



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

SUBSPACE-BASED EXPONENTIAL DATA FITTING USING LINEAR AND MULTILINEAR ALGEBRA

Promotor:
Prof. dr. ir. Sabine Van Huffel
Prof. dr. ir. Lieven De Lathauwer
Prof. dr. ir. Martine Wevers

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurwetenschappen
door
Jean-Michel PAPY

October 2005



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

SUBSPACE-BASED EXPONENTIAL DATA FITTING USING LINEAR AND MULTILINEAR ALGEBRA

Jury:

Prof. dr. ir. Herman Neuckermans, voorzitter
Prof. dr. ir. Sabine Van Huffel, promotor
Prof. dr. ir. Lieven De Lathauwer, promotor
Prof. dr. ir. Martine Wevers, promotor
Prof. dr. ir. Marc Moonen
Prof. dr. ir. Jan Engelen
Prof. dr. ir. Rodolphe Sepulchre (ULG Liège)
Prof. dr. ir. Karim Abed-Meraim (ENST Paris)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen
door

Jean-Michel PAPY

U.D.C. 519.6

U.D.C. 681.3*G12

U.D.C. 681.3*G13

October 2005

© Katholieke Universiteit Leuven – Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

D/2005/7515/82

ISBN 90-5682-648-4

Acknowledgement

The presented research study was carried out at the Department of Electrical Engineering (ESAT) of the Katholieke Universiteit Leuven and with support of the Department of Metallurgy and Material Engineering (MTM).

During my Ph.D. years, I have been privileged to be surrounded by competent, enthusiastic and supportive people. This thesis would not have been possible without their help and support.

First of all, I wish to express my sincere gratitude to Professor Sabine Van Huffel. She guided me throughout these five years from the moment on she accepted me as a research assistant, and introduced me into the world of subspace-based signal processing. She was always available, listening enthusiastically to our ideas and carefully to our difficulties, encouraging us, providing us with new ideas and congratulating our progress. I have been touched by her human qualities.

My thanks go to Professor Martine Wevers for her very helpful advices, support and guidance all along the PDT-Coil project. I have benefited a lot from her insights in the area of Material Engineering.

I am particularly grateful to Professor Lieven De Lathauwer who introduced me into the wonderland of multilinear algebra and without whom this work would not have been possible. He spared no effort to help me to solve numerous mathematical and numerical problems and to answer my countless questions about matrices and tensors. I also appreciated his way of seeing certain "things of life".

I want to acknowledge Professor Marc Moonen for his valuable comments on this thesis. I would like to thank him for having kindly driven me to the signal processing workshop PRORISC in Veldhoven.

I wish to acknowledge Professor Jan Engelen. The discussions I had with him helped me to understand laser sources and optical fibers technologies. I also thank him for having improved my XHTML skills and my English-French

translation skills.

I feel honoured that Rodophe Sepulchre from the Université Libre de Liège (ULG) and Karim Abed-Meraim from the Ecole Nationale Supérieure des Télécommunications (ENST) have accepted to be member of the Jury of this thesis.

I don't want to forget the help I received from Simon Doclo and Geert Rombouts who introduced me into the world of SVD-based and QR-based filtering and deepened my insights.

My thanks go also to Professor Rick Vaccaro from the University of Rhode Island (URI) for his wonderful lecture series in subspace-based signal processing.

I'm very grateful to my colleagues and friends Laurent Rippert (MTM) and Steve Vandenplas (MTM) for all the discussion we had about optical fibers.

I would like to mention the wonderful work of the systems managers Stef, Rik, Franck, Piet and Marc. Special thanks go to Stef for helping me to customize my Linux environment, for providing me with nice music and for the interesting discussions about music (and belgian beers).

I wish to thank Dr. Ir. Philippe Lemmerling for his precious help with the total least squares methods and also for the entertaining running sessions.

Furthermore, I thank the EU for financing my Ph.D. research via the PDT-coil project and its extension.

Hearty thanks are sent out to my colleagues in the Biomedical signal processing group for supplying an enjoyable work atmosphere. I especially want to acknowledge my roommates for the last five years, Chuan Lu, Kristiaan Pelckmans, Lieven De Lathauwer, Andy Devos, Qizheng Sheng, Pieter Pels, Wim Declercq, Lieveke Ameye, Geert Morren and Ben Van Calster. I don't forget my colleagues Patrick Glenisson and Bart Hamers, also my mentors in Aikibudo.

To my friends I met in Belgium, Anne, Olivier, Bernard, Cyril, Fabrice, Jacqueline, William, Yann, David, Steve, Laurent, Ghislaine, Pascal, Virginie, I send out a big thank you for their helping hand when needed and the nice moment we shared together. To my friends in France, Sebastien, Geraldine, Fabrice, Matthieu, Isabelle, Laurent, Nordine, Fabien, Michel, Marilou, I send my sincere thanks for their continuing friendship and support during my stay in Belgium, despite the long distance and my erratic communications. To my great regret, Fabrice passed away. See you friend ! I'm also very grateful to Linda and her family who provided me with an unconditional support.

There are no words to express my gratitude towards my parents for all the sacrifices and selfless effort they made. Robert, H el ene, Eliane, c'est grandement

grâce à vous que j'y suis arrivé. Merci ! Un énorme merci du fond du coeur à ma grand-mère, Denise, pour son sourire, sa générosité et sa gaieté de coeur.

And last but not least, Nadège. I met her some months ago and she has already changed my life. Thank you, Nadège, for your love, your patience and your understanding.

Abstract

The exponentially damped sinusoidal (EDS) model arises in numerous signal processing applications. It is therefore of great interest to have methods able to estimate the parameters of such a model in the single-channel as well as in the multi-channel case. Because such a model naturally lends itself to subspace representation, powerful matrix approaches like HTLS in the single-channel case, HTLSstack in the multi-channel case and HTLSDstack in the decimative case have been developed to estimate the parameters of the underlying EDS model. They basically consist in stacking the signal in Hankel (single-channel) or block Hankel (multi-channel) data matrices. Then, the signal subspace is estimated by means of the singular value decomposition (SVD). The parameters of the model, namely the amplitudes, the phases, the damping factors, and the frequencies, are estimated from this subspace. Note that the sample covariance matrix counterpart is called TLS-ESPRIT, multi-channel TLS-ESPRIT and decimative TLS-ESPRIT.

In these methods, the order of the model (i.e. the number of damped sinusoids) is assumed to be known. A variety of methods for estimating the model order exists. The recently developed method ESTER has been shown to outperform the existing Information Theoretic Criteria (ITC) based techniques. ESTER relies on the shift invariance property of the signal subspace. We propose an easy-to-implement SVD-based method which also exploits the same shift invariance property and outperforms the method ESTER.

As far as multi-channel signals are concerned, it may be of great interest to extract only the common sinusoids. This may be for instance the case in Electroencephalogram (EEG) monitoring or material health monitoring. So far, only techniques which extract the common damped sinusoids in the two-channel case have been described. We propose a flexible and accurate method that can be applied to an arbitrary number of channels.

The last part of the thesis deals with multilinear algebra, which is the algebra of higher-order tensors. Higher-order tensors can be seen as higher dimensional tables than can be addressed with more than two indices. First, we show

that the matrix approaches do not exploit all the structure which is present in the theoretical decomposition. This is especially true in the multi-channel and the decimative case. In a second step we demonstrate that a higher-order representation of the problem may help to take this structure into account. We derive the higher-order counterparts of the HTLS, HTLSstack and HTLSDstack methods for estimating the parameters of an EDS model, and show by means of a higher-order dimensionality reduction algorithm that the estimation of the signal subspace, and hence the parameters of the EDS model, may be more accurate than the one obtained via the matrix approaches.

Glossary

Mathematical Notation

a, b, \dots	scalars a, b, \dots
A, B, \dots	vectors A, B, \dots
$\mathbf{A}, \mathbf{B}, \dots$	matrices $\mathbf{A}, \mathbf{B}, \dots$
$\mathcal{A}, \mathcal{B}, \dots$	tensors $\mathcal{A}, \mathcal{B}, \dots$
A_i	i -th column vector of matrix \mathbf{A}
$a_{i_1 i_2 \dots i_N}$	entry on the tensor \mathcal{A} corresponding to the i_1 -th value of the mode-1 index, the i_2 -th value of the mode-2 index, \dots , and the i_N -th value of the mode- N index
$(M)_{i,j}, m_{i,j}$	element on the i -th row and the j -th column of matrix \mathbf{M}
M^\top	transpose of matrix \mathbf{M}
M^*	complex conjugate of matrix \mathbf{M}
$M^H = (M^*)^\top$	Hermitian transpose of matrix \mathbf{M}
M^{-1}	inverse of matrix \mathbf{M}
M^\dagger	pseudo-inverse of matrix \mathbf{M}
$\det \mathbf{M}$	determinant of matrix \mathbf{M}
$\text{diag}\{\mathbf{v}\}$	square diagonal matrix with vector \mathbf{v} as diagonal
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrix \mathbf{A} and \mathbf{B}
$\mathcal{A} \times_m \mathbf{U}$	m -mode product of a tensor \mathcal{A} by a matrix \mathbf{U}
$U \circ V$	outer product of vectors U and V
\mathbb{N}	the set of natural numbers
\mathbb{Z}	the set of integer numbers
\mathbb{R}	the set of real numbers
\mathbb{R}^+	the set of positive real numbers
\mathbb{C}	the set of complex numbers
\mathbb{R}^M	set of real M -dimensional vectors
\mathbb{C}^M	set of complex M -dimensional vectors
$\Re\{x\}$	real part of $x \in \mathbb{C}$
$\Im\{x\}$	imaginary part of $x \in \mathbb{C}$
a^*	complex conjugate of a

\mathbf{A}^*	complex conjugate of vector \mathbf{A}
$\text{conj}(\cdot)$	complex conjugation
\hat{x}	estimate of x
$ \cdot $	absolute value
$\ \cdot\ _2$	2-norm
$\ \cdot\ _F$	Frobenius norm
$\mathcal{E}\{\cdot\}$	expectation operator
$a \ll b$	a is much smaller than b
$a \gg b$	a is much larger than b
$a \approx b$	a is approximately equal to b

Fixed Symbols

$\exp\{\cdot\}$	exponential function
\mathbf{H}	Hankel matrix
\mathbf{U}, \mathbf{V}	singular vector matrices
N	number of samples
M	number of rows
L	number of columns
K	number of signal poles
g	greatest common divisor
n	discrete-time index
f	frequency-domain variable
$\omega = 2\pi f$	pulsation
z	z -domain variable
ν_s	sampling frequency
$\mathbf{0}$	zero vector or zero matrix
$\mathbf{0}_{M \times N}$	$M \times N$ zero matrix
\mathbf{I}_N	$N \times N$ identity matrix
j	$\sqrt{-1}$

Acronyms and Abbreviations

ALS	Alternating Least Squares
cf.	confer : compare with
DFT	Discrete Fourier Transform
EEG	Electroencephalography
e.g.	exempli gratia : for example
EDS	Exponentially Damped Sinusoids (e.g a sum of EDS)
EDS	Exponentially Damped Sinusoidal

	(e.g. the EDS model)
Eq.	equation
FFT	Fast Fourier Transform
Fig.	figure
FIR	Finite Impulse Response filter
GSVD	Generalized Singular Value Decomposition
HOSVD	Higher Order Singular Value Decomposition
HOVDMD	Higher Order Vandermonde Decomposition
i.e.	id est : that is
iff	if and only if
IFFT	Inverse Fast Fourier Transform
i.i.d	independent and identically distributed
IIR	Infinite Impulse Response filter
LMS	Least Mean Square adaptive filter
LS	Least Squares
MRS	Magnetic Resonance Spectroscopy
MRI	Magnetic Resonance Imaging
RQI	Rayleigh Quotient Iteration
MSE	Mean Square Error
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
TLS	Total Least Squares
VDMD	Vandermonde Decomposition
vs.	versus
WGN	white Gaussian noise
w.r.t.	with respect to

Contents

Acknowledgement	i
Abstract	v
Glossary	vii
Contents	xi
1 Introduction	1
1.1 Motivation and preliminary studies	1
1.2 The Exponentially Damped Sinusoidal (EDS) model	4
1.3 Purpose of the thesis	10
1.4 Chapter-by-chapter overview	10
2 Exponential data fitting	13
2.1 Introduction	13
2.2 Hankel total least squares method: HTLS	16
2.2.1 Subspace Representation	16
2.2.2 Shift-invariance property and harmonic retrieval via Total Least Squares Solution	17
2.3 Hankel Total Least Squares method for multi-channel signal: HTLSstack	20

2.4	Harmonic retrieval in the decimative case: HTLSDstack	22
2.5	Conclusion	26
3	A shift invariance-based order selection technique for the EDS model	27
3.1	Introduction	27
3.2	Subspace shift-invariance property and least squares approach .	28
3.3	SVD-based approach	29
3.3.1	Overestimation of the model order	30
3.3.2	Underestimation of the model order	32
3.3.3	Order determination	33
3.4	Results	35
3.5	Application examples	37
3.5.1	Application to dominant spindle frequency detection . .	38
3.5.2	Application to material health monitoring	38
3.6	Conclusions	41
4	Common pole estimation	43
4.1	Introduction	43
4.2	Problem statement	45
4.3	Determining a set of vectors spanning the common subspace . .	45
4.3.1	Discussion and algorithm	48
4.3.2	Complexity	49
4.4	Application to simulated signals	50
4.4.1	A five-channel example	51
4.4.2	A two-channel example	55
4.5	Application to EEG monitoring	59
4.6	Application to material health monitoring	63

<i>Contents</i>	xiii
4.7 Conclusion	64
5 A Short Introduction to Multilinear Algebra	67
5.1 Higher-order tensor	67
5.2 Basic definitions	68
5.2.1 Multiplication of a higher-order tensor by a matrix . . .	68
5.2.2 n -mode vectors and n -mode vector space	69
5.2.3 Matrix representation of a higher-order tensor	69
5.2.4 Scalar product, orthogonality and Frobenius-norm . . .	72
5.2.5 Rank properties of a higher-order tensor	72
5.3 The higher-order singular value decomposition	73
5.4 Best rank- (R_1, R_2, \dots, R_N) approximation	75
5.5 Summary	78
6 Exponential data fitting using multilinear algebra: the multi-channel case	79
6.1 Introduction	79
6.2 Tensor-based algorithms for harmonic retrieval	80
6.2.1 Single-channel signals	80
6.2.2 Multi-channel signals	84
6.3 Results	88
6.4 Conclusions	95
7 Exponential data fitting using multilinear algebra: the decimative case	97
7.1 Introduction	97
7.2 Higher-order Vandermonde structure of decimated data sequences	98
7.3 Pole extraction	101
7.4 Some comments on n -rank reduction	102

7.5	Results	105
7.5.1	Two-peak, undamped signal	105
7.5.2	Two-peak, damped signal	107
7.6	Conclusions	107
8	Conclusions and Further Research	111
8.1	Conclusions	111
8.2	Further research.	113
	Appendices	115
I	On-line detection method for transient waves applied to continuous health monitoring of carbon fiber reinforced polymer composites with embedded optical fibers	115
I.1	Introduction and Context	115
I.2	Signal characteristics and representation	117
I.3	Transient detection	123
I.3.1	SNR improvement	124
I.3.2	Noise level estimation	125
I.3.3	Spectral subtraction estimator – signal reconstruction	126
I.4	Results and discussion	127
I.5	Conclusions	133
	Bibliography	135
	List of publications	145
	Curriculum Vitae	147

Chapter 1

Introduction

1.1 Motivation and preliminary studies

This thesis originally started from the need of solving some signal processing problems related to structural health monitoring. The starting point was the EU PDT-Coil Project¹ in which the department of Electrical Engineering (ESAT) and the department of Material Sciences (MTM) of the Katholieke Universiteit Leuven were involved. In this project the problem to be solved was the following.

The larger hydrocarbon reservoirs in Europe are rapidly depleting. The remaining marginal fields can only be exploited commercially by the implementation of new intelligent technology, such as electric Coiled Tubing drilling or Intelligent Well Completions. Steel Coiled Tubing (CT) with an internal electrical wireline is the current standard for such operations. Steel CT suffers from corrosion and fatigue problems, which dramatically restrict the operational life. The horizontal reach of steel CT is limited due to its heavy weight. The inserted wireline results in major hydraulic power losses and is cumbersome to install.

The PDT-Coil project proposal aimed to solve these problems by investigating and developing a high-temperature, corrosion and fatigue resistant thermoplastic Intelligent Composite Coiled Tubing for electric drilling applications. This power and data transmission composite coiled tubing (PDT-CCT or PDT-Coil) contains embedded electrical power and data conductors and a fiber-optic sensing and monitoring system. Compared to conventional rotary drilling, such a

¹EU project number: ENK6/2000/00074, duration: 01.09.2000–28.02.2003 (phase 1). EU project number: NNE5/2001/887, duration: 01.01.2003–31.08.2005 (phase 2)

PDT-COIL will result in 45% less required workspace, 50% reduction of rig height, 60% noise reduction, 70% less rock cuttings and a power exhaust generation reduction of 85%. The PDT-COIL will result in 35% overall cost savings on the exploration of hydrocarbons. Time to first oil will be reduced due to the support of enhanced logging while drilling capabilities.

Therefore part of the work was to develop a reliable and sufficiently sensitive fiber-optic sensing and monitoring system.

It can be noticed that the industrial applications in aeronautics, automotive, etc. are numerous since more and more composite materials are used. Indeed, the long term behavior of composite materials, subject to environmental, thermal and/or mechanical loading is still not very well understood. This often leads industry to set high safety margins or to over-design. The advantage of composite materials is to be found in their low density and so high specific strength and stiffness. Hence, it is of major interest to study their resistance to damage. The embedment of optical fibers into composite materials and the use of intelligent data processing of the optical fiber signals enables the integration of a non-destructive testing (NDT) system into these complex materials.

For the PDT-Coil project, two intensity-modulated optical fiber sensors were investigated: one using multi-mode optical fibers and recently, one using a single-mode optical fiber. It was decided to use the polarimetric system working with single-mode optical fibers, which was much more sensitive than the previously developed multi-mode setup. Figure 1.1 shows a typical four point bending setup which aims to damage a composite tube and to measure damage signals. Our main task was related to damage detection. A set of propagating high-energy high-frequency acoustic waves is generated whenever permanent damages occurs in the composite material. This is roughly illustrated in Figure 1.2. These waves locally modify the optical and geometrical properties of the optical fiber. Consequently, the transfer function of the optical fiber is modified in a periodic way which results in the modulation of the output light intensity. In order to detect these transient signals, also called events, we developed fast advanced signal processing methods [60, 61].

Such a sensing optical fiber, embedded in a composite or other material, is a promising alternative for piezoelectric transducers used for acoustic emission monitoring. The latter are usually attached to the surface and are therefore most sensitive to surface waves. In contrast, the embedded optical fibers sense easily bulk waves. Some investigations have proven that the signal sensed by the piezoelectric transducers and by the optical fibers can be modelled by a sum of exponentially damped sinusoids (EDS) [67]. This idea brought us to the following kernel problem: are there any common components between the output signals from the piezoelectric transducers and the optical fibers? Is there a way to accurately catch these common components? For this purpose an efficient algorithm has been designed for a two-channel setup [90]. In Chapter 4,

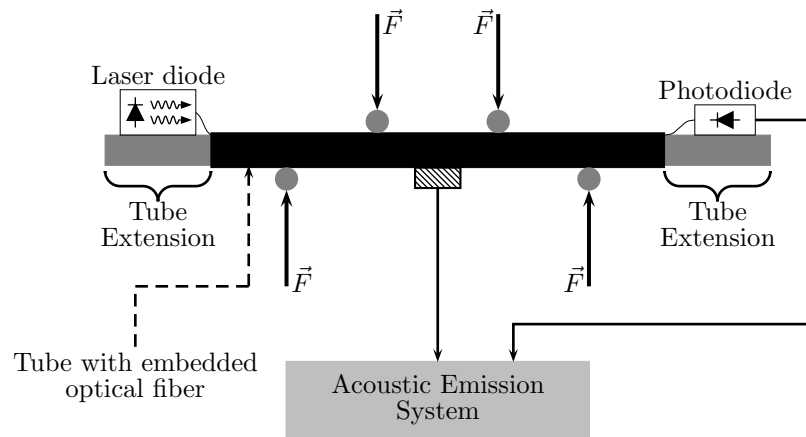


Figure 1.1: Four point bending test setup. The force applied will make the tube bend. Whenever damage occurs, transient acoustic waves can be detected by the optical fiber.

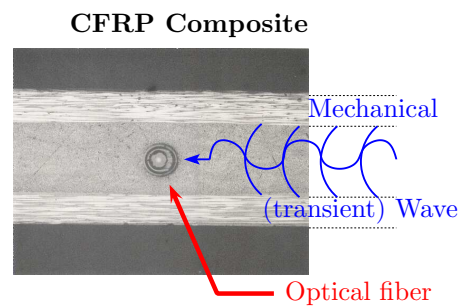


Figure 1.2: This image represents the cross-section of a composite material with an embedded optical fiber. It aims to illustrate what happens in case damage is occurring: a set of propagating transient elastic waves are released from the damage sites and will hit the optical fiber, thereby modifying its properties.

this algorithm has been further extended to an arbitrary number of channels.

1.2 The Exponentially Damped Sinusoidal (EDS) model

The main model which is dealt with in this thesis is the EDS model [80] because it is quite simple, and ubiquitous in numerous signal processing applications. We mainly focus on the, more mathematically tractable, complex exponential model which is the complex generalization of the EDS model. Most of the time this model is linked to physical phenomena that are driven by linear second order differential equations. A very simple example is, for instance, the propagation of time-harmonic uniform plane waves in a lossless unbounded medium. If we limit our attention to the case for which the electric field \mathcal{E} (V.m^{-1}) has only an x component and is only a function of the direction z , the solution to the time-harmonic Maxwell's equations [30] is of the form:

$$\mathcal{E}_x(z, t) = C_1 \cos(\omega t - \beta z) + C_2 \cos(\omega t + \beta z)$$

where ω is the angular frequency (rad.s^{-1}), β is the wave number (rad.m^{-1}), C_1 and C_2 are constants determined by boundary conditions. More generally, if these waves propagate in a lossy media, the electric field can be written as:

$$\mathcal{E}_x(z, t) = C_1 \exp\{-\alpha z\} \cos(\omega t - \beta z) + C_2 \exp\{+\alpha z\} \cos(\omega t + \beta z)$$

where α is called the attenuation constant. Often the solutions are used in complex form:

$$\mathbf{E}_x(z, t) = C_1 \exp\{-\alpha z + i(\omega t - \beta z)\} + C_2 \exp\{\alpha z + i(\omega t + \beta z)\}$$

where $i = \sqrt{-1}$, which is mathematically more tractable. The first term represents the propagation in the $+z$ direction and the second one the propagation in the $-z$ direction. In practical applications one might consider only one of these two terms. Moreover, more complex shapes can be considered like in direction-of-arrival (DOA) estimation where the electric field of a propagating wave impinging on a sensor ($z = 0$), is of the form [89]:

$$E(t) = C(t) \exp\{i(\omega t + \varphi(t))\}$$

where $\varphi(t)$ is the overall phase at time t and $C(t)$ is the shape of the wave. We assume that $C(t)$ and $\varphi(t)$ are slowly varying functions of time. This is called the narrow band assumption. Another well known phenomenon is the propagation of acoustic plane waves. The propagation of the pressure disturbance, in the 1-dimensional case (z direction) is given by:

$$\Delta P = \Delta P_0 \exp\{i\omega t - kz + \varphi\}.$$

where ΔP is the pressure disturbance (Pa) and k is called the medium wavenumber (rad.m^{-1}). Note that generally a medium has dispersion properties and therefore ΔP is exponentially damped.

In the previous section we introduced the concept of sensing propagating waves in composite materials by optical fibers. In material engineering these acoustic waves are typically detected by very simple methods (i.e. threshold set in the time-domain signal) and the analysis is typically performed by means of fast Fourier transform (FFT). In a composite material, several types of damage might occur and the acoustic emission waves might have different frequency content. Therefore it is very important to analyse and characterise the frequency (and amplitude) information in these exponentially damped signals since it will help to have information about damage occurring in materials [67]. However the possibility to use more refined tools like subspace-based exponential data modelling [86] is closely linked to the signal quality. While signals obtained from piezoelectric sensors have very high signal-to-noise ratio (SNR), those obtained from optical fibers have often very bad quality and a subspace-based parameter estimation becomes impossible. The time-domain signal model that is used in such applications after the acquisition process may be written as follows:

$$x_n = \sum_{k=1}^K a_k \sin\{(-\alpha_k + \omega_k)t_n + \varphi_k\} + e_n, \quad n = 0, \dots, N-1 \quad (1.1)$$

where x_n are the N samples of a real signal, e_n is additive white Gaussian noise (WGN), K is the number of different EDS, $t_n = n\Delta t$ is the time lapse between the time origin and the sample x_n and Δt is the sampling time interval. The parameters of this model are the amplitudes a_k , the phases φ_k , the damping factors α_k and the pulsations ω_k . The frequencies ν_k can be obtained from the pulsations by means of the equality : $\omega_k = 2\pi\nu_k$. So far a very limited number of exploitable signals have been obtained. That is the reason why it was not possible to perform a relevant study about the frequency content of the signals obtained by the optical fiber sensors. It shall be noticed that HTLS [86], a subspace-based exponential data modelling algorithm that will be described in the next chapter, has been successfully applied on acoustic emission signals obtained from piezoelectric transducers in order to characterize the type of damage occurring in composite materials [67].

One of the most important application in which the EDS model is used, is the magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) [85, 86, 62, 92, 91, 93, 12]. These are techniques used in fundamental research and in a clinical environment whose main asset is its ability to determine the concentration of chemical substances non-invasively. These applications are linked to the concept of nuclear magnetic resonance (NMR). Since our way to approach exponential data fitting has been greatly inspired by MRS, we first want to briefly describe the NMR principle before introducing

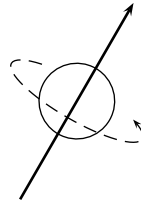


Figure 1.3: Intuitive representation of the spin of the nucleus. The nucleus can be represented as spinning around its own axis, thereby inducing a magnetic momentum

the complex version of the EDS model, which will be used in all the derivations of new techniques in this thesis.

The basic principle is to apply a static magnetic field \vec{B}_0 to a population of nuclei and then to excite this system by radio frequency pulses (oscillating magnetic field) in a plane perpendicular to the static field. For instance the ^1H nucleus has an intrinsic angular momentum which is called the spin to which a magnetic momentum is associated. The nature of this nucleus (one proton) implies that it can have only two different spin states: spin up and spin down. We can intuitively imagine the spin as a property induced by the spinning of the nucleus around its own axis. The associated magnetic momentum is parallel to the spinning axis as is illustrated in Fig. 1.3. In the field-free case the spins are randomly oriented. Whenever a static magnetic field is applied, the spins will start precessing around the direction of the static field. The precession frequency is called the Larmor frequency and depends on the magnetic field strength. Protons with spin up will align along the magnetic field, whereas protons with spin down will align along the opposite direction as illustrated in Fig.1.4. In the presence of the magnetic field and in equilibrium the number of particles with spin up and spin down is different. There are slightly more particles in a state parallel to the magnetic field. Because of this difference there will be a net magnetic field when combining all the spins. Notice that the magnetic moments are equally distributed on the cone and therefore the resulting magnetic field is parallel to the static magnetic field. The distribution between the two populations is given by the Boltzmann law which partly depends on the value of the static field. It is therefore possible to increase the net magnetization by increasing the magnetic field hence leading to a larger magnetic resonance (MR) signal. Sending photons having Larmor frequency will cause the absorption of their energy by the nuclei, which switch from parallel state to anti-parallel state. A certain duration of the radio frequency pulse will yield a saturation effect i.e. the number of protons having their spin aligned with

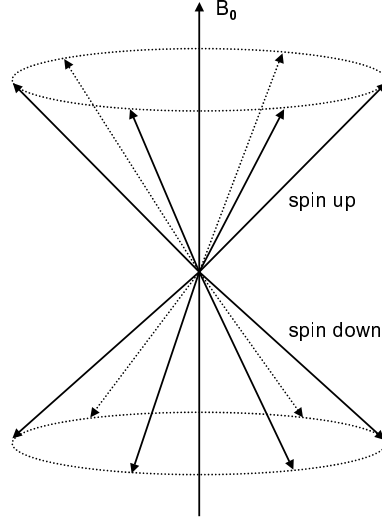


Figure 1.4: Illustration of the precession movement with two possible alignments

\vec{B}_0 will be equal to the number of protons having their spin aligned inversely. Moreover, the momenta turning around the \vec{B}_0 axis at Larmor frequency are in phase, therefore the resulting field component on the plane perpendicular to \vec{B}_0 is not zero. The return to equilibrium is therefore characterised by two relaxation times: T_1 which describes the recovery of the magnetization along the direction \vec{B}_0 , and T_2 which describes the decay of the magnetization oscillating at Larmor frequency in the plane perpendicular to \vec{B}_0 as is illustrated in Fig. 1.5. More explanations, can be found in [62]. This magnetization is detected in this plane by quadrature coils. The measured signal is the MR signal, also called free induction decay (FID). Actually T_2 has to be corrected due to field inhomogeneities. One can theoretically model the component of the FID in two perpendicular directions as a sum of EDS. After sampling on the MR scanner we have:

$$\begin{cases} x_n^1 = \sum_{k=1}^K a_k \sin\{(-\alpha_k + \omega_k)t_n + \varphi_k\} \\ x_n^2 = \sum_{k=1}^K a_k \cos\{(-\alpha_k + \omega_k)t_n + \varphi_k\} \end{cases} \quad (1.2)$$

which can be written in complex form as

$$x_n = \sum_{k=1}^K a_k \exp\{j\varphi_k\} \exp\{(-\alpha_k + j\omega_k)t_n\}, \quad n = 0, \dots, N-1 \quad (1.3)$$

where $j = \sqrt{-1}$ and K represents the number of different resonances. The amplitude a_k is related to the concentration of the metabolite and the frequency

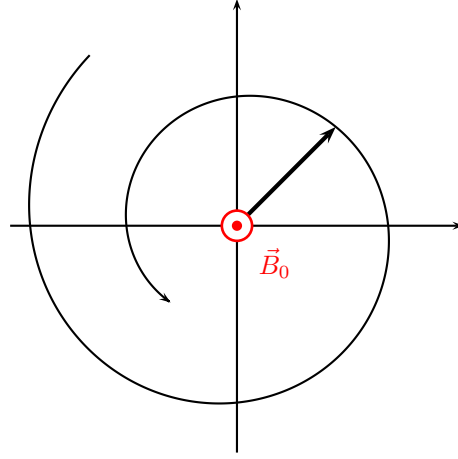


Figure 1.5: Illustration of the decay of the net magnetization in the plane perpendicular to \vec{B}_0 after the RF pulse. The projection of the magnetization onto the axis of the plane is a decaying sinusoidal function of time which has Larmor frequency.

ν_k characterize it. The damping α_k provides, among others, information about its mobility and molecular environment. In practice the MR signal has an exponential growth before the exponential decay. For simplicity we will not consider the raising part and we will assume the decaying part starting at $t_0 = 0$. We shall mention that a complex EDS can also be called a (weighted) discrete-time complex exponential.

The visual interpretation of MRS signal is much more evident in the frequency domain. Each discrete-time exponential will give rise to a bell-shaped curve in the spectrum modulus, whose peak is centered at the frequency ν_k . For this reason, in MRS, the representation of discrete-time exponential in the frequency domain is called a "peak". If we consider a continuous-time model function, its squared spectrum modulus is a sum of Lorentzian functions. For a discrete-time model, the frequency spectrum is no longer a sum of Lorentzian functions. However, in NMR this curve is called Lorentzian because it behaves like a Lorentzian function around the frequency ν_k . Let X_k be the discrete Fourier transform of x_n . We have:

$$X_l = \sum_{k=1}^K a_k \exp \left\{ \Gamma_{kl} \frac{(N-1)\Delta t}{2} \right\} \frac{\sinh \left(\frac{\Gamma_{kl} N \Delta t}{2} \right)}{\sinh \left(\frac{\Gamma_{kl} \Delta t}{2} \right)}, \quad (1.4)$$

where

$$\Gamma_{kl} = -\alpha_k + 2j\pi(\nu_k - \mu_l), \quad (1.5)$$

$$\mu_l = \frac{l}{N\Delta t}, \quad l = 0, \dots, N-1. \quad (1.6)$$

The maximal amplitude of $|X_l|$ and its width at the half height are a function of a_k and α_k , whereas the area under the curve is only related to a_k . To illustrate, Fig. 1.6 shows a single discrete-time model function and its Fourier transform modulus. Most of numerical examples used in this thesis are drawn from MRS.

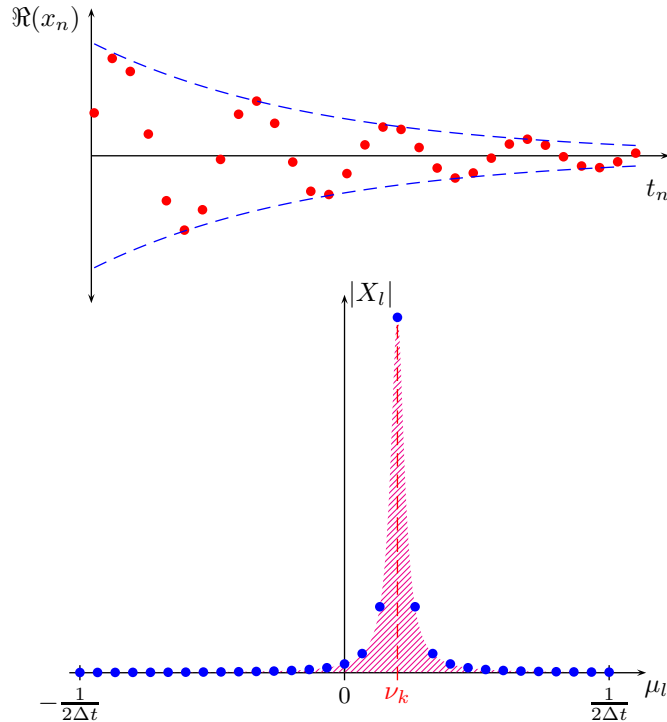


Figure 1.6: The top drawing represent the real part of a discrete-time complex exponential. The dashed line shows the exponential decay shape. The bottom drawing represents the Fourier transform of the top signal.

Other relevant applications might be mentioned. In seismology for instance, the investigation of the earthquake mechanism may require the use of the EDS model for the analysis of long-period seismograms [25].

In audio processing, the EDS model has been widely applied to the compression of transient signals generated by speech or percussive instruments [95, 3, 8, 54, 32, 84, 47, 9]. Note that in [8, 9], a delayed and damped sinusoidal model has been used to overcome the problem of delay between sinusoids, thereby allowing a lower compression rate compared to the standard EDS model. Other type of nonstationary signals such as in-vivo near infrared spectroscopy signals [52], blood pressure measurements [51], and electroencephalograms [17, 16, 18, 29]

can be modelled as a sum of EDS. Adding a damping helps to lower the model order when modelling transients (compared to undamped sinusoidal modelling). Also, we shall mention the application to hydrology and environmental sciences in which infiltrations, tails of hydrographs, and variograms can be modelled as one or a sum of exponential functions (no modulation) [65]. In the applications cited above, the use of the (complex) EDS model is directly linked to physics and especially to the oscillatory nature of studied phenomena and/or the dispersive nature of media. However other problems involving the parameter estimation of a sum of complex exponentials exist. For instance an overview of the problem of recovering a planar polygon from its measured complex moments, also called the shape-from-moments problem, is given in [24].

A last remark. Before being processed, the signals must be detected. In Appendix I we provide a very fast and efficient method that has been applied to material health monitoring. This method can be applied to any other field in which nonstationary signals (including EDS signals) have to be detected.

1.3 Purpose of the thesis

The goal of this thesis is basically to study and develop a complete set of subspace-based signal processing tools that can be applied to signals which can be modelled as a sum of complex exponentials.

Beyond the project, we have been particularly motivated by the problems arising in biomedical signal processing and especially in MRS. Thus, we will address the problem of detecting the number of complex exponentials in a given signal, the problem of parameter estimation in the single- and multi-channel case, as well as the possibility of finding common components between signals satisfying the EDS model.

1.4 Chapter-by-chapter overview

In Chapter 2 we introduce basic matrix techniques for parameter estimation of signals that can be modelled as a sum of complex exponentials. We address the single-channel, the multi-channel and the decimative case. This chapter might be seen as a basis for Chapter 6 and 7 and is of interest for the reader having no knowledge about subspace-based parameter estimation.

Chapter 3 presents a robust technique for the detection of the number of discrete-time exponentials present in a noisy signal that can be modelled as a sum of complex exponentials. This is a central chapter in the structure of this thesis in the sense that the number of discrete-time complex exponentials is a necessary parameter for all the subspace-based methods subsequently de-

veloped. The simplicity of this method and its robustness against noise, makes it very attractive to use. This novel technique has been published in *IEEE Signal Processing Letters* [59].

Chapter 4 tackles the important problem of extracting common information from multi-channel signals which can be modelled as a sum of complex exponentials. So far, no competing approach exists for a number of channels higher than 2. In this chapter we derive a novel, powerful and sufficiently flexible subspace technique to extract the common components (complex exponentials) and to extract their model parameters. This result has been published in *Signal Processing* [58].

Chapter 5 introduces the basics of multilinear algebra which is the algebra of higher-order tensors. This background material is necessary for the development of Chapters 6 and 7.

In Chapters 6 and 7 we present a multilinear generalisation of the matrix techniques described in Chapter 2. Those methods yield an improvement compared to the basic matrix techniques but also open new horizons for exponential data modelling. These multilinear algebra approaches are a first concrete step in the field of exponential data modelling. The related papers have been published in *Numerical Linear Algebra and its Applications* and *IEEE Transactions on Signal Processing*.

Chapter 8 summarizes the conclusions and outlines suggestions for further research.

Chapter 2

Exponential data fitting

The problem of estimating the parameters of EDS signals in the presence of noise is considered. More generally we consider the problem of harmonic retrieval using a matrix approach applied to discrete-time complex exponential signals. The extrapolation of the tackled techniques to EDS signals is straightforward. The EDS model arises in many applications such as radar, direction of arrival estimation, system identification, speech processing, audio processing, material engineering and MRS. The overview of the subspace-based techniques given in this chapter mainly aims to show the link between the exponential model and the subspace containing the information about the signal poles. Three cases are considered: the single-channel, the multi-channel, and the single-channel decimative case. This chapter summarizes a part of the background material needed to understand all the chapters of this thesis.

2.1 Introduction

Assume a single-channel noiseless time-domain signal containing N complex samples x_n , $n = 0, 1, \dots, N - 1$. Let this time series be modeled as a finite sum of K different exponentially damped complex sinusoids:

$$x_n = \sum_{k=1}^K a_k \exp\{j\varphi_k\} \exp\{(-\alpha_k + j\omega_k)t_n\} \quad (2.1)$$

where $j = \sqrt{-1}$, $t_n = n\Delta t$ is the time lapse between the time origin and the sample x_n and Δt is the sampling time interval. The parameters of this model are the amplitudes a_k , the phases φ_k , the damping factors α_k and the pulsations ω_k . The frequencies ν_k can be obtained from the pulsations by means of the

equality : $\omega_k = 2\pi\nu_k$. One can rewrite eq. (2.1) in a more compact form :

$$x_n = \sum_{k=1}^K c_k z_k^n \quad (2.2)$$

where $c_k = a_k \exp\{j\varphi_k\}$ denotes the k th complex amplitude including the phase, and $z_k = \exp\{(-\alpha_k + j\omega_k)\Delta t\}$ is called the k th pole of the signal with $z_k \neq z_{k'}$ for $k \neq k'$.

In the last decades, various techniques have been derived in order to estimate the signal poles z_k with respect to the model function (2.2). This problem is also referred to as harmonic retrieval. Among the most important ones, one can cite maximum likelihood methods [78], linear prediction methods [43], subspace-based methods [45, 7] and higher-order statistics [57]. Our work focuses on subspace-based methods which originate from Pisarenko's method [63]. In his work Pisarenko used the *eigenvalue decomposition* (EVD) of the sample covariance matrix in order to get the signal subspace, a set of vectors containing the relevant signal information. A famous and widely used subspace-based algorithm is ESPRIT, which exploits the rotational invariance of signal subspaces spanned by two temporally displaced data sequences. A first version of the ESPRIT algorithm was based on the least squares (LS) solution [74] and a second one appeared in the literature based on the total least squares (TLS) solution [73]. ESPRIT also uses the EVD of the sample covariance matrix. As far as the EDS model is concerned, there is an exact equivalence between the EVD of the sample covariance matrix and the *singular value decomposition* (SVD) of the Hankel data matrix. The data matrix equivalent of ESPRIT is known, in the MRS literature, as HSVD [4] and the data matrix equivalent of TLS-ESPRIT is known as HTLS [86, 85, 13]. The HTLS-algorithm, which is explained in Section 2.2 is a special case of the ESPRIT algorithm [73] and is a TLS variant of Kung et al.'s algorithm [45]. Although it is suboptimal [91], the HTLS algorithm has been proven to be robust, reliable and easy to implement. Basically, it consists of arranging the N samples of the signal x_n in a $L \times M$ Hankel matrix \mathbf{H} and computing the ordered Singular Value Decomposition (SVD) of \mathbf{H} . Then the K leftmost singular vectors of the left singular vector matrix are used to retrieve the model parameters.

This algorithm is typically applied to single-channel signals. However it is not rare to have a multi-channel setup from which several signals can be obtained at the same time. A particular case of multi-channel signals, is the case where all the signals contain the same poles. Processing them all together at the same time will give higher statistical accuracy. In order to achieve this goal, three extensions of the HTLS method, have been derived [11, 97] to handle the problem of multi-channel signals. The most interesting of them, HTLSstack [92], consists of stacking the Hankel matrices related to each channel in one big matrix and computing the SVD of this extended matrix in order to retrieve accurately the relevant subspace. This method is summarised in Section 2.3.

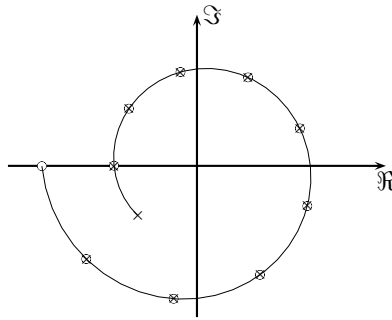


Figure 2.1: Each point on this spiral represents a sample of a discrete complex damped exponential $x(t_n) = A \cdot \exp\{(-\alpha + \omega)n\Delta t\}$ in the complex plane. The circles represent $x_1 = x(t_n)$ with $n = 0, 1, \dots, 9$, and the crosses $x_2 = x(t_n)$ with $n = 1, 2, \dots, 10$.

Another important issue is the behavior of these methods when two (or more) signal poles are very close to each other. As any other method, HTLS performs poorly when applied to closely spaced sinusoids. To get more accurate estimates, one may increase the sampling rate. However, as the sampling rate increases, the computational cost of the HTLS algorithm gets quickly prohibitive due to the large size of the data matrix \mathbf{H} . In [50, 49], three decimative versions of the HTLS method were derived, among which HTLSDstack turns out to be the most interesting method because it has approximately the same accuracy as HTLS while having a much lower computational cost. This is the data matrix counterpart of the decimative TLS-ESPRIT algorithm presented in [77] and [28]. The HTLSDstack algorithm uses a decimative approach to decrease the computational complexity while keeping an accuracy comparable to that of the HTLS method. In HTLSDstack the signal is decimated by a factor D with respect to the Nyquist theorem. The D downsampled sequences are arranged in separate Hankel matrices. Then, these Hankel matrices are stacked in a block Hankel matrix of which the SVD is computed. The rest of the algorithm is similar to the HTLS algorithm. In HTLSDstack the downsampled sequences can be considered as different channels that allow for an improved frequency separation. This algorithm, which is explained in Section 2.4, is, in essence, similar to the HTLSstack algorithm, which applies to a set of time series (channels) [97, 88].

It is amazing to see how the EDS model lends itself to subspace representation. For instance let us take a signal of the form $x(t_n) = A \cdot \exp\{(-\alpha + \omega)n\Delta t\} = A \cdot z^n$, $n = 0, 1, \dots, 10$. Let us consider $x_1 = x(t_n)$, $n = 0, 1, \dots, 9$ and $x_2 = x(t_n)$, $n = 1, 2, \dots, 10$. As show in Figure 2.1, x_1 and x_2 have eight points in common but the first and the last one are different due to the time shift Δt . However if we represent these signals as vectors, say $X_1 = [x(0)x(1)\dots x(9)]^T$

and $X_2 = [x(1)x(2)\dots x(10)]^\top$ we have that $X_2 = z.X_1$. In other words the two vectors X_1 and X_2 are linearly dependent and the multiplying coefficient is simply z , the signal pole. This link might explain the possible rank deficiency of the Hankel data matrix and what information is derived from such a property.

2.2 Hankel total least squares method: HTLS

2.2.1 Subspace Representation

First, given $x_n = \sum_{k=1}^K c_k \cdot z_k^n$, $n = 0, \dots, N-1$ and K , the data samples are arranged into a $L \times M$ Hankel matrix as follows :

$$\mathbf{H} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{M-1} \\ x_1 & x_2 & \ddots & \cdots & \vdots \\ x_2 & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2} \\ x_{L-1} & \cdots & \cdots & x_{N-2} & x_{N-1} \end{pmatrix} \quad (2.3)$$

with $\{L, M\}$ chosen such that $N = L + M - 1$ and $L > K$. This Hankel matrix can be straightforwardly decomposed in a product of three matrices :

$$\begin{aligned} \mathbf{H} = & \begin{pmatrix} 1 & \cdots & 1 \\ z_1^1 & \cdots & z_K^1 \\ z_1^2 & \cdots & z_K^2 \\ \vdots & \vdots & \vdots \\ z_1^{L-1} & \cdots & z_K^{L-1} \end{pmatrix} \begin{pmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_K \end{pmatrix} \begin{pmatrix} 1 & z_1^1 & z_1^2 & \cdots & z_1^{M-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & z_K^1 & z_K^2 & \cdots & z_K^{M-1} \end{pmatrix} \\ & = \mathbf{SCT}^\top. \quad (2.4) \end{aligned}$$

Equation (2.4) is called a *Vandermonde decomposition* (VDMD) in which the poles z_k are also called the *generators*. The column vectors of \mathbf{S} and \mathbf{T} are called Vandermonde vectors. The rank deficiency of \mathbf{H} is also reflected by the SVD defined as follows:

$$\mathbf{H} = \left(\hat{\mathbf{U}} \mathbf{U}_0 \right) \begin{pmatrix} \hat{\boldsymbol{\Sigma}} & 0 \\ 0 & \boldsymbol{\Sigma}_0 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{V}}^H \\ \mathbf{V}_0^H \end{pmatrix}, \quad (2.5)$$

where $\hat{\mathbf{U}} \in \mathbb{C}^{L \times K}$, $\mathbf{U}_0 \in \mathbb{C}^{L \times (L-K)}$, $\hat{\boldsymbol{\Sigma}} \in \mathbb{R}^{K \times K}$ contains the K non-zero singular values in decreasing order of magnitude, $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{(L-K) \times (M-K)}$, $\hat{\mathbf{V}} \in \mathbb{C}^{M \times K}$, $\mathbf{V}_0 \in \mathbb{C}^{K \times (M-K)}$. In the absence of noise $\boldsymbol{\Sigma}_0$ is a null matrix and the SVD of \mathbf{H} reduces to the product $\hat{\mathbf{U}} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{V}}^H$. The columns of \mathbf{S} (*resp.* \mathbf{T}) span

the same subspace as the columns of $\widehat{\mathbf{U}}$ (resp. $\widehat{\mathbf{V}}^*$). If the signal is corrupted by noise, $\mathbf{\Sigma}_0$ is full rank. In the presence of a reasonable level of noise, the truncated SVD of \mathbf{H} ,

$$\widehat{\mathbf{H}}_{L \times M} = \widehat{\mathbf{U}}_{L \times K} \widehat{\mathbf{\Sigma}}_{K \times K} \widehat{\mathbf{V}}_{M \times K}^H, \quad (2.6)$$

which is the best rank- K approximation of \mathbf{H} , can be considered as a denoised version of \mathbf{H} although it is not a Hankel matrix anymore. In this case, the column vectors of $\widehat{\mathbf{U}}$ yield a good approximation of the subspace spanned by the Vandermonde vectors.

2.2.2 Shift-invariance property and harmonic retrieval via Total Least Squares Solution

The matrices \mathbf{S} and \mathbf{T} possess a shift-invariance property that can be expressed as:

$$\begin{cases} \mathbf{S}_\downarrow \mathbf{Z} = \mathbf{S}^\uparrow \\ \mathbf{T}_\downarrow \mathbf{Z} = \mathbf{T}^\uparrow, \end{cases} \quad (2.7)$$

where the up (down) arrow placed behind a matrix stands for deleting the top (bottom) row of the considered matrix and $\mathbf{Z} = \text{diag}(z_1, z_2, \dots, z_K) \in \mathbb{C}^{K \times K}$. In the noise-free case, $\widehat{\mathbf{U}}$ equals \mathbf{S} up to a multiplication by a square non-singular matrix $\mathbf{Q} \in \mathbb{C}^{K \times K}$:

$$\widehat{\mathbf{U}} = \mathbf{S} \mathbf{Q}. \quad (2.8)$$

The matrices $\widehat{\mathbf{U}}^\uparrow$ (resp. $\widehat{\mathbf{U}}_\downarrow$) are related to the Vandermonde matrices \mathbf{S}^\uparrow (resp. \mathbf{S}_\downarrow) in the following way:

$$\begin{cases} \widehat{\mathbf{U}}^\uparrow = \mathbf{S}^\uparrow \mathbf{Q} \\ \widehat{\mathbf{U}}_\downarrow = \mathbf{S}_\downarrow \mathbf{Q}. \end{cases} \quad (2.9)$$

Combining equations (2.7) and (2.9) results in the shift-invariance property of $\widehat{\mathbf{U}}$:

$$\widehat{\mathbf{U}}^\uparrow = \widehat{\mathbf{U}}_\downarrow \mathbf{Q}^{-1} \mathbf{Z} \mathbf{Q} = \widehat{\mathbf{U}}_\downarrow \widetilde{\mathbf{Z}}. \quad (2.10)$$

Note that the same reasoning applies to the matrix $\widehat{\mathbf{V}}^*$. In the case of white noise, the equality does not hold anymore. The TLS solution of the overdetermined set of linear equations $\widehat{\mathbf{U}}^\uparrow \approx \widehat{\mathbf{U}}_\downarrow \widetilde{\mathbf{Z}}$, whose aim is to minimize the correction applied to the data in order to make the set compatible, is given by [87, 85]:

$$\widehat{\widetilde{\mathbf{Z}}} = -\mathbf{W}_{12} \mathbf{W}_{22}^{-1}, \quad (2.11)$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix} \quad (2.12)$$

is obtained from the SVD of the matrix $[\widehat{\mathbf{U}}_{\downarrow} \widehat{\mathbf{U}}^{\uparrow}]$:

$$[\widehat{\mathbf{U}}_{\downarrow} \widehat{\mathbf{U}}^{\uparrow}] \stackrel{\text{svd}}{=} \mathbf{Y} \mathbf{\Gamma} \mathbf{W}^H. \quad (2.13)$$

where $\mathbf{Y} \in \mathbb{C}^{(L-1) \times 2K}$ is the (semi-unitary) left singular vector matrix, $\mathbf{W} \in \mathbb{C}^{2K \times 2K}$ is the (unitary) right singular vector matrix, and $\mathbf{\Gamma} \in \mathbb{R}^{2K \times 2K}$ is the (diagonal) singular value matrix. Once $\widetilde{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of the signal poles:

$$\lambda_k = \hat{z}_k = \exp \{ (-\hat{\alpha}_k + 2\pi j \hat{\nu}_k) \Delta t \}. \quad (2.14)$$

Finally the complex amplitudes can be estimated by solving the set of equations:

$$\begin{pmatrix} 1 & \cdots & 1 \\ \hat{z}_1^1 & \cdots & \hat{z}_K^1 \\ \hat{z}_1^2 & \cdots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \cdots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix} \quad (2.15)$$

in the Least Squares (LS) sense. This summarizes the HTLS algorithm. Note that if the set of equations (2.10) is solved in the LS sense one obtains the HSVD algorithm. As the HTLS method will be used in Chapter 6, an outline of the related algorithm is given below.

Algorithm 2.2

HTLS/HSVD

Input: data samples x_n , $n = 0, \dots, N-1$ and model order K .

Output: \hat{a}_k , $\hat{\varphi}_k$, $\hat{\alpha}_k$, $\hat{\nu}_k$, $k = 1, \dots, K$.

Step 1: Arrange the data points x_n , $n = 0, \dots, N-1$ in a $(L \times M)$ -Hankel matrix \mathbf{H} , $N = L + M - 1$, $L > K$ as follows:

$$\mathbf{H} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{M-1} \\ x_1 & x_2 & \ddots & \cdots & \vdots \\ x_2 & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2} \\ x_{L-1} & \cdots & \cdots & x_{N-2} & x_{N-1} \end{pmatrix}$$

The best parameter accuracy is generally obtained with \mathbf{H} slightly rectangular. As it is impossible to predict the exact optimal matrix size, it is recommended to choose \mathbf{H} as square as possible which yields good results [85].

Step 2: Compute the SVD of \mathbf{H}

$$\mathbf{H} = \mathbf{U}_{L \times \min(L,M)} \mathbf{\Sigma}_{\min(L,M) \times \min(L,M)} \mathbf{V}_{M \times \min(L,M)}^H$$

Step 3: Truncate the SVD of \mathbf{H} in order to obtain its best rank- K' approximation

$$\mathbf{H} = \widehat{\mathbf{U}}_{L \times K'} \widehat{\mathbf{\Sigma}}_{K' \times K'} \widehat{\mathbf{V}}_{M \times K'}^H$$

The rank K' is chosen equal to the model order K which corresponds to the number of complex exponentials in the signal. If the signal is real then $K' = 2K$.

Step 4: Form the following overdetermined set of equations

$$\widehat{\mathbf{U}}^\uparrow \approx \widehat{\mathbf{U}}_\downarrow \widetilde{\mathbf{Z}}$$

$\widehat{\mathbf{U}}^\uparrow$ and $\widehat{\mathbf{U}}_\downarrow$ are derived from $\widehat{\mathbf{U}}$ by omitting its first and last row respectively.

- HSVD: compute an estimate of $\widetilde{\mathbf{Z}}$ by solving the above set of equations in the LS sense.
- HTLS: compute an estimate of $\widetilde{\mathbf{Z}}$ by solving the above set of equations in the TLS sense.

Once $\widetilde{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of the signal poles:

$$\lambda_k = \hat{z}_k = \exp \{ (-\hat{\alpha}_k + 2\pi j \hat{\nu}_k) \Delta t \} .$$

from which it is easy to obtain estimates of the damping factors α_k and frequencies ν_k .

Step 5: Eventually using the estimates \hat{z}_k , $k = 1, \dots, K$, and the signal sample x_n , $n = 0, \dots, N - 1$ compute the LS solution [80] $\hat{c}_k = \hat{a}_k \exp\{j\hat{\varphi}_k\}$ of

$$\begin{pmatrix} 1 & \dots & 1 \\ \hat{z}_1^1 & \dots & \hat{z}_K^1 \\ \hat{z}_1^2 & \dots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \dots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}$$

— Remark: the amplitudes and phases in step 5 can also be estimated using APES [79, 80].

2.3 Hankel Total Least Squares method for multi-channel signal: HTLSstack

Assume Q simultaneous one-dimensional noiseless time domain signals each consisting of N complex samples. Let this multi-channel signal be modelled as follows :

$$x_n^{(q)} = \sum_{k=1}^K a_k^{(q)} \exp\{j\varphi_k^{(q)}\} \exp\{(-\alpha_k + j\omega_k)t_n\} = \sum_{k=1}^K c_k^{(q)} z_k^n \quad (2.16)$$

in which q ($1 \leq q \leq Q$) denotes the channel number. First, the data samples of each channel are arranged into an $L \times M$ Hankel matrix as follows :

$$\mathbf{H}_q = \begin{pmatrix} x_0^{(q)} & x_1^{(q)} & x_2^{(q)} & \cdots & x_{M-1}^{(q)} \\ x_1^{(q)} & x_2^{(q)} & \ddots & \cdots & \vdots \\ x_2^{(q)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2}^{(q)} \\ x_{L-1}^{(q)} & \cdots & \cdots & x_{N-2}^{(q)} & x_{N-1}^{(q)} \end{pmatrix}. \quad (2.17)$$

In the noise-free case each matrix \mathbf{H}_q has a VDMD:

$$\mathbf{H}_q = \mathbf{S} \mathbf{C}_q \mathbf{T}^\top, \quad (2.18)$$

in which $\mathbf{S} \in \mathbb{C}^{L \times K}$, $\mathbf{T} \in \mathbb{C}^{K \times M}$ are the Vandermonde matrices defined in (2.4) and $\mathbf{C}_q = \text{diag}\{c_1^{(q)}, c_2^{(q)}, \dots, c_K^{(q)}\}$. Therefore all the matrices \mathbf{H}_q have the same column subspace and if the Q Hankel matrices are stacked in one block Hankel matrix:

$$\mathbf{H} = [\mathbf{H}_1 \mid \mathbf{H}_2 \mid \dots \mid \mathbf{H}_Q], \quad (2.19)$$

this matrix can be decomposed as follows:

$$\mathbf{H} = \mathbf{S} [\mathbf{C}_1 \mathbf{T}^\top \mid \mathbf{C}_2 \mathbf{T}^\top \mid \dots \mid \mathbf{C}_Q \mathbf{T}^\top] \quad (2.20)$$

Let $\widehat{\mathbf{U}}$ be defined by the best rank- K approximation of \mathbf{H} , in analogy with (2.28). Since the columns of $\widehat{\mathbf{U}}$ and \mathbf{S} span the same subspace, one can repeat the procedure described by equations (2.8)–(2.14) to retrieve the parameters-of-interest. This technique is particularly useful in the presence of noise because the repeated information which induces the rank deficiency of \mathbf{H} , will increase the gap between the singular values related to the signal subspace and the singular values related to the noise subspace. In other words the separation between the noise and the signal subspace will be enhanced as the number of channels will increase. We should mention that HTLSstack can be efficiently applied if all the signals $x_n^{(q)}$, $q = 1, \dots, Q$ have the same poles which implies

that they also have the same model order. As an example, if one of the signals contains an additional complex exponential, it is not sure that the resulting subspace $\widehat{\mathbf{U}}$ will contain the components related to this exponential. As the HTLSstack method will be used in Chapter 6, an outline of the related algorithm is given below.

Algorithm 2.3

HTLSstack

Input: data samples $x_n^{(q)}$, $n = 0, \dots, N - 1$, $q = 1, \dots, Q$ and model order K .

Output: $\widehat{a}_k^{(q)}$, $\widehat{\varphi}_k^{(q)}$, $\widehat{\alpha}_k$, $\widehat{\nu}_k$, $k = 1, \dots, K$, $q = 1, \dots, Q$.

Step 1: For every signal $x_n^{(q)}$, $q = 1, \dots, Q$, arrange the data points $x_n^{(q)}$, $n = 0, \dots, N - 1$ in a $(L \times M)$ -Hankel matrix \mathbf{H}_q , $N = L + M - 1$, $L > K$ as follows:

$$\mathbf{H}_q = \begin{pmatrix} x_0^{(q)} & x_1^{(q)} & x_2^{(q)} & \cdots & x_{M-1}^{(q)} \\ x_1^{(q)} & x_2^{(q)} & \ddots & \cdots & \vdots \\ x_2^{(q)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2}^{(q)} \\ x_{L-1}^{(q)} & \cdots & \cdots & x_{N-2}^{(q)} & x_{N-1}^{(q)} \end{pmatrix}.$$

As for the HTLS algorithm it is recommended to have \mathbf{H}_q as square as possible.

Step 2: Form a block Hankel matrix

$$\mathbf{H} = [\mathbf{H}_1 \mid \mathbf{H}_2 \mid \dots \mid \mathbf{H}_Q],$$

Step 3: Compute the SVD of \mathbf{H}

$$\mathbf{H} = \mathbf{U}_{L \times L} \mathbf{\Sigma}_{L \times L} \mathbf{V}_{Q.M \times L}^H$$

Note that if the matrices \mathbf{H}_q are almost square we have always $Q.M > L$ for all $Q \geq 2$, $Q \in \mathbb{N}$.

Step 4: Truncate the SVD of \mathbf{H} in order to obtain its best rank- K' approximation

$$\mathbf{H} = \widehat{\mathbf{U}}_{L \times K'} \widehat{\mathbf{\Sigma}}_{K' \times K'} \widehat{\mathbf{V}}_{M \times K'}^H$$

The rank K' is chosen equal to the model K order which corresponds to the number of complex exponentials in the signal. If the signal is real then $K' = 2K$.

Step 5: Form the following overdetermined set of equations

$$\widehat{\mathbf{U}}^\dagger \approx \widehat{\mathbf{U}}_\downarrow \widetilde{\mathbf{Z}}$$

and compute an estimate of $\widetilde{\mathbf{Z}}$ by solving the above set of equations in the TLS sense. $\widehat{\mathbf{U}}^\dagger$ and $\widehat{\mathbf{U}}_\downarrow$ are derived from $\widehat{\mathbf{U}}$ by omitting its first and last row respectively. Once $\widetilde{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of the signal poles:

$$\lambda_k = \hat{z}_k = \exp\{(-\hat{\alpha}_k + 2\pi j\hat{\nu}_k)\Delta t\}.$$

from which one can obtain estimates of the damping factors α_k and frequencies ν_k .

Step 6: Eventually, for each signal q , an estimate of the complex amplitudes $c_k^{(q)} = a_k^{(q)} \exp\{j\varphi_k^{(q)}\}$ is obtained from the LS solution of

$$\begin{pmatrix} 1 & \cdots & 1 \\ \hat{z}_1^1 & \cdots & \hat{z}_K^1 \\ \hat{z}_1^2 & \cdots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \cdots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1^{(q)} \\ c_2^{(q)} \\ \vdots \\ c_K^{(q)} \end{pmatrix} = \begin{pmatrix} x_0^{(q)} \\ x_1^{(q)} \\ x_2^{(q)} \\ \vdots \\ x_{N-1}^{(q)} \end{pmatrix}$$

2.4 Harmonic retrieval in the decimative case: HTLSDstack

In the case of closely spaced sinusoids, the HTLS algorithm, described in Section 2.2, might perform very poorly. To get more accurate estimates, a simple solution consists of increasing the sampling rate. However, as the sampling rate increases, the computational cost of the HTLS algorithm gets quickly prohibitive due to the large size of the matrix \mathbf{H} (2.3). The HTLSDstack algorithm overcomes this problem by using the downsampled sequences of an oversampled signal as different channels in which the frequency separation is artificially increased. Thus the signal of each channel is stacked in a Hankel matrix, and all the Hankel matrices are stacked in a block Hankel matrix. As for the HTLSstack algorithm described in Section 2.3, this will allow for a very good separation of the signal subspace from the noise subspace.

Let us consider x_n given by the model (2.2), and let us assume that we can decimate x_n by a factor D without aliasing effects. This means that, in practice, the frequencies higher than $\max(\nu_k)$ have to be filtered out. Moreover, we

will assume that N , the number of samples of x_n , is an integer multiple of D . Therefore we can generate D decimated sequences of $N_D = N/D$ samples:

$$x_n^{(d)} = x_{nD+d} = \sum_{k=1}^K c_k z_k^{nD+d} \quad n = 0, \dots, N_D - 1, \quad d = 0, \dots, D - 1, \quad (2.21)$$

each of which can be seen as a different channel. First, the data samples of each channel are arranged into a $L_D \times M_D$ Hankel matrix as follows :

$$\mathbf{H}_d = \begin{pmatrix} x_0^{(d)} & x_1^{(d)} & x_2^{(d)} & \cdots & x_{M_D-1}^{(d)} \\ x_1^{(d)} & x_2^{(d)} & \ddots & \cdots & \vdots \\ x_2^{(d)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N_D-2}^{(d)} \\ x_{L_D-1}^{(d)} & \cdots & \cdots & x_{N_D-2}^{(d)} & x_{N_D-1}^{(d)} \end{pmatrix}, \quad (2.22)$$

with $\{L_D, M_D\}$ chosen such that $N_D = L_D + M_D - 1$ and $L_D > K, M_D \cdot D \geq K$. The D Hankel matrices are stacked in a $(L_D \times D \cdot M_D)$ block Hankel matrix:

$$\mathbf{H} = [\mathbf{H}_0 \mid \mathbf{H}_1 \mid \dots \mid \mathbf{H}_{D-1}]. \quad (2.23)$$

This matrix can be decomposed as follows:

$$\mathbf{H} = \mathbf{S} [\mathbf{C}_0 \mathbf{T}^\top \mid \mathbf{C}_1 \mathbf{T}^\top \mid \dots \mid \mathbf{C}_{D-1} \mathbf{T}^\top] \quad (2.24)$$

in which

$$\mathbf{S} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1^D & \cdots & z_K^D \\ z_1^{2D} & \cdots & z_K^{2D} \\ \vdots & \vdots & \vdots \\ z_1^{(L_D-1)D} & \cdots & z_K^{(L_D-1)D} \end{pmatrix} \quad (2.25)$$

and

$$\mathbf{T} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1^D & \cdots & z_K^D \\ z_1^{2D} & \cdots & z_K^{2D} \\ \vdots & \vdots & \vdots \\ z_1^{(M_D-1)D} & \cdots & z_K^{(M_D-1)D} \end{pmatrix} \quad (2.26)$$

are respectively a $(L_D \times K)$ and a $(M_D \times K)$ Vandermonde matrix with z_1^D, \dots, z_K^D as generators, and in which

$$\mathbf{C}_d = \begin{pmatrix} c_1 z_1^d & & \\ & \ddots & \\ & & c_K z_K^d \end{pmatrix}. \quad (2.27)$$

Each factorized block $\mathbf{S}\mathbf{C}_d\mathbf{T}^\top$ is a VDMD. It appears from (2.24) that \mathbf{H} is a rank- K matrix; if $M_D \cdot D > K$, it is rank deficient. Hence, the dominant K -dimensional column subspace of the matrix \mathbf{H} contains the information that is relevant to retrieve the parameters of interest. This subspace is for instance spanned by the matrix of K leftmost singular vectors $\hat{\mathbf{U}}$ in the ordered SVD:

$$\hat{\mathbf{H}}_{L \times M} = \hat{\mathbf{U}}_{L \times K} \hat{\mathbf{\Sigma}}_{K \times K} \hat{\mathbf{V}}_{M \times K}^H \quad (2.28)$$

where $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are complex semi-unitary matrices and $\hat{\mathbf{\Sigma}}$ is a real diagonal matrix. In the noise-free case, the columns of $\hat{\mathbf{U}}$ span the same subspace as the columns of \mathbf{S} , while in the presence of noise they yield a LS approximation. The matrix \mathbf{S} possesses a shift-invariance property that can be expressed as:

$$\mathbf{S}_{\downarrow} \mathbf{Z} = \mathbf{S}^{\uparrow}, \quad (2.29)$$

where the up (down) arrow placed behind a matrix stands for deleting the top (bottom) row of the considered matrix and $\mathbf{Z} = \text{diag}(z_1^D, z_2^D, \dots, z_K^D) \in \mathbb{C}^{K \times K}$. In the case of white noise, $\hat{\mathbf{U}}$ equals \mathbf{S} up to a multiplication by a square non-singular matrix $\mathbf{Q} \in \mathbb{C}^{K \times K}$:

$$\hat{\mathbf{U}} = \mathbf{S}\mathbf{Q}. \quad (2.30)$$

Here again, since the columns of $\hat{\mathbf{U}}$ and \mathbf{S} span the same subspace, one can repeat the procedure described by equations (2.8)–(2.14) to retrieve the parameters-of-interest. Notice however that the eigenvalues λ_k of the matrix $\hat{\mathbf{Z}}$ yields an estimate of the decimated signal poles z_k^D :

$$\lambda_k = \exp\{(-\hat{\alpha}_k + 2\pi j\hat{\nu}_k)D\Delta t\} \quad (2.31)$$

where $\hat{\alpha}_k$ and $\hat{\nu}_k$ denote the estimates of α_k and ν_k respectively. As the HTLS-Dstack method will be used in Chapter 7, an outline of the related algorithm is given below.

Algorithm 2.4

HTLSDstack

Input: data samples x_n , $n = 0, \dots, N - 1$, decimation factor D , and model order K .

Output: $\hat{\alpha}_k, \hat{\varphi}_k, \hat{\alpha}_k, \hat{\nu}_k$, $k = 1, \dots, K$.

Step 1: Decimate x_n by a factor D making sure that no aliasing effect occurs (filter out frequencies higher than $\max \nu_k$) and obtain D decimated sequences $x_n^{(d)}$. For simplicity it is recommended that N is an integer multiple of D .

Step 2: For every signal $x_n^{(d)}$, $d = 1, \dots, D$, arrange the data points $x_n^{(d)}$, $n = 0, \dots, N/D - 1$ in a $(L_D \times M_D)$ -Hankel matrix \mathbf{H}_d , $N_D = L_D + M_D - 1$, $N_D = N/D$, $L_D > K$, $M_D \cdot D \geq K$ as follows:

$$\mathbf{H}_d = \begin{pmatrix} x_0^{(d)} & x_1^{(d)} & x_2^{(d)} & \cdots & x_{M_D-1}^{(d)} \\ x_1^{(d)} & x_2^{(d)} & \ddots & \cdots & \vdots \\ x_2^{(d)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N_D-2}^{(d)} \\ x_{L_D-1}^{(d)} & \cdots & \cdots & x_{N_D-2}^{(d)} & x_{N_D-1}^{(d)} \end{pmatrix},$$

It is recommended to have \mathbf{H}_d as square as possible [50, 49].

Step 3: Form a $(L_D \times D \cdot M_D)$ block Hankel matrix

$$\mathbf{H} = [\mathbf{H}_0 \mid \mathbf{H}_1 \mid \dots \mid \mathbf{H}_{D-1}].$$

Step 4: Compute the SVD of \mathbf{H}

$$\mathbf{H} = \mathbf{U}_{L_D \times L_D} \mathbf{\Sigma}_{L_D \times L_D} \mathbf{V}_{D \cdot M_D \times L_D}^H$$

Note that if the matrices \mathbf{H}_d are almost square we have always $D \cdot M_D > L_D$ for all $D \geq 2$, $D \in \mathbb{N}$.

Step 5: Truncate the SVD of \mathbf{H} in order to obtain its best rank- K' approximation

$$\mathbf{H} = \widehat{\mathbf{U}}_{L_D \times K'} \widehat{\mathbf{\Sigma}}_{K' \times K'} \widehat{\mathbf{V}}_{D \cdot M_D \times K'}^H$$

The rank K' is chosen equal to the model K order which corresponds to the number of complex exponentials in the signal. If the signal is real then $K' = 2K$.

Step 6: Form the following overdetermined set of equations

$$\widehat{\mathbf{U}}^\uparrow \approx \widehat{\mathbf{U}}_\downarrow \widetilde{\mathbf{Z}}$$

and compute an estimate of $\widetilde{\mathbf{Z}}$ by solving the above set of equations in the TLS sense. $\widehat{\mathbf{U}}^\uparrow$ and $\widehat{\mathbf{U}}_\downarrow$ are derived from $\widehat{\mathbf{U}}$ by omitting its first and last row respectively. Once $\widetilde{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of decimated the signal poles z_k^D :

$$\lambda_k = \hat{z}_k^D = \exp\{(-\hat{\alpha}_k + 2\pi j \hat{\nu}_k) D \Delta t\}$$

It is now easy to obtain estimates of the original (undecimated) damping factors α_k and frequencies ν_k .

Step 7: Finally using the estimates \hat{z}_k , $k = 1, \dots, K$, and the nondecimated signal sample x_n , $n = 0, \dots, N - 1$ compute the LS solution $\hat{c}_k = \hat{a}_k \exp\{j\hat{\varphi}_k\}$ of

$$\begin{pmatrix} 1 & \cdots & 1 \\ \hat{z}_1^1 & \cdots & \hat{z}_K^1 \\ \hat{z}_1^2 & \cdots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \cdots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}$$

2.5 Conclusion

In this chapter we reviewed three subspace-based parameter estimation methods and we also provided the corresponding algorithms. These methods aim to provide an estimate of parameters of signals satisfying the EDS model. We described the HTLS method which can be applied on single-channel signals, the HTLSstack method which is typically applicable to multi-channel signals and the HTLSDstack method applicable to single-channel signals that can be decimated. The latter is especially useful in the case of closely spaced sinusoids. We will often refer to these algorithms in the remainder of this thesis.

Chapter 3

A shift invariance-based order selection technique for the EDS model

This chapter presents a new subspace-based technique for automatic detection of the number of exponentially damped sinusoids. It consists in studying the shift-invariance of the dominant subspace of the Hankel data matrix. No threshold setting and no penalization terms are necessary. This model-based method, easy to implement, can be plugged in in most subspace-based harmonic retrieval algorithms.

3.1 Introduction

In most applications in which harmonic retrieval is a core problem [85, 86, 67, 95, 56], the model order K is assumed to be known. However, the model order is a very important parameter, without which no reliable estimation can be performed. This becomes obvious when looking at equation (2.8); the Vandermonde vectors (column vectors of \mathbf{S}) span a K -dimensional subspace and therefore $\hat{\mathbf{U}}$ must contain exactly K columns in order to span the subspace in which these Vandermonde vectors lie. While underestimation of the model order yields biased estimates [72], overestimation will result in fitting the noise to EDS's. Therefore, the determination of the model order is a key step in all high-resolution methods. A simple method consists of determining the numerical rank of \mathbf{H} (2.3) by looking at the gap between singular values. However this is a very rough method. More refined model order selection methods are

based on Information Theoretic Criteria (ITC), such as the Akaike Information Criterion (AIC) [102], the Minimum Description Length (MDL) [102] or the Efficient Detection Criterion (EDC) [103]. These so-called penalized likelihood methods assume that the noise is Gaussian and use this property to determine how many singular values are related to the noise subspace. Recently, a model-based method [72], called ESTER (ESTimation ERror), was developed to automatically find the order of a EDS model. Compared to the ITC, ESTER has the great advantage of not assuming that the noise is Gaussian. It does not require any penalization term since it is directly based on the underlying model. It merely assumes that the noise and the signal subspace are sufficiently separated, which implies that the noise has to be white and zero-mean. It has been proven to be robust against noise and outperforms the ITC cited above.

In this chapter we present a simple and novel SVD-based method, easy to implement, which is shown to be more general, flexible and robust than ESTER. It mainly relies on the shift-invariance property of \mathbf{S} . Moreover it does not assume gaussianity of the additive noise. As for ESTER, the main advantage of this method is that it does not require the setting of a threshold or penalization term. The input is the signal satisfying the EDS model and the output is the model order.

For sake of consistency we repeat two important properties:

- \mathbf{H} is a rank- K ($L \times M$)-Hankel matrix, defined by (2.3), and its column space is exactly spanned by the set of K Vandermonde vectors \mathbf{S} ,
- \mathbf{S} has the following shift-invariance property:

$$\mathbf{S}^\uparrow = \mathbf{S}_\downarrow \mathbf{D}, \quad (3.1)$$

where \uparrow (resp. \downarrow) means that the top row (resp. the bottom row) has been deleted and where $\mathbf{D} = \text{diag}\{z_1, \dots, z_K\}$.

The chapter is organized as follows. In Section 3.2 we summarize the ESTER method and we explain briefly some of its important properties. In Section 3.3 we derive our new method. In Section 3.4 the two methods are compared by means of Monte-Carlo simulations. Section 3.5 presents applications. In Section 3.6 we conclude the chapter.

3.2 Subspace shift-invariance property and least squares approach

For exponential data modelling the shift-invariance of the column subspace of \mathbf{H} is crucial. This property is induced by the shift-invariance of the Van-

dermonde matrix \mathbf{S} . In the noise-free case, the K first singular vectors of \mathbf{U} exactly span the same subspace as the vectors of \mathbf{S} and hence there exists a square non-singular matrix \mathbf{Q} such that:

$$\mathbf{S} = \mathbf{U}_K \mathbf{Q}, \quad (3.2)$$

where $\mathbf{U}_K = [U_1 U_2 \cdots U_K]$ is the semi-unitary matrix containing the K first columns of \mathbf{U} . From (3.1) it follows that

$$\mathbf{U}_K^\dagger = \mathbf{U}_{K\downarrow} \mathbf{\Phi}(K), \quad (3.3)$$

where $\mathbf{\Phi}(K) = \mathbf{Q} \mathbf{D} \mathbf{Q}^{-1}$ is an unknown square non-singular matrix whose eigenvalues are the signal poles. The equality in (3.3) only holds for the true order K . Considering a subspace of size $k \neq K$ leads to an inconsistent set of equations:

$$\mathbf{U}_k^\dagger \approx \mathbf{U}_{k\downarrow} \mathbf{\Phi}(k). \quad (3.4)$$

The ESTER method developed by Badeau, is based on this observation. ESTER consists of computing the residual error of the set $\mathbf{U}_k^\dagger = \mathbf{U}_{k\downarrow} \mathbf{\Phi}(k)$ in the least-squares (LS) sense:

$$E(k)^2 = \left\| \Delta \mathbf{U}_k^\dagger \right\|_2^2 = \left\| \mathbf{U}_{k\downarrow} \widehat{\mathbf{\Phi}}(k) - \mathbf{U}_k^\dagger \right\|_2^2 \quad (3.5)$$

with

$$\widehat{\mathbf{\Phi}}(k) = \left(\mathbf{U}_{k\downarrow}^H \mathbf{U}_{k\downarrow} \right)^\dagger \mathbf{U}_k^\dagger. \quad (3.6)$$

Indeed, in the noise-free case, the system has only an exact solution for $k = K$ and the error is theoretically 0. Moreover, it was shown by means of simulations that, in the presence of noise, the number of columns k for which $E(k)^2$ is minimal, is still K , the order of the EDS model. The estimation of K can then be formulated as follows:

$$K = \arg \max_k \frac{1}{E(k)^2}, \quad 1 \leq k < \min(L, M) \quad (3.7)$$

3.3 SVD-based approach

The least squares approach is inappropriate in the sense that it supposes that only \mathbf{U}_k^\dagger is perturbed. Whether it is due to the noise or due to an under/overestimation of the subspace size (i.e. $k \neq K$), always both $\mathbf{U}_{k\downarrow}$ and \mathbf{U}_k^\dagger are perturbed. Therefore we will jointly consider both the subspaces $\text{span}(\mathbf{U}_{k\downarrow})$ and $\text{span}(\mathbf{U}_k^\dagger)$ by starting from the following matrix:

$$\mathbf{U}_k^{\text{tb}} = \left[\mathbf{U}_k^\dagger | \mathbf{U}_{k\downarrow} \right]. \quad (3.8)$$

Indeed, this $(L-1) \times 2K$ matrix has some particular properties that are related to the shift-invariance property of \mathbf{S} . We first investigate the noise-free case and we will then discuss the presence of additive white noise. The subspace shift-invariance property of \mathbf{H} , described by equation (3.3), simply implies that the column vectors of \mathbf{U}_K^\uparrow span exactly the same subspace as the column vectors of \mathbf{U}_K^\downarrow . Therefore, in the special case where $k = K$, we have that \mathbf{U}_k^{tb} is a rank- K matrix. Consequently its last K singular values $\gamma_{K+1}, \dots, \gamma_{2K}$ are equal to zero. Assume $L-1 \leq 2K$ and let

$$\mathbf{U}_K^{\text{tb}} = \mathbf{Y}\mathbf{\Gamma}\mathbf{W}^{\text{H}} \quad (3.9)$$

be the SVD of \mathbf{U}_K^{tb} where $\mathbf{Y} \in \mathbb{C}^{(L-1) \times 2K}$, $\mathbf{W} \in \mathbb{C}^{2K \times 2K}$ and $\mathbf{\Gamma} = \text{diag}\{\gamma_1, \dots, \gamma_{2K}\}$, then we have $\{\gamma_i\}_{i=K+1}^{2K} = 0$. We will now investigate the case $k \neq K$ to see what happens to the rank in case of over- and underestimation. Our observations will lead to a new algorithm for the estimation of K .

3.3.1 Overestimation of the model order

This is the case when $k > K$. Consider the following stacked $((L-1) \times 2(K+p))$ -matrix:

$$\mathbf{U}_{K+p}^{\text{tb}} = \left[\mathbf{U}_K^\uparrow \mid U_{K+1}^\uparrow \dots U_{K+p}^\uparrow \mid \mathbf{U}_K^\downarrow \mid U_{K+1}^\downarrow \dots U_{K+p}^\downarrow \right]. \quad (3.10)$$

In general, the subspace spanned by U_{K+1}, \dots, U_{K+p} has no shift invariance properties. Neither does the subspace spanned by $U_{K+1}^\uparrow, \dots, U_{K+p}^\uparrow, U_{K+1}^\downarrow, \dots, U_{K+p}^\downarrow$ have a component in the column space of \mathbf{U}_K^\uparrow (or \mathbf{U}_K^\downarrow). This means that, in general, adding a column to \mathbf{U} increases the rank of \mathbf{U}^{tb} by two. Let us call a property generic when it holds with probability one when the parameters of the problem are drawn from a continuous probability density function. Then we have the following theorem:

Proposition 3.1 *Generically, $\mathbf{U}_{K+p}^{\text{tb}}$, defined by (3.10), with $p \in \mathbb{N}$, is a rank- $(K+2p)$ matrix for $K+2p \leq L-1$ and a rank- $(L-1)$ matrix for $K+2p \geq L-1$.*

Proof: We prove the theorem by showing that, for $p \in \mathbb{N}$, the matrix $\tilde{\mathbf{U}}_{K+p}^{\text{tb}} = \left[\mathbf{U}_{K+p}^\uparrow \mid U_{K+1}^\downarrow \mid \dots \mid U_{K+p}^\downarrow \right]$ is generically full rank.

The proof is by induction. For $p = 0$, we have that every $(K \times K)$ minor of $\tilde{\mathbf{U}}_K^{\text{tb}}$ is different from zero [37]. Now assume that, for $l = 0, 1, 2, \dots, p-1$, every $((K+2l) \times (K+2l))$ minor of $\tilde{\mathbf{U}}_{K+l}^{\text{tb}}$ is generically different from zero. Then we show that every $((K+2p) \times (K+2p))$ minor of $\tilde{\mathbf{U}}_{K+p}^{\text{tb}}$ is also generically different from zero if $K+2p \leq L-1$.

First we prove this for nonsingular, but not necessarily column-wise orthonormal, matrices \mathbf{U}_{K+p} of which the first K columns span the column space of \mathbf{S} . Assume that the opposite holds true. Without loss of generality we may assume that the upper $((K+2p) \times (K+2p))$ submatrix $\mathbf{A}(0,0)$ of $\tilde{\mathbf{U}}_{K+p}^{\text{tb}}$ is singular. Now let us perturb the entries $u_{1,K+p}$ and $u_{K+2p,K+p}$ of \mathbf{U}_{K+p} : $u_{1,K+p}(x) = u_{1,K+p}(0) + x$ and $u_{K+2p,K+p}(y) = u_{K+2p,K+p}(0) + y$. Let $\mathbf{A}(x,y)$ be defined accordingly. Then we have:

$$\det(\mathbf{A}(x,y)) = \det(\mathbf{A}(0,0)) + xC_x + yC_y + xyC_{xy},$$

in which C_x, C_y are the cofactors of $u_{1,K+p}$ and $u_{K+2p,K+p}$ in $\mathbf{A}(0,0)$, and in which C_{xy} is (possibly up to the sign) equal to the determinant of the submatrix of $\mathbf{A}(0,0)$ obtained by deleting the rows and columns of $u_{1,K+p}$ and $u_{K+2p,K+p}$. Since according to the induction hypothesis C_{xy} is generically different from zero, $\det(\mathbf{A}(x,y))$ is too.

Now let us take into account that \mathbf{U}_{K+p} is column-wise orthonormal. Assume again that the upper $((K+2p) \times (K+2p))$ submatrix $\mathbf{A}(0,0)$ of $\tilde{\mathbf{U}}_{K+p}^{\text{tb}}$ is singular. Perturbation of \mathbf{U}_{K+p} , as in the previous paragraph, leads to a submatrix $\mathbf{A}(x,y)$ that is generically nonsingular. Hence, the upper $((K+2p) \times (2K+2p))$ submatrix \mathbf{B} of $\mathbf{U}_{K+p}^{\text{tb}}$ is generically rank- $(K+2p)$. However, the perturbation of \mathbf{U}_{K+p} may have destroyed the orthonormality of the columns of \mathbf{U}_{K+p} . Let us restore the orthonormality by Gram-Schmidt orthogonalization:

$$\mathbf{U}'_{K+p} = [\mathbf{U}_{K+p-1} \mid \mathbf{U}_{K+p}(x,y)] \cdot \mathbf{R},$$

in which \mathbf{R} is nonsingular upper triangular and in which \mathbf{U}'_{K+p} is column-wise orthonormal. Consider matrices \mathbf{A}' and \mathbf{B}' , built from \mathbf{U}'_{K+p} rather than \mathbf{U}_{K+p} . We have:

$$\mathbf{B}' = \mathbf{B} \cdot \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}.$$

Since \mathbf{R} is nonsingular and \mathbf{B} is generically rank- $(K+2p)$, \mathbf{B}' is generically rank- $(K+2p)$. Since the submatrices \mathbf{U}_{K+1} and \mathbf{U}_{K+1} of \mathbf{B}' have the same column space, \mathbf{A}' is generically full rank. We conclude that a perturbation of \mathbf{U}_{K+p} leads to a column-wise orthonormal matrix \mathbf{U}'_{K+p} for which holds that every $((K+2p) \times (K+2p))$ minor of $\tilde{\mathbf{U}}_{K+p}^{\text{tb}}$ is different from zero, if $K+2p \leq L-1$. In other words, the $((K+2p) \times (K+2p))$ minors of $\tilde{\mathbf{U}}_{K+p}^{\text{tb}}$ are generically different from zero, if $K+2p \leq L-1$.

Finally, we consider the case $K+2p \geq L-1$. In this case, the rank of $\tilde{\mathbf{U}}_{K+p}^{\text{tb}}$ cannot be higher than $L-1$. If $L-1$ is even, then the maximal rank is reached for some value of p , as demonstrated above. For $L-1$ odd, a similar reasoning may be followed to obtain that generically $\text{rank}(\tilde{\mathbf{U}}_{K+p}^{\text{tb}}) = L-1$. \square

3.3.2 Underestimation of the model order

This is the case where $k < K$. In general, the k columns of \mathbf{U}_k contain contributions of all K Vandermonde vectors. This makes the columns of \mathbf{U}_k^\uparrow and $\mathbf{U}_{k\downarrow}$ span a $(2k)$ -dimensional subspace, if $2k \leq K$, or a K -dimensional subspace, if $2k \geq K$. We have the following theorem:

Proposition 3.2 *If $k < K$, the matrix \mathbf{U}_k^{tb} is generically full rank.*

Proof: We first prove that, if $k < K$ and $2k \leq K$, the matrix \mathbf{U}_k^{tb} is generically rank- $(2k)$, i.e., full rank. The proof is by induction.

Let us first check that $\begin{bmatrix} U_1^\uparrow & U_{1\downarrow} \end{bmatrix}$ is generically rank-2, for U_1 a vector in the column span of \mathbf{Y} . We have $U_1 = \mathbf{S}\mathbf{V}$, for some $\mathbf{V} = [v_1 v_2 \cdots v_K]^\top \in \mathbb{C}^K$. Hence we also have $U_{1\downarrow} = \mathbf{S}_\downarrow \mathbf{V}$ and $U_1^\uparrow = \mathbf{S}^\uparrow \mathbf{V} = \mathbf{S}_\downarrow \mathbf{D}\mathbf{V}$. U_1^\uparrow and $U_{1\downarrow}$ are linearly independent iff $\mathbf{D}\mathbf{V}$ and \mathbf{V} are linearly independent. Generic linear independence is guaranteed by the fact that the upper (2×2) minor of the matrix $[\mathbf{V} \ \mathbf{D}\mathbf{V}]$, equal to $v_1 v_2 (z_2 - z_1)$, is generically different from zero.

Now we prove the induction step, i.e., assuming that \mathbf{U}_l^{tb} is generically rank- $(2l)$ for $l = 1, 2, \dots, k-1$, we show that \mathbf{U}_k^{tb} is generically rank- $(2k)$ if $2k \leq K$. We first prove that this holds true if we don't take into account that \mathbf{U}_k is column-wise orthonormal. In other words, starting from the induction hypothesis, we prove that $\mathbf{W}_k^{\text{tb}} = \begin{bmatrix} \mathbf{W}_k^\uparrow & \mathbf{W}_{k\downarrow} \end{bmatrix}$ is generically rank- $(2k)$, where \mathbf{W}_k is a matrix of which the k linearly independent columns, which are not necessarily orthonormal, are in the column span of \mathbf{U}_k .

We have $\mathbf{W}_k = \mathbf{S}\mathbf{V}$, for some non-singular matrix $\mathbf{V} \in \mathbb{C}^{K \times k}$. Hence we also have $\mathbf{W}_{k\downarrow} = \mathbf{S}_\downarrow \mathbf{V}$ and $\mathbf{W}_k^\uparrow = \mathbf{S}^\uparrow \mathbf{V} = \mathbf{S}_\downarrow \mathbf{D}\mathbf{V}$. The matrix \mathbf{W}_k^{tb} is full rank iff the matrix $\mathbf{B}(0) = [\mathbf{V} \ \mathbf{D}\mathbf{V}]$ is full rank. Let us assume that the latter matrix is singular. Then its upper $(2k \times 2k)$ part $\mathbf{A}(0)$ is singular too. Let us perturb the entry v_{11} of \mathbf{V} : $v_{11}(x) = v_{11}(0) + x$. Let the matrix $\mathbf{A}(x)$ be defined accordingly. Then we have

$$\det(\mathbf{A}(x)) = \det(\mathbf{A}(0)) + x(C_{11} + z_1 C_{1k}) = x(C_{11} + z_1 C_{1k}),$$

in which C_{pq} is the cofactor of $(\mathbf{A}(0))_{pq}$. This determinant is generically different from zero under the induction hypothesis. In the odd case where it is zero, a small perturbation of one of the entries $(\mathbf{A}(x))_{21}, \dots, (\mathbf{A}(x))_{2k,1}$ changes C_{1k} , so that $\det(\mathbf{A}(x))$ becomes different from zero. This means that $\det(\mathbf{A}(x))$ is different from zero with probability one. Hence \mathbf{W}_k^{tb} is rank- $(2k)$ with probability one.

Now let us take into account that \mathbf{U}_k is column-wise orthonormal. Let us assume that \mathbf{U}_k^{tb} is singular, which is true iff $(\mathbf{S}_\downarrow)^\dagger \mathbf{U}_k^{\text{tb}}$ is singular. The latter

matrix takes the form $\mathbf{B}(0) = [\mathbf{V} \mathbf{D}\mathbf{V}]$, with \mathbf{V} some non-singular matrix in $\mathbb{C}^{K \times k}$. A generic perturbation, characterized by \mathbf{V}' , makes \mathbf{B} full rank, like in the previous paragraph. However, $\mathbf{S}\mathbf{V}'$ is not necessarily column-wise orthonormal anymore. Let us now define a matrix \mathbf{U}'_k , of which the columns form an orthonormal basis of the column span of $\mathbf{S}\mathbf{V}'$. \mathbf{B} is full rank iff \mathbf{U}'_k is. The column span of \mathbf{U}'_k is a perturbation of the column span of \mathbf{U}_k . We conclude that a perturbation of \mathbf{U}_k leads to a column-wise orthonormal matrix \mathbf{U}'_k for which \mathbf{U}'_k is full rank. Hence \mathbf{U}'_k is rank- $(2k)$ with probability one.

Finally we consider the case $k < K$ and $2k > K$. In this case the rank of \mathbf{U}_k cannot be higher than K . If K is even, then we have from the previous case that generically $\text{rank}(\mathbf{U}_k) = K$, by just considering a submatrix of \mathbf{U}_k that contains $K/2$ columns. If K is odd, then we can follow a similar reasoning as above to obtain that generically $\text{rank}(\mathbf{U}_k) = K$. \square

3.3.3 Order determination

We have seen that, in the noise-free case, the $(L-1) \times 2k$ matrix $[\mathbf{U}_k^\uparrow \mathbf{U}_k^\downarrow]$ is rank- k only for $k = K$. This implies that, for $k = K$, the k last singular values of \mathbf{U}_k are zero. However, in the presence of noise, the column vectors of \mathbf{U} are perturbed and the matrix $[\mathbf{U}_k^\uparrow \mathbf{U}_k^\downarrow]$ is never rank deficient, for any k . If the SVD of \mathbf{H} allows a good separation of the noise and the signal subspace then the k last singular values of $[\mathbf{U}_k^\uparrow \mathbf{U}_k^\downarrow]$ will still be small while the reasoning in the previous sections still holds. This means that K can still be estimated by inspection of the amplitude of the k smallest singular values. We propose the following model order estimator:

$$\hat{K} = \arg \max_k \frac{1}{E(k)}, \quad 1 \leq k < (L-1)/2 \quad (3.11)$$

with

$$E(k) = \frac{1}{k} \sum_{i=k+1}^{2k} \gamma_i \quad (3.12)$$

where γ_i are the singular values of $[\mathbf{U}_k^\uparrow \mathbf{U}_k^\downarrow]$. This method is called SAMOS standing for Subspace-based Automatic Order Selection. It is quite simple to implement as outlined below.

Algorithm 3.3

SAMOS

Input: Data samples x_n , $n = 0, \dots, N-1$.

Output: Estimate of Model order, \hat{K} .

Step 1: Arrange the data points x_n , $n = 0, \dots, N - 1$ in a $(L \times M)$ -Hankel matrix \mathbf{H} , $N = L + M - 1$, $L > K$ as follows:

$$\mathbf{H} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{M-1} \\ x_1 & x_2 & \ddots & \cdots & \vdots \\ x_2 & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2} \\ x_{L-1} & \cdots & \cdots & x_{N-2} & x_{N-1} \end{pmatrix}$$

In order to have a good approximation of the signal subspace it is recommended to take a Hankel matrix \mathbf{H} as square as possible [85].

Step 2: Compute the SVD of \mathbf{H}

$$\mathbf{H} = \mathbf{U}_{L \times \min(L,M)} \mathbf{\Sigma}_{\min(L,M) \times \min(L,M)} \mathbf{V}_{M \times \min(L,M)}^H$$

and get \mathbf{U}

Step 3: For k varying from 1 to $\frac{L-1}{2}$ with unit step

- Form the following matrix \mathbf{U}_k^{tb}

$$\mathbf{U}_k^{\text{tb}} = [\mathbf{U}_k^\uparrow | \mathbf{U}_k^\downarrow]$$

where $\mathbf{U}_k = [U_1 U_2 \cdots U_k]$ contains the k first left singular vectors of \mathbf{H} . The matrices \mathbf{U}_k^\uparrow and \mathbf{U}_k^\downarrow are derived from \mathbf{U}_k by omitting its first and last row respectively.

- Compute the SVD of \mathbf{U}_k^{tb}

$$\mathbf{U}_K^{\text{tb}} = \mathbf{Y} \mathbf{\Gamma} \mathbf{W}^H$$

- Compute the sum of the last K entries of the matrix $\mathbf{\Gamma} = \text{diag}\{\gamma_1, \dots, \gamma_{2k}\}$

$$E(k) = \frac{1}{k} \sum_{i=k+1}^{2k} \gamma_i$$

Step 4: Finally compute the estimation of the model order K

$$\hat{K} = \arg \max_k \frac{1}{E(k)}, \quad 1 \leq k < (L - 1)/2$$

	ν [Hz]	α [s ⁻¹]	a [a.u.]
Peak 1	0.2	0.01	1
Peak 2	0.3	0.02	1
Peak 3	-0.2	0.1	2
Peak 4	0.4	0.05	1
Peak 5	0.35	0.03	1

Table 3.1: Set of parameters for the five peak example. We refer to the set of parameters $\{\nu_k, \alpha_k, a_k\}$ as Peak k .

3.4 Results

In the first simulation we generated a signal of 129 samples based on a five poles model whose parameters are given in Table 3.1. For clarity, we recall that we can refer to a pole as a peak in the frequency domain as explained in Section 1.2. In order to have a good separation between the noise and the signal subspace it is preferable to use a Hankel matrix that is as square as possible [85]. Therefore we arranged the signal in a 65×65 Hankel matrix. Moreover the SVD performs a good separation between the noise and the signal subspace if the dimension of the signal subspace is reasonably small. For a $(L \times L)$ -Hankel data matrix it is reasonable not to have a dominant subspace larger than $L/4$ in order to have a good separation between the noise and the signal subspace. In our simulation we have $K = 5$ and we perform a search for $k = 1, \dots, 16$. The comparison is performed by means of Monte Carlo simulations consisting of 250 independent realizations for each noise level. The result is expressed in terms of the percentage of runs in which the model order was successfully determined as a function of the noise standard deviation σ . Figure 3.1 compares the performance of the two subspace-based algorithms in the case of additive WGN. It is clear that the new SVD-based rank detection algorithm yields a better performance; the probability of correct detection is higher over the full noise level range. To take a concrete example, ESTER achieves a 90 % success rate for $\sigma = 0.25$ while the new method SAMOS has the same performance for $\sigma = 0.35$, which represents a gain of about 3dB. We recall that ESTER already outperforms ITC's. In figure 3.2 we displayed the comparison in the case of non-Gaussian noise. The parameters are similar than these of the previous simulation except the noise which is drawn from a uniform distribution. Although the main remark is that SAMOS still performs better than ESTER, it is clear that SAMOS has approximately the same performance as for Gaussian noise while ESTER exhibits some performance degradation.

The next simulation compares the robustness of the two methods when two signal poles are getting very close. We use a two-pole model $x_n = \exp\{-0.02 +$

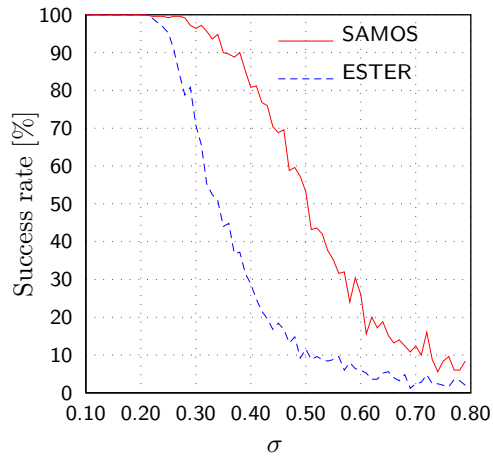


Figure 3.1: Comparison of the performance in terms of the probability of correct determination of the true model order ($K=5$) in the case of WGN for varying noise standard deviation σ .

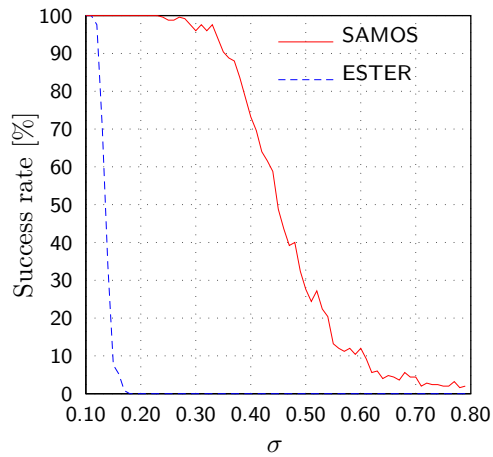


Figure 3.2: Comparison of the performance in terms of the probability of correct determination of the true model order ($K=5$) in the presence of non-Gaussian noise for varying noise standard deviation σ . The noise variable is i.i.d. and is drawn from a uniform distribution.

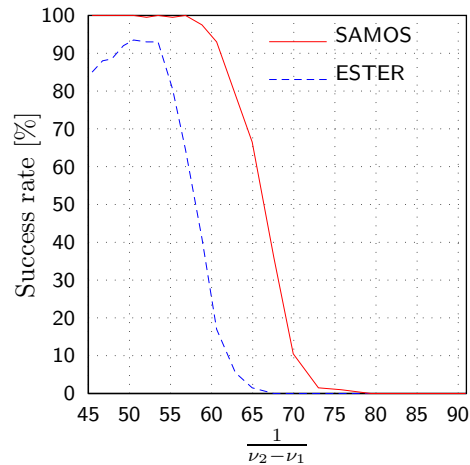


Figure 3.3: Comparison of the performance in terms of the probability of correct determination of the true model order ($K=2$) for varying frequency difference $\nu_2 - \nu_1$.

$2j\pi 0.2)n\} + \exp\{-0.02 + 2j\pi\nu_2)n\}$, $n = 0, \dots, 64$ ($\Delta t = 1$), where the frequency ν_2 of the second pole varies between 0.222 and 0.205. The noise has been chosen Gaussian in order to get the best performance of ESTER. The noise standard deviation is set to 0.4. The result is displayed in Figure 3.3. We observe again a higher robustness of the new method. Also, in this example, ESTER does never achieve a 100 % success rate.

3.5 Application examples

This method is applicable as far as the (complex) EDS model is concerned. Note that the purely (complex) sinusoidal model is a particular case of the (complex) EDS model (undamped), as such the method is perfectly applicable to signals containing pure sinusoids. Thus, as discussed in Chapter 1, there is a wide variety of potential applications. This technique is especially useful when subspace-based methods such as HTLS or TLS-ESPRIT are used. Let us consider the HTLS method. The model order K is assumed to be entered by the user since it is required in step 3 of Algorithm 2.2 (HTLS). However, as shown in steps 1-2 of Algorithm 2.2, the SVD of the data Hankel matrix \mathbf{H} is performed. Therefore we already have the matrix \mathbf{U} that is required for steps 3-4 of Algorithm 3.3 (SAMOS). Actually the two first steps of both algorithms are identical. As a result Algorithm 3.3 can be inserted between step 2 and step 3 of Algorithm 2.2.

3.5.1 Application to dominant spindle frequency detection

A nice application of the combination of both algorithms to biomedical signal processing has been investigated in [29]. In this paper the kernel problem is the determination of the dominant spindle frequency from EEG recordings. In order to monitor the sleep stage of patients, electroencephalograms (EEG) are recorded during night. In these sleep EEG's, some transient artifacts having a frequency ranging from 10.5 Hz to 16 Hz may appear. These very important sleep micro-events are called spindles. Depending on their frequency, specialists might be able to determine the sleep stage of a person based on which they can diagnose if this person suffers from sleep disorders. In an ideal situation, these detected artifacts have a single frequency. However, in real-world signals the spindles are corrupted by unwanted harmonics having slightly different frequencies. Therefore, before applying HTLS to determine the spindle frequency, an efficient determination of the number of sinusoids is necessary. Several simulations have been carried out in which a dominant sinusoid was corrupted either by noise or by a sinusoid whose frequency was close to the dominant one, or by both. The combination of SAMOS and HTLS has been successfully applied [29] allowing an accurate determination of the spindle frequencies.

3.5.2 Application to material health monitoring

In Figure 1.1, an experimental setup is shown whose aim is to detect damage by means of two types of sensors: a piezoelectric sensor and an optical fiber sensor. In figure 3.4 we show an example of a transient signal received simultaneously by these two sensors. Each signal consists of 751 samples and the sampling rate is 1 MHz. The optical signal is much more noisy than the ultrasonic one but, in both cases, it is rather difficult to accurately determine the number of peaks by just having a look in the frequency domain as shown in Figure 3.5. For instance it may be possible that a peak is generated by two closely spaced sinusoids resulting in a slightly wider peak. Moreover the noise also gives rise to some peaks in the frequency domain. We recall that in the noise-free case the number of peaks in the frequency domain is equal to the number of EDSs. Notice that the time-domain optical signal has an offset. To get rid of the high static component in the frequency domain resulting from this offset, we subtracted the mean of the signal.

A first observation is that the singular values of the (376×376) -Hankel data matrices exhibit no gap (see Figure 3.6). It indicates that there is no clear separation between the signal and the noise subspace, which is in general an issue in the estimation of the model order. In order to determine the model orders we applied SAMOS and ESTER to these two signals. The results are summarized in Table 3.2. Compared to SAMOS, ESTER clearly underestimates the model order. If we refer to the spectra in Figure 3.5, the model order of the optical

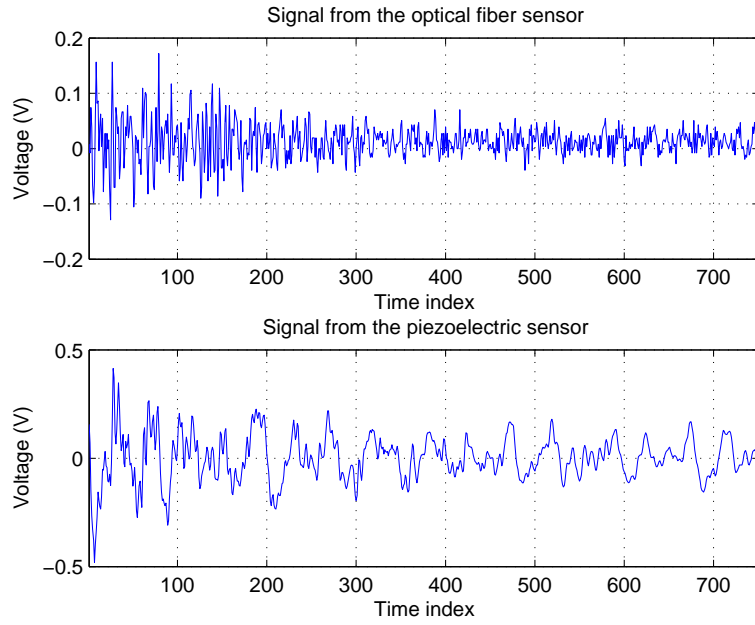


Figure 3.4: Example of signals obtained by means of the experimental setup depicted in Figure 1.1

	Optical signal	AE signal
SAMOS	15	7
ESTER	5	1

Table 3.2: Estimate of the model order of the optical and the AE signal yielded by SAMOS and ESTER. The corresponding actual value of \hat{K} (3.11) is twice the values indicated in this table.

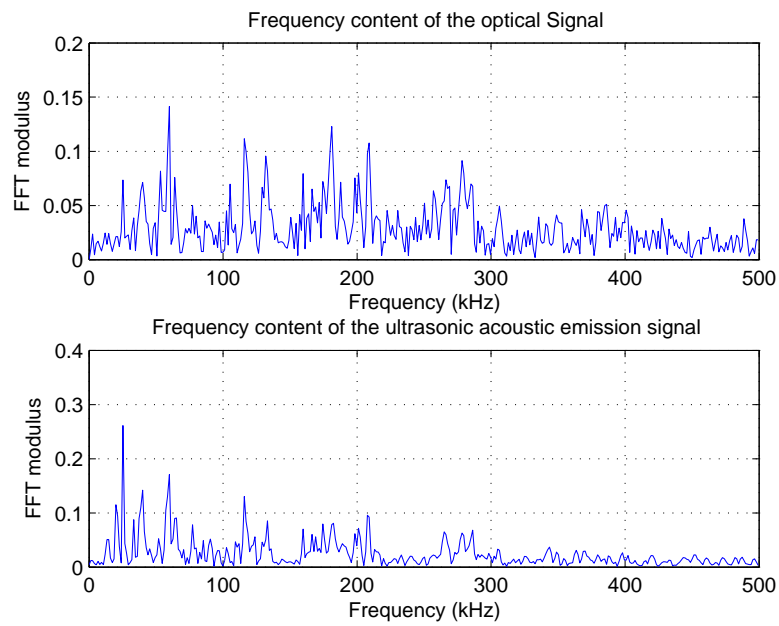


Figure 3.5: Normalized frequency spectra of the signals shown in Figure 3.4.

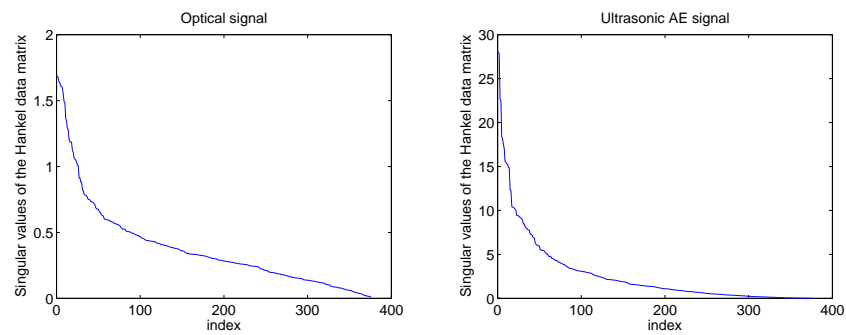


Figure 3.6: Scree plots of the singular values of the Hankel data matrices for the ultrasonic and the optical signal.

signal seems to be quite well estimated by SAMOS, whereas the model order of the AE signal seems to be largely underestimated. In the latter case, we may take the model order corresponding to the second highest value indicated by (3.12). We obtain a model order estimate of 22 ($\hat{K} = 44$) with SAMOS and 2 with ESTER. In figure 3.7 the reconstruction of these signals using HTLS is shown. In this example SAMOS gives more satisfactory results than ESTER but it still underestimates the model order, especially for the ultrasonic signal. It is likely that the dominant subspace of the matrix $\mathbf{U}_{(L-1)/2}^{\text{tb}}$ (3.9) is splitted into two subspaces that are (almost) orthogonal to each others. The first has dimension 14 and the second has dimension 30. Therefore the function $\frac{1}{E(k)}$ has a high value for $k = 14$ and $k = 44$. However the 14-dimensional subspace is less noisy than the 30-dimensional one. That is the reason why $\frac{1}{E(k)}$ has a higher value for $k = 14$ than for $k = 44$. Better results may be obtained by reducing the interval of search. However this prior knowledge is not always available.

3.6 Conclusions

We have presented a subspace-based method for determining the number of exponentially damped sinusoids present in a signal. In our approach we made no assumption about the gaussianity of the noise because we only use the signal subspace for the determination of the model order. This method, which merely assumes a good separation of the signal and noise subspace, is based on the shift invariance property of the dominant subspace of the Hankel data matrix. It has been successfully compared to the ESTER method, which was also based on the shift invariance property.

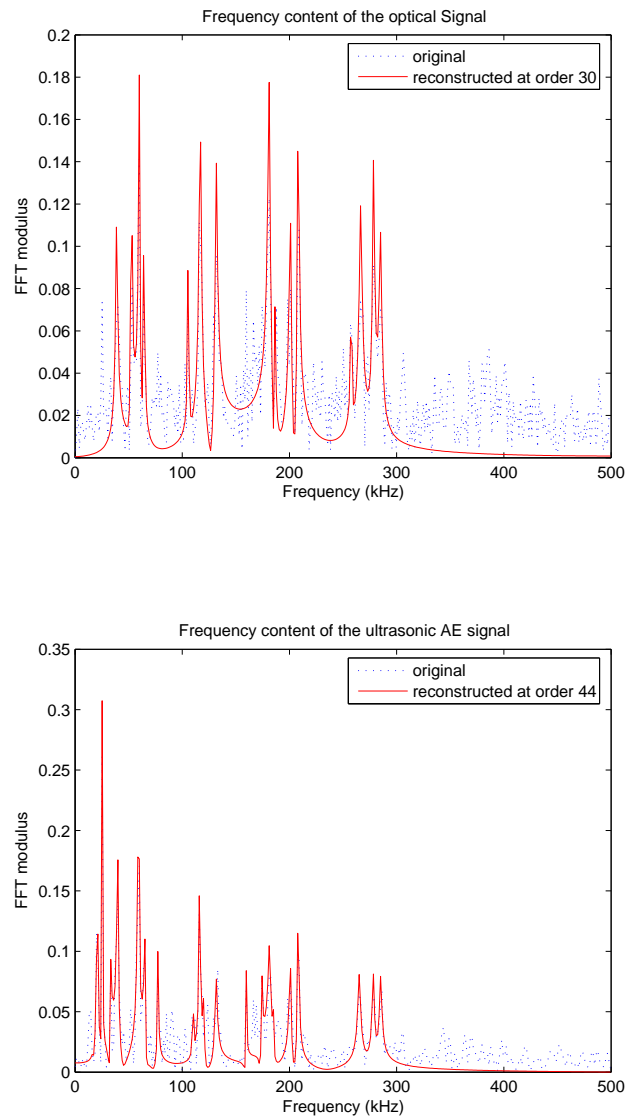


Figure 3.7: Reconstruction of the signal using the HTLS method.

Chapter 4

Common pole estimation

In this chapter we develop a technique for detecting and retrieving the common harmonics in a multi-channel setup. The different (complex) signals are arranged in a set of Hankel matrices. First we compute the singular value decomposition (SVD) of these matrices in order to denoise and normalize the relevant subspaces. Then a second SVD is applied to detect and select the common subspace. The poles of the harmonics are computed from the common subspace using the Total Least Squares (TLS) technique. This method is very flexible, robust and outperforms existing subspace-based methods. Moreover it can be applied to an arbitrary number of channels.

4.1 Introduction

Many real-world signals are naturally modelled as a sum of exponentially damped sinusoids. This is, for instance, the case in MRS where the quantification of the complex time-domain signals [85, 86, 13] is of great importance for brain tumor detection, in material health monitoring where the model parameters help to characterize the damage occurring in a structure [67], or in audio processing where such a model is used for low bit rate audio compression [9].

In general, this type of signals exhibits a rank deficient correlation matrix, and therefore lends itself to subspace-based processing. One of the most efficient techniques consists in first computing the *singular value decomposition* (SVD) of the Hankel data matrix; then a system of equations is built via a similarity transformation of the dominant subspace to a Vandermonde vector basis by means of the shift-invariance property; finally the Total Least Squares (TLS) solution of this system is computed and yields the frequencies and damping

factors. This method is called HTLS [85], as presented in Section 2.2. In the subspace-based processing framework, a multi-channel set-up makes it possible to extract the common information in a more accurate way if the signals are contaminated by noise. In this chapter we address the case in which we have several channels each of which can be modelled as a sum of discrete-time complex exponentials. In Section 2.3, we reviewed and gave an outline of a multi-channel harmonic retrieval algorithm, called HTLSstack. With the HTLSstack technique it is possible to estimate all the discrete-time complex exponential signals present in a multi-channel signal. However, it can be of great interest to quantify and extract only the common information present in the multi-channel signal in the case where the channels are contaminated by different undesired harmonics. For instance in biomedical engineering, the issue of removing undesirable artifact in multi-channel electroencephalography (EEG) recordings has been widely addressed [31, 33, 35, 39, 53, 96] This is possible by determining the subspace common to all the data matrices. In the two-channel case several methods already exist. For instance in [26], an algorithm based on the QR decomposition [26, 82] for determining the common subspace of two matrices via canonical angles is provided. It is possible, from this subspace, to retrieve the parameters of interest by computing the TLS solution. Alternatively, in [51] a method, called HTLS-SEP, based on SVD and TLS, is used to estimate the complex exponentials common to two signals. However, these two methods only apply in the two-channel case.

The method which is studied in this chapter applies to an arbitrary number of channels. It consists of using unitary matrices for determining a set of linearly independent vectors spanning the common space, if it exists. Then the parameters of the common complex exponentials are computed via TLS. This is achieved without setting a too severe constraint on the determination of the model order of each channel. This method has several big advantages. The main one is that, to our knowledge, there exists no competing approach for a number of channel higher than two. It has, as most of subspace-based methods, a virtually infinite resolution. Finally we shall mention its flexibility, simplicity of use and implementation.

In Section 4.2, a description of the problem is given. The underlying theory is explained in Section 4.3. Some practical issues are discussed and the algorithm is provided in Section 4.3.1. In Section 4.3.2, time complexity of the algorithm is briefly discussed. Then the algorithm is applied to simulated signals in Section 4.4, to EEG monitoring in Section 4.5 and to acoustic emission monitoring in Section 4.6. Finally we conclude the chapter in Section 4.7.

4.2 Problem statement

Let us first assume exact data. The influence of the noise will be discussed later on. Given a G -channel discrete-time signal $x_n^{(g)}$ where $g = 1, \dots, G$, denotes the channel number and $n = 0, \dots, N - 1$, is the time index. Let each signal $x_n^{(g)}$, $n = 0, \dots, N - 1$, be modelled as a finite sum of K_g weighted discrete-time complex exponentials (2.2), K_g being the order of the model in channel g , and let these models have κ common poles $\{z_k\}$, ($k = 1, \dots, \kappa$). For each channel g we can write:

$$x_n^{(g)} = \sum_{k=1}^{K_g} c_{g,k} z_{g,k}^n = \sum_{k=1}^{\kappa} \bar{c}_{g,k} z_k^n + \sum_{k=1}^{K_g - \kappa} \tilde{c}_{g,k} \tilde{z}_{g,k}^n \quad n = 0, \dots, N - 1, \quad (4.1)$$

with $\{c_{g,k}\} = \{\bar{c}_{g,k}, \tilde{c}_{g,k}\}$, $\{z_{g,k}\} = \{z_k, \tilde{z}_{g,k}\}$ and $\tilde{z}_{g,k} \neq \tilde{z}_{g',k'}$ for all $k \neq k'$ or $g \neq g'$. Notice that the complex amplitudes $\bar{c}_{g,k}$ related to the common poles z_k are not necessarily equal.

4.3 Determining a set of vectors spanning the common subspace

In the noise-free case we have, in analogy with (2.4), (2.5),

$$\mathbf{H}_g \stackrel{\text{VDMD}}{=} \mathbf{S}_g \mathbf{C}_g \mathbf{T}_g^T \stackrel{\text{SVD}}{=} \hat{\mathbf{U}}_g \hat{\mathbf{\Sigma}}_g \hat{\mathbf{V}}_g^H \quad g = 1, \dots, G, \quad (4.2)$$

where $\mathbf{H}_g \in \mathbb{C}^{L \times M_g}$, $\mathbf{S}_g \in \mathbb{C}^{L \times K_g}$, $\mathbf{C}_g \in \mathbb{C}^{K_g \times K_g}$, $\mathbf{T}_g \in \mathbb{C}^{M_g \times K_g}$, $\hat{\mathbf{U}}_g \in \mathbb{C}^{L \times K_g}$, $\hat{\mathbf{\Sigma}}_g \in \mathbb{R}^{K_g \times K_g}$, $\hat{\mathbf{V}}_g \in \mathbb{C}^{M_g \times K_g}$. The Vandermonde matrices \mathbf{S}_g , ($g=1, \dots, G$), have κ common vectors:

$$\mathbf{S}_g = \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ z_1 & \dots & z_\kappa & \tilde{z}_{g,1} & \dots & \tilde{z}_{g,K_g - \kappa} \\ z_1^2 & \dots & z_\kappa^2 & \tilde{z}_{g,1}^2 & \dots & \tilde{z}_{g,K_g - \kappa}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ z_1^{L-1} & \dots & z_\kappa^{L-1} & \tilde{z}_{g,1}^{L-1} & \dots & \tilde{z}_{g,K_g - \kappa}^{L-1} \end{pmatrix} = [\mathbf{S} \mid \tilde{\mathbf{S}}_g] \quad (4.3)$$

where the generators of \mathbf{S} are the common signal poles, and the generators of $\tilde{\mathbf{S}}_g$ are signal poles that are not present in all channels.

We recall that, generically, none of the column vectors of a given Vandermonde matrix \mathbf{S}_g can be expressed as a linear combination of the other ones when the generators are different. In other words, the column vectors of any \mathbf{S}_g can be seen as a nonorthogonal basis of a complex vector subspace of dimension

K_g belonging to a complex vector space of dimension L . It follows that the intersection of the G subspaces, spanned by the column vectors of each matrix \mathbf{S}_g , contains the column vectors of \mathbf{S} , but not more than that. Thus, the column vectors of \mathbf{S} form a nonorthogonal basis for the common subspace C . We formally write:

$$\bigcap_{g=1}^G \text{Span}(s_{g,1}, \dots, s_{g,K_g}) = C \quad (4.4)$$

with

$$\text{Dim}(C) = \kappa \quad (4.5)$$

where $s_{g,k}$ is the k th column vector of \mathbf{S}_g . Obviously, the column vectors of the semi-unitary matrix $\widehat{\mathbf{U}}_g$ and the column vectors of the Vandermonde matrix \mathbf{S}_g span exactly the same subspace. Therefore, property (4.4) also applies to the semi-unitary matrices $\widehat{\mathbf{U}}_g$:

$$\bigcap_{g=1}^G \text{Span}(u_{g,1}, \dots, u_{g,K_g}) = C \quad (4.6)$$

where $u_{g,k}$ is the k th column vector of $\widehat{\mathbf{U}}_g$. In what follows, we explain how to find the κ orthogonal vectors spanning the subspace C with a minimal number of operations and κ unknown. The process will rely on the proof of the existence of several unitary and semi-unitary matrices, that won't be explicitly written. From the equations (4.5) and (4.6) it follows that, by means of simple rotations of the K_g vectors $\{u_{g,k}\}$, it is possible to find a basis of κ vectors, common to the G subspaces. Therefore, if we stack the G matrices $\widehat{\mathbf{U}}_g$ ($g = 1, \dots, G$) in a matrix $\mathbf{U}_{\text{tot}} \in \mathbb{C}^{L \times \sum_{g=1}^G K_g}$:

$$\mathbf{U}_{\text{tot}} = \left[\widehat{\mathbf{U}}_1 \mid \widehat{\mathbf{U}}_2 \mid \dots \mid \widehat{\mathbf{U}}_G \right], \quad (4.7)$$

we can always find a set of G unitary matrices $\mathbf{Q}_g \in \mathbb{R}^{K_g \times K_g}$ ($g = 1, \dots, G$) such that:

$$\begin{aligned} \left[\widehat{\mathbf{U}}_1 \mathbf{Q}_1 \mid \dots \mid \widehat{\mathbf{U}}_G \mathbf{Q}_G \right] &= \left[\widehat{\mathbf{U}}_1 \mid \dots \mid \widehat{\mathbf{U}}_G \right] \mathbf{Q} \\ &= \left[\mathbf{U} \begin{array}{c} \vdots \\ \mathbf{U}_1^\perp \end{array} \mid \dots \mid \mathbf{U} \begin{array}{c} \vdots \\ \mathbf{U}_G^\perp \end{array} \right] \end{aligned} \quad (4.8)$$

where $\mathbf{Q} = \text{diag}\{\mathbf{Q}_1, \dots, \mathbf{Q}_G\}$, $\mathbf{U} \in \mathbb{C}^{L \times \kappa}$ contains the set of column vectors spanning the common subspace, and $\mathbf{U}_g^\perp \in \mathbb{C}^{L \times (K_g - \kappa)}$ span the orthogonal complement of \mathbf{U} in the column space of $\widehat{\mathbf{U}}_g$. By means of a permutation matrix $\mathbf{P} \in \mathbb{R}^{\sum_{g=1}^G K_g \times \sum_{g=1}^G K_g}$, the matrices \mathbf{U} can be moved to the left:

$$\left[\mathbf{U} \begin{array}{c} \vdots \\ \mathbf{U}_1^\perp \end{array} \mid \dots \mid \mathbf{U} \begin{array}{c} \vdots \\ \mathbf{U}_G^\perp \end{array} \right] = \left[\underbrace{\mathbf{U} \begin{array}{c} \vdots \\ \dots \\ \vdots \end{array}}_{G \text{ times}} \mid \underbrace{\mathbf{U}_1^\perp \begin{array}{c} \vdots \\ \dots \\ \vdots \end{array}}_{\mathbf{U}^\perp} \right] \mathbf{P} \quad (4.9)$$

Note that $\mathbf{U}^\perp \in \mathbb{C}^{L \times \sum_{g=1}^G (K_g - \kappa)}$ is a full rank matrix with semi-unitary blocks, that are mutually not necessarily orthogonal. If the SVD of \mathbf{U}^\perp is given by $\mathbf{V}\mathbf{\Sigma}\mathbf{W}^H$, then we obtain:

$$\left[\begin{array}{c|c} \mathbf{U} & \mathbf{U}^\perp \end{array} \right] = \underbrace{\mathbf{U} \cdot \sqrt{G} \mathbf{I} \cdot \left[\begin{array}{c|c} \mathbf{I}/\sqrt{G} & \mathbf{0}_1 \end{array} \right]}_{\text{rows normalisation}} + \mathbf{V} \cdot \mathbf{\Sigma} \cdot \left[\begin{array}{c|c} \mathbf{0}_2 & \mathbf{W}^H \end{array} \right] \quad (4.10)$$

where \mathbf{I} is a $(\kappa \times \kappa)$ -identity matrix, $\mathbf{0}_1 \in \mathbb{C}^{\kappa \times \sum_{g=1}^G (K_g - \kappa)}$, $\mathbf{0}_2 \in \mathbb{C}^{\sum_{g=1}^G (K_g - \kappa) \times \kappa}$, $\mathbf{0}_3 \in \mathbb{C}^{\sum_{g=1}^G (K_g - \kappa) \times G \cdot \kappa}$, $\mathbf{V} \in \mathbb{C}^{L \times \sum_{g=1}^G (K_g - \kappa)}$ is a semi-unitary matrix, $\mathbf{\Sigma} \in \mathbb{R}^{\sum_{g=1}^G (K_g - \kappa) \times \sum_{g=1}^G (K_g - \kappa)}$ is a diagonal matrix, and $\mathbf{W}^H \in \mathbb{C}^{\sum_{g=1}^G (K_g - \kappa) \times \sum_{g=1}^G (K_g - \kappa)}$ is a unitary matrix. Combining (4.8), (4.9), and (4.10), finally yields

$$\mathbf{U}_{\text{tot}} = \underbrace{\left[\begin{array}{c|c} \mathbf{U} & \mathbf{V} \end{array} \right]}_{\tilde{\mathbf{U}}} \underbrace{\left[\begin{array}{c|c} \sqrt{G} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{array} \right]}_{\tilde{\mathbf{\Sigma}}} \underbrace{\left[\begin{array}{c|c} \mathbf{I}/\sqrt{G} \quad \cdots \quad \mathbf{I}/\sqrt{G} & \mathbf{0}_1 \\ \mathbf{0}_3 & \mathbf{W}^H \end{array} \right]}_{\tilde{\mathbf{W}}^H} \mathbf{P}\mathbf{Q}^T \quad (4.11)$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{W}}^H$ are semi-unitary, $\tilde{\mathbf{\Sigma}}$ is diagonal and the diagonal core matrix $\mathbf{\Sigma} = \text{diag}\{\sigma_1, \dots, \sigma_{\sum (K_g - \kappa)}\}$ is such that:

$$\sqrt{G} > \sigma_1 \geq \dots \geq \sigma_{\sum (K_g - \kappa)} \quad (4.12)$$

Thus, we can draw a set of preliminary conclusions. First, it is clear that if there are κ poles common to the G channels, there exists a column subspace common to the G matrices \mathbf{H}_g and therefore the $L \times \sum_g K_g$ matrix \mathbf{U}_{tot} is of rank

$$K_{\mathbf{U}_{\text{tot}}} = \min(L, \sum_g K_g - (G - 1)\kappa). \quad (4.13)$$

Second, it is also clear that

$$\tilde{\mathbf{U}}_{\text{tot}} = \mathbf{U} \cdot \sqrt{G} \mathbf{I} \cdot \left[\begin{array}{c|c} \mathbf{I}/\sqrt{G} & \mathbf{0}_1 \end{array} \right] \mathbf{P}\mathbf{Q}^T \quad (4.14)$$

minimizes the function

$$\left\| \mathbf{U}_{\text{tot}} - \tilde{\mathbf{U}}_{\text{tot}} \right\|^2 \quad (4.15)$$

with

$$\left\| \tilde{\mathbf{U}}_{\text{tot}} \right\|^2 = \kappa \cdot G, \quad (4.16)$$

and hence a set of column vectors spanning the common subspace, and from which the common signal poles can be obtained, can be found in the best rank- κ approximation of \mathbf{U}_{tot} . Finally, the κ dominant singular values of \mathbf{U}_{tot} are all equal to \sqrt{G} .

4.3.1 Discussion and algorithm

In practice the data might be corrupted by noise, and the orders K_g might not be known. For the estimation of K_g one can refer to Chapter 3.

When the model orders are slightly overestimated, a good estimation of the common subspace is usually still possible. Indeed, if for some of the matrices \mathbf{H}_g an additional left singular vector u_{g,K_g+1} is taken into account, the lack of structure in these additional vectors ensures that the principle of derivation above still holds. Of course, when more vectors are added, the probability of creating an artificial intersection increases.

On the other hand, the influence of the noise tends to decrease the singular values that are theoretically equal to \sqrt{G} . Consequently, one has to set a threshold $\delta \in \mathbb{R}$ such that

$$\tilde{\sigma}_1, \dots, \tilde{\sigma}_\kappa \geq \sqrt{G} - \delta, \quad (4.17)$$

$\tilde{\sigma}_1, \dots, \tilde{\sigma}_\kappa$ being the set of singular values corresponding to the common subspace. Here again it might be possible to use the SAMOS algorithm described in Chapter 3 to automatically detect the first G components, provided that the gap between the G th singular value and $\tilde{\sigma}_1$ is high enough.

The matrices \mathbf{U} and \mathbf{V} must be interpreted with care. The columns of \mathbf{U} span the subspace in which lie the *common* Vandermonde vectors. The columns of $[\mathbf{U} \ \mathbf{V}]$ span the subspace in which lie *all* the Vandermonde vectors, the column span of \mathbf{V} contains only a component of the non-common poles.

To summarise, we have described how to retrieve a subspace common to a set of subspaces (4.7)–(4.11). The number of considered subspaces is arbitrary as long as the rank condition (4.13) is preserved. The use of semi-unitary matrices allows the separation and the quantification of the energy related to this common subspace. Also, we have proven that this theory can be easily applied to a set of EDS models having common poles. In this case, the considered subspaces are the dominant column spaces of Hankel matrices and the number of common poles is equal to the dimension of the common subspace (4.3)–(4.6). Since these matrices may be rank deficient we make use of the SVD to capture the dominant subspace (4.2).

This method is called MUSCLE standing for Multi-channel Subspace-based Common pole Estimation. An outline of the algorithm is given below.

Algorithm 4.3
MUSCLE

Input: Given G multi-channel signals $x_n^{(g)}$, $g = 1, \dots, G$ of N samples $n = 0, \dots, N - 1$ with corresponding model order K_g , L and M such that $N = L + M - 1$.

Output: $\hat{a}_k, \hat{\varphi}_k, \hat{\alpha}_k, \hat{\nu}_k, k = 1, \dots, K$.

Step 1: Map each signal into a Hankel matrix: $x_n^{(g)} \xrightarrow{\mathcal{H}_L^M(\cdot)} \mathcal{H}_L^M(x_n^{(g)}) = \mathbf{H}_g$, where $\mathcal{H}_L^M(\cdot)$ denotes the Hankel mapping operator.

Step 2: Compute the SVD of each \mathbf{H}_g separately: $\mathbf{H}_g = \mathbf{U}_g \mathbf{\Sigma}_g \mathbf{V}_g^H$,

Step 3: Retain the K_g leftmost singular vectors of the left singular vectors matrix \mathbf{U}_g : $\mathbf{U}_g (L \times M) \rightarrow \tilde{\mathbf{U}}_g (L \times K_g)$. If K_g is not given, it can for example be determined by using a rank determinator as presented in Chapter 3, a scree plot or an information criterion [44, 75, 102, 103],

Step 4: Stack the vectors obtained in the previous step in a $(L \times \sum_{g=1}^G K_g)$ -

$$\text{matrix: } \mathbf{U}_{\text{tot}} = \left[\tilde{\mathbf{U}}_1 | \tilde{\mathbf{U}}_2 | \dots | \tilde{\mathbf{U}}_G \right],$$

Step 5: Compute the SVD of \mathbf{U}_{tot} : $\mathbf{U}_{\text{tot}} = \tilde{\mathbf{V}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{W}}^H$,

Step 6: Display the entries of $\tilde{\mathbf{\Sigma}}$: if there exist κ common poles, the κ first singular values are equal to \sqrt{G} ,

Step 7: Form a matrix $\hat{\mathbf{U}}$ with the κ leftmost column vectors of $\tilde{\mathbf{V}}$ and use (2.11)–(2.14) to retrieve the parameters of interest.

This algorithm does not present any difficulty of implementation. However the main issue is to understand the key-steps 4, 5 and 6. Step 7 follows naturally from step 6 by using the well-known HTLS method described in section 2.2.

4.3.2 Complexity

Most of the computation complexity of this algorithm stems from the computation of the SVD. If we refer to the standard Golub-Reinsch SVD [26] algorithm, the cost for a SVD of a $L \times M$ matrix, including the computation of the three matrices \mathbf{U} , $\mathbf{\Sigma}$ and \mathbf{V} , is $4L^2M + 8LM^2 + 9M^3$ floating point operations (flops) [26, p.254]. In Step 2 of the algorithm we typically take a roughly

square Hankel matrix [85], i.e $L = M$ if N is odd and $L = M + 1$ if N is even. Let us take $L = M$ for simplicity. In this case each SVD algorithm involves a $O(M^3)$ process and the total cost is $O(GM^3)$. In step 4 of the algorithm, we compute the SVD of a $L \times \sum_{g=1}^G K_g$ matrix with $L \geq \sum_{g=1}^G K_g$. Therefore the complexity is $O(L^2 \cdot \sum_{g=1}^G K_g)$. Step 7 of the algorithm involve the SVD of an $(L - 1) \times 2K_c$ matrix as well as the eigenvalue decomposition of a $K_c \times K_c$ matrix. Since $K_c \ll (L - 1) \sim L$ the complexity of step 7 is $O(L^2 K_c)$.

As discussed earlier in this chapter there is no competing approach in the multi-channel case ($G > 2$). We can however compare the complexity of our algorithm with the one of HTLS-SEP in the two-channel case. The latter algorithm requires the SVD of two $L \times M$ matrices, the SVD of a $L \times 2M$ matrix and finally the SVD of a $L \times M$ matrix in order to get the common subspace. The method used for finding the signal poles is similar to our algorithm. Note that HTLS-SEP also requires the Hankel matrices to be roughly square ($L = M$). Finally we can see that HTLS-SEP requires one more SVD, which is the SVD of an $L \times 2M$ matrix. Moreover more matrix manipulations and multiplications are necessary. In conclusion, although both algorithms have roughly the same complexity ($O(L^3)$), HTLS-SEP is always slower than our new algorithm.

4.4 Application to simulated signals

In this section we illustrate the performance of the algorithm and we assess the reliability of the common subspace estimate. Complex circular symmetric WGN is added to the signal in each channel and we conduct Monte Carlo experiments consisting of $N_{\text{runs}} = 500$ runs. The performance is mainly expressed in terms of the RRMSE, defined by:

$$\text{RRMSE}(\gamma) = \left[\frac{1}{N_{\text{runs}}} \sum_{i=1}^{N_{\text{runs}}} |\gamma - \hat{\gamma}_i|^2 \right]^{\frac{1}{2}} \cdot \frac{100}{\gamma} [\%], \quad (4.18)$$

where $\hat{\gamma}_i$ is an estimate of the parameter γ in the i th run. We also refer to a weighted complex discrete-time exponential $c_k z_k^n$ as a peak, simply because in the frequency domain $c_k z_k^n$ has a bell shape. The number of peaks in the frequency domain is usually equal to the number of signal poles. As a measure of the performance, we will also use the measure of the average distance between two parameters $\hat{\gamma}$ and $\hat{\eta}$ which is expressed as follows:

$$d(\hat{\gamma}, \hat{\eta}) = \left[\frac{1}{N_{\text{runs}}} \sum_{i=1}^{N_{\text{runs}}} |\hat{\gamma}_i - \hat{\eta}_i|^2 \right]^{\frac{1}{2}} \quad (4.19)$$

where $\hat{\gamma}_i$ (resp. $\hat{\eta}_i$) is an estimate of the parameter γ (resp. η) in the i th run. If $\hat{\gamma} = \gamma$ is not anymore an estimate but a fixed value, then $d(\gamma, \hat{\eta})$ is

the deviation of the estimate $\hat{\eta}$ from the fixed value γ . Therefore, for instance $d(\eta, \hat{\eta})$ represents the RMSE of the parameter η .

4.4.1 A five-channel example

In this example we consider five signals, having two poles in common. Each signal is of order three, which means that there is one perturbing discrete-time complex exponential per channel. Figure 4.1 shows the signals in the frequency domain. Their relevant parameters are given in Table 4.1. All the phases φ are drawn from a uniform distribution $[0, 2\pi]$, and the amplitudes are perturbed by a WGN with standard deviation 0.1. Note that the third channel contains two closely spaced peaks. Applying steps 1 to 6 of Algorithm 4.3 for different lev-

Table 4.1: Set of parameters for the five-channel example.

		ν [Hz]	α [s ⁻¹]	a [a.u.]
Channel 1	Peak 1	-1379	208	6.1
	Peak 2	-353	117	2.8
	Peak 3	-685	256	9.9
Channel 2	Peak 1	-1379	208	6.1
	Peak 2	-353	117	2.8
	Peak 4	-271	197	6.0
Channel 3	Peak 1	-1379	208	6.1
	Peak 2	-353	117	2.8
	Peak 5	478	208	5.0
Channel 4	Peak 1	-1379	208	6.1
	Peak 2	-353	117	2.8
	Peak 6	815	319	7.3
Channel 5	Peak 1	-1379	208	6.1
	Peak 2	-353	117	2.8
	Peak 7	2162	808	17

els of noise yields the results shown in Figure 4.2. The simulation parameters were set as follows: $N = 128$, $L = 65$, $M = 64$, $G=5$, $\{K_g\}_{g=1}^5 = \{3, 3, 3, 3, 3\}$ and $\kappa = 2$. The subspace of interest in which lie all the Vandermonde vectors has dimension $\sum K_g - (G - 1) \cdot \kappa = 7$. The $\kappa = 2$ first singular values of \mathbf{U}_{tot} correspond to the common subspace and the following $\sum K_g - G\kappa = 5$ singular values correspond to a subspace in which the non-common Vandermonde vectors are projected. Finally the remaining $(G - 1)\kappa = 8$ singular values of \mathbf{U}_{tot} are related to the noise subspace. The singular values of the matrix \mathbf{U}_{tot} exhibit a quite robust behavior. First we notice that the gap between the weakest singular value related to the common subspace and the largest one related to the orthogonal subspace stays visible over a broad range of the noise

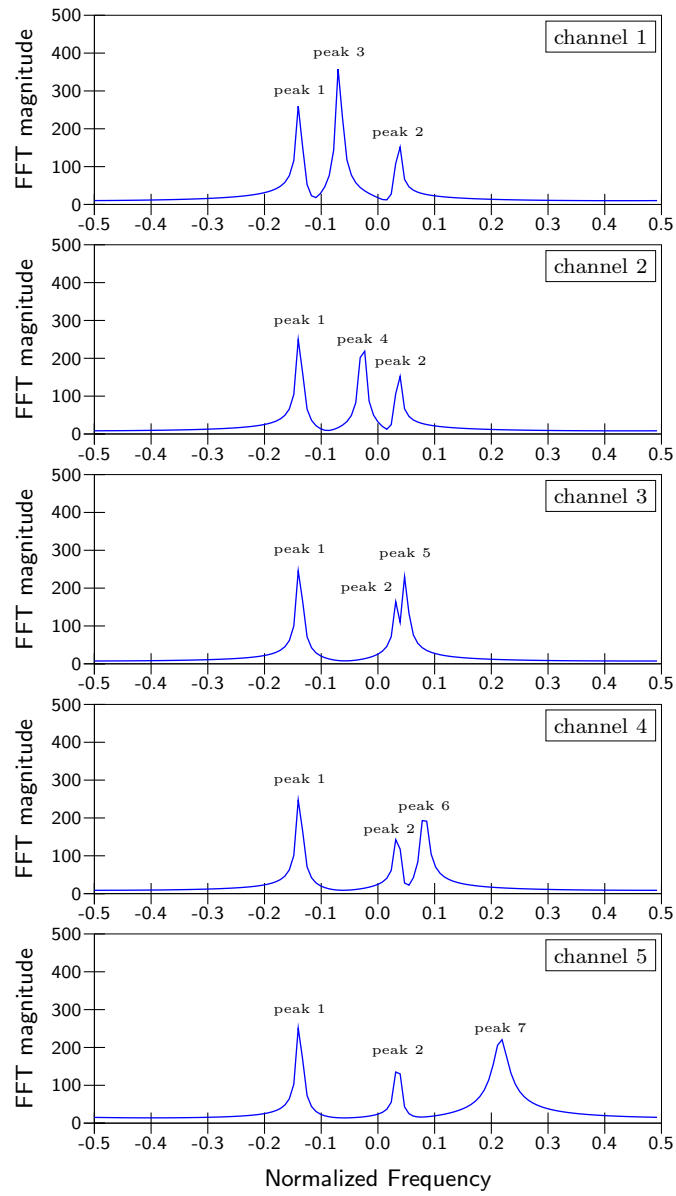


Figure 4.1: Frequency domain signals in the five-channel case.

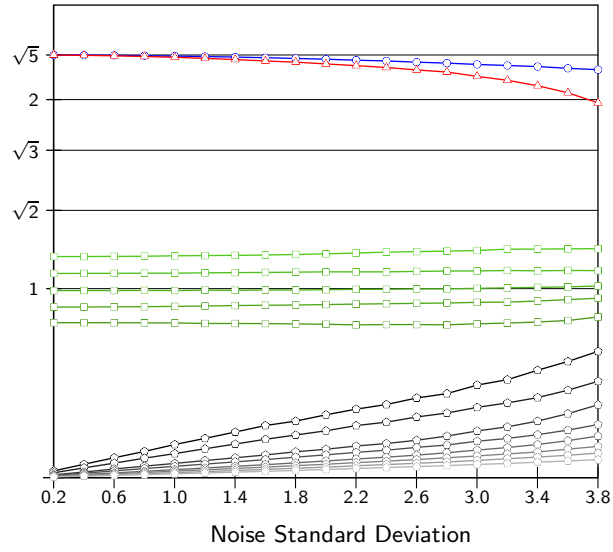


Figure 4.2: The $\sum_g^G K_g = 15$ singular values of \mathbf{U}_{tot} in the five-channel case. The two top singular values (\circ, \triangle) are related to the common subspace. The five following singular values (\square) are related to the subspace in which lie a projection of the undesired (not common) poles. The last eight singular values (\odot) are related to the noise subspace.

variance. The two largest singular values deviate very slowly from their initial value ($\sqrt{5}$). Moreover the five next singular values keep an almost constant value for any level of noise. Finally it is interesting to see that only the last singular values increase proportionally with the noise standard deviation. The best performance can be expected when the number of common poles and the number of perturbing harmonics are small. A noise standard deviation equal to 3.4 corresponds on average to a peak SNR of -5.2 dB for the smallest common peak and a peak SNR of 1.57 dB for common peak 1. At this noise level, the weakest singular value related to the common subspace crosses the value $\sqrt{4}$. At this point we must have the prior knowledge that there are no poles common to 4 channels. Figure 4.3 shows the estimates of the frequencies and damping factors of the common poles. We conclude that, over a wide SNR range, (1) it is possible to separate the signal subspace, in which the Vandermonde vectors related to all signal poles are lying, from the noise subspace, (2) within the signal subspace it is possible to distinguish the common subspace, and estimate the common poles. Our simulation consisted of 500 noise realizations per noise level, but actually the estimates are already very smooth for a few runs (i.e. 50–100), which means that this method does not suffer from outliers.

In most real-world situations it is very difficult to determine the exact number

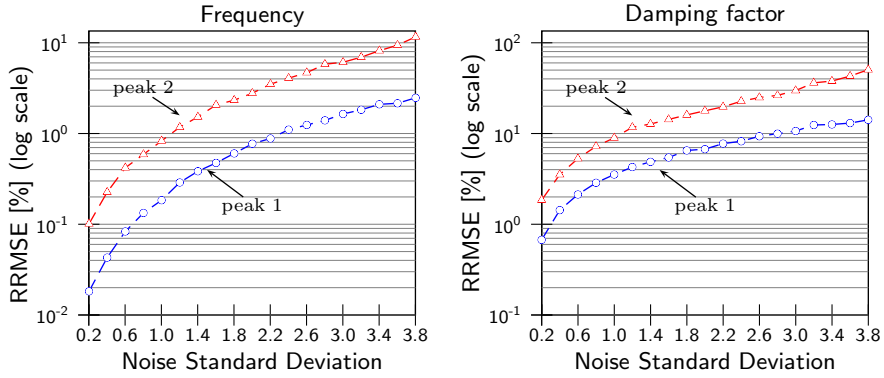


Figure 4.3: Estimation of the common frequencies and damping factors in the five-channel case.

of signal poles for each channel. Therefore we test the robustness of the algorithm against an overestimation of the model orders. In Figures 4.4 and 4.5 we plot the singular values of the matrix \mathbf{U}_{tot} in the case where the model order of each channel has been overestimated with respectively 1 and 2 poles. Except for the overestimation of K_g , the simulation parameters remained the same. We see that there is still a gap between the singular values related to the common subspace and the other singular values. We can observe two main effects. First, as explained in Section 4.3.1, by overestimating model orders, we include some more vectors that are orthogonal to the common subspace in

the sense that a smaller part is projected away. This increases the probability to get a new common subspace. This is reflected by an increase of the largest singular values related to the non-common subspace. On the other hand, the overestimation helps to catch more accurately the true common subspace. Indeed we can see in Figures 4.4 and 4.5 that the singular values related to the common subspace remain slightly closer to $\sqrt{5}$. This result is confirmed by

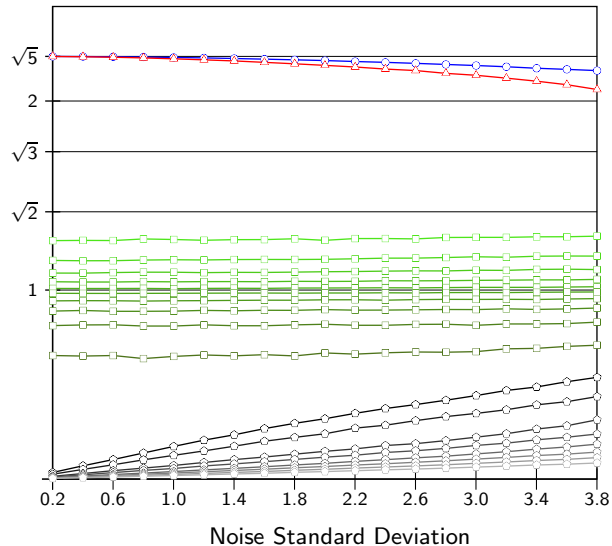


Figure 4.4: The $\sum_{g=1}^G K_g = 20$ singular values of \mathbf{U}_{tot} in the five-channel case with an overestimation of 1 pole for each channel. The two top singular values (\circ, \triangle) are related to the common subspace. The following five singular values (\square) are related to the subspace in which a projection of the undesired (not common) poles is lying. The last eight singular values (\diamond) are related to the noise space.

Figure 4.6 and Figure 4.7, which show the estimates of the common frequencies and damping factors.

4.4.2 A two-channel example

We also compare the new technique with the HTLS-SEP algorithm for two-channel data [51]. The two signals consist of two complex exponentials, one exponential being common to both signals. The parameters of the three peaks are given in Table 4.2. The sampling frequency is $\nu_s = 10^4$ Hz and the number of samples for each signal is $N = 128$. The frequency domain representation

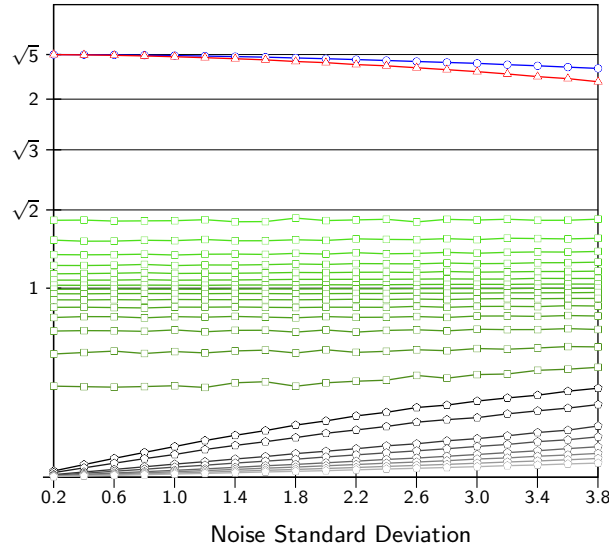


Figure 4.5: The $\sum_{g=1}^G K_g = 25$ singular values of \mathbf{U}_{tot} in the five-channel case with an overestimation of 2 poles for each channel. The two top singular values (\circ, \triangle) are related to the common subspace. The following five singular values (\square) are related to the subspace in which a projection of the undesired (not common) poles is lying. The last eight singular values (\circ) are related to the noise space.

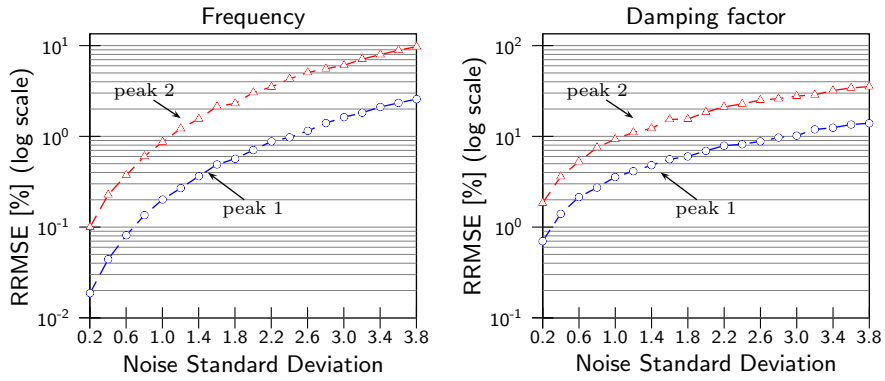


Figure 4.6: Estimation of the common frequencies and damping factors in the five-channel case. The model order of each channel has been increased by 1.

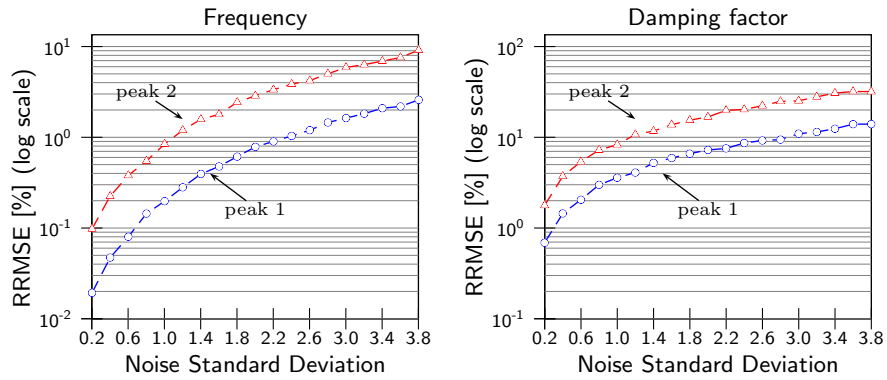


Figure 4.7: Estimation of the common frequencies and damping factors in the five-channel case. The model order of each channel has been increased by 2.

Table 4.2: Set of parameters for a three peaks example

	ν [Hz]	α [s^{-1}]	a [a.u]	φ [$^\circ$]
peak1	-1379	208	2.5	15
peak2	-685	256	9.9	15
peak3	-271	197	6.0	15

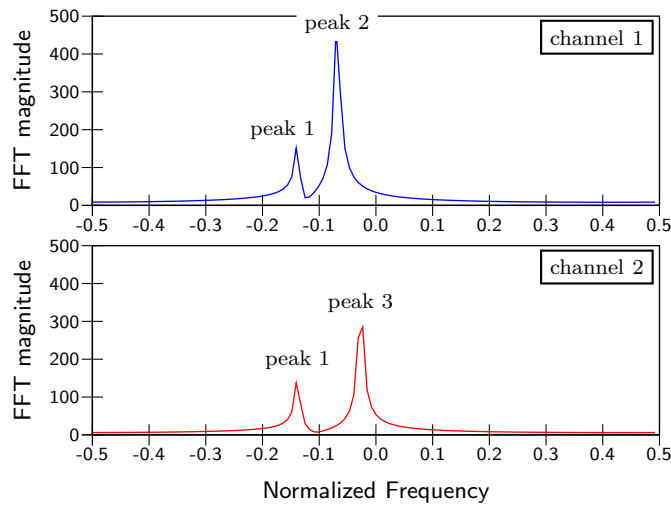


Figure 4.8: Frequency domain representation of the two signals, in the two-channels example. "Peak 1" is the common peak.

of this set of two signals is given in Figure 4.8. The data are processed with the following parameters: $L = 65$, $M_g = 64$, $G=2$, $K_1 = 2$, $K_2 = 2$ and $\kappa = 1$. Thus, as shown in Figure 4.9, the first singular value σ_1 slowly deviates from its theoretical value, which is $\sqrt{2}$ according to (4.11), until the noise standard deviation reaches 2. Within this range, σ_1 remains quite far away from σ_2 , the largest singular value related to the uncommon subspace. Detection that there is only one pole is clearly possible. The estimates of its frequency and

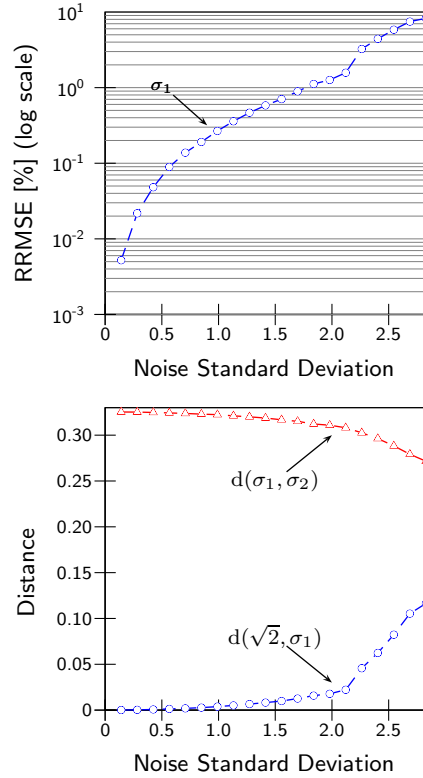


Figure 4.9: The left plot shows the RRMSE of the singular value related to the common signal pole in the two-channel case. The right plot shows the deviation of σ_1 from $\sqrt{2}$ (o) and the average distance from σ_1 to σ_2 (Δ).

the damping factor are shown in Figure 4.10. The results given of method are consistently better than the results of HTLS-SEP. First, the frequency estimate computed by HTLS-SEP becomes inconsistent when the noise standard deviation reaches 1.1, while for the new method, this happens when the noise standard deviation reaches 1.5. Second, in the noise range $[0.1 - 1.1]$, where both methods are consistent, the estimates of the frequency and the damping factor are about 30.39 % more accurate. The main reason is that we work with

normalized subspaces and therefore we get rid of the amplitudes of the peaks, thereby resulting into a very high robustness of the algorithm.

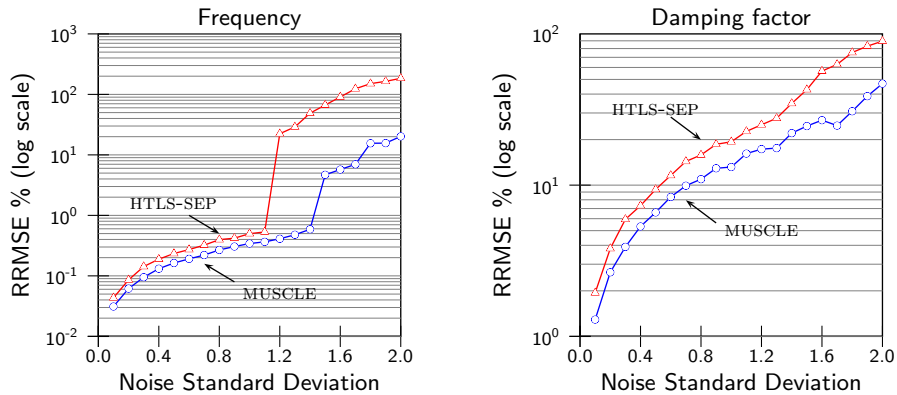


Figure 4.10: Estimates of the frequency and damping factor of the common pole in the two-channel case.

4.5 Application to EEG monitoring

The present application deals with real-life EEG recordings of patients with epilepsy. EEG is a tool to study the brain dynamics noninvasively, by measuring the electric potentials at several locations on the skull and thereby providing a continuous measure of cortical function over several channels. Typical waveforms appear in the EEG of patients with epilepsy. The epileptiform activity can be identified on the majority of the EEG channels. However, additional activity, typically due to the contribution of heart, muscle and eye-movement, appears on a limited number of channels in EEG recordings. Thus, the kernel problem is to remove artifacts and background EEG from the multichannel EEG in order to allow the correct interpretation of interictal activity (epileptic activity between seizures) and ictal activity (activity during a seizure). It has been shown that the method described in this chapter can be successfully applied to this kind of signals [16, 18, 17]

The first example is an application to real-life data recordings of interictal activity. In Figure 4.11, a comparison of the common pole estimation method with principal component analysis (PCA), which is a standard method in EEG signal processing [96], is shown. The following parameters were applied: $N = 70$, $L = 36$, $M = 35$, $G = 19$. The U_g matrices are truncated according to the

following rule:

$$\begin{cases} \sigma_g^2(K_g) & \geq d \cdot \sigma_g^2(1) \\ \sigma_g^2(K_g + 1) & < d \cdot \sigma_g^2(1) \end{cases} \quad (4.20)$$

with d equal to 0.01. We can clearly see (Figure 4.11 (a)) that the epileptic spikes activity in all the channels is corrupted by additive noise, especially in channels O1 and F3. Note that the epileptic activity that might be present in these two channels is completely obscured by muscle artifacts. Note that in this application, an intermediate step has been added after step 3 of the MUSCLE algorithm. First a rank- K_g matrix $\tilde{\mathbf{H}}_g$ is obtained by truncated SVD:

$$\tilde{\mathbf{H}}_g = \tilde{\mathbf{U}}_g \tilde{\mathbf{\Sigma}}_g \tilde{\mathbf{V}}_g^H \quad (4.21)$$

where $\tilde{\mathbf{U}}_g \in \mathbb{R}^{L \times K_g}$ contains the K_g first columns of the matrix of left singular vectors \mathbf{U}_g , $\tilde{\mathbf{V}}_g \in \mathbb{R}^{K_g \times K_g}$ contains the K_g first columns of the matrix of right singular vectors \mathbf{V}_g and where $\tilde{\mathbf{\Sigma}}_g = \text{diag}\{\sigma_{g,1}, \dots, \sigma_{g,K_g}\} \in \mathbb{R}^{K_g \times K_g}$ contains the first diagonal entries of $\mathbf{\Sigma}_g$. Since $\tilde{\mathbf{H}}_g$ has no Hankel structure, the intermediate reconstructed signals $\tilde{x}_n^{(g)}$ are obtained by averaging along the antidiagonals of $\tilde{\mathbf{H}}_g$. This step is typically a denoising procedure. This is illustrated in graph (b) of Figure 4.11. The noise is removed in all channels except channels O1 and F3 which still contain the muscle artifacts. This is because muscle artifacts are not broad band noise but can be modelled as a sum of (damped) sinusoids. Moreover these harmonics have high amplitudes. Therefore the signal subspace $\tilde{\mathbf{U}}_g$, $g = 16, 17$ (channels O1 and F3) contains the contribution of these harmonics. However, if we have a look to graph (c) of Figure 4.11, which shows the common dynamics, we can see that the muscle artifacts have been removed. Note that the algorithm perfectly conserves the phase and the amplitude of the common dynamics. Finally Figure 4.11 (d) shows the results of the PCA method. Typically, in our case, the PCA method consists in arranging each signal (channel) $x_n^{(g)}$, $n = 0, \dots, N - 1$, in a column vector $X^{(g)} = [x_0^{(g)} x_1^{(g)} \dots x_{N-1}^{(g)}]^T \in \mathbb{R}^{N \times 1}$ and to stack all the vectors in a matrix $\mathbf{X} = [X^{(1)} X^{(2)} \dots X^{(G)}] \in \mathbb{R}^{N \times G}$. Then the SVD is applied to this matrix:

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H, \quad (4.22)$$

where $\mathbf{U} = [U_1 \dots U_G] \in \mathbb{R}^{N \times G}$, $\mathbf{V}^H = [V_1 \dots V_G]^H \in \mathbb{R}^{G \times G}$ and $\mathbf{\Sigma} = \text{diag}\{\sigma_1, \dots, \sigma_G\}$. The first left singular vector (first column vector of \mathbf{U}) is called the first principal component and contains the dominant waveform. So, an estimate of the common waveform in each channel is obtained by reconstruction of the rank-1 matrix $\tilde{\mathbf{X}}$:

$$\tilde{\mathbf{X}} = \sigma_1 U_1 V_1^H. \quad (4.23)$$

The column vectors of $\tilde{\mathbf{X}} = [\tilde{X}^{(1)} \tilde{X}^{(2)} \dots \tilde{X}^{(G)}]$ are the reconstructed signals. One can see that PCA is less robust than the common pole estimation method against noise. Undesirable harmonics have been amplified on the reconstructed

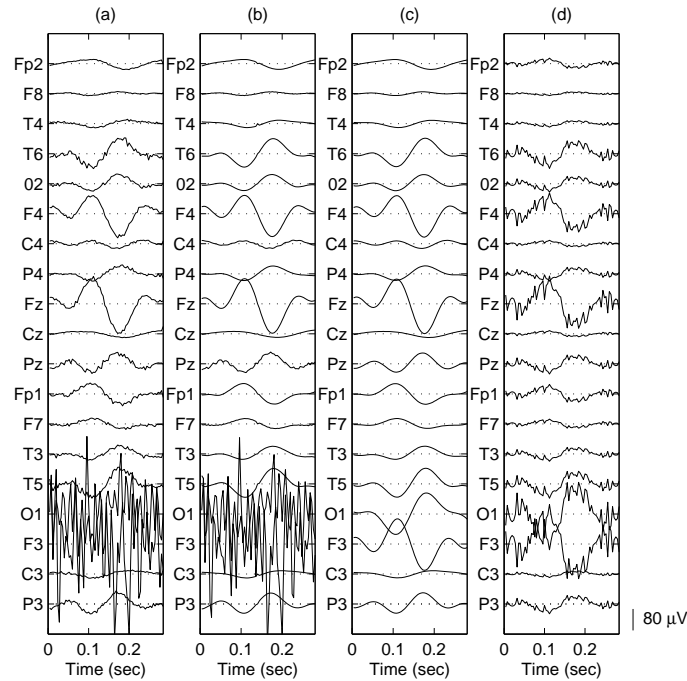


Figure 4.11: (a) Shows the original spike. The EEG after preprocessing each channel separately is shown in (b). The fit based on the common signal poles is shown in (c). The reconstructed EEG based on the first principal component is shown in (d).

channels. Figure 4.12 is a scree plot of the squared singular values. There are clearly 4 singular values in the interval $[\sqrt{18}, \sqrt{19}]$, which means that the dominant waveform is common to 19 channels (all channels) and is composed of 2 exponentially damped sinusoids.

The method has also been applied to real-life data recordings of ictal activity with similar performance as in the previous example from a signal processing point of view. In Figure 4.13 we show a comparison of our algorithm with PCA for 21-channel EEG recordings containing ictal activity. This set of signals was processed with the following parameters: $N = 1000$, $L = 501$, $M = 500$, $G = 21$, $d = 0.01$. The influence of noise tends to decrease the (squared) singular values related to the common subspace. In the noise-free case, these squared singular values are equal to 21. In the presence of a reasonable level of noise, they are generally between G and $G - 1$. So, the signal subspace related to the squared singular values above 20, has been selected. The original (noisy) signals are plotted in Figure 4.13 (a). The denoised version of these signals using

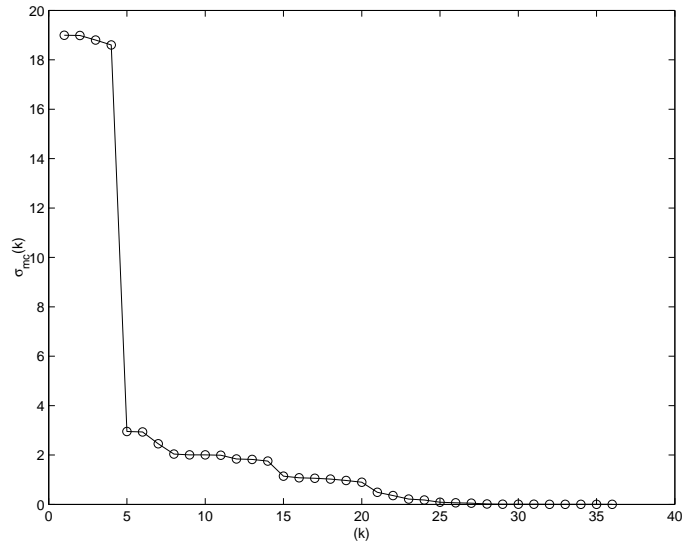


Figure 4.12: The singular values of the stacked matrix \mathbf{U}_{tot} (see step 6 of Algorithm 4.3) generated in the real-life example.

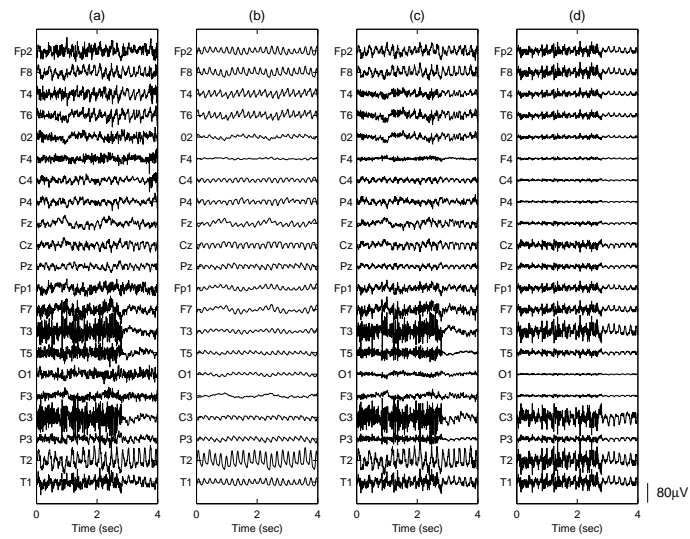


Figure 4.13: (a) 4 s EEG epoch containing ictal activity. (b) The reconstructed EEG based on the common signal poles. (c) The reconstructed EEG based on the three most energetic principal components. (d) The reconstructed EEG based on the first principal component.

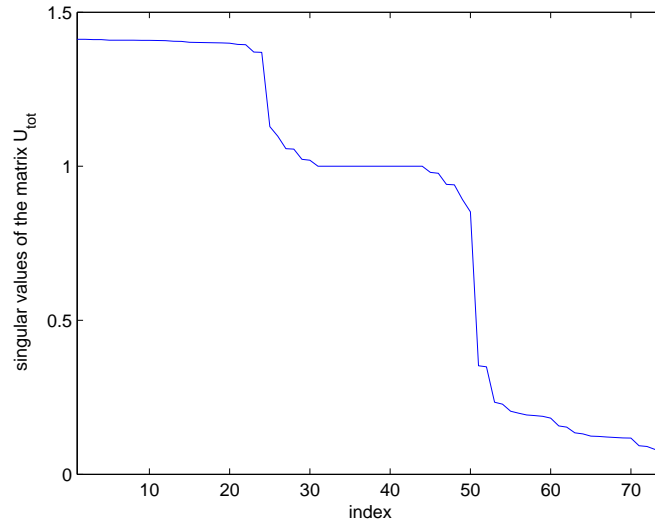


Figure 4.14: Singular values of the matrix $\mathbf{U}_{\text{tot}} = [\tilde{\mathbf{U}}_1 \tilde{\mathbf{U}}_2]$.

MUSCLE is shown in Figure 4.13 (b) while the outputs of the PCA method with three and one principal components are shown in Figure 4.13 (c) and (d). Here again, one can see that MUSCLE does a wonderful job in removing the background activity and thereby enhancing the common activity. It clearly outperforms the PCA method.

4.6 Application to material health monitoring

The application of MUSCLE to the signal obtained using the setup in Figure 1.1 is obvious. The signal received via the piezoelectric sensor may have common information with the signal obtained via the optical fiber sensor. The signals that we considered in this example are the same as in Figure 3.4. Let us call the optical signal x_1 and the ultrasonic signal x_2 . First, we set the following parameters : $N=751$, $L = 376$, $M = 376$, $K_1 = 30$, $K_2 = 44$. Figure 4.14, in which the singular values of $\mathbf{U}_{\text{tot}} = [\tilde{\mathbf{U}}_1 \tilde{\mathbf{U}}_2]$ are displayed, shows that there exists a subspace common to the Hankel matrices \mathbf{H}_1 and \mathbf{H}_2 . The size of this common subspace is equal to the number of singular values that are

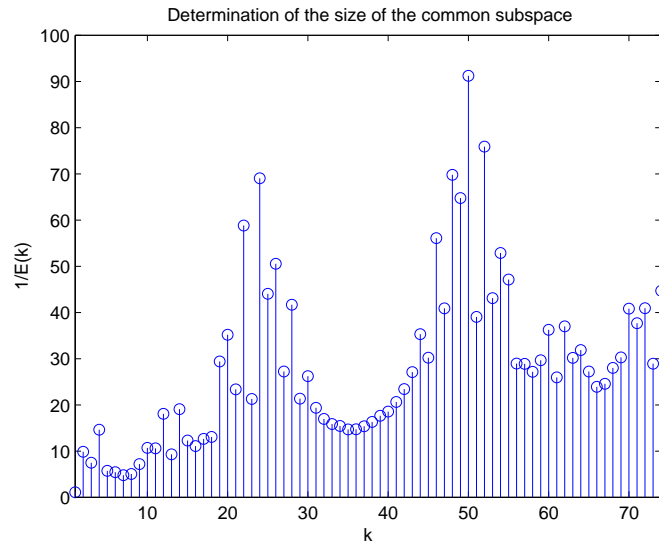


Figure 4.15: Application of SAMOS to the matrix \mathbf{U}_{tot} . The first peak shows that the common subspace is of size 24. The second indicated that the size of the subspace containing the noncommon information is $50 - 24 = 26$. The size of the noise subspace is $74 - 50 = 24$.

theoretically equal to $\sqrt{2}$. This size can also be determined by applying steps 3 and 4 of SAMOS to \mathbf{U}_{tot} . Figure 4.15 shows the values of the function $1/E(k)$ defined in (3.12). Clearly the two peaks delimit the three different subspaces: the common subspace, the subspace containing noncommon information, and the noise subspace. The size of the common subspace is 24. From this result we can conclude that the signals x_1 and x_2 have 12 common harmonics.

4.7 Conclusion

In this chapter, we have studied a novel subspace-based method for common pole estimation in the multi-channel case. We have proven that the use of semi-unitary matrices allows one to determine the number of common poles and get an accurate estimate of a subspace from which these poles can be computed. This method is easy to implement, very robust and can be applied to an arbitrary number of channels. Moreover, we have shown that a slight

overestimation of the individual model orders does not affect the accuracy of the common pole estimation. In the particular case in which only two channels are involved, this technique turns out to be more efficient than the HTLS-SEP technique developed for the two-channel case.

As a last remark, we shall mention that the kernel part of MUSCLE is the estimation of subspaces common to an arbitrary number of matrices. This kernel method can be applied to any type of matrices having the same column size.

Chapter 5

A Short Introduction to Multilinear Algebra

In this chapter we describe the tensor-algebraic material that will be needed in Chapter 6 and Chapter 7. We first give some basic definitions of multilinear algebra. Then we review a multilinear generalization of the singular value decomposition and the best rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors.

5.1 Higher-order tensor

Multilinear algebra is the algebra of higher-order tensors, which are the higher-order equivalents of vectors (first-order) and matrices (second-order), i.e., quantities of which the elements are addressed by more than two indices. Said in other words, higher-order tensors are the basic quantities of multilinear algebra. As such, it is important to give a clear definition of them. However, they may be approached in different ways depending on the goal for which they are used. This thesis relies essentially on the study of subspaces. In this context we can practically define an N th-order tensor as a higher-dimensional table that is addressed by N indices. They are also called N -way arrays. Figure 5.1 gives a visualization of the first three orders.

To facilitate the distinction between scalars, vectors, matrices and higher-order tensors in this chapter, we would like to provide the following additional details: first we recall that the type of a given quantity will be reflected by its representation: scalars are denoted by lower-case letters $(a, b, \dots; \alpha, \beta, \dots)$, vectors

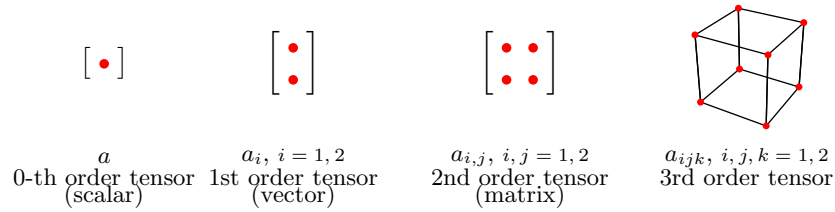


Figure 5.1: Visual representation of tensors up to order three.

are written as capitals (A, B, \dots), matrices correspond to bold-face capitals ($\mathbf{A}, \mathbf{B}, \dots$) and tensors are written as calligraphic letters ($\mathcal{A}, \mathcal{B}, \dots$).

This notation is consistently used for lower-order parts of a given structure. For example, the entry with row index i and column index j in a matrix \mathbf{A} , i.e., $(A)_{ij}$, is symbolized by a_{ij} (also $(A)_i = a_i$ and $(\mathcal{A})_{i_1 i_2 \dots i_N} = a_{i_1 i_2 \dots i_N}$). Furthermore, the i th column vector of a matrix \mathbf{A} is denoted as A_i , i.e., $\mathbf{A} = [A_1 A_2 \dots]$. To enhance the overall readability, we have made one exception to this rule: as we frequently use the characters i, r and n in the meaning of indices (counters), I, R and N will be reserved to denote the index upper bounds, unless stated otherwise.

First some useful basic definitions are provided in Section 5.2. These definitions are necessary for the reading of the subsequent sections. Then, a multilinear generalisation of the matrix SVD and a multilinear generalisation of the best rank- R approximation of a matrix are given in Section 5.3 and Section 5.4 respectively. We conclude the chapter in Section 5.5.

5.2 Basic definitions

In this section we introduce some elementary notations and definitions in multilinear algebra, which will be needed in the further developments.

5.2.1 Multiplication of a higher-order tensor by a matrix

We have the following definition.

Definition 5.1 (n -mode product [83]) The n -mode product of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ by a matrix $\mathbf{U} \in \mathbb{C}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is an $(I_1 \times I_2 \times \dots \times I_N)$ tensor.

$\cdots \times I_{n-1} \times J_n \times I_{n+1} \cdots \times I_N$)-tensor of which the entries are given by

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} \stackrel{\text{def}}{=} \sum_{i_n} a_{i_1 i_2 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} u_{j_n i_n}.$$

□

In this notation, the matrix product $\mathbf{G} = \mathbf{U} \cdot \mathbf{F} \cdot \mathbf{V}^\top$, involving matrices $\mathbf{F} \in \mathbb{R}^{I_1 \times I_2}$, $\mathbf{U} \in \mathbb{R}^{J_1 \times I_1}$, $\mathbf{V} \in \mathbb{R}^{J_2 \times I_2}$ and $\mathbf{G} \in \mathbb{R}^{J_1 \times J_2}$, is written as $\mathbf{G} = \mathbf{F} \times_1 \mathbf{U} \times_2 \mathbf{V}$. This is meaningful: the relationship between \mathbf{U} and \mathbf{F} and the relationship between \mathbf{V} (not \mathbf{V}^\top) and \mathbf{F} are in fact completely similar: in the same way as \mathbf{U} makes linear combinations of the rows of \mathbf{F} , \mathbf{V} makes linear combinations of the columns of \mathbf{F} ; in the same way as the columns of \mathbf{F} are multiplied by \mathbf{U} , the rows of \mathbf{F} are multiplied by \mathbf{V} ; in the same way as the columns of \mathbf{U} are associated with the column space of \mathbf{G} , the columns of \mathbf{V} are associated with the row space of \mathbf{G} . This typical relationship is denoted by means of the \times_n -symbol.

We have the following properties:

$$\begin{aligned} (\mathcal{A} \times_n \mathbf{U}) \times_m \mathbf{V} &= (\mathcal{A} \times_m \mathbf{V}) \times_n \mathbf{U} = \mathcal{A} \times_n \mathbf{U} \times_m \mathbf{V}, \\ (\mathcal{A} \times_n \mathbf{V}) \times_n \mathbf{U} &= \mathcal{A} \times_n (\mathbf{U}\mathbf{V}). \end{aligned}$$

Note that in general $\mathcal{A} \times_n \mathbf{U} \times_2 \mathbf{V} \neq \mathcal{A} \times_n (\mathbf{U} \cdot \mathbf{V}^\top)$.

5.2.2 n -mode vectors and n -mode vector space

An n -mode vector of an $(I_1 \times I_2 \times \cdots \times I_N)$ -tensor \mathcal{A} is an I_n -dimensional vector obtained from \mathcal{A} by varying the index i_n and keeping the other indices fixed. Figure 5.2 shows the n -mode vectors in the case of a third-order tensor. As such, the n -mode vectors are the generalization of column and row vectors in the matrix case. The subspace spanned by all the n -mode vectors, for a given value of n , is called the n -mode vector space.

5.2.3 Matrix representation of a higher-order tensor

Some of the results can be conveniently expressed in matrix terms. To this end, we must stack the elements of a higher-order tensor in a matrix. There are several ways to do so. One particular type of “matrix unfolding” will prove to be particularly useful, namely, the matrix representation of a given tensor in which all its column (row, ...) vectors are simply stacked one after another [83]. To avoid confusion, we will retain one particular ordering of the column (row, ...) vectors; for order three, these unfolding procedures are visualized in Figure 5.3. Notice that the definitions of the matrix unfoldings involve the tensor dimensions I_1, I_2, I_3 in a cyclic way and that, when dealing with an unfolding of dimensionality $I_c \times I_a I_b$, that the index i_a varies more slowly than i_b . In general, we define:

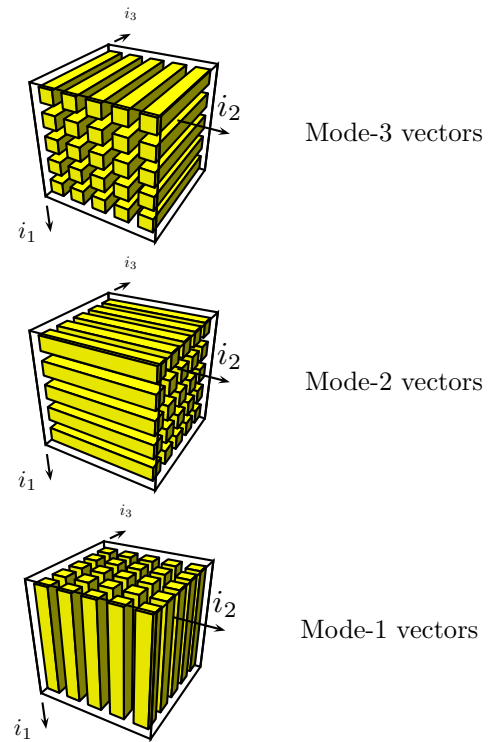


Figure 5.2: Visualization of the n -mode vectors of a third-order tensor.

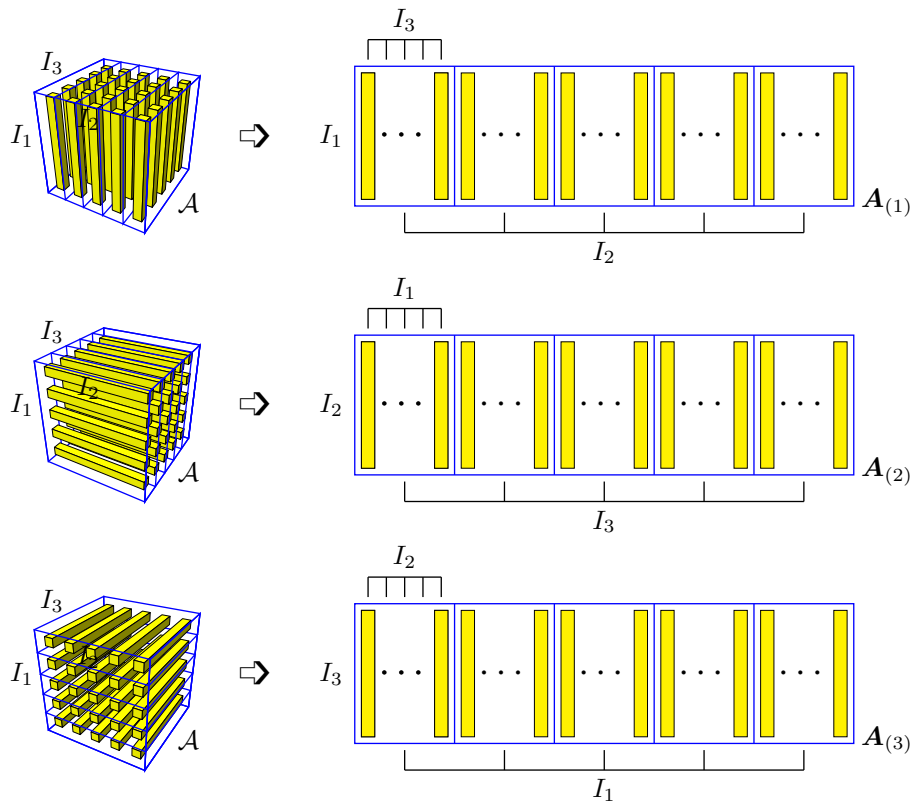


Figure 5.3: Unfolding of the $(I_1 \times I_2 \times I_3)$ -tensor \mathcal{A} to the $(I_1 \times I_2 I_3)$ -matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$ -matrix $\mathbf{A}_{(2)}$, and the $(I_3 \times I_1 I_2)$ -matrix $\mathbf{A}_{(3)}$.

Definition 5.2 Assume an N th-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$. The matrix unfolding $\mathbf{A}_{(n)} \in \mathbb{C}^{I_n \times (I_{n+1}I_{n+2} \cdots I_N I_1 I_2 \cdots I_{n-1})}$ contains the element $a_{i_1 i_2 \cdots i_N}$ at the position with row number i_n and column number equal to $(i_{n+1} - 1)I_{n+2}I_{n+3} \cdots I_N I_1 I_2 \cdots I_{n-1} + (i_{n+2} - 1)I_{n+3}I_{n+4} \cdots I_N I_1 I_2 \cdots I_{n-1} + \cdots + (i_N - 1)I_1 I_2 \cdots I_{n-1} + (i_1 - 1)I_2 I_3 \cdots I_{n-1} + (i_2 - 1)I_3 I_4 \cdots I_{n-1} + \cdots + i_{n-1}$. \square

5.2.4 Scalar product, orthogonality and Frobenius-norm

The scalar product $\langle \mathcal{A}, \mathcal{B} \rangle$ of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined in a straightforward way as $\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{\text{def}}{=} \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} a_{i_1 i_2 \cdots i_N} b_{i_1 i_2 \cdots i_N}^*$. The Frobenius-norm of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is then defined as $\|\mathcal{A}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$. Two tensors are called orthogonal when their scalar product is 0.

5.2.5 Rank properties of a higher-order tensor

There are major differences between matrices and higher-order tensors when rank properties are concerned. However we may start with an analogy with the matrix case. A N th-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is called rank-1 when it equals the outer product of N vectors $U^{(1)}, U^{(2)}, \dots, U^{(N)}, U^{(n)} \in \mathbb{C}^{I_n \times 1}$ [10, 22, 36]:

$$a_{i_1 i_2 \cdots i_N} = u_{i_1}^{(1)} u_{i_2}^{(2)} \cdots u_{i_N}^{(N)} \quad (5.1)$$

for all index values, which will be concisely written as:

$$\mathcal{A} = U^{(1)} \circ U^{(2)} \circ \cdots \circ U^{(N)}. \quad (5.2)$$

The n -rank of a higher-order tensor is the obvious generalization of the column (row) rank of matrices: it equals the dimension of the n -mode vector space. The dimension of this subspace is called the n -mode rank R_n . From the definition of a rank-one tensor (5.1)-(5.2), it follows that its n -mode vector space, is spanned by $U^{(n)}$, ($1 \leq n \leq N$). Therefore the n -mode ranks are equal to 1. An important difference between matrices and tensors is that, for general higher-order tensors, the different n -mode ranks are not necessarily the same, while for matrices, column and row rank are always the same. The n -rank will be denoted as $\text{rank}_n(\mathcal{A})$. A tensor of which the n -ranks are equal to R_n ($1 \leq n \leq N$) is called a rank- (R_1, R_2, \dots, R_N) tensor. A second difference between matrices and tensors is that, even when all the n -ranks of the tensor are the same, they can still be different from its rank, denoted as $\text{rank}(\mathcal{A})$; \mathcal{A} having rank R generally means that it can be decomposed in a sum of R , but not less than R , rank-1 terms (see e.g. [42]). From the definition of n -rank and rank follows that $R_n \leq R$ for all n .

Example 5.1 Consider the $(I_1 \times I_2 \times I_3)$ -tensor defined as follows:

$$\mathcal{A} = U^{(1)} \circ U^{(2)} \circ U^{(3)} + U^{(1)} \circ U^{(4)} \circ U^{(5)}$$

We have $\text{rank}(\mathcal{A})=2$. But we have also $\text{rank}_1(\mathcal{A}) = 1$ and $\text{rank}_2(\mathcal{A}) = \text{rank}_3(\mathcal{A}) = 2$. \triangle

The n -rank of a given tensor can be analyzed by means of matrix techniques:

Property 5.2.1 *The n -mode vectors of \mathcal{A} are the column vectors of the matrix unfolding $A_{(n)}$ and*

$$\text{rank}_n(\mathcal{A}) = \text{rank}(A_{(n)}).$$

Property 5.2.2 *Given a (R_1, R_2, \dots, R_N) -tensor we have the following inequality:*

$$R_n \leq \prod_{\substack{r=1 \\ r \neq n}}^N R_r$$

5.3 The higher-order singular value decomposition

In [21, 83] the following tensor decomposition was discussed.

Theorem 5.3 (*Nth-order singular value decomposition*). *Every complex $(I_1 \times I_2 \times \dots \times I_N)$ -tensor A can be written as the product*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)}, \quad (5.3)$$

in which:

- $\mathbf{U}^{(n)} = [U_1^{(n)} U_2^{(n)} \dots U_{I_n}^{(n)}]$ is a unitary $(I_n \times I_n)$ -matrix,
- \mathcal{S} is a complex $(I_1 \times I_2 \times \dots \times I_N)$ -tensor of which the subtensors $\mathcal{S}_{i_n} = \alpha$, obtained by fixing the n -th index to α , have the properties of:
 - all-orthogonality: two subtensors $\mathcal{S}_{i_n} = \alpha$ and $\mathcal{S}_{i_n} = \beta$ are orthogonal for all possible values of n , α and β subject to $\alpha \neq \beta$:

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when } \alpha \neq \beta, \quad (5.4)$$

- ordering:

$$\|\mathcal{S}_{i_n=1}\| \geq \|\mathcal{S}_{i_n=2}\| \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\| \geq 0, \quad (5.5)$$

for all possible values of n .

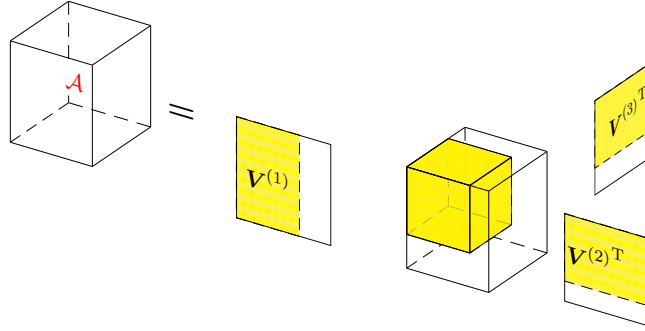


Figure 5.4: Visualization of the HOSVD of a third-order tensor \mathcal{A} . If \mathcal{A} is n -mode rank deficient, only the shaded part of the core-tensor contains entries different from zero.

The Frobenius-norms $\|\mathcal{S}_{i_n=i}\|$, symbolized by $\sigma_i^{(n)}$, are n -mode singular values of \mathcal{A} and the vector $U_i^{(n)}$ is an i th n -mode singular vector. The decomposition is visualized for third-order tensors in Figure 5.4.

Applied to a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, Theorem 5.3 says that it is always possible to find orthogonal transformations of the column, row and 3-mode space such that $\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^{(1)\text{H}} \times_2 \mathbf{U}^{(2)\text{H}} \times_3 \mathbf{U}^{(3)\text{H}}$ is all-orthogonal and ordered (the new basis vectors are the columns of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$). All-orthogonality means that the different “horizontal matrices” of \mathcal{S} (the first index i_1 is kept fixed, whilst the two other indices, i_2 and i_3 , are free) are mutually orthogonal with respect to the scalar product of matrices (i.e. the sum of the products of the corresponding entries vanishes); at the same time, the different “frontal” matrices (i_2 fixed) and the different “vertical” matrices (i_3 fixed) should be mutually orthogonal as well. The ordering constraint imposes that the Frobenius-norm of the horizontal (frontal resp. vertical) matrices does not increase as the index i_1 (i_2 resp. i_3) is increased. This decomposition is clearly a generalization of the matrix SVD. Note, w.r.t. the condition of all-orthogonality on \mathcal{S} , that in the matrix case the matrix of singular values is all-orthogonal as well: due to its diagonal structure, the scalar product of two different rows or columns also vanishes. On the other hand, one cannot define a higher-order SVD by imposing diagonality on the core tensor \mathcal{S} : in general, it is impossible to reduce higher-order tensors to a diagonal form by means of unitary transformations, because the number of degrees of freedom in such a hypothetical decomposition is smaller than the number of entries of the tensor that has to be decomposed. For convenience, we will refer to the decomposition in Theorem 5.3 as the “higher-order SVD” (HOSVD). This is substantiated by the many striking analogies with the matrix SVD, established in [21]. Nevertheless, one should be

aware that focusing on different properties of the matrix SVD can lead to the definition of different (formally less striking) multilinear SVD-generalizations. One possibility is to look for unitary transformations that make the core tensor as diagonal as possible (in a least squares sense) [14]. Another alternative is the decomposition of the tensor in a minimal number of rank-1 terms; this decomposition is often called the canonical decomposition (CANDECOMP) or parallel factors decomposition (PARAFAC) [10, 15]. Imposing orthogonality constraints on these rank-1 terms may lead to yet other generalizations [38]. The matrix of n -mode singular vectors $\mathbf{U}^{(n)}$ can directly be found as the matrix of left singular vectors of the matrix unfolding $\mathbf{A}^{(n)}$. The n -mode singular values correspond to the singular values of this matrix unfolding. The core tensor \mathcal{S} can then be computed by bringing the matrices of singular vectors to the left side of Eq. (5.3):

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^{(1)\text{H}} \times_2 \mathbf{U}^{(2)\text{H}} \dots \times_N \mathbf{U}^{(N)\text{H}}. \quad (5.6)$$

The fact that the number of non-vanishing singular values of a given matrix equals its (column/row) rank, carries over to the n -mode singular values and the n -rank values of a given tensor [21]. The fact that we have N different sets of n -mode singular values is conform the fact that we also have N different n -ranks. Like for matrices, this link even holds in a numerical sense: the number of significant n -mode singular values of a given tensor equals its numerical n -rank.

5.4 Best rank- (R_1, R_2, \dots, R_N) approximation

In this section we consider a multilinear generalization of the best rank- R approximation of a given matrix. Formally, the problem we want to solve, is the following:

Given an N th-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, find a tensor $\hat{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, with $\text{rank}_1(\hat{\mathcal{A}}) = R_1$, $\text{rank}_2(\hat{\mathcal{A}}) = R_2$, \dots , $\text{rank}_N(\hat{\mathcal{A}}) = R_N$, that minimizes the least-squares cost function

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2. \quad (5.7)$$

The mathematical formulation for the matrix counterpart is similar. In the matrix case a lower rank approximation can be readily obtained by truncated SVD [26]. The n -rank conditions imply that $\hat{\mathcal{A}}$ can be decomposed as

$$\hat{\mathcal{A}} = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)}, \quad (5.8)$$

in which $\mathbf{U}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\mathbf{U}^{(2)} \in \mathbb{C}^{I_2 \times R_2}$, \dots , $\mathbf{U}^{(N)} \in \mathbb{C}^{I_N \times R_N}$ each having orthonormal columns and $\mathcal{B} \in \mathbb{C}^{R_1 \times R_2 \times \dots \times R_N}$. Similarly to the second-order

case, where the best approximation of a given matrix $\mathbf{A} \in \mathbb{C}^{I_1 \times I_2}$ by a matrix $\widehat{\mathbf{A}} = \mathbf{U}^{(1)} \cdot \mathbf{B} \cdot \mathbf{U}^{(2)\text{H}}$, with $\mathbf{U}^{(1)} \in \mathbb{C}^{I_1 \times R}$ and $\mathbf{U}^{(2)} \in \mathbb{C}^{I_2 \times R}$ column-wise orthonormal, is equivalent to the maximization of $\mathbf{U}^{(1)\text{H}} \cdot \mathbf{A} \cdot \mathbf{U}^{(2)}$, we have that the minimization of f is equivalent to the maximization of

$$g(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}) = \left\| \mathcal{A} \times_1 \mathbf{U}^{(1)\text{H}} \times_2 \mathbf{U}^{(2)\text{H}} \dots \times_N \mathbf{U}^{(N)\text{H}} \right\|^2. \quad (5.9)$$

The optimal core tensor follows from

$$\mathcal{B} = \mathcal{A} \times_1 \mathbf{U}^{(1)\text{H}} \times_2 \mathbf{U}^{(2)\text{H}} \dots \times_N \mathbf{U}^{(N)\text{H}}. \quad (5.10)$$

It is natural to ask whether the best rank- (R_1, R_2, \dots, R_N) approximation of a higher-order tensor can be obtained by truncation of the HOSVD. The situation turns out to be quite different for tensors here [40, 22, 23]. By discarding the smallest n -mode singular values, one obtains a tensor $\widehat{\mathcal{A}}$ that is in general not the best possible approximation under the given n -mode rank constraints (see e.g. [23, Example 3]). Nevertheless, the ordering assumption (5.5) implies that the "energy" of \mathcal{A} is mainly concentrated in the part corresponding to low values of i_1, i_2, \dots, i_N . Consequently, if $\sigma_{R_n}^{(n)} \gg \sigma_{R_{n+1}}^{(n)}$, $\widehat{\mathcal{A}}$ is still to be considered as a good approximation of \mathcal{A} . The error is bounded as follows [21].

Property 5.4.1 *Let the HOSVD of \mathcal{A} be given as in Theorem 5.3 and let the n -mode rank of \mathcal{A} be equal to \tilde{R}_n ($1 \leq n \leq N$). Define a tensor $\widehat{\mathcal{A}}$ by discarding the smallest n -mode singular values $\sigma_{R_{n+1}}^{(n)}, \sigma_{R_{n+2}}^{(n)}, \dots, \sigma_{\tilde{R}_n}^{(n)}$, for given values of R_n ($1 \leq n \leq N$), i.e., set the corresponding parts of \mathcal{S} equal to zero. Then we have*

$$\left\| \mathcal{A} - \widehat{\mathcal{A}} \right\|^2 \leq \sum_{i_1=R_1+1}^{\tilde{R}_1} \sigma_{i_1}^{(1)2} + \sum_{i_2=R_2+1}^{\tilde{R}_2} \sigma_{i_2}^{(2)2} + \dots + \sum_{i_N=R_N+1}^{\tilde{R}_N} \sigma_{i_N}^{(N)2}. \quad (5.11)$$

(For completeness, we mention that the truncation of \mathcal{S} may destroy its all-orthogonality. In this sense, the components of the truncated HOSVD of \mathcal{A} may be different from the components of the HOSVD of $\widehat{\mathcal{A}}$, as opposed to the matrix case.) In [22, 41, 40, 81] the following approach was followed for the computation of the best rank- (R_1, R_2, \dots, R_N) approximation. Imagine that the matrices $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(n-1)}, \mathbf{U}^{(n+1)}, \dots, \mathbf{U}^{(N)}$ are fixed and that the only unknown is the column-wise orthonormal matrix $\mathbf{U}^{(n)}$. We have:

$$g = \left\| \widetilde{\mathcal{A}}^{(n)} \times_n \mathbf{U}^{(n)\text{H}} \right\|^2, \quad (5.12)$$

in which

$$\widetilde{\mathcal{A}}^{(n)} \stackrel{\text{def}}{=} \mathcal{A} \times_1 \mathbf{U}^{(1)\text{H}} \dots \times_{n-1} \mathbf{U}^{(n-1)\text{H}} \times_{n+1} \mathbf{U}^{(n+1)\text{H}} \dots \times_N \mathbf{U}^{(N)\text{H}}. \quad (5.13)$$

In a matrix format we have:

$$g = \left\| \mathbf{U}^{(n)\text{H}} \cdot \tilde{\mathbf{A}}_{(n)}^{(n)} \right\|^2, \quad (5.14)$$

with

$$\tilde{\mathbf{A}}_{(n)}^{(n)} = \mathbf{A}_{(n)} \cdot \left(\mathbf{U}^{(n+1)} \otimes \dots \otimes \mathbf{U}^{(N)} \otimes \mathbf{U}^{(1)} \otimes \dots \otimes \mathbf{U}^{(n-1)} \right), \quad (5.15)$$

in which \otimes represents the Kronecker product. Hence the columns of $\mathbf{U}^{(n)}$ can be found as an orthonormal basis for the dominant subspace of the n -mode vector space of $\tilde{\mathbf{A}}_{(n)}^{(n)}$. Repeating this procedure for different mode numbers leads to a ALS algorithm for the (local) maximization of $f(\tilde{\mathbf{A}})$: in each step the estimate of one of the matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(N)}$ is optimized, while the other matrix estimates are kept constant. This technique is a higher-order extension of the orthogonal iteration for matrices [27]. An outline of the Higher-order iteration algorithm (HOOI) is given below. It makes sense to initialize the higher-order orthogonal iteration with the truncated HOSVD. The HOSVD-estimate usually belongs to the attraction region of the best rank- (R_1, R_2, \dots, R_N) approximation, although there is no absolute guarantee of convergence to the global optimum [22].

Algorithm 5.4

HOOI

Step 1: Initial values: $\mathbf{U}_0^{(n)} \in \mathbb{C}^{I_n \times R_n}$ of which the columns form a basis for the dominant R_n -dimensional left singular subspace of $\mathbf{A}_{(n)}$ ($2 \leq n \leq N$); or repeat the algorithm for several different initial values.

Step 2: Iterate until convergence:

$$\begin{aligned} & - \tilde{\mathbf{U}}_{k+1}^{(1)} = \mathcal{A} \times_2 \mathbf{U}_k^{(2)\text{H}} \times_3 \mathbf{U}_k^{(3)\text{H}} \dots \times_N \mathbf{U}_k^{(N)\text{H}}; \\ & \quad \text{Maximize over } \mathbf{U}^{(1)} \in \mathbb{C}^{I_1 \times R_1} \text{ with } \mathbf{U}^{(1)\text{H}} \mathbf{U}^{(1)} = \mathbf{I}: \\ & \quad h(\mathbf{U}^{(1)}) = \left\| \mathbf{U}^{(1)\text{H}} \cdot \tilde{\mathbf{U}}_{k+1}^{(1)} \right\|; \\ & \quad \max \left(h(\mathbf{U}^{(1)}) \right) = h(\mathbf{U}_{\max}^{(1)}) = \lambda_{k+1}^{(1)}; \\ & \quad \mathbf{U}_{k+1}^{(1)} = \mathbf{U}_{\max}^{(1)}; \\ & - \tilde{\mathbf{U}}_{k+1}^{(2)} = \mathcal{A} \times_1 \mathbf{U}_{k+1}^{(1)\text{H}} \times_3 \mathbf{U}_k^{(3)\text{H}} \dots \times_N \mathbf{U}_k^{(N)\text{H}}; \end{aligned}$$

Maximize over $\mathbf{U}^{(2)} \in \mathbb{C}^{I_2 \times R_2}$ with $\mathbf{U}^{(2)\text{H}}\mathbf{U}^{(2)} = \mathbf{I}$:

$$h\left(\mathbf{U}^{(2)}\right) = \left\| \mathbf{U}^{(2)\text{H}} \cdot \tilde{\mathbf{U}}_{k+1}^{(2)} \right\|;$$

$$\max\left(h\left(\mathbf{U}^{(2)}\right)\right) = h\left(\mathbf{U}_{\max}^{(2)}\right) = \lambda_{k+1}^{(2)};$$

$$\mathbf{U}_{k+1}^{(2)} = \mathbf{U}_{\max}^{(2)};$$

– ...

$$\tilde{\mathbf{U}}_{k+1}^{(N)} = \mathcal{A} \times_1 \mathbf{U}_{k+1}^{(1)\text{H}} \times_2 \mathbf{U}_{k+1}^{(2)\text{H}} \cdots \times_{N-1} \mathbf{U}_{k+1}^{(N-1)\text{H}};$$

Maximize over $\mathbf{U}^{(N)} \in \mathbb{C}^{I_N \times R_N}$ with $\mathbf{U}^{(N)\text{H}}\mathbf{U}^{(N)} = \mathbf{I}$:

$$h\left(\mathbf{U}^{(N)}\right) = \left\| \mathbf{U}^{(N)\text{H}} \cdot \tilde{\mathbf{U}}_{k+1}^{(N)} \right\|;$$

$$\max\left(h\left(\mathbf{U}^{(N)}\right)\right) = h\left(\mathbf{U}_{\max}^{(N)}\right) = \lambda_{k+1}^{(N)};$$

$$\mathbf{U}_{k+1}^{(N)} = \mathbf{U}_{\max}^{(N)}.$$

5.5 Summary

In this chapter, we briefly described the tensor-algebraic material that will be needed in the further derivations. Interested readers are referred to [46, 21, 22, 23] for more details.

Chapter 6

Exponential data fitting using multilinear algebra: the multi-channel case

In this chapter we present the higher-order counterpart of the HTLS and HTLSstack methods described in Chapter 2. The method typically consists of arranging single-channel or multi-channel data in a higher-order tensor and performing a rank reduction of this tensor. The parameters of the signals are estimated using the TLS method.

6.1 Introduction

In the introduction to the HTLS and HTLSstack methods in Chapter 2, it has been shown that the EDS model lends itself to a subspace-based representation. In the single-channel case, the typical structure of the noise-free Hankel data matrix (2.3) allows a Vandermonde decomposition (2.4), while in the multi-channel case, the block Hankel data matrix (2.23) will allow a block Vandermonde decomposition (2.20). In both cases, the (left) Vandermonde vectors characterize the model: the generators are the signal poles, and the number of Vandermonde vectors is equal to the model order, K . However, we have seen that one cannot directly access the Vandermonde vectors but only the subspace spanned by them. In the matrix case this subspace is typically obtained by an SVD of the (block) Hankel data matrix. When noise is added to the signal, the K -dimensional dominant subspace of the data matrix (K being equal to the number of underlying Vandermonde vectors) still yields a

good approximation of the noise-free subspace spanned by the Vandermonde vectors. Generally this dominant subspace can be obtained by truncating the SVD of the data matrix.

A deeper analysis of the *multi*-channel matrix approach reveals that the structure of the columns is only partially reflected by the SVD. In the single-channel case there is an equivalence between the VDMD (2.4) and the SVD (2.28) that can be expressed as follows:

$$\begin{aligned}\text{span}(\mathbf{S}) &= \text{span}(\mathbf{U}) \\ \text{span}(\mathbf{T}) &= \text{span}(\mathbf{V}^*)\end{aligned}\tag{6.1}$$

In the multi-channel case we only have $\text{span}(\mathbf{S}) = \text{span}(\mathbf{U})$. Obviously, the repetition of matrix \mathbf{T}^T in (2.20) is not taken into account by the matrix SVD. This structure into account using multilinear algebra, thereby improving the results so far obtained.

Although in the *single*-channel case the complete structure is already taken into account by the matrix approach. However, we will show that a higher-order representation is possible and that the resulting subspace estimation *may* be more accurate than the one obtained by the matrix approach. Only the third-order case will be addressed. The generalisation to orders higher than three is straightforward.

In Section 6.2 we derive two tensor based methods for harmonic retrieval: the single-channel case is addressed in Section 6.2.1 and the multi-channel case is addressed in Section 6.2.2. Then in Section 6.3 the performance of these algorithms is illustrated by means of simulations. Finally in Section 6.4 we conclude the chapter.

6.2 Tensor-based algorithms for harmonic retrieval

6.2.1 Single-channel signals

First, we stack the data in a third-order tensor as follows. A first $(I_1 \times I_2)$ -Hankel matrix is built using the P ($P < N$) first samples of this signal x_0, x_1, \dots, x_{P-1} . A second matrix is built using the segment x_1, x_2, \dots, x_P , obtained by shifting the previous segment over one sample, and so on until the end of the signal. Then these Hankel matrices are stacked one behind the other in a $(I_1 \times I_2 \times I_3)$ -Hankel tensor \mathcal{H} as visualized in figure 6.1. An element

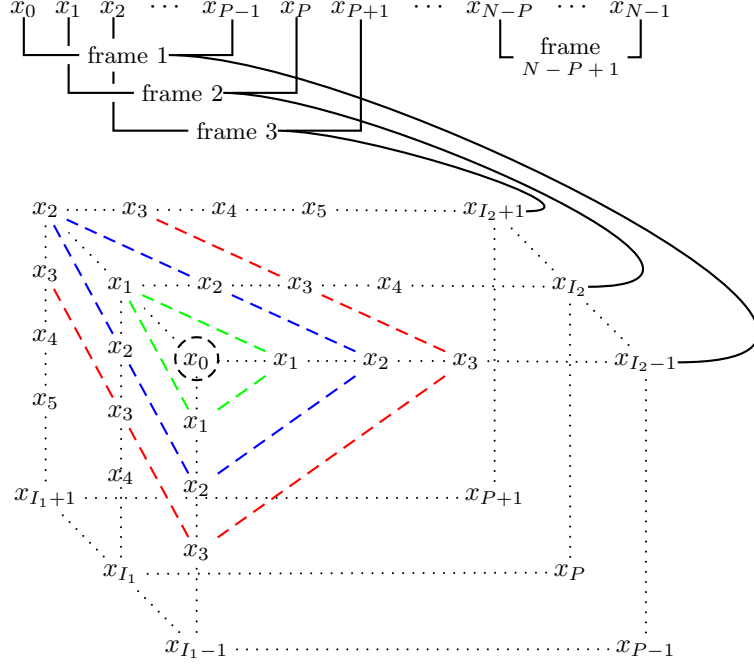


Figure 6.1: Segmentation of the signal and construction of a tensor with Hankel matrices. The dotted lines delimit the tensor while the dashed lines show his Hankel structure in the three directions. The dashed lines show the diagonal tensor slices on which all entries are equal.

$h_{i_1 i_2 i_3}$ of this tensor is given by:

$$h_{i_1 i_2 i_3} = x_{(i_1-1)+(i_2-1)+(i_3-1)} = x_{i_1+i_2+i_3-3} \quad (6.2)$$

where $1 \leq i_1 \leq I_1$, $1 \leq i_2 \leq I_2$, $1 \leq i_3 \leq I_3$ and $I_1 + I_2 + I_3 = N + 2$. As long as the latter constraint is verified, the dimensions of the tensor may be chosen by the user. In what follows, we assume, $I_1 > K$. Let the one-dimensional complex signal x_n be a K pole time domain signal modeled by equation (2.2). Then we have:

$$h_{i_1 i_2 i_3} = \sum_{k=1}^K c_k (z_k^{i_1-1} \cdot z_k^{i_2-1} \cdot z_k^{i_3-1}) . \quad (6.3)$$

In other words, the tensor \mathcal{H} is a weighted sum of 3rd-order rank-1 tensors, consisting of the outer product of three Vandermonde vectors:

$$\mathcal{H} = \sum_{k=1}^K c_k \cdot \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{I'_1} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{I'_2} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{I'_3} \end{pmatrix}, \quad (6.4)$$

where $I'_n = I_n - 1$. This can also be written as:

$$\mathcal{H} = \mathcal{C} \times_1 \mathbf{S}^{(1)} \times_2 \mathbf{S}^{(2)} \times_3 \mathbf{S}^{(3)}, \quad (6.5)$$

in which \mathcal{C} is the pseudo-diagonal $(K \times K \times K)$ -core-tensor containing the K complex amplitudes c_k ; $\mathbf{S}^{(1)} \in \mathbb{C}^{K \times I_1}$, $\mathbf{S}^{(2)} \in \mathbb{C}^{K \times I_2}$ and $\mathbf{S}^{(3)} \in \mathbb{C}^{K \times I_3}$ are Vandermonde matrices. In analogy with the VDMD (2.4), the decomposition (6.5) is called a Higher-Order Vandermonde Decomposition (HOVDMD). It is visualized in figure 6.2. By way of comparison, note that in the matrix case we can write:

$$\begin{aligned} \mathbf{H} &= \sum_{k=1}^K c_k \cdot \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{L'} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{M'} \end{pmatrix}, \\ &= \mathbf{C} \times_1 \mathbf{S} \times_2 \mathbf{T}, \end{aligned} \quad (6.6)$$

where $M' = M - 1$ and $L' = L - 1$. The three matrices \mathbf{C} , \mathbf{S} and \mathbf{T} are defined in (2.4). From the structure of the HOVDMD (6.5) it follows that the n -ranks of \mathcal{H} are all equal to the number of signal poles K . A decomposition reflecting the column/row rank deficiency of \mathcal{H} can be used to retrieve the parameters of interest. Note that, the n -mode vector space of \mathcal{H} equals the column space of $\mathbf{S}^{(n)}$ ($1 \leq n \leq 3$).

The structure induced by (6.5) implies that, in the absence of noise, the HOSVD of \mathcal{H} takes the following form :

$$\mathcal{H} = \widehat{\mathcal{D}} \times_1 \widehat{\mathbf{U}}^{(1)} \times_2 \widehat{\mathbf{U}}^{(2)} \times_3 \widehat{\mathbf{U}}^{(3)}, \quad (6.7)$$

in which $\widehat{\mathcal{D}}$ is an all-orthogonal, ordered, complex $(K \times K \times K)$ -tensor and $\widehat{\mathbf{U}}^{(n)} = [U_1^{(n)}, \dots, U_K^{(n)}]$ is a complex $(I_n \times K)$ -matrix whose orthonormal columns span the column space of $\mathbf{S}^{(n)}$. In the presence of noise, \mathcal{H} will be a full n -rank tensor. Just like in the matrix case, it makes sense to proceed with the HOSVD-components of the best rank- (R_1, R_2, R_3) approximation of

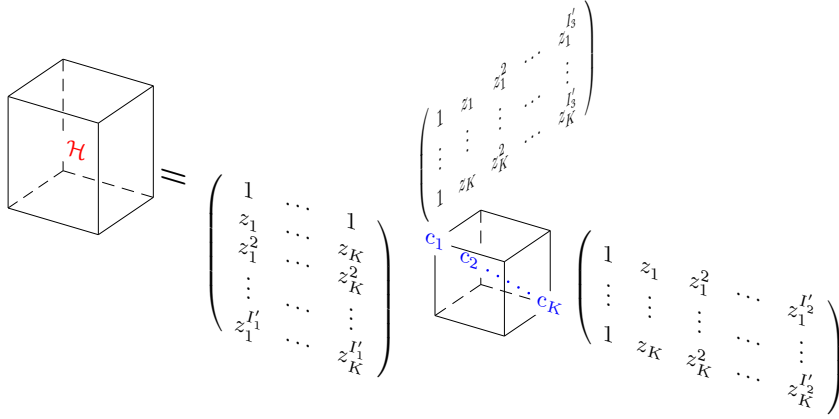


Figure 6.2: Visualization of the HOVDMD for a third order Hankel tensor

\mathcal{H} . Like in the matrix case, we claim that $\widehat{\mathbf{U}}^{(n)}$ equals $\mathbf{S}^{(n)}$ ($n = 1, 2, 3$) up to a multiplication by a square non-singular matrix $\mathbf{Q} \in \mathbb{C}^{K \times K}$:

$$\widehat{\mathbf{U}}^{(n)} = \mathbf{S}^{(n)} \mathbf{Q}. \quad (6.8)$$

On the other hand, the shift-invariance property still holds in the tensor case:

$$\mathbf{S}_{\downarrow}^{(n)} \mathbf{Z} = \mathbf{S}^{(n)\dagger}, \quad (6.9)$$

with $\mathbf{Z} = \text{diag}\{z_1, z_2, \dots, z_K\}$. Combining (6.8) and (6.9) yields the following matrix equation:

$$\widehat{\mathbf{U}}^{(n)\dagger} = \widehat{\mathbf{U}}_{\downarrow}^{(n)} \bar{\mathbf{Z}}, \quad (6.10)$$

with $\bar{\mathbf{Z}} = \mathbf{Q}^{-1} \mathbf{Z} \mathbf{Q}$. This equation is similar to (2.10) and can be processed in the same way.

To conclude, we outline a tensor-based algorithm for the estimate of the K signal poles of x_n (2.1). This algorithm is called HO-HTLS standing for Higher-Order HTLS.

Algorithm 6.2.1

HO-HTLS

Input: data samples x_n , $n = 0, \dots, N - 1$ and model order K .

Output: $\hat{a}_k, \hat{\varphi}_k, \hat{\alpha}_k, \hat{\nu}_k$, $k = 1, \dots, K$.

Step 1: map x_n on a third-order tensor \mathcal{H} as in figure 6.1

Step 2: find the best rank- (K', K', K') approximation of \mathcal{H} (for instance, by applying the HOOI algorithm [22], Section 5.4), and get the matrices $\hat{\mathbf{U}}^{(1)}$, $\hat{\mathbf{U}}^{(2)}$ and $\hat{\mathbf{U}}^{(3)}$. The rank K' is chosen equal to the model order K which corresponds to the number of complex exponentials in the signal. If the signal is real then $K' = 2K$.

Step 3: Choose one of the three n -mode singular vector matrix $\hat{\mathbf{U}}^{(n)}$, $n = 1, 2, 3$ and form the following overdetermined set of equations

$$\hat{\mathbf{U}}^{(n)\uparrow} \approx \hat{\mathbf{U}}_{\downarrow}^{(n)} \bar{\mathbf{Z}},$$

with $\bar{\mathbf{Z}} = \mathbf{Q}^{-1} \mathbf{Z} \mathbf{Q}$. $\hat{\mathbf{U}}^{(n)\uparrow}$ and $\hat{\mathbf{U}}_{\downarrow}^{(n)}$ are derived from $\hat{\mathbf{U}}^{(n)}$ by omitting its first and last row respectively.

Step 4: compute an estimate of $\bar{\mathbf{Z}}$ by solving the above set of equations in the TLS sense.

Step 5: compute the eigenvalues of $\bar{\mathbf{Z}}$. Once $\bar{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of the signal poles:

$$\lambda_k = \hat{z}_k = \exp \{(-\hat{\alpha}_k + 2\pi j \hat{\nu}_k) \Delta t\}.$$

from which it is easy to obtain estimates of the damping factors α_k and frequencies ν_k .

Step 6: Eventually using the estimates \hat{z}_k , $k = 1, \dots, K$, and the signal sample x_n , $n = 0, \dots, N - 1$ compute the LS solution $\hat{c}_k = \hat{a}_k \exp\{j\hat{\phi}_k\}$ of

$$\begin{pmatrix} 1 & \cdots & 1 \\ \hat{z}_1^1 & \cdots & \hat{z}_K^1 \\ \hat{z}_1^2 & \cdots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \cdots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}$$

6.2.2 Multi-channel signals

We now consider the multi-channel signal as described by equation (2.16). In this case, a structure similar to the one in (6.4) can be obtained in a simple way. We now form a third-order tensor \mathcal{H} by stacking the Q Hankel matrices \mathbf{H}_q in (2.22) one behind the other. This is visualized in figure 6.3. An element of \mathcal{H} is given by :

$$h_{i_1 i_2 i_3} = x_{(i_1-1)+(i_2-1)}^{(i_3)} = x_{i_1+i_2-2}^{(i_3)}, \quad (6.11)$$

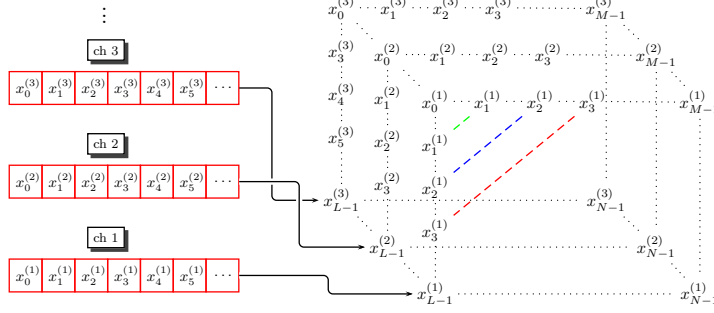


Figure 6.3: Construction of a tensor from Hankel matrices representing a multi-channel signal. The dotted lines delimit the tensor while the dashed lines show its partial Hankel structure.

where $1 \leq i_1 \leq L$, $1 \leq i_2 \leq M$, $1 \leq i_3 \leq Q$, $\min\{L, M\} > K$, and $L + M = N + 1$. Expressing each element of \mathcal{H} with respect to the model (2.16) yield:

$$h_{i_1 i_2 i_3} = \sum_{k=1}^K c_k^{(i_3)} (z_k^{i_1-1} \cdot z_k^{i_2-1}) . \quad (6.12)$$

In the most general case, the coefficients $c_k^{(a)}$ are all different. One can decompose \mathcal{H} as follows:

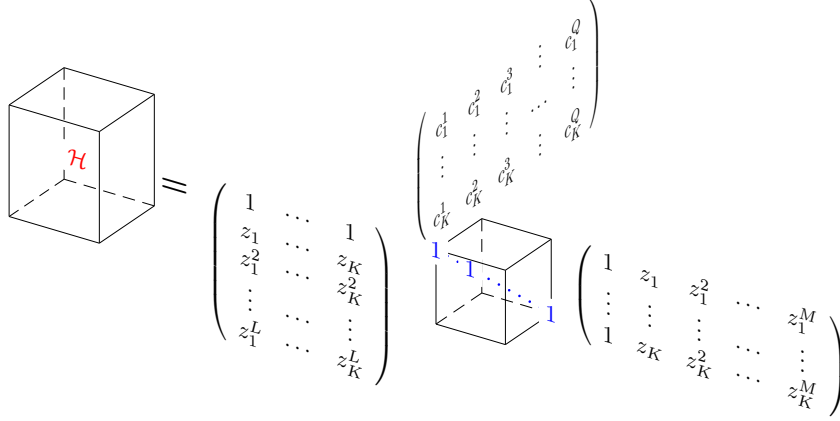
$$\mathcal{H} = \sum_k \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^L \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^M \end{pmatrix} \circ \begin{pmatrix} c_k^{(1)} \\ c_k^{(2)} \\ c_k^{(3)} \\ \vdots \\ c_k^{(Q)} \end{pmatrix} . \quad (6.13)$$

This might compactly be written as:

$$\mathcal{H} = \mathcal{I} \times_1 \mathbf{S}^{(1)} \times_2 \mathbf{S}^{(2)} \times_3 \mathbf{C} , \quad (6.14)$$

in which \mathcal{I} is a pseudo-diagonal $(K \times K \times K)$ -tensor with ones on its diagonal (or unit tensor), $\mathbf{S}^{(1)} \in \mathbb{C}^{L \times K}$ and $\mathbf{S}^{(2)} \in \mathbb{C}^{M \times K}$ are Vandermonde matrices, and $\mathbf{C} \in \mathbb{C}^{K \times Q}$ is a matrix containing the complex amplitudes. This decomposition is visualized in figure 6.4. The structure induced by (6.14) implies that, in the absence of noise, the HOSVD of \mathcal{H} takes the following form :

$$\mathcal{H} = \widehat{\mathcal{F}} \times_1 \widehat{\mathbf{U}}^{(1)} \times_2 \widehat{\mathbf{U}}^{(2)} \times_3 \widehat{\mathbf{W}} , \quad (6.15)$$

Figure 6.4: Decomposition of the tensor \mathcal{H} in the multi-channel case

in which $\widehat{\mathcal{F}}$ is an all-orthogonal, ordered, complex $(K \times K \times K)$ -tensor, $\widehat{\mathbf{U}}^{(1)} = [U_1^{(1)}, \dots, U_K^{(1)}]$ is a complex $(L \times K)$ -matrix whose orthonormal columns span the column space of $\mathbf{S}^{(1)}$, $\widehat{\mathbf{U}}^{(2)} = [U_1^{(2)}, \dots, U_K^{(2)}]$ is a complex $(M \times K)$ -matrix whose orthonormal columns span the column space of $\mathbf{S}^{(2)}$, and $\widehat{\mathbf{W}}$ is a complex $(Q \times K)$ -matrix whose orthonormal columns span the column space of \mathbf{C} . In the presence of noise, as \mathcal{H} becomes n -mode full rank, we can proceed with the HOSVD components of the best rank- (K, K, K) approximation of \mathcal{H} . Like in equation (6.8), we can claim that $\widehat{\mathbf{U}}^{(n)}$ equals $\mathbf{S}^{(n)}$ ($n = 1, 2$) up to a multiplication by some square non-singular matrix. The shift-invariance property of $\mathbf{S}^{(n)}$ can be exploited in the same way as in (6.9), (6.10) and (2.10). Note that the algebraic structure reflected by (6.13) arises very naturally from the structure of the data. If the number of exponentials K is smaller than the number of channels Q , the tensor is mode-3 rank deficient. This property, which is not exploited in the matrix approach, makes the tensor approach more accurate. Above, we outline the algorithm in the multi-channel case. This algorithm is called HO-HTLSstack standing for Higher-Order HTLSstack.

Algorithm 6.2.2

HO-HTLSstack

Input: data samples $x_n^{(q)}$, $n = 0, \dots, N - 1$, $q = 1, \dots, Q$ and model order K .

Output: $\hat{a}_k^{(q)}$, $\hat{\varphi}_k^{(q)}$, $\hat{\alpha}_k$, $\hat{\nu}_k$, $k = 1, \dots, K$, $q = 1, \dots, Q$.

Step 1: For every signal $x_n^{(q)}$, $q = 1, \dots, Q$, arrange the data points $x_n^{(q)}$, $n = 0, \dots, N - 1$ in a $(L \times M)$ -Hankel matrix \mathbf{H}_q , $N = L + M - 1$,

$L > K$ as follows:

$$\mathbf{H}_q = \begin{pmatrix} x_0^{(q)} & x_1^{(q)} & x_2^{(q)} & \cdots & x_{M-1}^{(q)} \\ x_1^{(q)} & x_2^{(q)} & \ddots & \cdots & \vdots \\ x_2^{(q)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2}^{(q)} \\ x_{L-1}^{(q)} & \cdots & \cdots & x_{N-2}^{(q)} & x_{N-1}^{(q)} \end{pmatrix}.$$

As for the HTLSstack algorithm it is recommended to have \mathbf{H}_q as square as possible.

Step 2: form a $(L \times M \times Q)$ -tensor \mathcal{H} by stacking the Q Hankel matrices \mathbf{H}_q one behind the other, as indicated by figure 6.3.

Step 3: find the best rank- $(K', K', \min(K', Q))$ approximation of \mathcal{H} (for instance, by applying the HOOI algorithm [22], Section 5.4), and get the matrices $\hat{\mathbf{U}}^{(1)}$ and $\hat{\mathbf{U}}^{(2)}$ ($\hat{\mathbf{U}}^{(3)}$ does not contain any relevant information for our algorithm). The rank K' is chosen equal to the model order K which corresponds to the number of complex exponentials in the signal. If the signal is real then $K' = 2K$.

Step 4: Choose one of the two n -mode singular vector matrix $\hat{\mathbf{U}}^{(n)}$, $n = 1, 2$ and form the following overdetermined set of equations

$$\hat{\mathbf{U}}^{(n)\dagger} \approx \hat{\mathbf{U}}_{\downarrow}^{(n)} \bar{\mathbf{Z}},$$

with $\bar{\mathbf{Z}} = \mathbf{Q}^{-1} \mathbf{Z} \mathbf{Q}$. The matrices $\hat{\mathbf{U}}^{(n)\dagger}$ and $\hat{\mathbf{U}}_{\downarrow}^{(n)}$ are derived from $\hat{\mathbf{U}}^{(n)}$ by omitting its first and last row respectively.

Step 5: compute an estimate of $\bar{\mathbf{Z}}$ by solving the above set of equations in the TLS sense.

Step 6: compute the eigenvalues of $\bar{\mathbf{Z}}$. Once $\bar{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of the signal poles:

$$\lambda_k = \hat{z}_k = \exp \{(-\hat{\alpha}_k + 2\pi j \hat{\nu}_k) \Delta t\}.$$

from which it is easy to obtain estimates of the damping factors α_k and frequencies ν_k .

Step 7: Eventually, for each signal q , an estimate of the complex amplitudes $c_k^{(q)} = a_k^{(q)} \exp\{j\varphi_k^{(q)}\}$ is obtained from the LS solution of

$$\begin{pmatrix} 1 & \cdots & 1 \\ \hat{z}_1^1 & \cdots & \hat{z}_K^1 \\ \hat{z}_1^2 & \cdots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \cdots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1^{(q)} \\ c_2^{(q)} \\ \vdots \\ c_K^{(q)} \end{pmatrix} = \begin{pmatrix} x_0^{(q)} \\ x_1^{(q)} \\ x_2^{(q)} \\ \vdots \\ x_{N-1}^{(q)} \end{pmatrix}$$

6.3 Results

For single-channel data, the accuracy of the matrix approach is compared in this section to the accuracy of the tensor approach by means of the two peaks example from [85]:

$$\begin{aligned} y_n &= x_n + e_n \\ &= \underbrace{\exp\{-0.01 + 2j\pi 0.2\} n}_{\text{pole 1}} + \underbrace{\exp\{-0.02 + 2j\pi 0.22\} n}_{\text{pole 2}} + \underbrace{e_n}_{\text{noise}}, \quad (6.16) \\ & \quad n = 0, 1, \dots, N = 24 \end{aligned}$$

in which e_n is a complex circular symmetric WGN. The reason for the choice of this problem is related to the difficulty of its resolution: first, the signal is composed with a few number of points and then the two peaks are very closely spaced. For the matrix approach, we work with the matrix $\hat{\mathbf{U}}$ containing the 2 dominant left singular vectors of the Hankel matrix \mathbf{H} . For the tensor approach, we work with the matrix $\hat{\mathbf{U}}^{(3)}$ in the HOSVD (5.8) of the best rank-(2, 2, 2) approximation of \mathcal{H} . We conduct Monte Carlo experiments consisting of $N_{\text{runs}} = 1000$ runs. The performance is expressed in terms of the RRMSE:

$$RRMSE = \left[\frac{1}{N_{\text{runs}}} \sum_{i=1}^{N_{\text{runs}}} |\gamma - \hat{\gamma}_i|^2 \right]^{\frac{1}{2}} \cdot \frac{100}{\gamma} \quad [\%] \quad (6.17)$$

where $\hat{\gamma}_i$ is the estimate of the parameter γ in the i th run. It is a fact that, when dealing with subspace-based processing, the dimension of the initial data matrix is very important. But to our knowledge, no strong theoretical foundation exists which enables to find easily the optimal matrix size, and we have unfortunately the same problem for the tensor approach. However we wish to compare the matrix and tensor approach in a consistent way, which means that we must start with the optimal size for both the data matrix \mathbf{H} and tensor \mathcal{H} . As far as the matrix approach is concerned, we could explore all the possible

matrix sizes and determine for which one the performance of the algorithm is optimal. This idea is used in [86] where it is explained that the HTLS algorithm applied to a $i \times (N + 1 - i)$ Hankel matrix yields very good results when $i \in [N/3, 2N/3]$. This is confirmed by figure 6.5. To compare with the

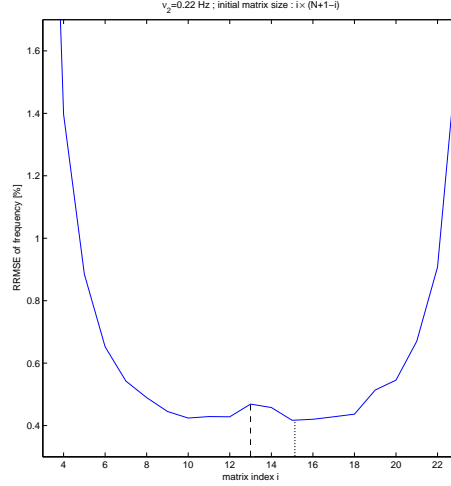


Figure 6.5: RRMSE of ν_2 at SNR=30 dB obtained for matrices of dimension $i \times (N + 1 - i)$. The index corresponding to the highest accuracy is $i = 15$. The performance curve for the other parameters exhibits exactly the same behavior.

tensor approach, we chose the optimal dimensionality (15×11). Following [86], for the tensor approach, all possible values of I_1 , I_2 have been tried with I_3 constrained:

$$I_3 = N + 2 - I_1 - I_2 = 27 - I_1 - I_2, \quad (6.18)$$

and the result is shown in figure 6.6. The best results were obtained for tensors that consist of a small number of vertical “slices” ($I_3 = 4, 5, 6$). For these values the results obtained by the tensor approach are consistently more accurate than the best matrix results over a large range of values for I_1 and I_2 . Like in the matrix case, the best result is obtained for vertical slices that are slightly rectangular. In this example we retained the optimal dimensionality ($14 \times 8 \times 5$). Matrix and tensor results are compared, for varying SNR, in figure 6.7. The average improvement in accuracy is given in table 6.1.

	Frequency	Damping factor
pole 1	4.2 %	5.5 %
pole 2	5%	4.1 %

Table 6.1: Average improvement of the tensor approach compared to the matrix approach in the single-channel case.

For the multi-channel case model (6.16) has been slightly modified as follows:

$$\begin{aligned}
y_n^{(q)} &= x_n^{(q)} + e_n^{(q)} \\
&= c_1^{(q)} \exp\{(-0.01 + 2j\pi 0.2) n\} + c_2^{(q)} \exp\{(-0.02 + 2j\pi 0.22) n\} + e_n^{(q)}, \\
&\quad \text{pole 1} \qquad \qquad \qquad \text{pole 2} \qquad \qquad \qquad \text{noise} \\
n &= 0, 1, \dots, N = 24 \\
q &= 1, 2, \dots, Q = 12
\end{aligned}
\tag{6.19}$$

in which $e_n^{(q)}$ is a different complex circular symmetric WGN sequence for each channel q . The complex amplitudes $c_k^{(q)}$ are drawn from a Gaussian distribution with zero-mean and unit-variance for $1 \leq k \leq K = 2$ and $1 \leq q \leq Q = 12$. For the matrix approach, we work with the matrix $\hat{\mathbf{U}}$ containing the 2 dominant left singular vectors of the Hankel matrix \mathbf{H} . For the tensor approach, we work with the matrix $\hat{\mathbf{U}}^{(1)}$ in the HOSVD (5.8) of the best rank-(2, 2, 2) approximation of \mathcal{H} . Note that the philosophy of the multi-channel case is very different from the single-channel case: the data are arranged in a different way but the matrices H_q (the ‘‘slices’’) are not modified and therefore the dimension of the subspace of interest remains unchanged. In this case there is a theoretical foundation for the values of I_3 . For all dimensionalities the tensor algorithm performs better than the matrix algorithm. The figures have been drawn for the dimensionalities that lead to the most accurate results for the matrix algorithm. We worked with a $(13 \times 13 \times 12)$ -matrix \mathbf{H} and a $(13 \times 13 \times 12)$ -tensor \mathcal{H} . Figure 6.8 shows that the tensor approach consistently yields better results. Moreover the improvement increases as the noise level increases as shown in figure 6.9. In the curves related to ν_1 and ν_2 , this trend continues for $\sigma > 0.4$, although α_1 and α_2 can no longer be accurately estimated.

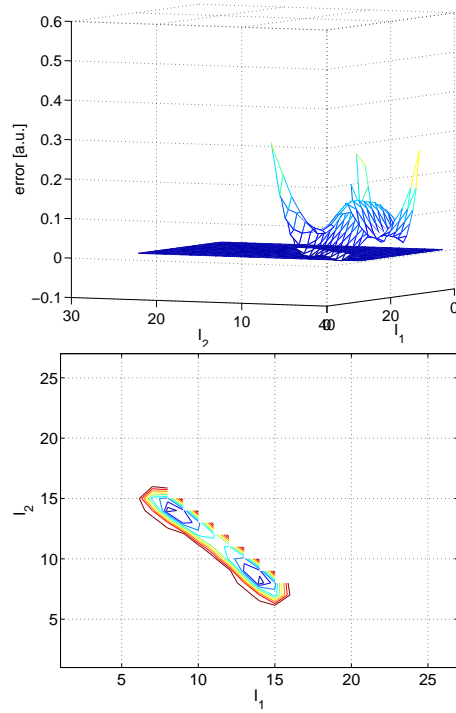


Figure 6.6: Single-channel case. The upper plot is the error map of ν_2 at SNR=30 dB. It shows the RRMSE obtained for tensors of dimension $I_1 \times I_2 \times I_3$ from which is subtracted the smallest RRMSE obtained for the matrix approach (see figure 6.5). Therefore the indices for which the error is negative (below the plane) correspond to a better performance of the tensor approach. The right contour plot represents the 2-D projection of the surface for which the values are negative. The indices corresponding to the highest accuracy are $I_1 = 14$ and $I_2 = 8$. The performance curve for the other parameters exhibit exactly the same behavior.

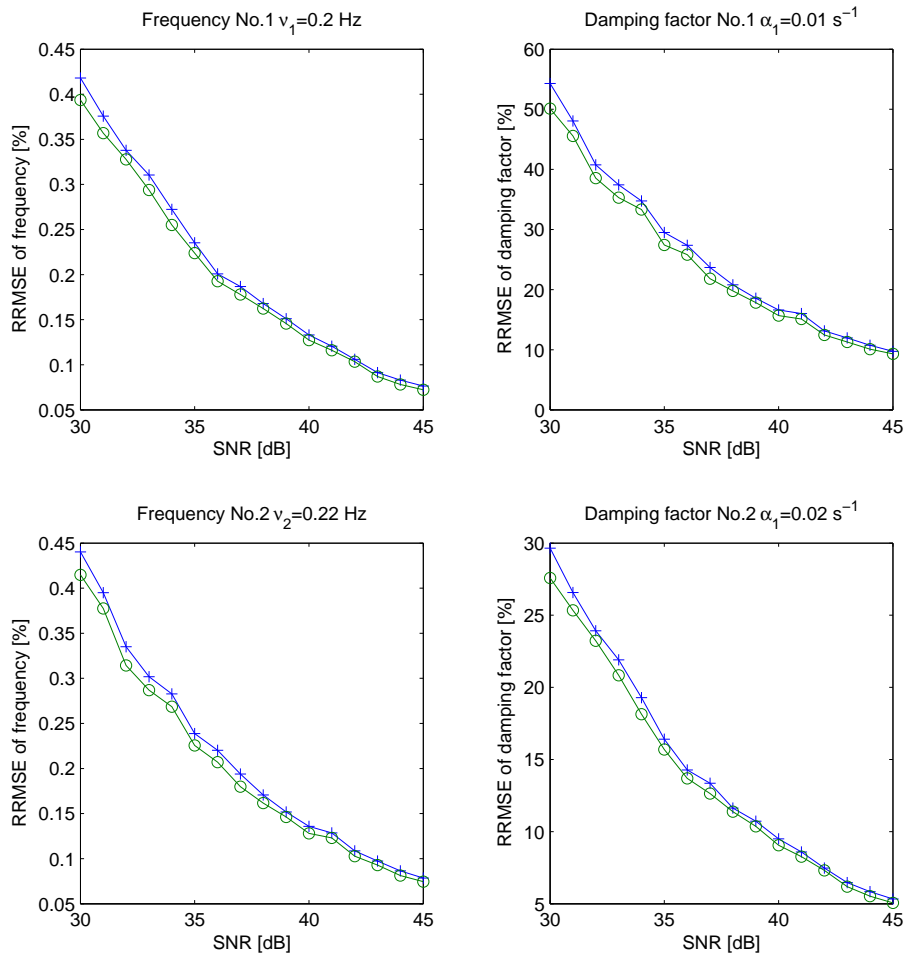


Figure 6.7: Comparison of the results from the matrix approach (+) and the tensor approach (o) in the single-channel case

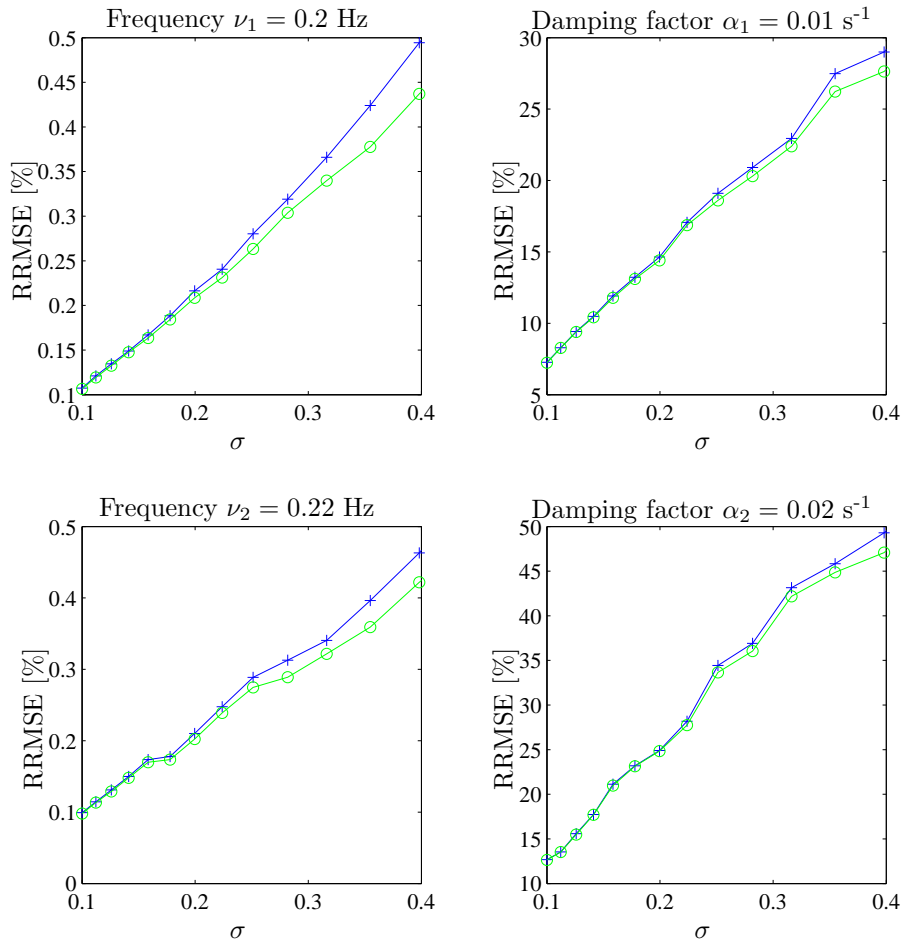


Figure 6.8: Plot of the RRMSE versus noise standard deviation σ in the multi-channel case. The matrix approach is displayed with "+" and the tensor is displayed with "o".

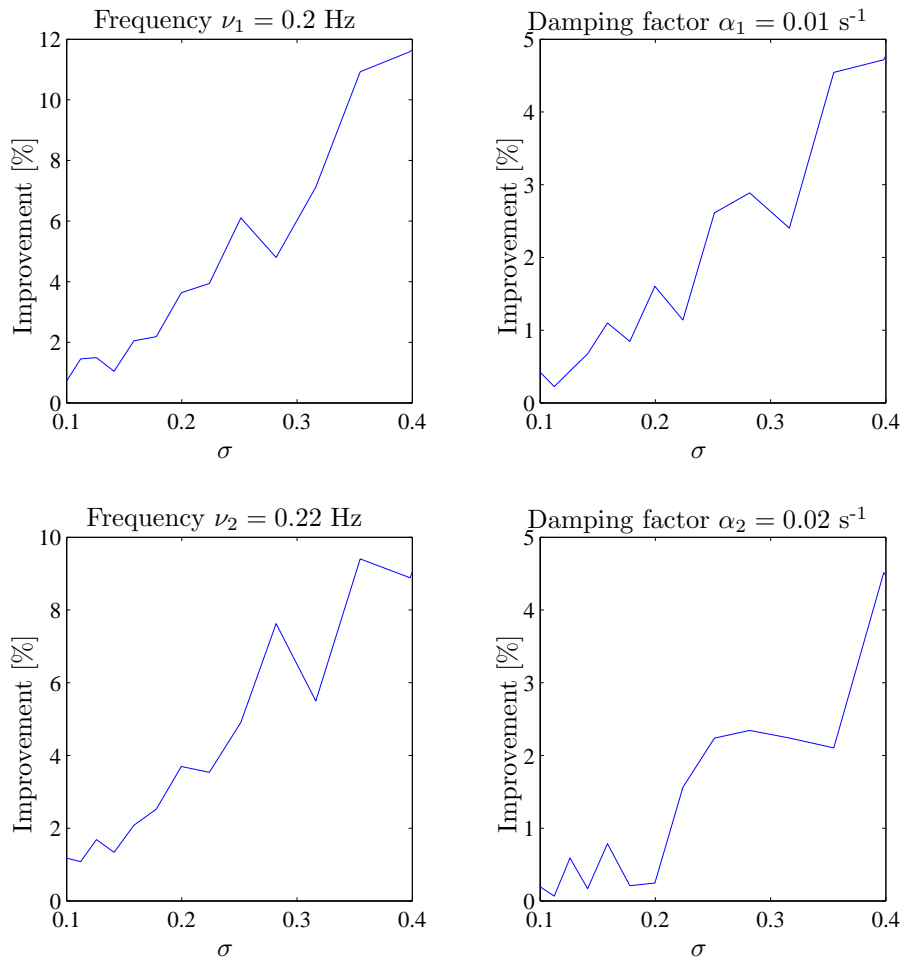


Figure 6.9: Plot of the improvement factor versus noise standard deviation σ in the multi-channel case.

6.4 Conclusions

In this chapter we have introduced the basics of a multilinear algebra based approach to harmonic retrieval in the single-channel and the multi-channel case.

The basis of the approach to single-channel data is the fact that the multiplicative (Vandermonde) structure of an harmonic signal allows to represent this signal as a higher-order rank-1 tensor. We have demonstrated by means of an example that the tensor approach may yield more accurate results than its matrix counterpart. That is, it is in principle *possible* to increase the accuracy. In practice one has to choose proper dimensions of the data tensor. How to choose optimal dimensions is left as an open problem. However, the gain in accuracy is quite modest. We conclude that, in its present form, the single-channel tensor approach is rather of theoretical importance.

For multi-channel data it is very natural to work in a multilinear framework. One simply stacks the different Hankel matrices, corresponding to the different channels, in a third-order tensor, instead of putting them one after the other in a big matrix. The tensor algorithm allows to exploit properties of the data structure that are not being used in matrix techniques. From a computational point of view, the iteration that computes the tensor approximation, is more expensive than current algorithms for the estimation of the dominant subspace of a matrix.

The idea presented in this chapter can be used to derive a tensor version of the matrix approach to decimation-based harmonic analysis of oversampled data [50]. This result is presented in the next chapter.

Chapter 7

Exponential data fitting using multilinear algebra: the decimative case

The HTLSDstack method described in Chapter 2 can be applied to signals satisfying the EDS model that can be decimated (downsampled). This approach is particularly useful when signal poles are closely spaced because the decimated sequences of such a signal can be seen as independent channels that allow for an improved frequency separation. In this context a higher-order approach is very useful to estimate the signal poles. In this chapter we show that the concepts introduced in Chapter 5 can be used to derive tensor-based algorithms for harmonic retrieval in the decimative case and we propose a higher-order counterpart of the HTLSDstack algorithm, based on these concepts from multilinear algebra. Basically, it consists of decimating an oversampled signal by a factor D , making sure that no aliasing occurs. Each sequence is arranged in a Hankel matrix and the D Hankel matrices are stacked in a third-order tensor. A dimensionality reduction algorithm is applied to the tensor and the resulting subspace is used to retrieve the signal poles. It will be shown that this approach more accurate than its matrix counterpart.

7.1 Introduction

The particular case of signals that can be decimated has been addressed in Chapter 2, when introducing the HTLSDstack method. In this method, the signal is decimated and each decimated sequence is considered as an indepen-

dent channel. Each generated downsampled sequence $x_n^{(d)}$ is arranged in a Hankel matrix and these Hankel matrices are stacked in a big (block) Hankel matrix. The K dimensional dominant subspace of this matrix spans the subspace in which lie the Vandermonde vectors whose generators are the signal poles. In essence, the HTLSDstack and the HTLSstack techniques are similar. This corresponds to the fact that the theoretical decomposition of \mathbf{H} in (2.24) is similar to that in (2.20). As noticed in Chapter 6, the repetition of the Vandermonde matrix \mathbf{V}^T in each block of \mathbf{H} in (2.20), cannot be reflected by a simple matrix SVD. It is therefore useful to work in a tensor framework. The same argument applies to decimated data.

It will also become clear that for oversampled data, an unsymmetric tensor approximation allows to deal with effect of ill-conditioning, which is an additional advantage of working in the framework of multilinear algebra.

In Section 7.2 a multilinear approach which allows one to take this property into account is described. Then in Section 7.3 we show how the subspace of interest can be estimated by means of dimensionality reduction of the data tensor. Some additional comments on n -rank reduction of tensors are provided in Section 7.4. An application of the tensor-based algorithm to simulated data is given in Section 7.5 and finally Section 7.6 conclude the chapter.

7.2 Higher-order Vandermonde structure of decimated data sequences

Let us form a $(L_D \times M_D \times D)$ tensor \mathcal{H} by stacking the D Hankel matrices \mathbf{H}_d in (2.22) one behind the other. This is visualized in Figure 7.1. An element $h_{i_1 i_2 i_3}$ of this tensor is given by:

$$h_{i_1 i_2 i_3} = x_{(i_1-1)D+(i_2-1)D+(i_3-1)}, \quad (7.1)$$

where $1 \leq i_1 \leq L_D$, $1 \leq i_2 \leq M_D$, $1 \leq i_3 \leq D$, $L_D + M_D - 1 = N_D$ or $(L_D + M_D - 1)D = N$. As long as the latter constraint is verified, the dimensions of the tensor may be chosen by the user. Let the one-dimensional noise-free complex signal \hat{x}_n be a K -pole time domain signal modeled by equation (2.21). In what follows, we assume that $L_D > K$. If we replace each sample in Figure 7.1 by this model, an element of the tensor can be expressed as follows:

$$h_{i_1 i_2 i_3} = \sum_{k=1}^K c_k \left(z_k^{(i_1-1)D} \cdot z_k^{(i_2-1)D} \cdot z_k^{i_3-1} \right) \quad (7.2)$$

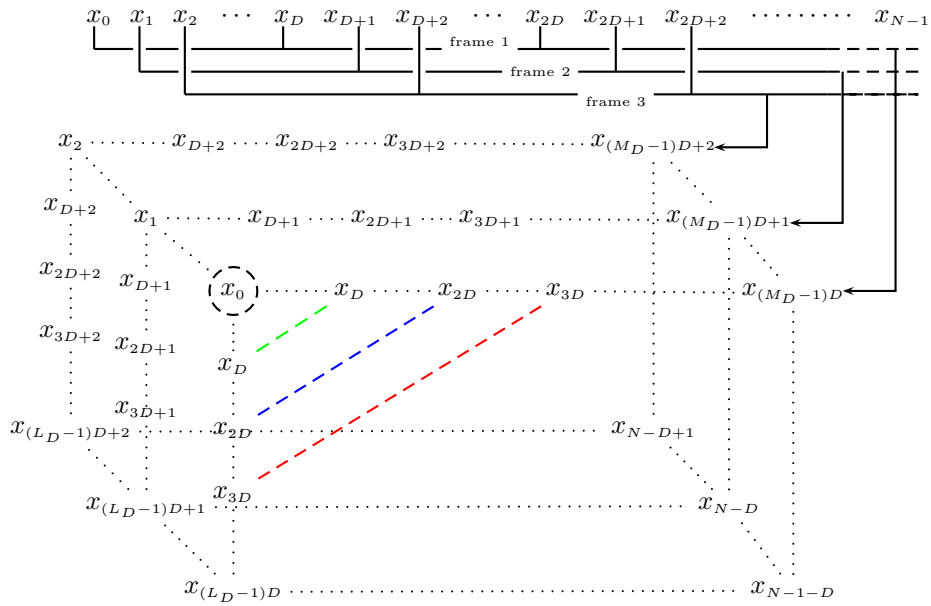


Figure 7.1: Construction of a tensor from Hankel matrices representing a decimated signal. The dotted lines delimit the tensor while the dashed lines show its partial Hankel structure.

and therefore the tensor \mathcal{H} is a weighted sum of third-order rank-1 tensors, consisting of the outer product of three Vandermonde vectors:

$$\mathcal{H} = \sum_{k=1}^K c_k \cdot \begin{pmatrix} 1 \\ (z_k^D)^1 \\ (z_k^D)^2 \\ \vdots \\ (z_k^D)^{L_D-1} \end{pmatrix} \circ \begin{pmatrix} 1 \\ (z_k^D)^1 \\ (z_k^D)^2 \\ \vdots \\ (z_k^D)^{M_D-1} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{D-1} \end{pmatrix} \quad (7.3)$$

This can also be written as:

$$\mathcal{H} = \mathcal{C} \times_1 \mathbf{S}^{(1)} \times_2 \mathbf{S}^{(2)} \times_3 \mathbf{S}^{(3)} \quad (7.4)$$

in which \mathcal{C} is the pseudo-diagonal $(K \times K \times K)$ -core-tensor containing the K complex amplitudes c_k ; $\mathbf{S}^{(1)} \in \mathbb{C}^{K \times L_D}$, $\mathbf{S}^{(2)} \in \mathbb{C}^{K \times M_D}$ and $\mathbf{S}^{(3)} \in \mathbb{C}^{K \times D}$ are Vandermonde matrices. Recall that the VDMD of a $(L \times M)$ Hankel matrix \mathbf{H} is given by:

$$\mathbf{H} = \sum_{k=1}^K c_k \cdot \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{L-1} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{M-1} \end{pmatrix}, \quad (7.5)$$

$$= \mathbf{C} \times_1 \mathbf{S} \times_2 \mathbf{T}, \quad (7.6)$$

in which

$$\mathbf{S} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1^1 & \cdots & z_K^1 \\ z_1^2 & \cdots & z_K^2 \\ \vdots & \cdots & \vdots \\ z_1^{L-1} & \cdots & z_K^{L-1} \end{pmatrix}, \quad (7.7)$$

$$\mathbf{T} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1^1 & \cdots & z_K^1 \\ z_1^2 & \cdots & z_K^2 \\ \vdots & \cdots & \vdots \\ z_1^{M-1} & \cdots & z_K^{M-1} \end{pmatrix}, \quad (7.8)$$

$$\mathbf{C} = \text{diag}(c_1, \dots, c_K). \quad (7.9)$$

In analogy with (7.6), decomposition (7.4) is called a HOVDMD. It is visualized in Figure 7.2.

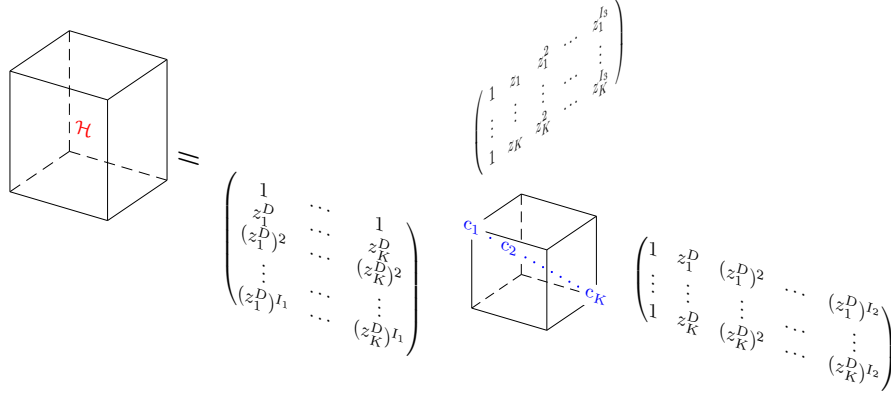


Figure 7.2: Visualization of the HOVDMD of \mathcal{H} . This representation is equivalent to the formal writing (7.3) for $I_1 = L_D - 1$, $I_2 = M_D - 1$ and $I_3 = D - 1$.

7.3 Pole extraction

Notice that the 1-mode vectors of \mathcal{H} are generated as linear combinations of the Vandermonde vectors that form the columns of $\mathbf{S}^{(1)}$ in (7.4). On the other hand, the 1-mode vectors are also generated as linear combinations of the K 1-mode singular vectors of \mathcal{H} that correspond to the non-zero 1-mode singular values. Let the part of the HOSVD of \mathcal{H} that corresponds to the non-zero part of its core tensor be given by:

$$\mathcal{H} = \tilde{\mathcal{B}} \times \tilde{\mathbf{V}}^{(1)} \times \tilde{\mathbf{V}}^{(2)} \times \tilde{\mathbf{V}}^{(3)}, \quad (7.10)$$

in which $\tilde{\mathcal{B}}$ is a complex $(K \times \min(K, M_D) \times \min(K, D))$ all-orthogonal (core) tensor, $\tilde{\mathbf{V}}^{(1)}$, $\tilde{\mathbf{V}}^{(2)}$ and $\tilde{\mathbf{V}}^{(3)}$ are respectively a $(L_D \times K)$, $(M_D \times \min(K, M_D))$ and $(D \times \min(K, D))$ column-wise orthonormal matrix. We now have that

$$\tilde{\mathbf{V}}^{(1)} = \mathbf{S}^{(1)} \mathbf{Q}, \quad (7.11)$$

in which \mathbf{Q} is a $(K \times K)$ nonsingular matrix. Equation (7.11) is the higher-order equivalent of equation (2.30). One can proceed in the same way as in Section 2.4 to compute the signal poles from $\tilde{\mathbf{V}}^{(1)}$. However, one comment has to be made. The derivation so far assumed the absence of noise. If noise is present, tensor \mathcal{H} will not be exactly rank- $(K \times \min(K, M_D) \times \min(K, D))$. As explained in Chapter 5, the best rank- $(K \times \min(K, M_D) \times \min(K, D))$ approximation of \mathcal{H} and hence its dominant 1-mode subspace cannot in general be obtained by mere truncation of the HOSVD. Instead we compute the best

rank- $(K \times \min(K, M_D) \times \min(K, D))$ approximation

$$\widehat{\mathcal{H}} = \widehat{\mathcal{B}} \times_1 \widehat{\mathbf{V}}^{(1)} \times_2 \widehat{\mathbf{V}}^{(2)} \times_3 \widehat{\mathbf{V}}^{(3)}, \quad (7.12)$$

in which $\widehat{\mathcal{B}}$ is a complex $(K \times \min(K, M_D) \times \min(K, D))$ all-orthogonal (core) tensor, $\widehat{\mathbf{V}}^{(1)}$, $\widehat{\mathbf{V}}^{(2)}$ and $\widehat{\mathbf{V}}^{(3)}$ are respectively a $(L_D \times K)$, $(M_D \times \min(K, M_D))$ $(D \times \min(K, D))$ column-wise orthonormal matrix, and we estimate the signal poles from $\widehat{\mathbf{V}}^{(1)}$.

7.4 Some comments on n -rank reduction

In the preceding section, we showed that it is natural to compute the signal poles from the best rank- $(K, \min(K, M_D), \min(K, D))$ approximation of \mathcal{H} . However, only the mode-1 K -dimensional dominant subspace is of interest. In this case, the following theorem shows that it is not necessary to confine ourselves to a rank- $(K, \min(K, M_D), \min(K, D))$ approximation of tensor \mathcal{H} :

Theorem 7.1 *Consider a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ that is rank- (R_1, R_2, R_3) . Let the HOSVD of \mathcal{A} be given by*

$$\mathcal{A} = \mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \times_3 \mathbf{V}^{(3)}.$$

Then the best rank- (R_1, R_2, \tilde{R}_3) approximation of \mathcal{A} , with $\tilde{R}_3 < R_3$, is obtained by truncation of \mathcal{B} and $\mathbf{V}^{(3)}$.

Proof: Let the best rank- (R_1, R_2, \tilde{R}_3) approximation of \mathcal{A} be equal to $\tilde{\mathcal{A}}$, and let its HOSVD be given by

$$\tilde{\mathcal{A}} = \tilde{\mathcal{B}} \times_1 \tilde{\mathbf{V}}^{(1)} \times_2 \tilde{\mathbf{V}}^{(2)} \times_3 \tilde{\mathbf{V}}^{(3)}. \quad (7.13)$$

By definition, the unitary matrices $\tilde{\mathbf{V}}^{(1)}$, $\tilde{\mathbf{V}}^{(2)}$, $\tilde{\mathbf{V}}^{(3)}$ maximize [22]:

$$g(\tilde{\mathbf{V}}^{(1)}, \tilde{\mathbf{V}}^{(2)}, \tilde{\mathbf{V}}^{(3)}) = \left\| \mathcal{A} \times_1 \tilde{\mathbf{V}}^{(1)\text{H}} \times_2 \tilde{\mathbf{V}}^{(2)\text{H}} \times_3 \tilde{\mathbf{V}}^{(3)\text{H}} \right\|^2 \quad (7.14)$$

$$= \left\| \left(\mathcal{B} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \times_3 \mathbf{V}^{(3)} \right) \times_1 \tilde{\mathbf{V}}^{(1)\text{H}} \times_2 \tilde{\mathbf{V}}^{(2)\text{H}} \times_3 \tilde{\mathbf{V}}^{(3)\text{H}} \right\|^2 \quad (7.15)$$

$$= \left\| \mathcal{B} \times_1 \left(\tilde{\mathbf{V}}^{(1)\text{H}} \cdot \mathbf{V}^{(1)} \right) \times_2 \left(\tilde{\mathbf{V}}^{(2)\text{H}} \cdot \mathbf{V}^{(2)} \right) \right\|^2$$

$$\times_3 \left(\tilde{\mathbf{V}}^{(3)H} \cdot \mathbf{V}^{(3)} \right) \Big\| \Big|^2. \quad (7.16)$$

For any given choice of $\mathbf{V}^{(3)}$, the function g reaches its maximal value for $\tilde{\mathbf{V}}^{(1)} = \mathbf{V}^{(1)}$, $\tilde{\mathbf{V}}^{(2)} = \mathbf{V}^{(2)}$. Substitution in (7.13) yields:

$$\max_{\tilde{\mathbf{V}}^{(1)}, \tilde{\mathbf{V}}^{(2)}} g \left(\tilde{\mathbf{V}}^{(1)}, \tilde{\mathbf{V}}^{(2)}, \tilde{\mathbf{V}}^{(3)} \right) = \left\| \mathcal{B} \times_3 \left(\tilde{\mathbf{V}}^{(3)H} \cdot \mathbf{V}^{(3)} \right) \right\|^2 \quad (7.17)$$

$$= \left\| \left(\mathcal{B} \times_3 \mathbf{V}^{(3)} \right) \times_3 \tilde{\mathbf{V}}^{(3)H} \right\|^2. \quad (7.18)$$

In (7.18), $\mathcal{B} \times_3 \mathbf{V}^{(3)}$ is itself an HOSVD. From the properties of the HOSVD [21, Prop. 9, p. 1267], it follows that (7.18) is maximal when $\tilde{\mathbf{V}}^{(3)}$ is taken equal to the left $(I_3 \times \tilde{R}_3)$ submatrix of $\mathbf{V}^{(3)}$. In that case we have

$$\tilde{\mathbf{V}}^{(3)H} \cdot \mathbf{V}^{(3)} = \begin{bmatrix} \mathbf{I}_{\tilde{R}_3 \times \tilde{R}_3} & \mathbf{0}_{\tilde{R}_3 \times (R_3 - \tilde{R}_3)} \end{bmatrix}, \quad (7.19)$$

in which $\mathbf{I}_{\tilde{R}_3 \times \tilde{R}_3}$ and $\mathbf{0}_{\tilde{R}_3 \times (R_3 - \tilde{R}_3)}$ are the $(\tilde{R}_3 \times \tilde{R}_3)$ identity matrix and the $(\tilde{R}_3 \times (R_3 - \tilde{R}_3))$ zero matrix, respectively. The corresponding core tensor $\tilde{\mathcal{B}}$ is obtained from

$$\tilde{\mathcal{B}} = \mathcal{A} \times_1 \tilde{\mathbf{V}}^{(1)H} \times_2 \tilde{\mathbf{V}}^{(2)H} \times_3 \tilde{\mathbf{V}}^{(3)H} \quad (7.20)$$

$$= \mathcal{B} \times_1 \left(\tilde{\mathbf{V}}^{(1)H} \cdot \mathbf{V}^{(1)} \right) \times_2 \left(\tilde{\mathbf{V}}^{(2)H} \cdot \mathbf{V}^{(2)} \right) \times_3 \left(\tilde{\mathbf{V}}^{(3)H} \cdot \mathbf{V}^{(3)} \right) \quad (7.21)$$

$$= \mathcal{B} \times_3 \begin{bmatrix} \mathbf{I}_{\tilde{R}_3 \times \tilde{R}_3} & \mathbf{0}_{\tilde{R}_3 \times (R_3 - \tilde{R}_3)} \end{bmatrix}, \quad (7.22)$$

i.e., it is simply obtained by truncation of \mathcal{B} . ■

Taking a value for \tilde{R}_3 that is strictly lower than R_3 is particularly useful when the smallest mode-3 singular values of \mathcal{A} are in the order of magnitude of the noise level. It is then numerically preferable to extract the dominant \tilde{R}_3 -dimensional mode-3 vector space.

Let us now turn back to the problem. The decimative approach is most often used in the case where poles are very close. This means that, in equation (7.4), the Vandermonde matrix $\mathbf{S}^{(3)}$ is very ill-conditioned because its generators are very close in the complex z -plane. The matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are less ill-conditioned because of the decimation: the generators $\{z_k^D\}$ are less close than the generators $\{z_k\}$ when the damping factors are not too large. The condition of the problem can be easily evaluated by inspecting the mode- n singular values in the HOSVD of \mathcal{H} . If, for instance, there is a large gap between the 3-mode singular values, it is better to compute the best rank- $(K, \min(K, M_D), K')$ approximation of \mathcal{H} with $K' < \min(K, D)$. Concentrating on the dominant part of \mathcal{H} increases the robustness.

Below, we outline the algorithm in the decimative case. This algorithm is called HO-HTLSDstack standing for Higher-Order HTLSDstack.

Algorithm 7.4

HO-HTLSDstack

Input: data samples x_n , $n = 0, \dots, N - 1$, decimation factor D , and model order K .

Output: $\hat{a}_k, \hat{\varphi}_k, \hat{\alpha}_k, \hat{\nu}_k$, $k = 1, \dots, K$.

Step 1: Decimate x_n by a factor D making sure that no aliasing effect occurs (filter out frequencies higher than $\max \nu_k$) and obtain D decimated sequences $x_n^{(d)}$. For simplicity it is recommended that N is an integer multiple of D .

Step 2: For every signal $x_n^{(d)}$, $d = 1, \dots, D$, arrange the data points $x_n^{(d)}$, $n = 0, \dots, N/D - 1$ in a $(L_D \times M_D)$ -Hankel matrix \mathbf{H}_d , $N_D = L_D + M_D - 1$, $N_D = N/D$, $L_D > K$, $M_D \cdot D \geq K$ as follows:

$$\mathbf{H}_d = \begin{pmatrix} x_0^{(d)} & x_1^{(d)} & x_2^{(d)} & \cdots & x_{M_D-1}^{(d)} \\ x_1^{(d)} & x_2^{(d)} & \ddots & \cdots & \vdots \\ x_2^{(d)} & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N_D-2}^{(d)} \\ x_{L_D-1}^{(d)} & \cdots & \cdots & x_{N_D-2}^{(d)} & x_{N_D-1}^{(d)} \end{pmatrix},$$

It is recommended to have \mathbf{H}_d as square as possible.

Step 3: map $\hat{x}_n^{(d)}$ to a $(L_D \times M_D \times D)$ -tensor \mathcal{H} as in figure 7.1 (see equation (7.1))

Step 4: find the best rank- $(K, \min(K, M_D), K')$ approximation of \mathcal{H} with $K' \leq \min(K, D)$ (for instance, by applying the HOOI algorithm [22, Sec. 4.1]), and let the columns of $\hat{\mathbf{V}}^{(1)}$ form a basis for the dominant K -dimensional subspace

Step 5: form the following overdetermined set of equations

$$\hat{\mathbf{V}}^{(1)\uparrow} \approx \hat{\mathbf{V}}_{\downarrow}^{(1)} \bar{\mathbf{Z}},$$

with $\bar{\mathbf{Z}} = \mathbf{Q}^{-1} \mathbf{Z} \mathbf{Q}$. The matrices $\hat{\mathbf{V}}^{(1)\uparrow}$ and $\hat{\mathbf{V}}_{\downarrow}^{(1)}$ are derived from $\hat{\mathbf{V}}^{(1)}$ by omitting its first and last row respectively.

Step 6: compute an estimate of $\bar{\mathbf{Z}}$ by solving the above set of equations in the TLS sense.

Step 7: compute the eigenvalues of $\bar{\mathbf{Z}}$. Once $\bar{\mathbf{Z}}$ is estimated, its eigenvalues λ_k yield an estimate of the decimated signal poles z_k^D :

$$\lambda_k = \hat{z}_k^D = \exp\{(-\hat{\alpha}_k + 2\pi j\hat{\nu}_k)D\Delta t\}$$

Step 8: Finally using the estimates \hat{z}_k , $k = 1, \dots, K$, and the nondecimated signal sample x_n , $n = 0, \dots, N - 1$ compute the LS solution $\hat{c}_k = \hat{a}_k \exp\{j\hat{\varphi}_k\}$ of

$$\begin{pmatrix} 1 & \cdots & 1 \\ \hat{z}_1^1 & \cdots & \hat{z}_K^1 \\ \hat{z}_1^2 & \cdots & \hat{z}_K^2 \\ \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \cdots & \hat{z}_K^{N-1} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix}$$

7.5 Results

7.5.1 Two-peak, undamped signal

This example is taken from [50]. The signal consists of two undamped complex exponentials ($K = 2$) contaminated by complex circularly symmetric WGN. The true parameters are: $\nu_1 = 0.2$ Hz, $\nu_2 = 0.205$ Hz, $a_1 = a_2 = 1$, $\alpha_1 = \alpha_2 = 0$, $\varphi_1 = \varphi_2 = 0$, $\Delta t = 0.1$ and $n = 1, \dots, 1000$. The normalized frequencies are 0.02 and 0.0205, respectively, which allows for a decimation factor $D = 10$ without risk of aliasing. A Monte-Carlo simulation consisting of 100 independent runs is carried out. The results shown in Figure 7.3, are formulated in terms of a relative measure of the RMSE with respect to the noise standard deviation. This measure is calculated as follows:

$$RRMSE(\nu_k) = \frac{1}{Nruns} \cdot \sqrt{\sum_{i=1}^{Nruns} (\hat{\nu}_{k,i} - \nu_k)^2} \cdot \frac{100}{0.5 \times |\nu_2 - \nu_1|} \quad (7.23)$$

where $\hat{\nu}_{k,i}$ is the estimate of ν_k at run i and $Nruns$ is the number of trials. For instance, $RRMSE(\nu_1) < 100\%$ means that the frequency ν_1 is on the average in the interval $[\nu_1 - |\nu_2 - \nu_1|/2, \nu_1 + |\nu_2 - \nu_1|/2]$. Note that this value does not mean that the estimates are not consistent since the RRMSE is an average value. The measure is only a way to compare the accuracy of the results to $|\nu_2 - \nu_1|/2$; the value 100% can be considered as a conservative threshold. For the matrix approach, the parameters are computed from the dominant column subspace $\mathbf{U} \in \mathbb{C}^{50 \times 2}$ of the matrix $\mathbf{H} \in \mathbb{C}^{50 \times 500}$ (see equation (2.24)). Concerning the tensor approach, we performed a best rank-(2, 2, 2) and a best rank-(2, 2, 1)

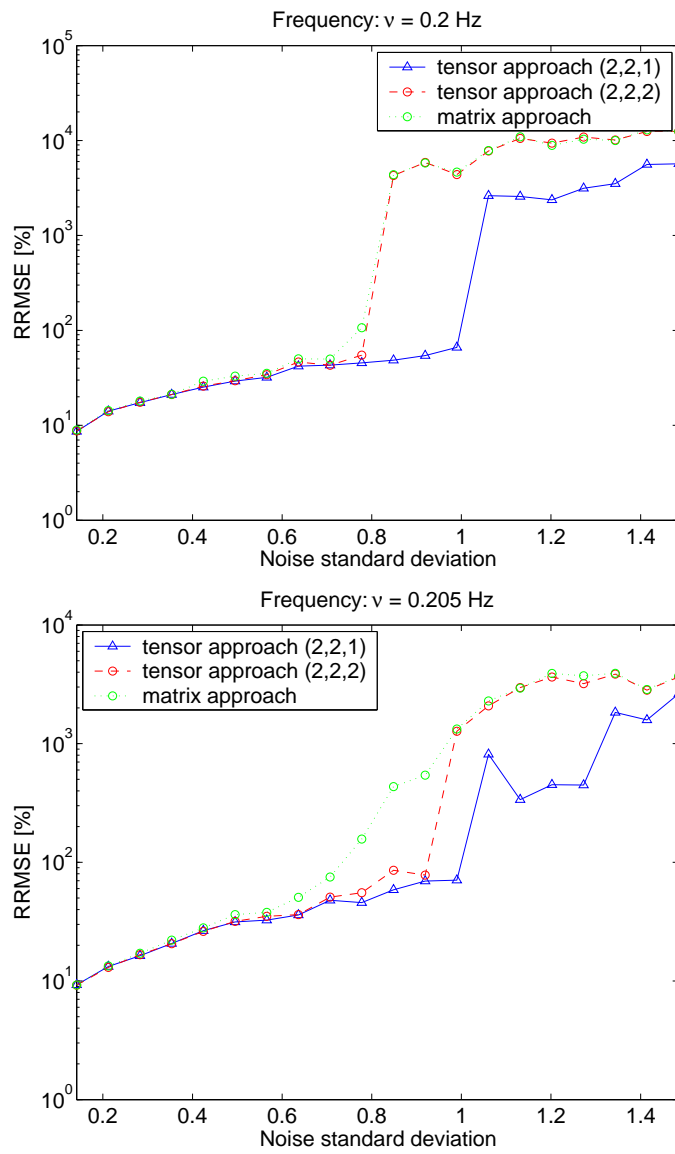


Figure 7.3: Comparison of the tensor and the matrix approach for the two-peak, undamped example, in decimative pole estimation.

approximation of the third order tensor $\mathcal{H} \in \mathbb{C}^{50 \times 50 \times 10}$. Then we used the 1-mode dominant subspace $\widehat{\mathbf{V}}^{(1)} \in \mathbb{C}^{50 \times 2}$ to estimate the parameters. We used square submatrices which leads to the best performance of HTLSDstack [50]. From Figure 7.3 it is clear that for high SNR the tensor approach and the matrix approach perform nearly the same although the tensor approach is slightly more accurate. However, the tensor approach is more robust when the SNR decreases. Moreover Figure 7.3 shows that the best rank-(2, 2, 2) approximation is less reliable than the best rank-(2, 2, 1) approximation. The reason is that mode-3 subspace is ill-conditioned.

7.5.2 Two-peak, damped signal

We consider two closely spaced peaks whose parameters are: $\nu_1 = 0.2$, $\nu_2 = 0.22$, $a_1 = a_2 = 1$, $\alpha_1 = 0.01$, $\alpha_2 = 0.02$, $\varphi_1 = \varphi_2 = 0$, $\Delta t = 0.04$ and $n = 0, \dots, 624$. The decimation factor is $D = 25$. The Monte-Carlo simulation consists of 100 trials. In this example we used the best rank-2 approximation of the (13×325) matrix \mathbf{H} and the best rank-(2, 2, 1) approximation of the $(13 \times 13 \times 25)$ -tensor \mathcal{H} . The results, shown in Figure 7.4, show that the tensor approach is clearly more accurate than the matrix approach when the noise standard deviation higher is than 0.1. Below this value, the improvement given by the tensor approach is not significant.

7.6 Conclusions

In this chapter we have considered harmonic analysis of a signal whose poles are very close. When the signal can be downsampled it is possible to use so-called decimative methods. Currently, an effective matrix-based technique is the HTLSDstack algorithm. Starting from this approach, we have shown that we can store the data in a third-order tensor instead of a matrix. The tensor formalism allows one to take into account a structure in the third mode of the tensor that is not exploited in the matrix-based algorithms.

We have derived a higher-order version of HTLSDstack and we have shown that this tensor technique is more robust against noise than its matrix counterpart. To some extent, this is due to the fact that the tensor framework offers more versatility in terms of the choice of dimensions of the approximation. This allows one to take into account the ill-conditioning in the third mode of the tensor.

The price has to be paid in terms of heavier computations. HTLSDstack only requires the SVD of the stacked data matrix \mathbf{H} . On the other hand, in the HOOI algorithm each iteration step implies 3 SVDs (of smaller matrices). In

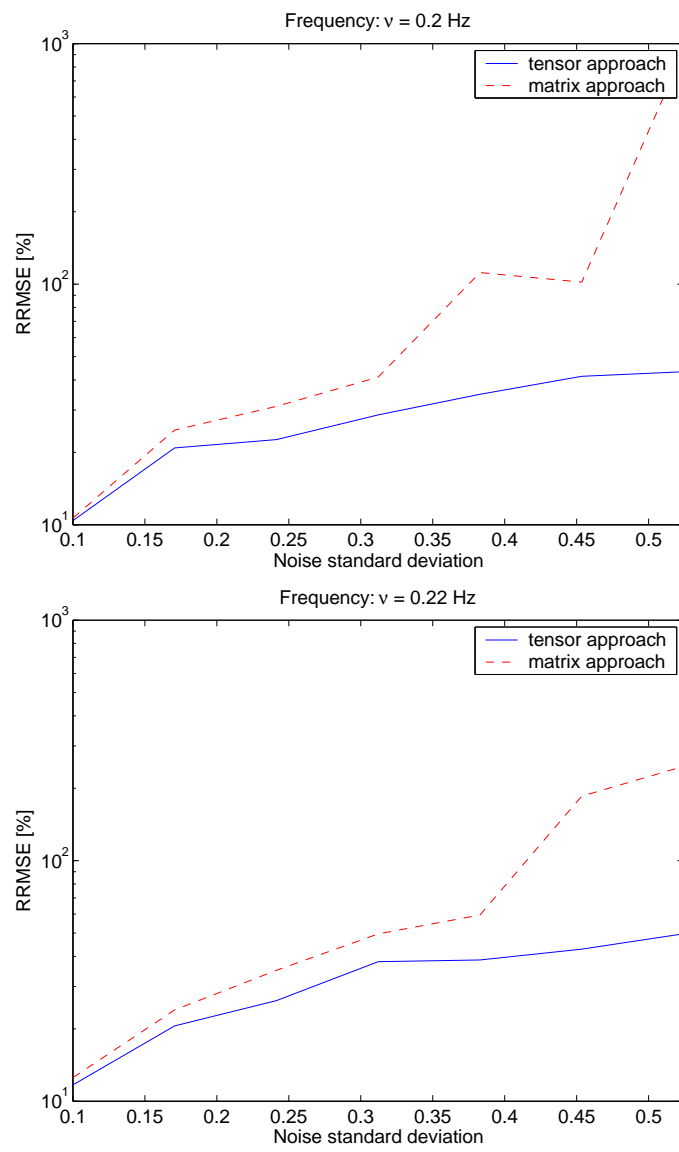


Figure 7.4: Comparison of the tensor and the matrix approach for the two-peak, damped example, in decimative pole estimation.

our simulations, this ALS procedure typically required between 3 and 15 loops to converge, depending on the level of noise.

As a higher-order extension of HTLSDstack, the HO-HTLSDstack method can be applied in the case where signals have closely spaced sinusoids. Typically this kind of signals can be found in MRS [50, 49]. However as long as the possibility is given to increase the sampling rate when signals are acquired, the HO-HTLSDstack can be used. Typical applications, for which the technique may be useful include audio processing, material health monitoring, EEG monitoring, seismology.

Chapter 8

Conclusions and Further Research

In this last chapter, findings and conclusions of the presented work are summarized. As research is a never ending story, some possible directions for further research are suggested. We proceed on a topic-by-topic basis.

8.1 Conclusions

In this thesis we developed a consistent set of signal processing tools for exponential data modeling. We only considered subspace-based methods. First we derived a novel algorithm for model order detection based on the shift-invariance property of the signal subspace. The second contribution tackles the important problem of extracting common information from multi-channel signals. Then three higher-order counterparts of matrix-based methods have been derived for the estimation of the model parameters.

Model order detection

Model order selection is a necessary step in exponential data fitting. We have developed a new subspace-based method called SAMOS for the model order detection. This method does not require the setting of any penalization term or threshold. Moreover it is not assumed that noise is gaussian. Like ESTER [72], this method is based on the shift-invariance property of the signal subspace. To our knowledge, SAMOS and ESTER are the only methods that exploit such

a property in order to provide an estimate of the model order. We have seen in simulations involving synthetic signals, that the detection rate is much better than that of ESTER. The latter was already outperforming existing ITC-based approaches.

Common pole estimation

We have studied a novel subspace-based method, called MUSCLE, for common pole estimation in the multi-channel case. This technique has been proven to be useful in EEG monitoring and material health monitoring, but is not restricted to these fields. The method is easy to implement, very robust and can be applied to an arbitrary number of channels. Moreover, to our knowledge, no competing method exist for a number of channels higher than two. Accurate estimation of the model order for each channel is sometimes difficult. However, we have shown, in the new method, that a slight overestimation of the individual model orders does not affect the accuracy of the common pole estimation. In the two-channel case, it is possible to compare MUSCLE to the HTLS-SEP technique. Monte Carlo simulations demonstrated that the new method was much more robust and yet less complex.

Exponential data fitting using multilinear algebra

The separation of complex exponential signals can be addressed in a multi-way framework. We have mapped the signals to third-order tensors and extracted the signal subspace by means of a higher-order dimensionality reduction algorithm. We could estimate the model parameters in a more accurate way than with the conventional matrix approach.

1) The single-channel case

In the single-channel case, we stacked the signal samples in a third-order tensor having a full Hankel structure. The theoretical decomposition of this tensor showed a higher-order Vandermonde structure. Making use of the shift invariance property of the subspaces we could extract the model parameters via TLS. Our goal was to show that (i) the single-channel problem can be formulated in a multilinear framework, and that (ii) for certain dimensions, the tensor technique may consistently yield better results than the matrix algorithm. Therefore, in its present form, the result is rather of theoretical importance.

2) The multi-channel case

In the multi-channel case, defining the optimal dimensions was not a problem. It was already proven that in the matrix case, stacking square Hankel matrices yields to very good results. In the tensor-based method, the Hankel matrices were simply stacked in a third-order tensor. The performance of the tensor approach in the multi-channel case is always better than the matrix-based

approach.

3) *The decimative case*

We have studied harmonic analysis in the case of a signal whose poles are very close. When the signal can be downsampled it is possible to use so-called decimative methods. Decimating a signal amounts to artificially generate several channels. Currently, an effective matrix-based technique is the HTLSDstack algorithm.

In this context it is possible to use the same method as for the multi-channel case. With the tensor formalism, we could take into account the structure that was not exploited by the HTLSDstack algorithm. Moreover, we have used the rank properties of tensors in order to circumvent the problem of ill-conditioning occurring in the mode-3 subspace. We have derived a higher-order version of HTLSDstack and we have proven that this tensor technique is more robust against noise than its matrix counterpart.

8.2 Further research.

Model order detection

One of the key-steps in SAMOS is the use of the noise-related singular values of a stacked matrix in order to determine the model order. These singular values should be zero. However in the presence of noise they are not. So far our detection criterion simply consists of the sum of these singular values. However, this may not be the best criterion. A following step in this research would be to investigate whether exploiting knowledge of the noise distribution leads to an even better detection criterion.

Another issue is the computational complexity of this algorithm. Although it is quite simple to implement, it requires the computation of SVDs of matrices that incrementally become bigger and bigger. Updating the SVD as soon as a new column is added, would be computationally less expensive.

Common pole estimation

MUSCLE benefits from a slight overestimation of the true model order in each channel, especially at high levels of noise. It would be interesting to check whether there exists an overestimation factor yielding optimal results. Another idea is to combine SAMOS and MUSCLE in order to have an algorithm able to automatically determine the number of common poles.

Exponential data fitting using multilinear algebra

In our study of exponential data fitting using multilinear algebra, a key-step is the dimensionality reduction of the data tensor. This reduction was achieved

using the HOOI algorithm, which is of the ALS type and can be computationally expensive. Further research is necessary to decrease the computational cost of the rank reduction algorithm. A Rayleigh Quotient Iteration (RQI) method for tensor n -mode rank reduction has already been derived [19]. The use of this algorithm, or other numerically more sophisticated procedures, would reduce the computation time.

For the single-channel case, we are currently investigating whether the optimal tensor dimensionality may be predicted by first-order perturbation analysis [20]. Like in the decimative case, it may be advantageous to consider unsymmetric tensor approximations, where one of the mode- n ranks is strictly smaller than the number of exponentials. If this is the case, then the choice of the optimal dimensionality would be less crucial.

Appendix I

On-line detection method for transient waves applied to continuous health monitoring of carbon fiber reinforced polymer composites with embedded optical fibers

This appendix describes a fast algorithm for the detection of transient signals. This method has been developed in the material health monitoring framework. However it can be applied in any other field where transient signals, such as EDS signals, have to be rapidly detected.

I.1 Introduction and Context

The long term behaviour of composite materials, subject to environmental, thermal and/or mechanical loading is still not very well known. This often leads industry to set high safety margins or to over-design. The advantage of composite materials is to be found in their low weight and hence high spe-

cific strength and stiffness. Therefore, it is of major interest to study their mechanical resistance to damage.

Damage processes inside composite materials are very complex, and the occurrence of damage modes is difficult to foresee. For carbon fiber reinforced polymer (CFRP) composite materials, a nondestructive test (NDT) technique of stress and strain sensing has been worked out based on the microbending concept[6]. It consists in embedding optical fibers into the material. A set of propagating waves¹ is generated whenever damage occurs in the composite material. These waves can be detected by the optical fibers as a transient signal that modulates the light intensity.

An assessment of this theoretical framework is given in [69, 70, 71, 68] where thin plates of CFRP composite materials (with embedded optical fibers) have been damaged during a tensile test up to the complete fracture. A laser light with constant power is put on one side of the embedded optical fiber and the overall light intensity is measured on the other side. The conclusions of these investigations show that some modulations of the optical signal intensity that coincide in time with damage-related waves can be detected during damage development.

Thus it is of considerable interest to identify short-duration nonstationarities in the observed signal. Indeed our problem is the detection of single or multiple narrowband transient signals buried in ambient nonstationary colored noise plus possibly some discrete components due to tonals. This is a typical problem of damage detection where "rare" events carrying information about damage have to be detected within a time-domain signal. In the case of transient signal detection, making an optimal time detector assumes at least knowledge of the model of the noise and the signal. Unfortunately, the signal parameters, such as time of occurrence, shape or frequency content as well as the noise behaviour, are not *a priori* known. In such conditions, making use of the cumulative sum (CUSUM) or the generalized likelihood ratio (GLR) algorithms[5] for detecting changes, for instance in the autoregressive (AR) coefficients and the input variance of the signal, becomes very difficult if the signal-to-noise ratio (SNR) is low. Processing the data in other domains, such as time-scale or time-frequency, can improve the detection rate. For instance, an efficient wavelet-based detector using a high order statistics(HOS)-based criterion[66] has been investigated in stationary white gaussian noise conditions. The detection of transient signals in unknown but stationary colored noise is addressed in a general way in the wavelet domain, as well as in the time-frequency domain[98, 99, 100, 101, 34]. This detector is based on Nuttall's power-law detector[55] and shows very nice properties concerning the detection of short-duration transient signals having an unknown structure. Therefore the goal of this study is to adapt this constant

¹In the case of thin plates, as used here, two main modes are generated : the flexural mode and the longitudinal mode

false alarm rate (CFAR) power-law detector to slowly nonstationary colored noise. By slowly-nonstationary is meant that the noise level in each discrete Fourier transform (DFT) bin varies slowly compared to the local nonstationarities due to the events. Making a robust detector requires that the detection criterion adapts itself to the noise level and remains stable. Z.Wang's CFAR power-law detector uses indeed an estimate of the noise level which is optimal only in the case of stationary noise and which is not valid anymore in the presence of the transient signal. Noise level estimation[94, 48, 1] has been widely and intensively investigated particularly in the framework of speech enhancement. In general, noise estimation is processed in the time-frequency domain by filtering the so-called time-trajectories[2]. That is the principle which is used in this appendix : the noise is estimated according to Wiener filter theory on a bin-by-bin basis and is recursively smoothed taking into account the previous estimates[76]. This spectral subtraction method is used here in a dual way : a pure signal enhancement and a self-adaptive estimate of the noise level in each frequency bin. The main point is that the noise estimation continues even during the "activity" period. Indeed the novelty is to derive new versions from the improved CFAR power-law detector[99] and to apply these in a successful way to a one-dimensional signal containing a high level of slowly-nonstationary additive colored noise, as shown in picture I.1.

This appendix is organized as follows. Section I.2 describes the main characteristics of a transient signal and the preprocessing step to be applied to the time-domain input signal. Section I.3 explains the principle of the power-law detector and section I.3.1 gives tools to improve the sensitivity of this detector. Eventually, section I.4 gives the formulation of three different new detectors and compares their respective sensitivity.

I.2 Signal characteristics and representation

Whenever damage occurs, a set of propagating waves is produced inside the material and on its surface. The passage of these waves over the embedded fiber locally modifies the optical and geometrical fiber properties, inducing a fluctuation of light intensity. This is what is measured as a transient signal and is called an *event*. But although the mechanical waves are exponentially damped, the optical events can take other shapes. This is due to quite complex deformation phenomena induced by the mechanical waves propagating through the optical fiber (OF). It turns out that there is a very high level of background noise and it is difficult to visually see where all the interesting transients are. Moreover the frequency content can be different for different events (see figure I.2). Transient signals could in the most general case be defined as follows. They :

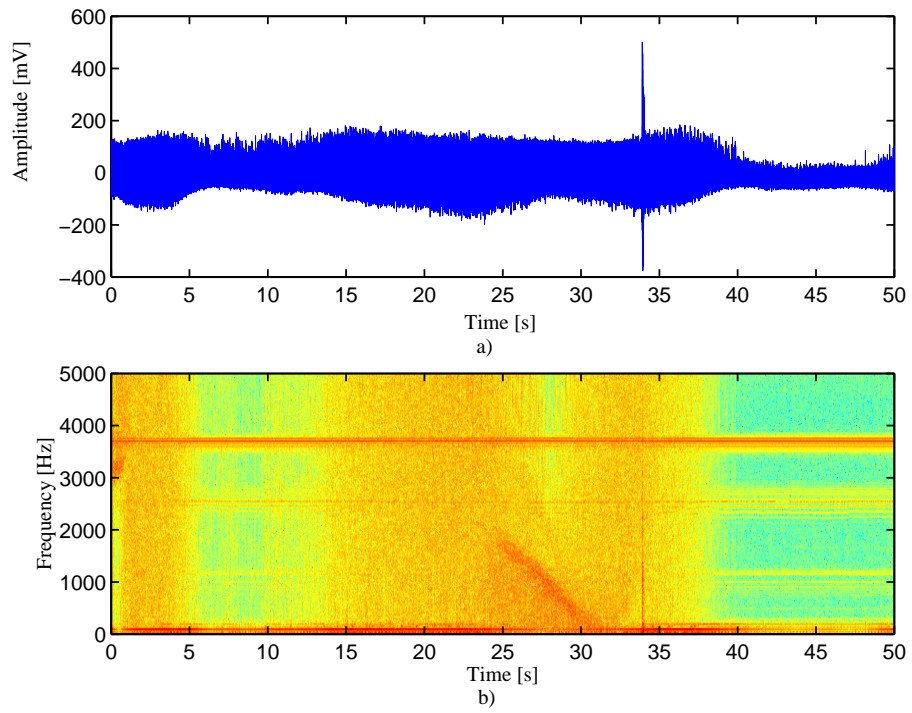


Figure I.1: The optical signal buried by a high level of non-stationary colored noise. a) time-domain representation, b) time-frequency representation.

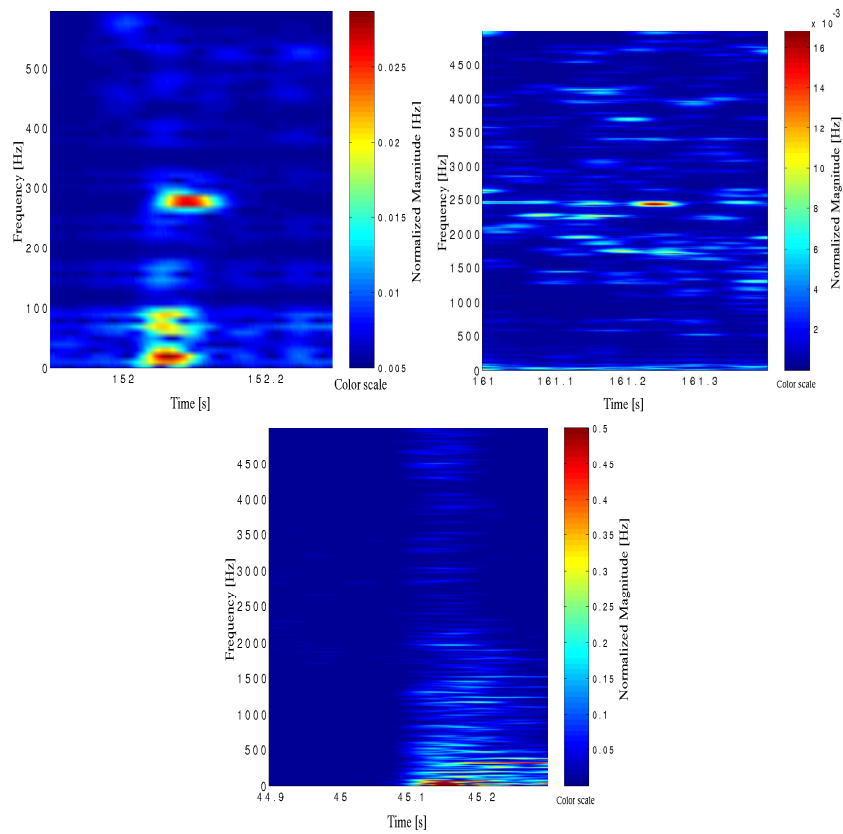


Figure I.2: Spectral content for 3 different events

- have very short duration in time compared to the observation period,
- have an effect, at the time of occurrence, on the signal properties or state. These changes occur sufficiently rapidly, so they are not confused with the slow nonstationarities that might be present in the noise.

In the real world these changes are expressed as a fast energy rise in the time-domain signal, which decreases as soon as it has reached its maximum. It is most likely exponentially decaying and the rising front may be so abrupt that a truncated pure decaying exponential (I.1) can be a convenient model[66, 99] as depicted in figure I.3.

$$s_1(t) = \begin{cases} Ae^{-\alpha(t-t_0)} & t \geq t_0 \\ 0 & \text{elsewhere} \end{cases} \quad (\text{I.1})$$

α is the so-called damping factor, and $A_M = A$ is the (maximum) amplitude

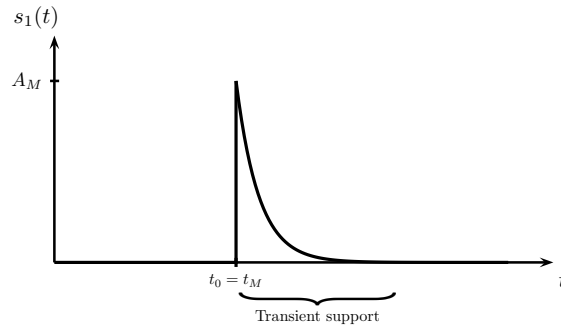


Figure I.3: The damped exponential model

at $t = t_0$, the beginning of the transient. As an alternative, the sum of two damped exponentials can also be considered :

$$s_2(t) = A \frac{e^{-\alpha t} - e^{-\beta t}}{\beta - \alpha}, \quad t \geq 0 \quad (\text{I.2})$$

The shape depicted in figure I.4 is more realistic in the case of a sufficiently high sampling rate of the observed signal.

There are several reasons for which we processed the signals in the frequency domain :

- signals can be processed in real-time,

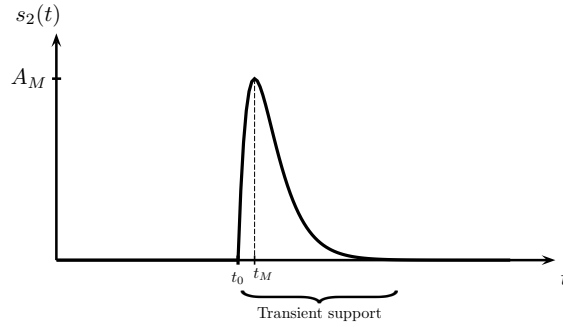


Figure I.4: The sum of two damped exponentials, $t_M = (\ln \beta - \ln \alpha) / (\beta - \alpha)$.

- interpretation in this transform domain is quite simple,
- several smooth and robust noise estimators exist,
- abrupt changes in the time-domain signal can also be better emphasized in the frequency domain,
- power-law detectors work remarkably well when applied to frequency components.

In what follows, we consider a bandlimited, sampled noisy optical signal $x(i)$. i denotes the sampling time index and we compute the discrete short-time Fourier transform as follows. The signal $x(i)$ is transformed into the frequency domain by applying a window function $w(i)$ to a frame of L consecutive samples of $x(i)$ and by computing the fast Fourier transform (FFT) of size L of the windowed signals. Before the next FFT computation the window is shifted by I samples. This sliding window FFT, depicted in figure I.5, results in a set of frequency-domain signals which can be represented as :

$$X(\lambda, k) = \sum_{\mu=-\frac{L}{2}}^{\frac{L}{2}} x(\lambda I + \mu) w(\mu) e^{-2j\pi \frac{k\mu}{L}} \quad (\text{I.3})$$

where λ is the subsampled time index centered on the middle of the window, $\lambda \in \mathbb{Z}$, and k is the frequency bin index, which is related to the normalized center frequency ω_k by $\omega_k = 2\pi k/L$. Every time interval $\Delta t = t_k - t_{k-1} = I/f_s$, a new set of frequency components $X(\lambda, k)$ is rowwise stored into a matrix. It is very well known that applying an apodization window different from a rectangular window limits the influence of the high frequency components on their nearest neighbours. This influence has a very small impact in the presence

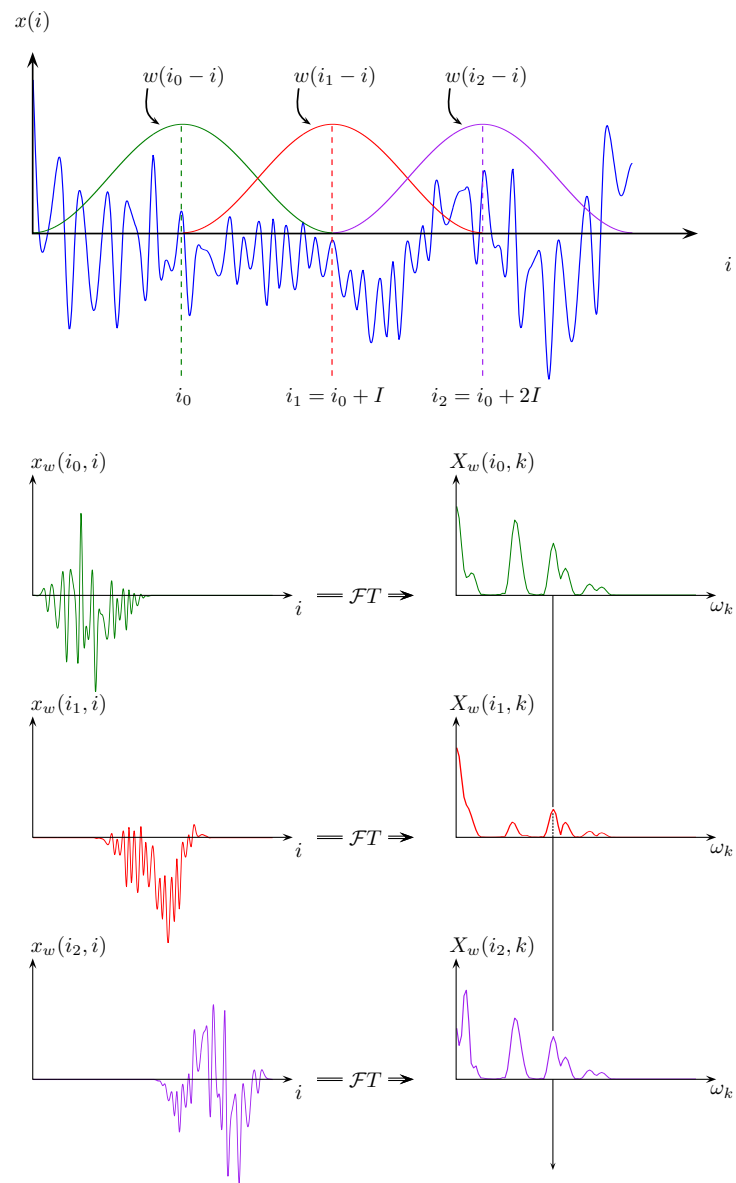


Figure I.5: Principle of the sliding window FFT

of the noise only but the independence of frequency components can be locally lost in case of a sudden rise of a frequency component. Preserving maximal independence may help the power-law detector described in section I.3 to reach optimal performances. Typically we use a Hanning window. In the sequel of the appendix a dual notation will be used for X . The time index λ is written as a subscript while the frequency bin index k within the brackets is replaced by its related normalized center frequency ω_k : $X(k, \lambda) \equiv X_\lambda(\omega_k)$.

I.3 Transient detection

First let us assume that the noise $n(i)$ is stationary white and Gaussian. Under the signal-absent hypothesis \mathcal{H}_0 , pre-processing and squaring the discrete Fourier magnitude components yields independent and identically distributed (iid) exponential random variates[64]. Under the signal-present hypothesis \mathcal{H}_1 , the transient signal occupies an arbitrary set of M frequency bins amongst a total number of K and Nuttall[55] assumes that the signal-bearing set S of size M is also exponentially distributed but with a different parameter $\beta_1 \neq \beta_0$:

$$\begin{aligned} \mathcal{H}_0 : p(X_\lambda^2) &= \prod_{k=1}^K \frac{1}{\beta_0} e^{-\frac{X_\lambda^2(\omega_k)}{\beta_0}} \\ \mathcal{H}_1 : p(X_\lambda^2) &= \prod_{k \notin S} \frac{1}{\beta_0} e^{-\frac{X_\lambda^2(\omega_k)}{\beta_0}} \cdot \prod_{k \in S} \frac{1}{\beta_1} e^{-\frac{X_\lambda^2(\omega_k)}{\beta_1}} \end{aligned} \quad (\text{I.4})$$

According to this assumption a power-law detection criterion has been developed :

$$T(X_\lambda^2) = \sum_{k=1}^K (X_\lambda^2(\omega_k))^\nu \quad (\text{I.5})$$

with ν an adjustable exponent and its CFAR version is introduced by the same author :

$$T(X_\lambda^2) = \frac{\sum_{k=1}^K (X_\lambda^2(\omega_k))^\nu}{\left(\sum_{k=1}^K X_\lambda^2(\omega_k) \right)^\nu} \quad (\text{I.6})$$

Although by construction (I.6) is not affected by a scale factor it performs worse than (I.5) even if only stationary white Gaussian (WG) noise is used. Since the detector (I.5) works for stationary WG noise with known noise power, it has been extended by Wang[99] to stationary colored noise with unknown power :

$$T(X_\lambda^2) = \sum_{k=1}^K \left(\frac{X_\lambda^2(\omega_k)}{\overline{X_{\lambda-1}^2(\omega_k)}} \right)^\nu \quad (\text{I.7})$$

with

$$\overline{X_{\lambda-1}^2(\omega_k)} = \frac{1}{\lambda-1} \sum_{l=1}^{\lambda-1} X_l^2(\omega_k) \quad (\text{I.8})$$

$\overline{X_{\lambda-1}^2(\omega_k)}$ represents a smoothed estimation of the mean value of X^2 in the frequency bin ω_k , up to the time index $\lambda - 1$. If we assume that at time index $\lambda - 1$, $X^2(\omega_k)$ is still under hypothesis \mathcal{H}_0 , i.e. exponentially distributed with parameter $1/\beta_0$ then the log-likelihood function yields an estimator of the mean $\beta_0 = \overline{X_{\lambda-1}^2(\omega_k)}$ defined in (I.8). Thus since the variance of such a variate is β_0^2 , then $X^2(\omega_k)/\overline{X_{\lambda-1}^2(\omega_k)}$ is approximately exponentially distributed with parameter 1. It is clear that (I.7) is hence a self-normalizing version of (I.5). This is very important because preprocessing colored noise requires a self-normalization since the variance and the mean strongly depend on the location of the frequency bin : the lower the frequency, the higher the mean and the variance of $X^2(\omega_k)$. In our signal, no *a priori* information is known about the frequency content of transients. It turns out that $1.5 < \nu < 2$ is a good choice when information about the number of active frequency bins M is completely unknown. But as a matter of fact, the sensitivity of $T(X)$ can be improved. Actually, $X_{\lambda}^2(\omega_k)$ oscillates around $\overline{X_{\lambda-1}^2(\omega_k)}$ so the ratio has a mean equal to 1. In the presence of noise-only it can be assumed that $\overline{X_{\lambda-1}^2(\omega_k)} \approx \overline{X_{\lambda}^2(\omega_k)}$. When a transient is present at time index λ then the ratio, and hence $T(X)$, increases abruptly because the estimate $\overline{X_{\lambda-1}^2(\omega_k)}$ is very smooth while the variance and mean of $X_{\lambda}^2(\omega_k)$ increase abruptly. It is now straightforward to see that the lower $\overline{X_{\lambda-1}^2(\omega_k)}$ the higher the sensitivity of T to the presence of an additional signal :

$$\mathcal{H}_1, \lambda \implies T(X_{\lambda}^2) = \sum_{k=1}^K \left(\frac{N_{\lambda}^2(\omega_k) + \Delta(\omega_k)}{\overline{X_{\lambda-1}^2(\omega_k)}} \right) \quad (\text{I.9})$$

where $\Delta(\omega_k)$ is a supplementary amount of energy in the frequency bin k due to the sudden presence of a transient. $\Delta(\omega_k)/\overline{X_{\lambda-1}^2(\omega_k)}$ is a smoothed SNR which influences directly the sensitivity of $T(X)$. This result is normal but it must be emphasized here that a low residual signal could lead to a higher variance of $T(X)$ if X is not correctly smoothed and hence that could lead to a higher false detection rate. In the next section it is shown that under some assumptions the noise can be recursively estimated and updated frame-to-frame. A new signal with a higher SNR can be reconstructed after spectral subtraction.

I.3.1 SNR improvement

The goal is to emphasize the abrupt changes that can possibly occur in a given frequency bin by estimating continuously the noise level within it. Practically, the more the bursts are emphasized, the more robust the final detection criterion. In this section we have to make several assumptions. A quick observation of the signal shows that most of the time the signal consists of pure noise. That means the largest amount of information we get is from the noise itself. In other words the probability that a data segment contains only noise is much higher than the probability that this segment also contains a transient. Although this

assumption is not satisfied close to the final fracture we will consider it is true no matter what occurs close to the fracture because most of final events are high energy events and consequently easily detectable. Two possible situations are to be considered :

$$\begin{aligned}\mathcal{H}_0 : x(i) &= n(i) \\ \mathcal{H}_1 : x(i) &= s(i) + n(i)\end{aligned}\quad (\text{I.10})$$

where $x(i)$ is the processed signal, $n(i)$ the noise, and $s(i)$ the "clean" signal. \mathcal{H}_0 is referred to as the signal-absent hypothesis, and \mathcal{H}_1 is the signal-present one as already defined in section I.3. In this context, under \mathcal{H}_1 it is crucial to assume additivity of noise with the transient. In any hypothesis, \mathcal{H}_0 or \mathcal{H}_1 , we must assume the quasi-stationarity of $n(i)$ within each processed time data-segment so that it makes sense to associate with this variate a linear transformation such as the Fourier transform (ergodicity is assumed). If $n = [n(1)...n(L)]$ is a realization of a stationary process then $\mathcal{TF}\{n(i)\}^2 = N^2(\omega_k)$, where $\mathcal{TF}\{\cdot\}$ stands for the Fourier transform operator, is at least a rough approximation of the true power spectral density (PSD) of the noise $P_N(\omega_k) = \mathcal{TF}\{C_N\}$ where C_N is the true correlation function of the noise. The noise $n(i)$ and the transient $s(i)$ must be assumed uncorrelated so that we can claim the additivity of the PSD's :

$$P_X(\omega_k) = P_S(\omega_k) + P_N(\omega_k) \quad (\text{I.11})$$

The changes due to the transients present in the frequency components must be much more abrupt than the change due to nonstationarities present in the noise. That amounts to assuming the quasi-stationarity of the noise during the presence of the signal-of-interest. In the sequel a noise estimator is described which has been developed by Sovka[76] according to these assumptions.

I.3.2 Noise level estimation

In the case of uncorrelated signals n and s the linear estimator of n which minimizes the mean-square error is :

$$\hat{N}(\omega_k) = H_N(\omega_k) \cdot X(\omega_k) = \sqrt{\frac{P_N(\omega_k)}{P_S(\omega_k) + P_N(\omega_k)}} \cdot X(\omega_k) \quad (\text{I.12})$$

H_N is the so-called Wiener filter. The use of these complex power spectra is not very handy and one often uses the magnitude spectra approximation. Moreover it is only possible to have a reliable estimation of the noise using the past samples. Hence, the noise estimation at time index λ will be given by :

$$|\hat{N}_{\lambda+1}(\omega_k)| = H_{N_\lambda}(\omega_k) \cdot |X_{\lambda+1}(\omega_k)| = \left(\frac{\overline{|N_\lambda(\omega_k)|^2}}{\overline{|S_\lambda(\omega_k)|^2} + \overline{|N_\lambda(\omega_k)|^2}} \right)^{\frac{1}{2}} \cdot X_{\lambda+1}(\omega_k) \quad (\text{I.13})$$

where $|\cdot|$ denotes the modulus, $|\overline{N_{\lambda-1}(\omega_k)}|$ is a smoothed estimate of the noise spectrum, and $|\overline{S_{\lambda-1}(\omega_k)}|$ is a smoothed estimate of the transient spectrum calculated as follows :

$$|\overline{S_{\lambda-1}(\omega_k)}| = |X_{\lambda-1}(\omega_k)| - |\overline{N_{\lambda-1}(\omega_k)}| \quad (\text{I.14})$$

where

$$|\overline{N_{\lambda}(\omega_k)}| = r \cdot |\overline{N_{\lambda-1}(\omega_k)}| + (1-r) \cdot |\widehat{N}_{\lambda}(\omega_k)| \quad (\text{I.15})$$

r is the so-called exponential forgetting factor with $0 < r < 1$. This is strictly equivalent to a first order IIR filter whose unit sample response is :

$$h(\lambda) = (1-r)r^{\lambda} \cdot u(\lambda) \quad \text{with} \quad u(\lambda) = \begin{cases} 1 & \lambda \geq 0 \\ 0 & \lambda < 0 \end{cases} \quad (\text{I.16})$$

Although it is sufficient to have an estimate of the noise to implement the proposed detector, it can be useful to reconstruct the signal so that the general improvement over all frequencies can be visually appreciated. The reconstructed signal can also be further analyzed off-line with other signal processing tools.

I.3.3 Spectral subtraction estimator – signal reconstruction

According to the additive noise model and the linearity of the Fourier transform, we have :

$$\begin{aligned} x(i) &= n(i) + s(i) \\ &\quad \downarrow \\ \underbrace{|X(\omega_k)|e^{j\theta_X(\omega_k)}}_{X(\omega_k)} &= \underbrace{|N(\omega_k)|e^{j\theta_N(\omega_k)}}_{N(\omega_k)} + \underbrace{|S(\omega_k)|e^{j\theta_S(\omega_k)}}_{S(\omega_k)} \end{aligned} \quad (\text{I.17})$$

The spectral subtraction filter $H(\omega_k)$ is calculated by replacing the noise spectrum $N(\omega_k)$ which can be readily measured. The magnitude $|N(\omega_k)|$ of $N(\omega_k)$ is replaced by its estimate $|\widehat{N}(\omega_k)|$ updated during online processing according to the method depicted in section I.3.2 and the phase $e^{j\theta_N(\omega_k)}$ of $N(\omega_k)$ is replaced by the phase $e^{j\theta_X(\omega_k)}$ of $X(\omega_k)$. It results in the spectral subtraction estimator :

$$\widehat{S}(\omega_k) = H(\omega_k) \cdot |X(\omega_k)|e^{j\theta_X(\omega_k)} = [|X(\omega_k)| - |\widehat{N}(\omega_k)|]e^{j\theta_X(\omega_k)} \quad (\text{I.18})$$

which can be rewritten in term of the noise estimator H_N :

$$\widehat{S}(\omega_k) = \{|X(\omega_k)| - H_N(\omega_k) \cdot |X(\omega_k)|\}e^{j\theta_X(\omega_k)} \quad (\text{I.19})$$

and therefore

$$H(\omega_k) = 1 - H_N(\omega_k) \quad \forall \omega_k \quad (\text{I.20})$$

At this step, the denoised signal $\hat{s}(i)$ can be reconstructed using the overlap-add (OLA) method [2].

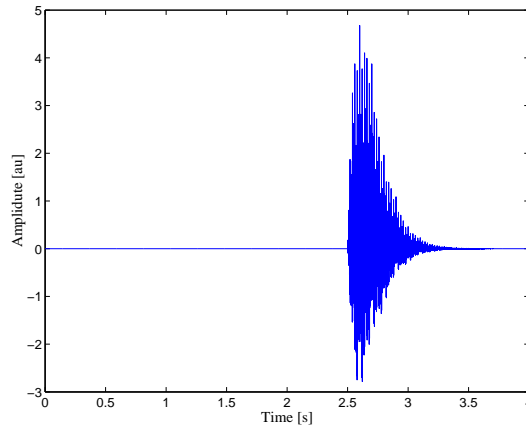


Figure I.6: Theoretical transient signal

I.4 Results and discussion

The proposed method is not really automatic : it requires a special study of its parameters for each application. The first parameter to be tuned is the FFT window length. The whole study is based on the abrupt variation of the frequency components when a transient is being processed. Embedding a transient in only one window is a good criterion for the segmentation of the time-domain signal because this choice typically leads to the most abrupt change in the spectral components. However, it is not always possible since the transient can have various lengths. It turns out that most transients with low SNR have a length lower than 300 samples while high SNR events are likely spread out over 1000 samples in length. According to this observation a window length of about 256 samples leads to a very good segmentation . But using such a small window affects the accuracy of the Fourier transform resulting into crude estimates of the frequency content of the signal which leads to an unavoidable decrease of the sensitivity of the power-law detector. Moreover a too short segmentation of the signal yields too noisy frequency components which are very difficult to smooth correctly. A last aspect concerns the nonstationarities of the noise : in fact the length of the Fourier window should be negligible compared to the length of a nonstationarity. The best compromise that has been found is a window length of 512 samples and an overlap of 256 samples. It is not necessary to have a higher overlap unless more precision is needed in the localization of the wanted signals.

Of importance is the exponential forgetting factor r . A simulated transient signal (*c.f.* figure I.6) generated according to the model (I.2) has been added to a high level of colored Gaussian noise as depicted in figure I.7. In or-

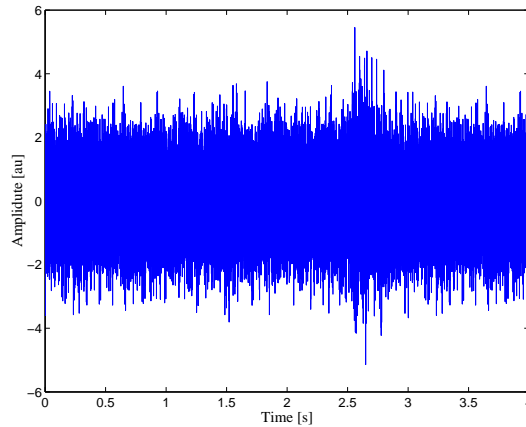


Figure I.7: A high level of colored noise is added to the signal of figure (I.6). $\text{SNR} \approx -7\text{dB}$.

der to illustrate the influence of the factor r , the denoising procedure depicted in section I.3.1 has been performed using a set of values of r , starting from $r = 0.6$, with a step of $\Delta r = 0.05$. Of course the last value is not 1, because as demonstrated in (I.15), the algorithm is no longer adaptive for $r = 1$. Figure I.8 shows that $0.9 < r < 1$ is a very reasonable choice yielding a small signal distortion but a high noise reduction. In figure I.9 which displays the evolution of a frequency component over time, it is shown that a value of r around 0.9 yields a bad smoothing of the noise while a value close to 1 leads to a delay between the smoothed estimate and the true behavior of the noise. $r = 0.95$ has been found to be the best compromise.

Eventually, the last parameter to be set is the exponent ν on which the detector relies. Actually, given that neither the transient signals nor the behaviour of the noise are *a priori* known it is impossible at this moment to make receiver operating characteristic (ROC) curves. Nevertheless, ν has been set to 1.7 according to the performance analysis made by Wang [99] on transient signals with unknown frequencies in stationary colored gaussian noise. In sections I.3.2 and I.3.3 several useful estimates have been derived from the spectrum $X(\omega_k)$:

- an estimate of the noise modulus $|\widehat{N}(\omega_k)|$,
- a smoothed estimate of the noise modulus $|\overline{N(\omega_k)}|$,
- an estimate of the denoised signal $\widehat{N}(\omega_k)$.

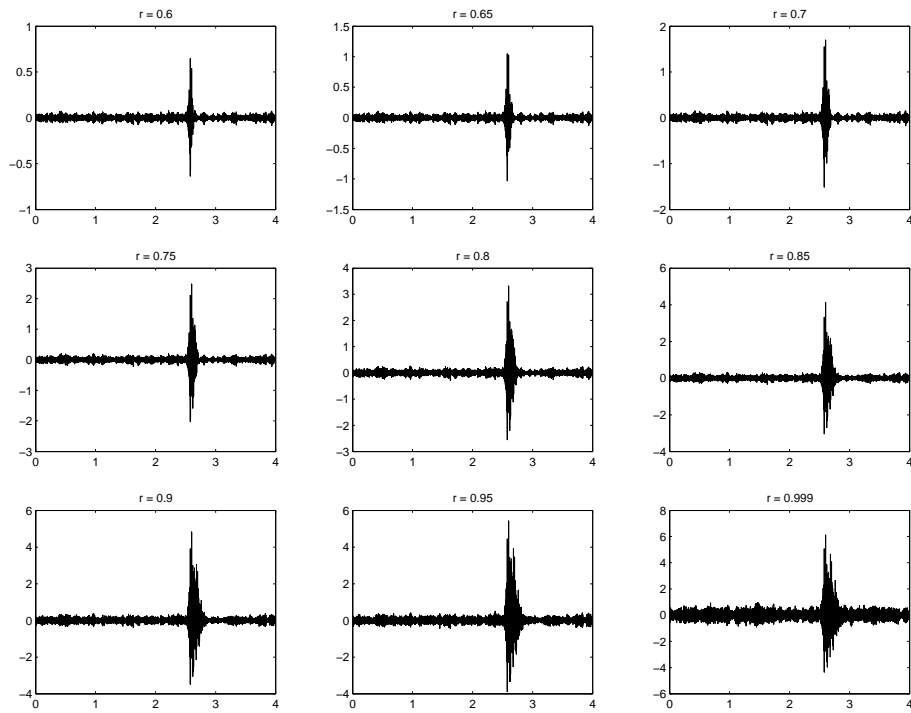


Figure I.8: Influence of the forgetting factor r on the distortion of the reconstructed signal $\hat{s}(i)$

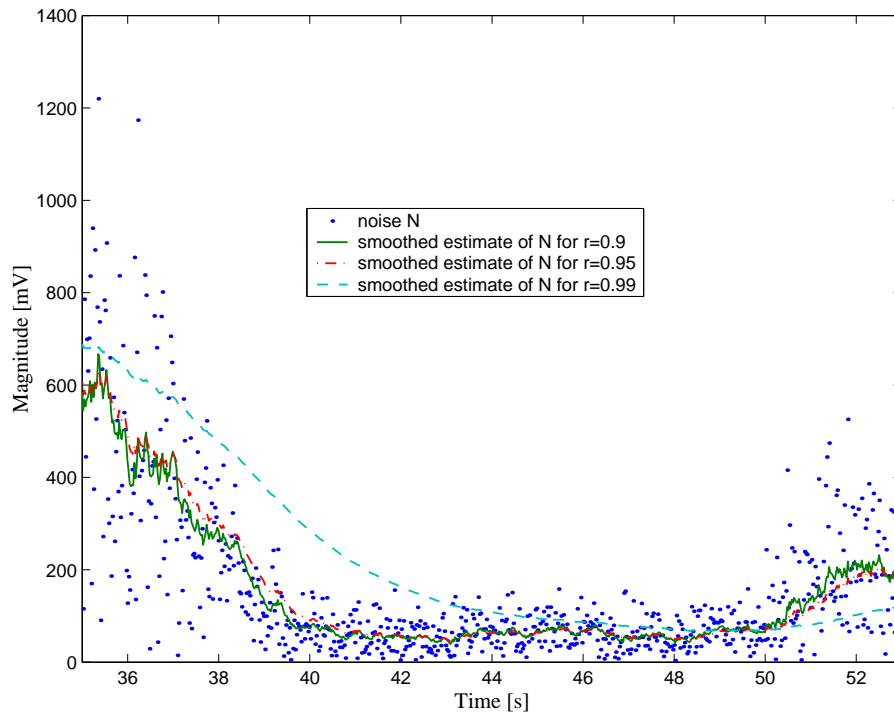


Figure I.9: Influence of r on the smoothed estimate of one frequency component.

Hence new power-law detectors have been derived from the improved power-law detector (I.7) :

$$\begin{aligned}
T_1(X_\lambda) &= \sum_{k=1}^K \left(\frac{|X_\lambda(\omega_k)|^2}{|N_\lambda(\omega_k)|^2} \right)^\nu \\
T_2(X_\lambda) &= \sum_{k=1}^K \left(\frac{|X_\lambda(\omega_k)|^2}{|N_{\lambda-1}(\omega_k)|^2} \right)^\nu \\
T_3(X_\lambda) &= \sum_{k=1}^K \left(\frac{|\widehat{S}_\lambda(\omega_k)|^2}{|\widehat{S}_{\lambda-1}(\omega_k)|^2} \right)^\nu
\end{aligned} \tag{I.21}$$

and compared to :

$$\begin{aligned}
T_4(X_\lambda) &= \sum_{k=1}^K \left(\frac{|X_\lambda(\omega_k)|^2}{|X_{\lambda-1}(\omega_k)|} \right)^\nu \\
E_1(X_\lambda) &= \sum_{k=1}^K |X_\lambda(\omega_k)|^2 \\
E_2(X_\lambda) &= \sum_{k=1}^K |\widehat{S}_\lambda(\omega_k)|^2
\end{aligned} \tag{I.22}$$

where T_4 is the improved power-law detector [99], E_2 is the energy detector computed from the signal enhanced by the denoising method described in section I.3.1, and E_1 is the energy detector computed from the original signal. What characterizes the sensitivity of the detection criterion is the importance of its deviation in the presence of a transient signal compared to its mean value during the noise-only period. But at the same time, what characterizes the robustness of a detector is its low standard deviation (std) during noise periods. The first requirement is to be evaluated in relation with the misdetection rate while the second is closely related to the false alarms rate and these two points are contradictory. Figure I.10 illustrates the respective efficiency of the detectors T_1 , T_2 , T_3 , Wang's improved detector T_4 , and the energy detector before denoising E_1 and after denoising E_2 . They have been plotted in decibel (dB) scale, so they can be displayed on the same graph. It is clear that some detectors have a quite high standard deviation but show a big change when a transient is present while others have a smaller global standard deviation but with a less emphasized change. Therefore a good sensitivity criterion is the ratio of the maximum deviation during the detection to the standard deviation of the detector in the presence of noise only. From the simulation results of figure I.10, the mean, std and the ratio between the maximum value of the deviation under \mathcal{H}_1 to the standard deviation under \mathcal{H}_0 have been computed and results have been gathered in table I.1. We see clearly that T_1 and T_2 have the best sensitivity while their maximum deviation is not that high compared to T_4 or E_1 . This is because they have a very low variance compared to other detectors. Finally, real-world signals containing nonstationary colored noise have

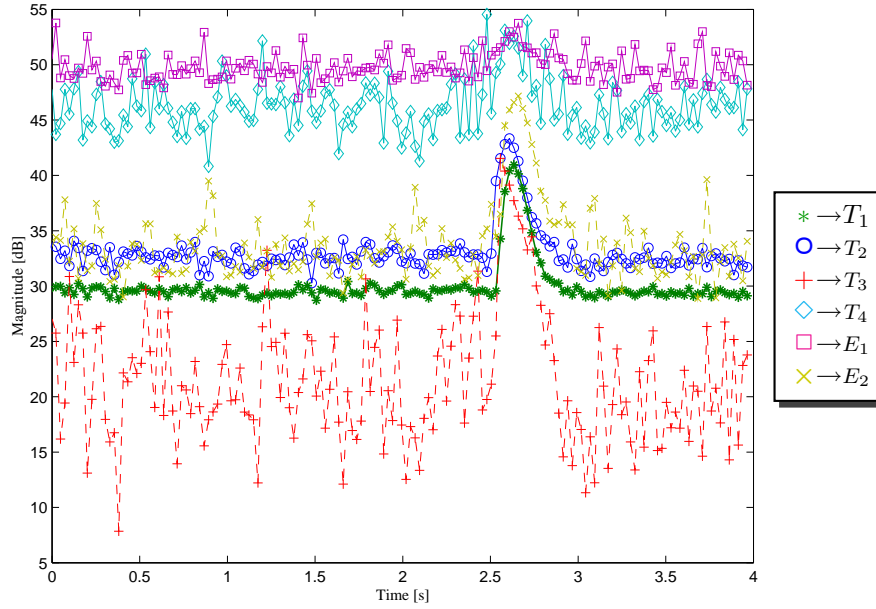


Figure I.10: Five detector criteria have been applied to the simulated signal of figure I.7

	Mean	Standard deviation ²	Maximum value ³	$(\frac{maxvalue}{std})_{dB}$
T_1	900	75	1.25×10^4	51.16
T_2	1.87×10^3	333	2.16×10^4	41.68
T_3	250	329	1.58×10^4	38.71
T_4	4.8×10^4	3.6×10^4	2.85×10^5	20.8
E_1	9.5×10^4	3×10^4	2.38×10^5	20.7
E_2	2.3×10^3	1.4×10^3	5.21×10^4	35.95

Table I.1: Detector statistics

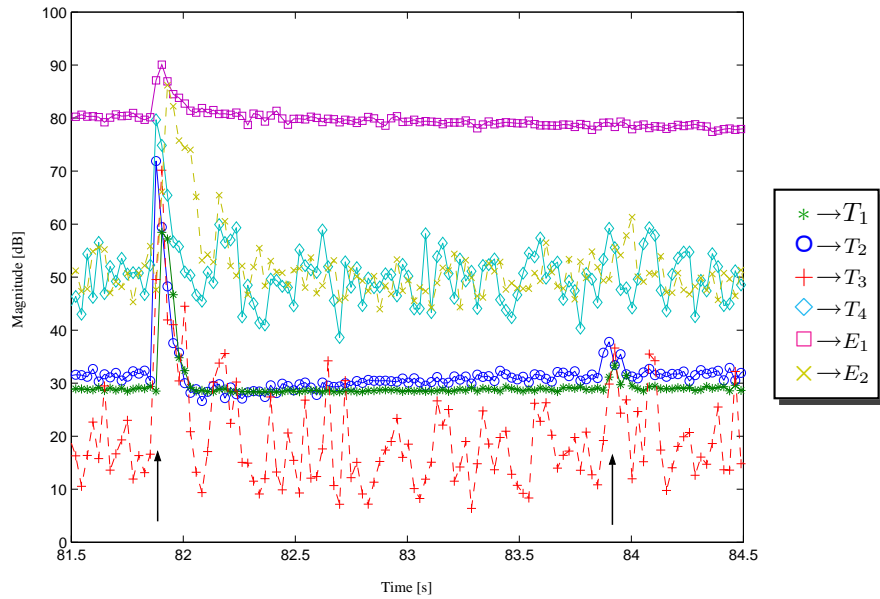


Figure I.11: Detectors output during part of a tensile test. The arrows indicate the occurrence of transients.

also been processed. For instance, figure I.11 shows these detectors applied to a digitized optical signal acquired during a tensile test which consisted in damaging a CFRP composite plate with embedded optical fibers. The sampling rate was set to 10^4 Hz, $N = 512$ samples, $I = 256$ samples, $r = 0.95$, $\nu = 1.7$. It is clearly shown that T_1 and T_2 have the best sensitivity for high and low SNR events while T_3 , T_4 and E_2 perform quite poorly in low SNR conditions. However there is a slight difference between T_1 and T_2 about the detection time accuracy. As a matter of fact, T_2 seems to detect a transient faster than T_1 . However, T_2 remains more "noisy" than T_1 .

I.5 Conclusions

In this appendix a noise estimation method based on spectral subtraction and Wiener filtering has been combined with an improved power-law detector to

derive two new sensitive and robust detector criteria. The first, T_1 , is based on an estimate of the current noise, which confers to the latter a low variance. The second, T_2 , relies on a very smooth estimate of the mean of the current noise. They have been applied to digitized optical signals acquired during tensile tests and compared with three other detectors including the original power-law detector T_4 (I.7). As first results, these two detectors show a real sensitivity while they have a very small variance, which enable them to threshold the data more easily than others. Future work includes the setting and the optimization of an adaptive threshold to overcome the possibility of important and permanent changes in the background noise level which could induce a change in the detector variance.

Bibliography

- [1] Levent Arslan, Alan McCree, and Vishu Viswanathan. New methods for adaptive noise suppression. In *International Conference on Acoustics Speech and Signal Processing (ICASSP-95)*, Detroit, USA, May 1995.
- [2] Carlos Avendanõ. *Temporal Processing of Speech in a Time Feature Space*. PhD thesis, Oregon Graduate Institute of Science & Technology, 1997.
- [3] Roland Badeau, Rémy Boyer, and Bertrand David. Eds parametric modeling and tracking of audio signals. In *Proceedings of the 5th International conference on Digital Audio Effects (DAFx-02)*, pages 139–144, Hamburg, Germany, September 26–28 2002. Online article: <http://www.dfax.de/>.
- [4] H. Barkhuijsen, R. de Beer, and D. van Ormondt. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *Journal of Magnetic Resonance*, 73:553–557, 1987.
- [5] M. Basseville and Igor V. Nikiforov. *Detection of abrupt changes*. NJ: Prentice-Hall, 1993.
- [6] J.W. Berthold. Historical review of microbend fiber-optic sensors. *J Lightwave Technology*, 13(7):1193–1199, 1995.
- [7] D.V. Bhaskar Rao and K.S. Arun. Proceeeding ieee. In *Model-based processing of signals: a state-space approach*, volume 80, 1992.
- [8] Rémy Boyer and Karim Abed-Meraim. Efficient parametric modeling for audio transients. In *Proceedings of the 5th International conference on Digital Audio Effects (DAFx-02)*, pages 1–4, Hamburg, Germany, September 26–28 2002. Online article: <http://www.dfax.de/>.
- [9] Rémy Boyer and Karim Abed-Meraim. Audio modeling based on delayed sinusoids. *IEEE Transactions on Speech and Audio Processing*, 2004.

- [10] Rasmus Bro. Tutorial & applications. In *2nd Internet Conf. in Chemometrics (INCINC96)*, *Chemom. Intell. Lab. Syst.*, volume 38, pages 149–171, 1997. (special issue).
- [11] H. Chen, S. Van Huffel, A.J.W. Van den Boom, and Van den Bosch P.P.J. Extended htl methods for parameter estimation of multiple data sets modeled as sums of exponentials. In *13th International Conference on Digital Signal Processing (DSP'97)*, pages 1035–1038, Santorini, Greece, July 1997.
- [12] Hua Chen. *Subspace-based parameter estimation of exponentially damped sinusoids with application to nuclear magnetic resonance spectroscopy data*. PhD thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, 1996.
- [13] Hua Chen, Sabine Van Huffel, and Joos Vandewalle. Bandpass prefiltering for exponential data fitting with known frequency region of interest. *Signal Processing*, 48:135–154, 1996.
- [14] Pierre Comon. Independent component analysis—a new concept? *Signal Processing*, 36(3):287–314, 1994.
- [15] Pierre Comon. Tensor decompositions. In I.K. Proudler (Eds.) J.G. McWhirter, editor, *Mathematics in Signal Processing V*. Clarendon Press, Oxford, UK, 2002.
- [16] Wim De Clercq. *Advanced Preprocessing Techniques and Nonlinear Signal Analysis Applied To Scalp Electroencephalograms For The Prediction Of Epileptic Seizures*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, May 2005.
- [17] Wim De Clercq, Jean-Michel Papy, Wim Van Paesschen, and Sabine Van Huffel. Modelling common dynamics in multichannel signals with applications to artifact and background removal in eeg recordings. *IEEE Transactions on Biomedical Engineering*, 2005. Accepted for publication.
- [18] Wim De Clercq, Bart Vanrumste, Jean-Michel Papy, Anneleen Vergult, Wim Van Paesschen, and Sabine Van Huffel. Removing artifacts and background activity in multichannel electroencephalograms by enhancing common activity. In *Proceedings of the 2005 IEEE Engineering in Medicine and biology 27th Annual Conference*, Shanghai, China, September 1-4 2005.
- [19] Lieven De Lathauwer, Luc Hoegaerts, and Joos Vandewalle. A grassmann-rayleigh quotient iteration for dimensionality reduction in ica. In *Fifth International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, pages 335–342, Granada, Spain, September 2004.

- [20] Lieven DE LATHAUWER. First-order perturbation analysis of the best rank- (r_1, r_2, r_3) approximation in multilinear algebra. *Journal of Chemometrics*, 18(1):2–1, Jan. 2004. TRICAP Special Issue (B. Rayens, Ed.).
- [21] Lieven DE LATHAUWER, Bart DE MOOR, and Joos VANDEWALLE. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, April 2000.
- [22] Lieven DE LATHAUWER, Bart DE MOOR, and Joos VANDEWALLE. On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, April 2000.
- [23] Lieven DE LATHAUWER and Joos VANDEWALLE. Dimensionality reduction in higher-order signal processing and rank- (r_1, r_2, \dots, r_n) reduction in multilinear algebra. *Linear Algebra and its Applications*, 391:31–55, Nov 2004. Special Issue on Linear Algebra in Signal and Image Processing.
- [24] Michael Elad, Peyman Milanfar, and Gene H. Golub. Shape from moments – an estimation theory perspective. *IEEE Transactions on Signal Processing*, 52(7):1814–1829, July 2004.
- [25] Freeman Gilbert. Vandermonde matrix analysis of long-period seismograms. *Geophysical Journal International*, 146:843–849, September 2001.
- [26] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, third edition, 1996.
- [27] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.
- [28] B. Halder and T. Kailath. Efficient estimation of closely spaced sinusoidal frequencies using subspace-based methods. *IEEE Signal Processing Letters*, 4(2):49–51, February 1997.
- [29] E. Huupponen, W. De Clercq, G. Herrero, A. Saastamoinen, K. Egiazarian, A. Varri, B. Vanrumste, A. Vergult, S. Van Huffel, W. Van Paesschen, J. Hasan, , and S-L. Himanen. Determination of dominant simulated spindle frequency with different methods. Technical report, Faculty of Engineering, Katholieke Universiteit Leuven, 2005.
- [30] Umran S. Inan and Aziz S. Inan. *Electromagnetic Waves*. Prentice Hall, 2000.
- [31] C.J. James and O. J. Gibson. Temporally constrained ica: An application to artifact rejection in electromagnetic brain signal analysis. *IEEE Transactions on Biomedical Engineering*, 50(9):1108–1116, September 2003.

- [32] J. Jensen. *Sinusoidal Models for Speech Signal Processing*. PhD thesis, Aalborg University, Denmark, Denmark 2000.
- [33] T.P. Jung, S. Makeig, C. Humphries, T.W. Lee, M.J. McKeown, V. Iragui, and T.J. Sejnowski. Removing encephalographic artifacts by blind source separation. *Psychophysiology*, 2000.
- [34] Ivars P. Kirsteins, Sanjay K. Metha, and John Fay. Power-law processors for detecting unknown signals in colored noise. In *1997 International Conference on Acoustics, Speech and Signal Processing (ICASSP 97)*, volume 1, page 483, Munich, Germany, April 21-24 1997.
- [35] K. Kobayashi, C.J. James, T. Nakahori, T. Akiyama, and J. Gotman. Isolation of epileptiform discharges from unaveraged eeg by independent component analysis. *Clinical Neurophysiology*, 110(10):1755–1763, October 1999.
- [36] E. Kofidis and P. Regalia. On the best rank-1 approximation of higher order supersymmetric tensors. *SIAM J. Matrix Anal. Appl.*, 23:863–884, 2002.
- [37] Pascal Koiran, Natacha Portier, and Gilles Villard. A rank theorem for vandermonde matrices. *Linear Algebra and its Applications*, 378:99–107, 2004.
- [38] Tamara G. Kolda. Orthogonal tensor decompositions. *SIAM J. Matrix Anal. Appl.*, 23:243–255, 2001.
- [39] Z.J. Koles, J.C. Lind, and A.C. Soong. Spatio-temporal decomposition of the eeg: a general approach to the isolation and localization of sources. *Electroencephalography and Clinical Neurophysiology*, 1995.
- [40] P.M. Kroonenberg. *Three-Mode Principal Component Analysis*. DSWO Press, Leiden, 1983.
- [41] P.M. Kroonenberg and J. de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45:69–97, 1980.
- [42] J.B. Kruskal. Three-way arrays : rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Applications*, 18(2):95–138, 1977.
- [43] R. Kumareasan and D.W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustic, Speech and Signal processing*, 30(6):833–840, 1982.
- [44] Debasis Kundu and Amit Mitra. Estimating the number of signals of the damped exponential models. *Computational Statistics & Data Analysis*, 36(2):245–256, April 2001.

- [45] S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *J. Opt. Soc. Amer.*, 73(12):1799–1811, 1983.
- [46] Lieven De Lathauwer. *Signal processing based on multilinear algebra*. PhD thesis, Katholieke Universiteit Leuven, 1997.
- [47] Phillippe Lemmerling, Ioannis Dologlou, and Sabine Van Huffel. Speech compression based on exact modeling and structured total least norm optimization. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 1998.
- [48] Rainer Martin. Spectral subtraction based on minimum statistics. In *European Signal Processing Conference (EUSIPCO-94)*, volume 2, pages 1182–1185, Edinburgh, Scotland, September 1994.
- [49] Geert Morren. *Advanced signal processing applied to in-vivo spectroscopy and heart rate variability*. PhD thesis, Faculty of Engineering, K.U.Leuven, Leuven, Belgium, May 2004.
- [50] Geert Morren, Philippe Lemmerling, and Sabine Van Huffel. Decimative subspace-based parameter estimation technique. *Signal Processing*, 83:1025–1033, 2003.
- [51] Geert Morren, Philippe Lemmerling, Sabine Van Huffel, Gert Naulaers, H. Devlieger, and P. Casaer. Detection of autoregulation in the brain of premature infants using a novel subspace-based technique. In *Proceeding of the 23rd Annual International Conference of the IEEE Engineering in Medicine And Biology Society (EMBC2001)*, Istanbul, Turkey, October 2001.
- [52] Geert Morren, Sabine Van Huffel, I. Helon, Gert Naulaers, H. Daniels, H. Devlieger, and P. Casaer. Effect of non-nutritive sucking on heart rate, respiration and oxygenation: a model-based signal processing approach. *Comparative Biochemistry and Physiology – Part A*, 132, 2002.
- [53] H. Nam, T.G. Yim, S.K. Han, J.B. Oh, and S.K. Lee. Independent component analysis of ictal eeg in medial temporal lobe epilepsy. *Epilepsia*. 2002, 2002.
- [54] J. Nieuwenhuijse, R. Heusdens, and E. F. Deprettere. Robust exponential modeling of audio signal. In *Proceedings of International Conference on Acoustic, Speech and Signal Processing*, volume 6, 1998.
- [55] Albert Nuttall. Near-optimum detection of random signals of unknown location, struture, extent and strength. In *Oceans'95. MTS/IEEE. Challenges of Our Changing Global Environment. Conference Proceedings*, volume 3, pages 1659–64, 1995.

- [56] Ali Olfat and Said Nader-Esfahani. A new signal subspace processing for DOA estimation. *Signal Processing*, 84:721–728, 2004.
- [57] C.K. Papadopoulos and C.L. Nikias. parameter estimation of exponentially damped sinusoids using higher-order statistics. *IEEE Transactions on Acoustic, Speech and Signal processing*, 38(8):1424–1436, 1990.
- [58] Jean-Michel Papy, Lieven De Lathauwer, and Sabine Van Huffel. Common pole estimation in the multi-channel exponential data modelling. *Signal Processing*, 2005. In press.
- [59] Jean-Michel Papy, Lieven De Lathauwer, and Sabine Van Huffel. A shift invariance-based order selection technique for the exponential data modelling. *IEEE letters on Signal Processing*, 2005. Accepted.
- [60] Jean-Michel Papy, Sabine Van Huffel, Laurent Rippert, and Martine Wevers. Spectral subtraction method applied to damage detection in composite materials with embedded optical fibers. In *ProRISC Benelux Workshop on Circuits, Systems and Signal Processing (ProRISC2001)*, pages 542–549, Veldhoven, The Netherlands, 2001.
- [61] Jean-Michel Papy, Sabine Van Huffel, Laurent Rippert, and Martine Wevers. On-line detection method for transient waves applied to continuous health monitoring of carbon fiber reinforced polymer composites with embedded optical fibers. In Ralph C. Smith, editor, *SPIE's 10th Annual International Symposium on Smart Structures and Materials : Modeling, signal processing and control (SSM2003)*, volume 5049, San Diego, California, USA, March 2003.
- [62] Pieter Pels. *Analysis and Improvement of Quantification Algorithms for Magnetic Resonance Spectroscopy*. PhD thesis, Katholieke Universiteit Leuven, Belgium, January 2005.
- [63] V. Pisarenko. The retrieval from harmonics from a covariance function. *Geophys. J. R. Astron. Soc.*, 33:347–366, 1973.
- [64] M.B. Priestley. *Spectral analysis and time series*. Academic Press, 1981.
- [65] José A. Ramos. Exponential data fitting applied to environmental data. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004.
- [66] Philippe Ravier. *Détection de transitoires par ondelettes adaptées – Critères d'adaptation fondés sur les statistiques d'ordre supérieur*. PhD thesis, Institut National Polytechnique de Grenoble, 1998.
- [67] Laurent Rippert. *Optical fiber for damage monitoring in carbon fiber reinforced plastic composite materials*. PhD thesis, Katholieke universiteit Leuven, March 2003.

- [68] Laurent Rippert, Jean-Michel Papy, Martine Wevers, and Sabine Van Huffel. Fiber optic sensor for continuous health monitoring in cfrp composite materials. In *Smart Structures and Materials 2002: Modeling, Signal Processing, and Control*, pages 312–323, San Diego, California, USA, March 2002.
- [69] Laurent Rippert, Martine Wevers, and Sabine Van Huffel. Optical and acoustic damage detection in laminated CFRP composite materials. *Composites Science and Technology*, 60:2713–2724, June 2000.
- [70] Laurent Rippert, Martine Wevers, and Sabine Van Huffel. Optical sensor for continuous damage monitoring in composite materials. In *4th International Conference and exhibition on Optoelectronics, Optical Sensors & Measurement Techniques (OPTO 2000)*, pages 107–112, Erfurt Fair, Germany, May 9–11 2000.
- [71] Laurent Rippert, Martine Wevers, and Sabine Van Huffel. Optical sensor for continuous damage monitoring in composite materials. In *24th European Conference on Acoustic Emission Testing (EWGAE 2000)*, pages 69–76, Senlis, France, May 24–26 2000.
- [72] Bertrand David Roland Badeau and Gaël Richard. Selecting the modeling order for the esprit high resolution method: an alternative approach. In *2004 International Conference on Acoustics, Speech, and Signal Processing ICASSP04*, volume II, pages 1025–1028, Montréal, Québec, Canada, 17–21 May 2004.
- [73] R. Roy and T. Kailath. Esprit—estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust., Speech, Signal Processing*, 37(7):984–995, 1989.
- [74] R. Roy, A. Paulraj, and T. Kailath. Esprit—a subspace rotation approach for to estimation of parameters of cisoids in noise. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(4):1340–1342, 1986.
- [75] Abd-Krim Seghouane and De Lathauwer Lieven. A bootstrap model selection criterion based on kullback’s symmetric divergence’. In *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP’03)*, St. Louis, Missouri, September 2003.
- [76] P. Sovka, P. Pollák, and J. Kybic. Extended spectral subtraction. In *European Signal Processing Conference (EUSIPCO-96)*, Trieste, Italy, September 1996.
- [77] P. Stoica and A.E. Nordsjö. Subspace-based frequency estimation in the presence of moving-average noise using decimation. *Signal Processing*, 63:221–220, 1997.

- [78] P. Stoica and K.C. Sharmen. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoustic, Speech and Signal processing*, 38(7):1132–1143, 1990.
- [79] P. Stoica and T. Sundin. Nonparametric nmr spectroscopy. *Journal of Magnetic Resonance*, 2001.
- [80] Petre Stoica and Randolph Moses. *Spectral Analysis of Signals*. Pearson Prentice Hall, 2005.
- [81] J. ten Berghe, J. de Leeuw, and P.M. Kroonenberg. Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 52:183–191, 1987.
- [82] Lloyd N. Trefethen and David Bau. *Numerical linear algebra*. SIAM, 3rd edition, 1997.
- [83] L R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.
- [84] R. Vafin, R. Heusdens, S. Van De Par, and W. B. Kleijn. Improved modeling of audio signals by modifying transient locations. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '01)*, New Paltz, NY, 2001.
- [85] Sabine Van Huffel. Enhanced resolution based on minimum variance estimation and exponential data modeling. *Signal Processing*, 33(3):333–355, 1993.
- [86] Sabine Van Huffel, Hua Chen, Caroline Decanniere, and Paul Van Hecke. Algorithm for time-domain NMR data fitting based on total least squares. *Journal of Magnetic Resonance, Series A* 110:228–237, 1994.
- [87] Sabine Van Huffel and Joos Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*, volume 9 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, Pennsylvania, 1991.
- [88] Sabine Van Huffel, Yu Wang, Leen Vanhamme, and Paul Van Hecke. Automatic Frequency Alignment and Quantitation of Single Resonances Spectra via Complex Principal Component Analysis. *Journal of Magnetic Resonance*, 158:1–14, 2002.
- [89] H. L. Van Trees. *Detection, Estimation and modulation theory*. Wiley, New York, 1971.
- [90] Steve Vandenplas, Jean-Michel Papy, Martine Wevers, and Sabine Van Huffel. Acoustic emission monitoring using a multimode optical fiber sensor. In *SPIE Proceedings Smart Structures/NDE Joint Conferences*, March 2004.

- [91] Leen Vanhamme. *Advanced time-domain methods for nuclear magnetic resonance spectroscopy data analysis*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, November 1999.
- [92] Leen Vanhamme and Sabine Van Huffel. Multichannel quantification of biomedical magnetic resonance spectroscopy signals. In F.T. Luk, editor, *Proceedings of SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations VIII*, volume 3461, pages 237–248, San Diego, California, 19–24 July 1998.
- [93] Leentje Vanhamme, T. Sundin, Paul Van Hecke, and Sabine Van Huffel. Mr spectroscopy quantitation: a review of time-domain methods. *NMR Biomed.*, 2001.
- [94] Peter Vary. Noise suppression by spectral magnitude estimation — mechanism and theoretical limits. In *Signal Processing*, volume 8, pages 387–400, July 1985.
- [95] Werner Verhelst, Kris Hermus, Lemmerling Philippe, Patrick Wambacq, and Sabine Van Huffel. Modeling audio of damped sinusoids using total least squares algorithms. In Sabine Van Huffel and Philippe Lemmerling, editors, *Total Least Squares and Errors-in-Variables Modeling : Analysis, Algorithms and Applications*, pages 331–340. Kluwer Academic Publishers, 2002.
- [96] G.L. Wallstrom, R.E. Kass, A. Miller, J.F. Cohn, and N.A. Fox. Automatic correction of ocular artifacts in the eeg: a comparison of regression-based and component-based methods. *International Journal of Psychophysiology*, 2004.
- [97] Yu Wang, Sabine Van Huffel, Leen Vanhamme, Nicola Mastronardi, and Paul Van Hecke. Advanced signal processing methods for quantitation of resonances in magnetic resonance spectra. In *Proc. of the Thirteenth IEEE Symposium on Computer-Based Medical Systems (CBMS2000)*, pages 63–68, Houston, USA, June 2000.
- [98] Zhen Wang and Peter Willett. A performance study of some transient detectors. *IEEE Transactions on Signal Processing*, 48(9):2682–86, September 2000.
- [99] Zhen Wang and Peter Willett. All-purpose and plug in power-law detector for transient signals. *IEEE Transaction on Signal Processing*, pages 2454–2466, November 2001.
- [100] Zhen Wang and Peter Willett. Improved power-law detection of transients. In *International Conference on Acoustics, Speech and Signal Processing*, volume 5, Salt Lake City, USA, May 7-11 2001.

- [101] Zhen Wang, Peter Willett, and R. Streit. Wavelets in the frequency domain for narrowband process detection. In *International Conference on Acoustics, Speech and Signal Processing*, volume 5, Salt Lake City, USA, May 7-11 2001.
- [102] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. on Acoustics, Speech, Signal Processing*, 33(2):387–392, April 1985.
- [103] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals in presence of white noise. *Journal of Multivariate Analysis*, 20(1):1–25, 1986.

List of publications

Publications in International Journals

- **Jean-Michel Papy**, Lieven De Lathauwer and Sabine Van Huffel. Exponential data fitting using multilinear algebra: the single- and the multi-channel case. *Numerical Linear Algebra with Applications*, 2005. In Press.
- Wim De Clercq, Bart Vanrumste, **Jean-Michel Papy**, Wim Van Paesschen, Sabine Van Huffel. Modelling Common Dynamics in Multichannel Signals with Applications to Artifact and Background Removal in EEG Recordings. To appear In *IEEE Transactions on Biomedical Engineering*, December 2005.
- **Jean-Michel Papy**, Lieven De Lathauwer and Sabine Van Huffel. Common pole estimation in the multi-channel exponential data modelling. Accepted for publication In *Signal Processing*, March 2005.
- **Jean-Michel Papy**, Lieven De Lathauwer and Sabine Van Huffel. Exponential data fitting using multilinear algebra: the decimative case. Accepted for publication In *IEEE Transactions on Signal Processing*, September 2005.
- **Jean-Michel Papy**, Lieven De Lathauwer and Sabine Van Huffel. A shift invariance-based model order selection technique. Accepted for publication In *IEEE Signal Processing Letters*, July 2005.
- Steve Vandenplas, **Jean-Michel Papy**, Martine Wevers and Sabine Van Huffel. Acoustic emission monitoring using a multimode optical fiber sensor. *Insight*, 46(4):203–209, April 2004.

Publications on International Conferences

- **Jean-Michel Papy**, Sabine Van Huffel, Laurent Rippert and Martine Wevers. Spectral Subtraction Method applied to Damage Detection in Composite Materials with Embedded Optical Fibers. In Proceedings of the *ProRISC Benelux Workshop on Circuits, Systems and Signal Processing*, Veldhoven, The Netherlands, November 29-30, 2001.
- Laurent Rippert, **Jean-Michel Papy**, Martine Wevers and Sabine Van Huffel. Fiber optic sensor for continuous health monitoring in CFRP composite materials. In Proceedings of the *SPIE International Symposium on smart structures and Materials and NDE for health monitoring and Diagnostics*, San Diego, California, USA, March 17-21, 2002.
- **Jean-Michel Papy**, Laurent Rippert, Sabine Van Huffel and Martine Wevers. On-line detection for transient waves applied to continuous monitoring of CFRP composite materials with embedded optical fibers. In Proceedings of the *SPIE International Symposium on smart structures and Materials and NDE for health monitoring and Diagnostics*, San Diego, California, U.S.A., March 2-6, 2003.
- Steve Vandenplas, **Jean-Michel Papy**, Martine Wevers and Sabine Van Huffel. Acoustic emission monitoring using a multimode optical fiber sensor. In Proceedings of the *SPIE Smart Structures/NDE Joint Conferences (SPIE SS/NDE 2004)*, San Diego, CA, March 2004.
- Steve Vandenplas, **Jean-Michel Papy**, Martine Wevers, Sabine Van Huffel, and Daniele Inaudi. Acoustic emission monitoring using a polarimetric single mode optical fiber sensor. In Proceedings of the *17th International Conference of Optical Fibre Sensors (OFS-17)*, Bruges, Belgium, May 23-27, 2005.
- Wim De Clercq, Bart Vanrumste, **Jean-Michel Papy**, Wim Van Paesschen, Sabine Van Huffel. Modelling Common Dynamics in Multichannel Signals with Applications to Artifact and Background Removal in EEG Recordings. In Proceedings of the *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, Shanghai (China), September 1-4, 2005.

Curriculum Vitae

Jean-Michel Papy was born in Pamiers (France) on September 2, 1974. He received the M.Sc. degree in Signal, Image and Acoustics from the Université Paul Sabatier (Toulouse, France) in 2000. His main topic was Acoustics and his Master's thesis under the supervision of Prof. Yannick Deville and Dr. René Gamba was devoted to “Atténuation aux bruits d'impacts–Isolement acoustique des planchers”. This work was realized in collaboration with the company Acoustique Gamba et Associés.

In January 2001, he started working as a researcher at the Electrical Engineering Department at ESAT in the Laboratory of Signals, Identification, System Theory, and Automation (SCD/SISTA) at the Katholieke Universiteit Leuven (Belgium) and in January 2002 he started his PhD. During this period he was supported by the PDT-Coil Project (EU project number: ENK6/2000/00074, duration: 01.09.2000–28.02.2003 (phase 1). EU project number: NNE5/2001/887, duration: 01.01.2003–31.08.2005 (phase 2)).

