)
\mathcal{A}	tensor (Chapter 3)

$a_{i,j,k}$	(i, j, k) -th element of a third order tensor \mathcal{A} (Chapter 3)
$[\mathcal{A}]_{(r)} \in \mathbb{C}^{I_r \times (I_1 \dots I_{r-1} I_{r+1} \dots I_R)}$	r -mode unfolding of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_R}$ (Chapter 3)
$\mathcal{A} \times_r \mathbf{U} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times J_r \times \dots \times I_R}$	r -mode product of a tensor \mathcal{A} and matrix $\mathbf{U} \in \mathbb{C}^{J_r \times I_r}$ (Chapter 3)
\sqcup_r	concatenation operation along mode r (Chapter 3)
\hat{d}	estimated number of sources impinging onto an array (Chapter 3)
$\alpha^{(G)}$	total number of sequentially defined eigenvalues (Chapter 3)
$a^{(G)}(p)$	arithmetic mean of the smallest $p = K - k$ global eigenvalues (Chapter 3)
$g^{(G)}(p)$	geometric mean of the smallest $p = K - k$ global eigenvalues (Chapter 3)
$\hat{\mathbf{R}}_{XX}^{(r)}$	sample covariance matrix of the r -mode unfolding of the data tensor \mathcal{X} (Chapter 3)
$\ \text{IF}(\mathbf{z}, \mathbf{g}_1, F)\ $	absolute value of the influence function at the bivariate complex valued standard Gaussian distribution F for a complex-valued outlier \mathbf{z} , given in terms of the eigenvector functional \mathbf{g}_1 corresponding to the eigenvalue λ_1 (Chapter 3)
R -D AIC_{rob}	R -dimensional robust extension of the AIC (Chapter 3)
R -D MDL_{rob}	R -dimensional robust extension of the MDL (Chapter 3)
$a_{\text{rob}}^{(G)}(p)$	computed analogously to $a^{(G)}(p)$ with the difference that the eigenvalues are estimated using the robust MM-covariance matrix estimator (Chapter 3)
$g_{\text{rob}}^{(G)}(p)$	computed analogously to $g^{(G)}(p)$ with the difference that the eigenvalues are estimated using the robust MM-covariance matrix estimator (Chapter 3)
$\hat{\mathbf{R}}_{\tilde{X}\tilde{X},\text{robust}}^{(r)}$	robust covariance matrix of the r -mode unfolding of the data tensor \mathcal{X} (Chapter 3)
$\mathbf{y}_m \in \mathbb{C}^N$	observation vector at sensor m (Chapter 3)
$\mathcal{N}_c(x; \mu, \sigma^2)$	complex-valued Gaussian distribution with mean μ and variance σ^2 (Chapter 3)
\mathcal{H}_c	contamination density, which is only assumed to be symmetric, and centered around zero (Chapter 3)
ω_k	known frequency of a sinusoid (Chapter 3)
$\log f_Y(y \tilde{\boldsymbol{\theta}})$	log-likelihood function given data y parametrized by $\boldsymbol{\theta}$ (Chapter 3)
$\varphi(\cdot)$	location score function of the ML-estimator (Chapter 3)

$\hat{\varphi}(\cdot)$	estimate of the location score function of the ML-estimator (Chapter 3)
$m(v, \lambda)$	modulus transformation of v with parameter λ (Chapter 3)
$\hat{f}_W(w, h)$	estimated density of W with global bandwidth parameter h (Chapter 3)
$\mathcal{K}(\cdot)$	standard Gaussian kernel density function (Chapter 3)
ν	step size of the iterative Newton-Raphson procedure (Chapter 3)
$\mathbf{a}(\theta_k)$	$M \times 1$ array steering vector for a source impinging from direction θ_k (Chapter 3)
$s_k(t)$	source waveform of the k^{th} signal at time t (Chapter 3)
$\mu_p(t), \mu_q(t)$	time-varying means at the p^{th} and the q^{th} sensors (Chapter 3)
$\mathbf{D}_{\mathbf{x}\mathbf{x}}(t, f)$	STFD matrix of the signal $\mathbf{x}(t)$ (Chapter 3)
$\hat{D}_{x_p x_q}(t, f)$	a bilinear time-frequency distribution estimate of Cohen's class (Chapter 3)
L	window length of the PVWD (Chapter 3)
$d(t, l)$	complex-valued distance (Chapter 3)
$\bar{D}_{x_p x_q}(t, l, z_p, z_q)$	Monte Carlo estimate of $E_{F_\epsilon}[\hat{D}_{x_p x_q}(t, f)]$ (Chapter 3)
$\text{IF}(z_p, z_q; \hat{D}_{x_p x_q}(t, f), F)$	influence function of the STFD estimator $\hat{D}_{x_p x_q}(t, f)$, evaluated at the nominal distribution F for outliers z_p, z_q (Chapter 3)
$\text{EIF}(z_p, z_q; \hat{D}_{x_p x_q}(t, f), F)$	empirical influence function of the STFD estimator $\hat{D}_{x_p x_q}(t, f)$, evaluated at the nominal distribution F for outliers z_p, z_q (Chapter 3)
$A_p(z)$	polynomial representation of autoregressive model of order p (Chapter 4)
$B_q(z)$	polynomial representation of moving-average model of order q (Chapter 4)
$B^\infty(z)$	polynomial representation of moving-average model of infinite order (Chapter 4)
p_0	order of long autoregressive model that approximates the ARMA model (Chapter 4)
$\hat{c}_{XX}(\kappa)$	estimate of the covariance function of X (Chapter 4)
ξ_n^ϵ	binary series that indicates the presence (or absence) of an outlier (Chapter 4)
$\hat{\mathbf{S}}_{\text{cleaned}}$	artifact-cleaned ECG signal (Chapter 4)
$\mathbf{S}_{\text{recorded}}$	recorded ECG signal (Chapter 4)

$\hat{\mathbf{S}}_{\text{outliers}}$	additive outlier component in recorded ECG signal (Chapter 4)
\mathbf{x}	non-stationary ICP measurement (Chapter 4)
$\hat{\mathbf{S}}_R$	reconstructed ICP signal (Chapter 4)
$\hat{\mathbf{S}}_F$	ICP signal forecast (Chapter 4)
$\text{IF}(F_{X,\xi,W}^\epsilon; \hat{\boldsymbol{\theta}})$	dependent data influence functional, i.e., functional derivative at F_X (Chapter 4)
$F_{X,\xi,W}^\epsilon$	joint distribution of the processes X_n , ξ_n and W_n (Chapter 4)
$\hat{\mathbf{X}}_{n n}$	state vector from state recursion of the robust filter (Chapter 4)
Y_n^f	estimate of X_n obtained with a robust filter, given contaminated observations $\mathbf{Y}_n = (Y_n, Y_{n-1}, \dots, Y_{n-p+1})^\top$ (Chapter 4)
Σ_n^1	first column of the prediction error covariance matrix Σ_n (Chapter 4)
\mathbf{P}_n	filtering error covariance matrix (Chapter 4)
$W(x) = \frac{\psi(x)}{x}$	weighting function of the filter (Chapter 4)
$\hat{Y}_{n n-1}$	robust one step ahead prediction of Y_n (Chapter 4)
$\mathbf{Z}^f(\boldsymbol{\theta})$	filtered innovation estimate (Chapter 4)
Median BC(ε)	median bias curve, represents the MBC obtained in 50 % of the cases as a function of ε (Chapter 4)
$Q_x\text{BC}(\varepsilon)$	quantile bias curve, represents the MBC obtained in x % of the cases as a function of ε (Chapter 4)
$\mathbf{Z}^{\text{BIP}}(\boldsymbol{\theta})$	BIP-ARMA innovations series (Chapter 4)
$\mathbf{Z}^{\text{BIP}(m)}(\zeta)$	memory- m BIP-AR(m) innovations estimate (Chapter 4)
$\hat{a}_{m,m}$	memory- m BIP-AR(m) parameter estimate (Chapter 4)
ζ_0	coarse search grid for the grid search to minimize the τ -estimate of scale (Chapter 4)
ζ_{fine}	fine search grid for the grid search to minimize the τ -estimate of scale (Chapter 4)
$\hat{\boldsymbol{\theta}}_\tau^{(\text{AR})}$	initial “long autoregressive” estimator of the ARMA parameters, given the long autoregressive approximation (Chapter 4)
$\hat{\boldsymbol{\theta}}_\tau^{(\text{MA})}$	initial “long moving-average” estimator of the ARMA parameters, given the long autoregressive approximation (Chapter 4)

$\hat{\boldsymbol{\theta}}_{\tau}^{(\text{CO})}$	initial “long covariance” estimator of the ARMA parameters, given the long autoregressive approximation (Chapter 4)
$\hat{\boldsymbol{\theta}}_{\tau}^{(\text{RI})}$	initial “long inverse correlation” estimator of the ARMA parameters, given the long autoregressive approximation (Chapter 4)
$\hat{\boldsymbol{\theta}}_{\tau}$	final BIP τ -estimate of the ARMA parameters (Chapter 4)
$\mathbf{Z}_{\tau}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau})$	final BIP τ -estimate of the innovations series (Chapter 4)
$\hat{\sigma}_{\tau}(\mathbf{Z}_{\tau}^{\text{BIP}}(\hat{\boldsymbol{\theta}}_{\tau}))$	final BIP τ -estimate of the scale of the innovations series (Chapter 4)
$\hat{\sigma}_{ML}^2(\mathbf{Z}(\hat{\boldsymbol{\theta}}_{ML}))$	ML-estimate of the scale of the innovations series (Chapter 4)
$\hat{C}_{XX}^{\text{BIP-ARMA}}(e^{j\omega}, \hat{\boldsymbol{\theta}}_{\tau})$	BIP τ spectrum estimate based on the robustly estimated ARMA-parameters (Chapter 4)
$\alpha(p, q, N)$	penalty term of information criteria, as a function of the model orders p, q and the data length N (Chapter 4)
RAIC	AR model order selection criterion based on M-estimation (Chapter 4)
$\log(f(\mathbf{y} p, \hat{\boldsymbol{\theta}}^f))$	robust log-likelihood of an AR(p) model obtained with a robustly-filtered estimate (Chapter 4)
SIC^f	Schwarz-type information criterion based on robustly-filtered-estimate (Chapter 4)
AIC^f	Akaike-type information criterion based on robustly-filtered-estimate (Chapter 4)
$\text{AIC}^{f\tau}$	Akaike-type information criterion based on the τ -estimate of scale obtained from a robustly-filtered τ -estimate (Chapter 4)
WAIC	information criterion based on a weighted likelihood (Chapter 4)
$f(\mathbf{y} p, q, \hat{\boldsymbol{\theta}}_{\tau})$	robust likelihood for an ARMA(p, q) model obtained with the BIP τ -estimator (Chapter 4)
$\text{ICR}_{\text{LR}}^{\text{BIP}\tau}$	information criterion obtained from estimating a robust likelihood with the BIP τ -estimator (Chapter 4)
$\text{SICR}_{\text{LR}}^{\text{BIP}\tau}$	Schwarz-type information criterion obtained from estimating a robust likelihood with the BIP τ -estimator (Chapter 4)

$\text{AICR}_{\text{LR}}^{\text{BIP}\tau}$	Akaike-type information criterion obtained from estimating a robust likelihood with the BIP τ -estimator (Chapter 4)
$P(\text{ARMA}(p, q) \mathbf{y})$	approximation of the posterior probability of the $\text{ARMA}(p, q)$ model (Chapter 4)
$\text{RAIC}_{\text{BIP}\tau}$	extension of the Akaike-type model order selection criterion based on M-estimation to τ -estimation (Chapter 4)
$\text{RSIC}_{\text{BIP}\tau}$	extension of the Schwarz-type model order selection criterion based on M-estimation to τ -estimation (Chapter 4)
$\text{IC}_{\hat{\sigma}_\tau}^{\text{BIP}\tau}$	information criterion based on the τ -estimate of scale obtained from a BIP τ -estimate (Chapter 4)
$\text{SIC}_{\hat{\sigma}_\tau}^{\text{BIP}\tau}$	Schwarz-type information criterion based on the τ -estimate of scale obtained from a BIP τ -estimate (Chapter 4)
$\text{WAIC}_{\text{BIP}\tau}$	weighted likelihood Akaike-type criterion based on the residuals obtained with the BIP τ -estimate (Chapter 4)
$\text{WSIC}_{\text{BIP}\tau}$	weighted likelihood Schwarz-type criterion based on the residuals obtained with the BIP τ -estimate (Chapter 4)
$\text{SIC}_{\hat{\sigma}_{RA}}$	information criterion based on the normalized MAD-estimate of scale of the innovations estimate obtained from a ranks-based-estimate (Chapter 4)

Bibliography

- [1] S. M. Stigler, “Simon Newcomb, Percy Daniell and the history of robust estimation 1885-1920,” *J. Am. Stat. Assoc.*, vol. 68, no. 344, pp. 872–879, 1973.
- [2] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, vol. 2, John Wiley & Sons, Inc., Publication, 2009.
- [3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics, The Approach Based On Influence Functions*, John Wiley & Sons, 1985.
- [4] S. A. Kassam and V. Poor, “Robust techniques for signal processing: a survey,” *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, Mar 1985.
- [5] S. A. Kassam., *Signal detection in non-Gaussian noise*, Springer Texts in Electrical Engineering. Springer-Verlag, New York,, 1988.
- [6] H. V. Poor., *An introduction to signal detection and estimation*, Springer Verlag, 1994.
- [7] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, “Robust estimation in signal processing: a tutorial-style treatment of fundamental concepts,” *IEEE Signal Proc. Magazine*, vol. 29, no. 4, pp. 61–80, 2012.
- [8] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics, Theory and Methods*, John Wiley & Sons, Ltd, 2006.
- [9] V.J. Yohai and R.H. Zamar, “High breakdown-point estimates of regression by means of the minimization of an efficient scale,” *J. Amer. Statist. Assoc.*, vol. 83, no. 402, pp. 406–413, 1988.
- [10] M. Salibian-Barrera, G. Willems, and R. Zamar, “The fast τ -estimator for regression.,” *J. Comput. Graph. Stat.*, vol. 17, no. 3, pp. 659–682, 2008.
- [11] E. Ronchetti, “Robust model selection in regression,” *Statistics & Probability Letters*, vol. 3, no. 1, pp. 21–23, 1985.
- [12] P. Shi and C.L. Tsai, “A note on the unification of the Akaike information criterion.,” *J. Royal Statist. Soc. B*, vol. 60, no. 3, pp. 551–558, 1998.
- [13] C. M. Hurvich and C. L. Tsai, “Regression and time series model selection in small samples,” *Biometrika*, vol. 76, no. 2, pp. 297, 1989.
- [14] M. Muma and A. M. Zoubir, “Robust model selection for corneal height data based on τ -estimation,” in *In Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP) 2011 in Prague, Czech Republic*, May 2011, pp. 4096–4099.
- [15] J. Schwiegerling, J.E. Greivenkamp, and J.M. Miller, “Representation of videokeratoscopic height data with Zernike polynomials.,” *J. Opt. Soc. Amer., A*, vol. 12, no. 10, pp. 2105–2113, 1995.

- [16] D. R. Iskander, M. J. Collins, and B. Davis, "Optimal modeling of corneal surfaces with Zernike Polynomials," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 87–95, Jan. 2001.
- [17] W. Alkhalidi, D. R. Iskander, and A. M. Zoubir, "Model-order selection in Zernike polynomial expansion of corneal surfaces using the efficient detection criterion.," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2429 – 2437, Oct. 2010.
- [18] A. Schnall, "Robust model order selection for corneal height data," Studienarbeit, Technische Universität Darmstadt, Oct. 2011.
- [19] X. Wang and H. V. Poor, "Robust multiuser detection in non-Gaussian channels," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 289–305, Feb. 1999.
- [20] A. M. Zoubir and R. Breich, "Multiuser detection in heavy tailed noise," *Digit. Signal Process.: Rev. J.*, vol. 12, no. 2-3, pp. 262–273, Apr.-Jul. 2002.
- [21] T. A. Kumar and K. D. Rau, "A new M-estimator based robust multiuser detection in flat-fading non-Gaussian channels," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1908–1913, Jul. 2009.
- [22] U. Hammes, Wolsztynski E., and A. M. Zoubir, "Robust tracking and geolocation for wireless networks in NLOS environments," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 889–901, Oct. 2009.
- [23] I. Guvenc and C. C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Commun. Surveys Tutorials*, vol. 11, no. 3, pp. 107–124, Mar. 2009.
- [24] J. C. Prieto, C. Croux, and A. R. Jiménez, "RoPEUS: a new robust algorithm for static positioning in ultrasonic systems," *Sensors*, vol. 9, no. 6, pp. 4211–4229, 2009.
- [25] C. V. Stewart, "Robust parameter estimation in computer vision," *Siam Review*, vol. 41, no. 3, pp. 513–537, 1999.
- [26] M. Ye, R. M. Haralick, and L. G. Shapiro, "Estimating piecewise-smooth optical flow with global matching and graduated optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1625–1630, Dec. 2003.
- [27] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 41, no. 5, pp. 349–358, May 2002.
- [28] L. Thomas and L. Mili, "A robust GM-estimator for the automated detection of external defects on barked hardwood logs and stems," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3568–3576, Jul 2007.
- [29] C. G. Bénar, D. Schön, S. Grimault, B. Nazarian, B. Burle, Roth M., J. M. Badier, P. Marquis, C-. Liegeois-Chauvel, and J. L. Anton, "Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI," *Human Brain Mapp.*, vol. 28, no. 7, pp. 602–613, 2007.

- [30] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964.
- [31] P. J. Rousseeuw, "Least median of squares regression," *J. Am. Statist. Assoc.*, vol. 79, pp. 871–880, 1984.
- [32] P. J. Rousseeuw and V. J. Yohai, "Robust regression by means of S-estimators," *Lect. Notes Stat.*, vol. 26, pp. 256–272, 1984.
- [33] O. Hössjer, "On the optimality of S-estimators," *Stat. Prob. Letters*, vol. 14, pp. 413–419, 1992.
- [34] H. Linhart and W. Zucchini, *Model Selection*, John Wiley and Sons Inc., 1986.
- [35] A. D. R. McQuarrie and C.-L. Tsai, *Regression and Time Series Model Selection*, World Scientific Publishing Co. Pte. Ltd., 1998.
- [36] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Proc. Magazine*, vol. 21, no. 4, pp. 36–47, Mar. 2004.
- [37] H. Akaike, "Statistical predictor identification," *Ann. Math. Statist.*, vol. 22, pp. 203–217, 1969.
- [38] C. L. Mallows, "Some comments on Cp," *Technom.*, vol. 15, pp. 661–675, 1973.
- [39] H. Akaike, "Information theory and an extension of the maximum likelihood principle," *Int. Sympos. Inf. Theory*, vol. 2, pp. 267–281, 1973.
- [40] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
- [41] H. Akaike, "A Bayesian analysis of the minimum AIC procedure," *Ann. Math. Statist.*, vol. 30, pp. 9–14, 1978.
- [42] G. Schwarz, "Estimating the dimensions of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.
- [43] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
- [44] E. J. Hannan and Quinn B. G., "The determination of the order of an autoregression," *J. Royal Statist. Soc. B*, vol. 41, pp. 190–195, 1978.
- [45] R. Shibata, "Asymptotic efficient selection of the order of the model for estimating the parameters of a linear process," *Ann. Statist.*, vol. 8, pp. 147–164, 1980.
- [46] C. M. Hurvich and C. L. Tsai, "Model selection for least absolute deviations regression in small samples," *Statist. Probab. Lett.*, vol. 9, pp. 259–265, 1990.
- [47] J. A. F. Machado, "Robust model selection and M-estimation," *Econ. Theory*, vol. 9, pp. 478–493, 1993.

- [48] E. Ronchetti and R. G. Staudte, “Robust version of Mallows’s C_p ,” *Amer. Stat. Assoc.*, vol. 89, no. 426, pp. 550–559, Jun. 1994.
- [49] S. Sommer and R. G. Staudte, “Robust variable selection in regression in the presence of outliers and leverage points,” *Austral. Stat.*, vol. 37, pp. 323–336, 1995.
- [50] S. Sommer and R. M. Huggins, “Variables selection using the wald test and a robust c_p ,” *Appl. Statist.*, vol. 45, pp. 15–29, 1996.
- [51] G. Qian and H. R. Künsch, “On model selection in robust linear regression,” *Research Report, ETH Zürich*, vol. 80, pp. 1–33, 1996.
- [52] G. Qian and H. R. Künsch, “On model selection via stochastic complexity in robust linear regression,” *Stat. Plann. Inference*, vol. 75, no. 1, pp. 91–116, Nov. 1998.
- [53] C. Agostinelli, “Robust model selection in regression via weighted likelihood methodology,” *Stat. Prob. Lett.*, vol. 56, pp. 289–300, 2002.
- [54] S. Kay, “Conditional model order estimation,” *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1910–1917, Sept. 2001.
- [55] M. Muma, D.R. Iskander, and M.J. Collins, “The role of cardiopulmonary signals in the dynamics of eye’s wavefront aberrations,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 373 – 383, Feb. 2010.
- [56] L. N. Thibos, R. A. Applegate, Schwiegerling J. T., and R. Webb, “Standards for reporting the optical aberrations of eyes,” *J. Refractive Surgery*, vol. 18, pp. 652–660, 2002.
- [57] D. R. Iskander, W. Alkhaldi, and A. M. Zoubir, “On the computer intensive methods in model selection,” in *In Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP) 2008 in Las Vegas, NV, USA*, 2008, pp. 3461–3464.
- [58] U. Hammes, E. Wolsztynski, and A. M. Zoubir, “Transformation based robust semiparametric estimation,” *IEEE Signal Process. Lett.*, vol. 15, pp. 845–848, 2008.
- [59] M. Muma, Y. Cheng, Roemer F., M. Haardt, and Zoubir A. M., “Robust source number enumeration for R-dimensional arrays in case of brief sensor failures,” in *In Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP) 2012 in Kyoto, Japan.*, Mar. 2012, pp. 3709–3712.
- [60] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.
- [61] Z. Lu and A.M. Zoubir, “Flexible detection criterion for source enumeration in array processing,” *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1303–1314, Mar. 2013.

- [62] P. Chen, T.-J. Wu, and J. Yang, "A comparative study of model selection criteria for the number of signals," *IET Radar Sonar Navig.*, vol. 2, no. 3, pp. 180–188, Jun. 2008.
- [63] Q.-T. Zhang, K. M. Wong, P.C. Yip, and Reilly J.P., "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.
- [64] E. Fishler, M. Grossmann, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027–1036, May 2002.
- [65] J. P. C. L. da Costa, F. Roemer, M. Haardt, and R. T. de Sousa Jr., "Multi-dimensional model order selection," *EURASIP J Adv Signal Process.*, vol. 1, no. 1, pp. 1–26, 2011.
- [66] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: new methods and results for class A and class B noise models," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999.
- [67] Y. I. Abramovich and P. Turcaj, "Impulsive noise mitigation in spatial and temporal domains for surface-wave over-the-horizon radar," *DTIC Document, Cooperative Research Centre for Sensor Signal and Information Processing, Mawson Lakes, Australia*, vol. 1, pp. 19–24, 1999.
- [68] P. C. Etter, *Underwater Acoustic Modeling and Simulation*, Taylor & Francis, New York, 2003.
- [69] K. L. Blackard, T.-S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sept 1993.
- [70] T. K. Blankenship, D. M. Kriztman, and T. S. Rappaport, "Measurements and simulation of radio frequency impulsive noise in hospitals and clinics," in *Proc. IEEE 47th Vehicular Technology Conf.*, 1997, vol. 3, pp. 1942–1946.
- [71] D. C. Alexander, G. J. Barker, and S. R. Arridge, "Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data," *Magn. Reson. Med.*, vol. 48, no. 2, pp. 331–340, 2002.
- [72] Y. Zhang, N. Meratnia, and Havinga P., "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surv. & Tut.*, vol. 12, no. 2, pp. 159–170, 2010.
- [73] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J Matrix Anal. A.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [74] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," in *IEEE Int Workshop Statist. Signal Processing (SSP2003)*, 2003, pp. 462–425.

- [75] H. P. Lopuhaa, "Multivariate τ -estimator for location and scatter," *Canad. J. Statist.*, vol. 19, pp. 307–321, 1991.
- [76] H. P. Lopuhaa, "Highly efficient estimators of multivariate location with high breakdown point," *Ann. Statist.*, vol. 20, pp. 398–413, 1992.
- [77] V. J. Yohai, "High breakdown-point and high efficiency estimates for regression," *Ann. Statist.*, vol. 15, pp. 642–656, 1987.
- [78] K. S. Tatsuoaka and D. E. Tyler, "On the uniqueness of S-functionals and constrained M-functionals under non-elliptical distributions," *Ann. Statist.*, vol. 28, no. 1, pp. 1219–1243, 2000.
- [79] M. Salibian-Barrera, S. Van Aelst, and G. Willems, "Principal components analysis based on multivariate MM estimators with fast and robust bootstrap," *J. Am. Statist. Assoc.*, vol. 101, no. 475, pp. 1198–1211, 2006.
- [80] M. Muma, U. Hammes, and A. M. Zoubir, "Robust semiparametric amplitude estimation of sinusoidal signals: the multi-sensor case," in *Proc. 3rd IEEE Int. Workshop Comp. Adv. Multi-Sensor Adapt. Proces. (CAMSAP 09) Aruba, Dutch Antilles 2009*, 2009.
- [81] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: survey, new results, and an application," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 338–352, Feb. 2000.
- [82] P. J. Bickel, C. A. J. Klaassen, Ritov Y. R., and Wellner J. A., *Efficient and Adaptive Estimation for Semiparametric Models*, Springer, New York, 1993.
- [83] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, U.K., 1986.
- [84] M. P. Wand, J. S. Marron, and D. Ruppert, "Transformations in density estimation," *J. Amer. Statist. Assoc.*, vol. 86, no. 414, pp. 343–353, 1991.
- [85] J. A. John and N. R. Draper, "An alternative family of transformations," *Appl. Statist.*, vol. 29, pp. 190–197, 1980.
- [86] R. J. Carrol, "A robust method for testing transformations to achieve approximate normality," *J. R. Statist. Soc., Series B*, vol. 42, no. 1, pp. 71–78, 1980.
- [87] W. Sharif, M. Muma, and A. M. Zoubir, "Robustness analysis of spatial time-frequency distributions based on the influence function," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1958–1971, Aug. 2013.
- [88] B. Boashash, *Time-frequency signal analysis and processing - a comprehensive reference*, Elsevier, 2003.
- [89] A. M. Zoubir and D. R. Iskander, "Bootstrap modeling of a class of nonstationary signals," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 399–408, Feb. 2000.

- [90] Y. Zhang and M. G. Amin, "Spatial averaging of time-frequency distributions for signal recovery in uniform linear arrays," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2892–2902, Oct. 2000.
- [91] Y. Zhang, W. Mu, and M. G. Amin, "Subspace analysis of spatial time-frequency distribution matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 747–759, Apr. 2001.
- [92] Y. Zhang, W. Mu, and M. G. Amin, "Maximum likelihood methods for array processing based on time-frequency distributions," *Advanced signal processing algorithms, architectures, and implementations IX*, vol. 1, no. 1, pp. 502–513, 1999.
- [93] P. Heidenreich, L. A. Cirillo, and A. M. Zoubir, "Morphological image processing for FM source detection and localization," *Signal Process.*, vol. 89, no. 6, pp. 1070–1080, Jun. 2009.
- [94] H. Wiese, H. Claussen, and J. Rosca, "Particle filter based DOA estimation for multiple source tracking (MUST)," in *Proc. 45th Asilomar Conference on Signals, Systems and Computers*, 2011, pp. 624–628.
- [95] W. Sharif, Y. Chakhchoukh, and A. M. Zoubir, "Robust spatial time-frequency distribution matrix estimation with application to direction-of-arrival estimation," *Signal Process.*, vol. 91, no. 11, pp. 2630–2638, Nov. 2011.
- [96] M. G. Amin and Y. Zhang, "Direction finding based on spatial time-frequency distribution matrices," *Digital Signal Process.*, vol. 10, no. 4, pp. 325–359, Apr. 2000.
- [97] W. Sharif, P. Heidenreich, and A.M. Zoubir, "Robust direction-of-arrival estimation for FM sources in the presence of impulsive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 3662–3665.
- [98] W. Sharif, Y. Chakhchoukh, and A. M. Zoubir, "Direction-of-arrival estimation of FM sources based on robust spatial time-frequency distribution matrices," in *In Proc. IEEE Workshop Statist. Signal Proces. (SSP 2011)*, 2011, pp. 537–540.
- [99] Y. Zhang and M. G. Amin, "MIMO radar for direction finding with exploitation of time-frequency representations," in *In Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 2760–2763.
- [100] Y. Zhang, M.G. Amin, and B. Himed, "Joint DOD/DOA estimation in MIMO radar exploiting time-frequency signal representations," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 102–112, 2012.
- [101] R. Mengqi and X. Z. Yue, "A novel multiple sparse source localization using triangular pyramid microphone array," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 83–86, 2012.
- [102] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2888–2897, Nov. 1998.

-
- [103] L. A. Cirillo, A. M. Zoubir, and M. G. Amin, "Parameter estimation for locally linear FM signals using a time-frequency Hough transform," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4162–4175, Sept. 2008.
- [104] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.
- [105] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir, "Blind separation of nonstationary sources," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 605–608, Jul. 2004.
- [106] L. A. Cirillo, A. M. Zoubir, and M. G. Amin, "Blind source separation in the time-frequency domain based on multiple hypothesis testing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2267–2279, Jun. 2008.
- [107] X. Dong, H. Weng, D. G. Beetner, T. H. Hubing, D. C. Wunsch, M. Noll, H. Goksu, and B. Moss, "Detection and identification of vehicles based on their unintended electromagnetic emissions," *IEEE Trans. Electromagn. Compat.*, vol. 48, no. 4, pp. 752–759, Apr. 2006.
- [108] M. Djeddi and M. Benidir, "Robust polynomial Wigner-Ville distribution for the analysis of polynomial phase signals in α -stable noise," in *In Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2004)*, 2004.
- [109] B. Barkat and L. Stankovic, "Analysis of polynomial FM signals corrupted by heavy-tailed noise," *Signal Proces.*, vol. 84, pp. 69–75, 2004.
- [110] M. Sahmoudi and K. Abed-Meraim, "A robust time-frequency distribution for multicomponent non-stationary FM signals analysis in impulsive alpha-stable noise," in *Proc. IEEE Workshop on Statist. Signal Proces. (SSP 2005)*, 2005.
- [111] I. Djurovic, L. Stankovic, and J. F. Böhme, "Robust L-estimation based forms of signal transforms and time-frequency representations," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1753–1761, Jul. 2003.
- [112] N. Zaric, I. Orovic, and S. Stankovic, "Robust time-frequency distributions with complex-lag argument," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–10, 2010.
- [113] C. I. Ioana, Y. Zhang, M. G. Amin, F. Ahmad, and B. Himed, "Time-frequency analysis of multipath Doppler signatures of maneuvering targets," in *In Proc. IEEE Int. Conf. Acoust., Speech Signal Process*, 2012, pp. 3397–3400.
- [114] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications," *Proc. IEEE*, vol. 80, no. 4, pp. 540–568, 1992.
- [115] Leon Cohen, *Time-Frequency Analysis*, Prentice-Hall Signal Processing. Pearson Education POD, 1 edition, Jan. 1995.

- [116] R. D. Martin and D. J. Thomson, "Robust-resistant spectrum estimation," *Proc. IEEE*, vol. 70, no. 9, pp. 1097–1115, 1982.
- [117] O. H. Bustos and V. J. Yohai, "Robust estimates for ARMA models," *J. Am. Statist. Assoc.*, vol. 81, no. 393, pp. 155–168, 1986.
- [118] Y. Yang, H. He, and G. Xu, "Adaptively robust filtering for kinematic geodetic positioning," *J. Geodesy*, vol. 75, no. 2, pp. 109–116, 2001.
- [119] T. C. Aysal and K. E. Barner, "Meridian filtering for robust signal processing," *IEEE Trans. S*, vol. 55, no. 8, pp. 3349–3962, Aug. 2007.
- [120] B. Spangl and R. Dutter, "Estimating spectral density functions robustly," *REVSTAT-Statst. J.*, vol. 5, no. 1, pp. 41–61, 2007.
- [121] B. Andrews, "Rank-based estimation for autoregressive moving average time series models," *J. Time Ser. Anal.*, vol. 29, no. 1, pp. 51–73, 2008.
- [122] K. Liang, X. Wang, and T. H. Li, "Robust discovery of periodically expressed genes using the Laplace periodogram," *BMC Bioinform.*, vol. 10, no. 1, pp. 1–15, 2009.
- [123] N. Muler, D. Peña, and V. J. Yohai, "Robust estimation for ARMA models," *Ann. Statist.*, vol. 37, no. 2, pp. 816–840, 2009.
- [124] Y. Chakhchoukh, P. Panciatici, and P. Bondon, "Robust estimation of SARIMA models: Application to short-term load forecasting," in *In Proc. IEEE Workshop Statist. Signal Proces. (SSP 2009)*, Cardiff, UK, Aug. 2009.
- [125] H. Dong, Z. Wang, and H. Gao, "Robust H_∞ filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 1957–1966, Apr. 2010.
- [126] T. H. Li, "A nonlinear method for robust spectral analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2466–2474, May 2010.
- [127] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.
- [128] Y. Chakhchoukh, *Contribution à l'estimation robuste des modèles SARIMA*, Ph.D. thesis, L'Université Paris-Sud 11, 2010.
- [129] Y. Chakhchoukh, P. Panciatici, and L. Mili, "Electric load forecasting based on statistical robust methods," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 982–991, Mar. 2010.
- [130] F. Strasser, M. Muma, and A. M. Zoubir, "Motion artifact removal in ECG signals using multi-resolution thresholding," in *In Proc. European Signal Processing Conference (EUSIPCO) 2012 in Bucharest, Romania*, Aug. 2012, pp. 899–903.

- [131] B. Han, M. Muma, M. Feng, and A. M. Zoubir, "An online approach for intracranial pressure forecasting based on signal decomposition and robust statistics," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2013)*, in Vancouver, Canada, May 2013, pp. 6239–6243.
- [132] Muma. M. and A. M. Zoubir, "A new robust estimation and model order selection method for ARMA processes," *to be submitted to IEEE Trans. Signal Proces.*, 2013.
- [133] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*, Wiley Series in Probability and Statistics, 2011.
- [134] P. Stoica and R. L. Moses, *Spectral analysis of signals*, New Jersey: Prentice hall, 2005.
- [135] J. E. Davidson, "Problems with the estimation of moving average processes," *J. Econom.*, vol. 16, no. 3, pp. 295–310, 1981.
- [136] P. M. T. Broersen, *Automatic autocorrelation and spectral analysis*, Springer-Verlag London Limited, 2006.
- [137] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, vol. 46, no. 3/4, pp. 306–316, 1959.
- [138] J. Durbin, "The fitting of time-series models.," *Revue Inst. Int. de Stat.*, vol. 28, no. 3, pp. 233244, 1960.
- [139] Krause D. J. Graupe, D. and J. Moore, "Identification of autoregressive moving-average parameters of time series," *IEEE Trans. Autom. Control*, vol. 20, no. 1, pp. 104–107, 1975.
- [140] M. B. Priestley, *Spectral analysis and time series*, London: Academic Press, 1981.
- [141] P. M. Broersen and S. de Waele, "Automatic identification of time-series models from long autoregressive models," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 5, pp. 1862–1868, 2005.
- [142] R. A. Davis, K. Knight, and J. Liu, "M-estimation for autoregressions with infinite variance," *Stoch. Proc. Appl.*, vol. 40, no. 1, pp. 145–180, 1992.
- [143] S.G. Mallat, *A Wavelet Tour of Signal Processing*, Elsevier Academic Press, 3 edition, 2009.
- [144] M. Feng and Z. Zhang, "iSyNCC: An intelligent system for patient monitoring and clinical decision support in neuro-critical-care," in *In Proc. IEEE Ann. Int. Conf. Eng. Medicine Biol. Soc. (EMBC 2011)*, Jul. 2011, pp. 6426–6429.
- [145] Feng, L. Y. Loy, F. Zhang, and C. Guan, "Artifact removal for intracranial pressure monitoring signals: A robust solution with signal decomposition," in *In Proc. IEEE Ann. Int. Conf. Eng. Medicine Biol. Soc. (EMBC 2011)*, 2011, pp. 7111–7114.

- [146] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis,” *Proc. R. Soc. A*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
- [147] R. D. Martin and V. J. Yohai, “Influence functionals for time series,” *Ann. Statist.*, vol. 14, no. 3, pp. 781–818, 1986.
- [148] H. Künsch, “Infinitesimal robustness for autoregressive processes,” *Ann. Statist.*, vol. 12, no. 3, pp. 843–863, 1984.
- [149] M. G. Genton and A. Lucas, “Comprehensive definitions of breakdown points for independent and dependent observations,” *J. Royal Statist. Soc. B*, vol. 65, no. 1, pp. 81–94, 2003.
- [150] R. H. Jones, “Maximum likelihood fitting of ARMA models to time series with missing observations,” *Technometrics*, vol. 22, no. 3, pp. 389–395, 1980.
- [151] C. Masreliez, “Approximate non-Gaussian filtering with linear state and observation relations,” *IEEE Trans. Autom. Control*, vol. 20, no. 1, pp. 107–110, 1975.
- [152] P. M. T. Broersen, “Autoregressive model orders for Durbin’s MA and ARMA estimators,” *IEEE Trans. Signal Proces.*, vol. 48, no. 8, pp. 2454–2457, Aug. 2000.
- [153] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*, Springer, 2006.
- [154] V. Katkovnik, “Robust M-periodogram,” *IEEE Trans. Signal Proces.*, vol. 46, no. 11, pp. 3104–3109, 1998.
- [155] Abraham B. Gould A. & Robinson L. (1985). de Gooijer, J. G., “Methods for determining the order of an autoregressive-moving average process: A survey,” *Int. Stat. Review*, vol. 53, no. 3, pp. 301–329, 1985.
- [156] R. D. Martin, “Robust estimation of autoregressive models,” *Directions in Time Series*, vol. 1, pp. 228–262, 1980.
- [157] J. Behrens, *Robuste Ordnungswahl für autoregressive Prozesse*, Ph.D. thesis, University of Kaiserslautern, Germany, 1991.
- [158] Raftery A. E. Le, N. D. and R. D. Martin, “Robust Bayesian model selection for autoregressive processes with additive outliers,” *J. Amer. Stat. Assoc.*, vol. 91, no. 433, pp. 123–131, 1996.
- [159] E. Ronchetti, “Robustness aspects of model choice,” *Statistica Sinica*, vol. 7, pp. 327–338, 1997.
- [160] C. Agostinelli, “Robust Akaike information criterion for ARMA models,” *Rendiconti per gli Studi Economici Quantitativi*, vol. 1, pp. 1–14, 2004.

- [161] S. Vlaski, M. Muma, and A. M. Zoubir, “Robust bootstrap methods with an application to geolocation in harsh LOS/NLOS environments,” in *Submitted to Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP) 2014 in Florence, Italy*, 2014.
- [162] J. Dagdagan, M. Muma, and A. M. Zoubir, “Robust testing for stationarity in the presence of outliers,” in *Submitted to Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP) 2014 in Florence, Italy*, 2014.
- [163] T. Schäck, M. Muma, T. Kelava, M. Schmidt, Feng. M., and A. M. Zoubir, “Robust causality analysis of non-stationary multivariate time series,” *to be submitted to IEEE Trans. Biomed. Eng.*, 2014.
- [164] A. Kelava, M. Muma, M. Schmidt, and A. M. Zoubir, “A new approach for quantifying the coherence of multivariate non-stationary data with an application to psychophysiological measures during emotion regulation,” *under revision in Psychometrica*, 2013.
- [165] C. Croux and Joossens K., “Robust estimation of the vector autoregressive model by a least trimmed squares procedure,” in *In Proc. Comp. Statist. (COMPSTAT 2008)*, P. Brito, Ed., 2008, pp. 489–501.
- [166] E. Martinez, N. Muler, and V. J. Yohai, “Robust estimates for vector autoregressive models,” *Comp. Statist. Data Anal.*, vol. 65, no. 1, pp. 68–79, Sept. 2013.
- [167] S. Barbarossa and G. Scutari, “Bio-inspired sensor network design,” *IEEE Signal Proc. Magazine*, vol. 24, no. 3, pp. 26–35, 2007.
- [168] F.S. Cattivelli and A.H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, 2010.
- [169] S.-Y. Tu and A.H. Sayed, “Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, 2012.
- [170] P. Di Lorenzo and S. Barbarossa, “Decentralized estimation and control of algebraic connectivity of random ad-hoc networks,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2013)*, in Vancouver, Canada, 2013, pp. 4474–4478.
- [171] S. Al-Sayed, A.M. Zoubir, and A.H. Sayed, “An optimal error nonlinearity for robust adaptation against impulsive noise,” in *In Proc. 14th IEEE Workshop Signal Process. Adv. Wireless Comm. (SPAWC)*, Darmstadt, Germany., 2013, pp. 410–414.
- [172] V. Ugrinovskii, “Distributed robust estimation over randomly switching networks using consensus,” *Automatica*, vol. 49, no. 1, pp. 160–168, 2013.

Curriculum Vitae

Name: Michael Muma
 Date of birth: 28.04.1981
 Place of birth: Mannheim (Germany)
 Family status: Married, one child

Education

09/2003 - 04/2009 Technische Universität Darmstadt (Darmstadt, Germany)
 Electrical Engineering and Information Technology,
 Telecommunications
 Diplom-Ingenieur (Dipl.-Ing)

04/2009 Diploma Thesis: “A Time-Frequency Coherence Analysis of the Role of Cardiopulmonary Signals in the Dynamics of Eye’s Wavefront Aberrations”
 This research was done at the School of Optometry, Queensland University of Technology in Brisbane, Australia.

08/2008 Student Thesis: “Eyelid Detection in Videokeratographic Images”

06/2000 High school degree (Abitur) at Goethe Gymnasium, Bensheim, Germany

Work experience

since 04/2009 Research Associate at Signal Processing Group
 Technische Universität Darmstadt

06/2003 - 07/2003 Employed at Sirona The Dental Company, Bensheim, Germany as a Mechatronics Engineer in the Department for the Assembly of Dental CAD/CAM-Systems

09/2001 - 06/2003 Apprenticeship as Mechatronics Engineer at Sirona The Dental Company, Bensheim, Germany.

09/2000 - 08/2001 Civilian Service at AWO Nursing Home, Bensheim, Germany.

Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

