

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT DER TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee)

ADVANCED TIME-DOMAIN METHODS FOR NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY DATA ANALYSIS

Promotor:
Prof. dr. ir. S. Van Huffel

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Leentje VANHAMME

November 1999



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT DER TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee)

ADVANCED TIME-DOMAIN METHODS FOR NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY DATA ANALYSIS

Jury:
Prof. dr. ir. J. Delrue, voorzitter
Prof. dr. ir. S. Van Huffel, promotor
Dr. ir. I. Dologlou (ILSP, Greece)
Prof. dr. ir. R. Pintelon (ELEC, VUB)
Prof. dr. ir. D. Vandermeulen
Prof. dr. ir. J. Vandewalle
Prof. dr. P. Van Hecke (Faculteit Geneeskunde)
Dr. ir. D. van Ormondt (T.U. Delft)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Leentje VANHAMME

U.D.C. 543.422.25 November 1999

©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/1999/7515/48

ISBN 90-5682-218-7

Acknowledgments

During my Ph.D. years I have been privileged to be surrounded by competent, enthusiastic and supportive people. Time has come to send out a big "thank you!" to them all.

Prof. Sabine Van Huffel has been an instrumental figure throughout this undertaking. My decision to take the first step of this journey was largely due to her initial support and encouragement. When I look back at the past years and the traveled roads, I can only conclude that it has been a great and mind-expanding experience. Along its course, she provided me with a continuous stream of valuable research suggestions and made herself always available to discuss matters or to proofread my papers.

This work would not have been possible without the presence of Prof. Paul Van Hecke of the Faculty of Medicine, whose expert knowledge on all practical MRS aspects was immeasurable. He spared no effort to provide me "with yet another" set of MRS data or to answer my countless questions. It has been a great pleasure to collaborate with him.

I would like to thank dr. ir. Dirk van Ormondt of the T.U. Delft for the clarity with which he explained various MRS aspects to me throughout the different stages of my Ph.D. work as well as for the vivid interest he has always shown in my work.

My thanks go to dr. ir. Yannis Dologlou of ILSP, Greece, for his inspiring discussions on any number of subjects and for agreeing to be a member of my jury.

I am especially grateful to Prof. Rik Pintelon of the VUB for providing me with valuable insights in frequency-domain methods.

I would like to express my gratitude to Prof. Joos Vandewalle for giving me the opportunity to start my Ph.D. as well as for participating as a member of my reading committee.

Prof. Jan Delrue is gratefully acknowledged for chairing the jury. I thank Prof. Dirk Vandermeulen for agreeing to be a member of my jury.

During his stay at SISTA-COSIC and the University Hospital, Aad van den Boogaart passed on to me part of his seemingly boundless enthusiasm. It is with great pleasure that I reflect upon those days. I want to thank him for the enjoyable cooperation from which I learned a lot as well as for his efforts to keep in touch with me over the years.

Special thanks go to Tomas Sundin, with whom I had the pleasure to work together during the last year and a half. My research work no doubt benefited greatly from his enlightening views on filters as well as many other signal processing areas.

My thanks also go to all the members of the TMR network with whom I met at least on a yearly basis. It has always been great fun to work with them all and our get-togethers were always very inspiring. Sophie Cavassila, Prof. Danielle Graveron-Demilly, Alain Coron and all others; thanks for the great times!

The thought of the Nijmegen crew always makes me smile. They have an inimitable way of combining extreme fun with seriousness. I especially would like to thank René in't Zandt for his critical views on my work and for providing me with help on various occasions.

A hearty thanks is sent out to all the colleagues and friends in SISTA-COSIC for supplying a rich and rewarding work atmosphere. I greatly appreciate the efforts of Bart Motmans and his unique ability to smoothly take care of a diverse breadth of matters. A special word of thank goes to Ida Tassens for arranging any number of practical issues. I thank system manager Eric Olemans for all the help he provided over the years. Thanks also to librarian Guy Hermans for the swift handling of all my paper and book requests.

On a monetary note, I thank the K.U.L. and the I.W.T. for financing my Ph.D. research.

My parents have been my best mentors, preparing me for life in its various aspects. There are no words to express my gratitude for all the selfless efforts they made. I thank my mother for her love, friendship and encouragement; especially in difficult times. To my great sadness, my father did not live to see this day. I am extremely grateful, however, for all the wisdom that he passed on to me and for the heartwarming memories. It is to my father with my deepest love and respect that I dedicate this thesis.

Leentje Vanhamme
October 1999

Abstract

Over the past years magnetic resonance spectroscopy (MRS) has been of significant importance both as a fundamental research technique in different fields, as well as a diagnostic tool in medical environments. With MRS, for example, spectroscopic information, such as the concentrations of chemical substances, can be determined non-invasively. To that end, the signals are first modeled by an appropriate model function and mathematical techniques are subsequently applied to determine the model parameters. In this thesis, signal processing algorithms are developed to quantify in-vivo and ex-vivo MRS signals. These are usually characterized by a poor signal-to-noise ratio, overlapping peaks, deviations from the model function and in some cases the presence of disturbing components (e.g. the residual water in proton spectra). The work presented in this thesis addresses a part of the total effort to provide accurate, efficient and automatic data analysis of MRS signals.

The various methods that exist to extract the parameters from a MRS signal are explained and connections between methods are pointed out. Typically, time-domain methods are divided into black-box and interactive methods. An often used interactive method is VARPRO. HSVD and HTLS are well-known examples of black-box methods. Interactive methods fit a model function to the data and minimize a cost function. Linear prediction and state-space methods are two variants of the black-box methods which do not involve the minimization of a cost function. Algorithms which fall into this latter class of methods only approximately solve the linear prediction or state-space problem setting. It is shown here how the linear prediction and state-space problems can be solved without introducing approximations leading to maximum likelihood (ML) parameter estimates. However, then again the problems reduce to the solution of a minimization problem. The connection between time- and frequency-domain methods is also pointed out.

In the class of interactive time-domain methods, AMARES (*a*dvanced *m*ethod for *a*ccurate, *r*obust and *e*fficient *s*pectral fitting) is developed and shown to outperform the popular VARPRO method in terms of accuracy, robustness and flexibility.

In biochemical studies MRS signals are in many cases acquired consecutively to monitor metabolic changes over time. Often information concerning the time evolution of some of the parameters in these time series is present. It is shown how the interactive and black-box time-domain methods can be extended to process multiple signals simultaneously. This leads to improved and more robust parameter estimates compared to the estimates obtained by processing the signals individually.

Various techniques that have been used in the past to obtain estimates of parameters of selected peaks in the presence of unknown or uninteresting spectral features separated in frequency from the metabolites of interest are investigated and compared with each other. This problem is denoted by frequency-selective (FS) parameter estimation. In this context, a new technique based on the use of maximum-phase FIR filters in combination with AMARES is developed and compared with the existing methods. In terms of final parameter accuracy this new method outperforms the others. The method is also very easy to use and has a low computational complexity.

The removal of residual water in proton spectra is treated as a special case of FS parameter estimation. A fast version of a singular value decomposition (SVD)-based method is presented and found to be less performant than the filter-based approach in terms of efficiency, accuracy and ease of use.

Notation

The current section lists the symbols and acronyms that occur frequently in this thesis. The last column refers to the page on which the symbol or acronym is first used and/or defined.

Remark. The notation used in this thesis allows one to distinguish between scalars, vectors and matrices. Lower-case characters represent scalar values. Boldface lower-case characters are used for vectors. Boldface capitals represent matrices.

List of symbols

a^*	complex conjugate of a	
a_i	ith element of vector $\mathbf{a}$	
a_k	amplitude of the kth exponential	page 4
$\hat{a}$	estimate of parameter a, a is e.g. a_k, d_k, f_k,...	page 14
$\|a\|$	absolute value of $a \in \mathbb{R}$; modulus of $a \in \mathbb{C}$	
$\|\mathbf{a}\|$	vector norm of $\mathbf{a}$	
$\mathbf{a}$	vector	
$\mathbf{A}$	matrix	
$\mathbf{A}^T$	transpose of $\mathbf{A}$	
$\mathbf{A}^H$	hermitian transpose of $\mathbf{A}$	
$\mathbf{A}^{-1}$	inverse of $\mathbf{A}$	
$\mathbf{A}^{\uparrow}$	matrix obtained by deleting top row of $\mathbf{A}$	page 30
$\mathbf{A}_{\downarrow}$	matrix obtained by deleting bottom row of $\mathbf{A}$	page 30
$\|\mathbf{A}\|_F$	Frobenius norm of $\mathbf{A}$	page 26
$\mathbf{A}^{\dagger}$	Moore-Penrose pseudo-inverse of $\mathbf{A}$	page 21
c_k	complex amplitude $a_k e^{j\phi_k}$ of the kth exponential	page 21
$\mathbb{C}^{n\times m}$	set of all $n \times m$ complex-valued matrices	
d_k	damping of the kth exponential	page 4
e_n	the nth noise sample	page 4
$E(.)$	expected value operator	page 22

$eig(.)$	eigenvalues of (.)	page 30
f_k	frequency of the kth exponential	page 4
$\mathbf{F}$	Fisher information matrix	page 22
$\mathbf{g}(f)$	Fourier vector	page 112
$\mathbf{h}$	filter coefficients vector	page 111
$\mathbf{H}_{lp}, \mathbf{h}_{lp}$	matrix, vector associated with linear prediction framework, composed of signal samples	page 25
$\bar{\mathbf{H}}_{lp}, \bar{\mathbf{h}}_{lp}$	matrix, vector associated with linear prediction framework, composed of noise-free signal samples	page 25
$\mathbf{H}_s$	matrix associated with state-space framework, composed of signal samples	page 31
$\bar{\mathbf{H}}_s$	matrix associated with state-space framework, composed of noise-free signal samples	page 29
$Im(a)$	imaginary part of a	page 20
$\mathbf{I}_n$	$n \times n$ identity matrix	page 45
j	$\sqrt{-1}$	page 4
$\mathbf{J}(\mathbf{x})$	Jacobian matrix	page 43
$\min_x f$	minimize f with respect to x	page 26
$\mathbf{0}$	zero matrix/vector	
$\mathcal{O}$	order of magnitude	page 143
$Re(a)$	real part of a	page 20
$\mathbb{R}^{n \times m}$	set of all $n \times m$ real-valued matrices	
$sign(a)$	signum function	page 45
t_n	the nth time sample	page 5
Δt	sampling time interval	page 5
y_n	the nth measured signal sample	page 4
$\bar{y}_n$	the nth noise-free signal sample	page 4
y_{ns}	the nth measured signal sample of signal s of the time series	page 74
y_{f_n}	nth sample of filtered measured signal	page 110
$\bar{y}_{f_n}$	nth sample of filtered noise-free signal	page 111
$\mathcal{Y}_l$	the lth element of the Fourier transformed data vector	page 35
$\overline{\mathcal{Y}}_l$	the the lth element of the Fourier transformed noise-free data vector	page 6
z_k	signal pole $e^{(-d_k + j2\pi f_k)\Delta t}$ of the kth exponential	page 24
σ	standard deviation of the noise on the real and imaginary part	page 14
ϕ_k	phase of the kth exponential	page 4
$\frac{\delta f}{\delta x}$	partial derivative of f with respect to x	page 22
$\nabla f(\mathbf{x})$	gradient of f evaluated in $\mathbf{x}$	page 40
$\nabla^2 f(\mathbf{x})$	Hessian of f evaluated in $\mathbf{x}$	page 40

List of acronyms and abbreviations

a.u.	arbitrary unit	page 56
AR	auto regressive	page 25
ARMA	auto regressive moving average	page 25
ADP	adenosine diphosphate	page 11
ATP	adenosine triphosphate	page 8
Cho	choline	page 11
Cr	creatine	page 11
CRB	Cramér-Rao bound	page 13
CSI	chemical shift imaging	page 2
CTLS	constrained total least squares	page 27
dB	decibel	page 14
DFT	discrete fourier transform	page 14
ES	external standard	page 11
FIR	finite impulse response	page 15
flop	floating point operation	page 45
FS	frequency selective	page 15
FT	fourier transform	page 101
Glc	glucose	page 11
Glu	glutamate	page 11
Gln	glutamine	page 11
Hz	Hertz	
IQML	iterative quadratic maximum likelihood	page 29
kHz	kiloHertz	
Lac	lactate	page 11
LP	linear prediction	page 23
LS	least squares	page 21
MDL	minimum description length	page 106
mI	myo-inositol	page 11
ML	maximum likelihood	page 19
mM	millimolar	
MRS	magnetic resonance spectroscopy	page 1
MRSI	magnetic resonance spectroscopy imaging	page 2
ms	milliseconds	
NAA	N-acetyl-aspartate	page 11
NMR	nuclear magnetic resonance	page 1
NLLS	nonlinear least squares	page 20
PCr	phosphocreatine	page 11
PDE	phosphodiester	page 11
P_i	inorganic phosphate	page 11
PME	phosphomonoester	page 11
RF	radio frequency	page 4
(R)RMSE	(relative) root mean squared error	page 13

s	seconds	
SNR	signal-to-noise ratio	page 2
STLN	structured total least norm	page 27
SVD	singular value decomposition	page 25
T	Tesla	
TLS	total least squares	page 26
w.r.t.	with respect to	page 15
i.i.d.	independently identically distributed	page 26

Contents

Acknowledgments i

Abstract iii

Notation v

Contents ix

Summary in Dutch xv

1 Introduction 1

1.1 Situation 2

1.2 Basic principles of MRS 3

1.3 Mathematical description of the MRS signal 4

1.3.1 FID signal 4

1.3.2 Echo signal 5

1.4 Visualizing the NMR signals 6

1.5 The concept of prior knowledge 8

1.6 ^{31}P and ^{1}H MRS signals 11

1.7 Goals of the thesis 13

1.8 Monte-Carlo simulation procedure 13

1.9 Chapter-by-chapter overview 14
1.10 Conclusions 16

2 Overview of parameter estimation methods 19

2.1 Time-domain methods 19
2.1.1 Interactive methods 20
2.1.2 Black-box methods 23
2.1.3 Remarks 33
2.1.4 Summary 34
2.2 Frequency-domain methods 35
2.3 Other methods 37
2.4 Conclusions 37

3 VARPRO and AMARES 39

3.1 Unconstrained minimization 40
3.1.1 General purpose unconstrained minimization 40
3.1.2 NLLS minimization 42
3.1.3 Examples of practical NLLS algorithms 44
3.2 Specific implementation issues 47
3.2.1 General functional 48
3.2.2 Variable projection functional 50
3.3 Numerical evaluation 55
3.3.1 Influence of different problem formulations to ensure that the dampings are positive 57
3.3.2 Influence of starting values on the minimization of the functionals 58
3.3.3 Influence of prior knowledge on the minimization of the functionals 61
3.3.4 General findings 64

3.4 Possibilities of imposing prior knowledge in VARPRO and AMARES . . . 65
3.4.1 VARPRO: multiplet approach . . . 65
3.4.2 AMARES: singlet approach . . . 66
3.4.3 Example . . . 67
3.5 AMARES: an overview . . . 70
3.6 Conclusions . . . 72

4 Quantification of series of spectra **73**
4.1 Extensions of AMARES to quantitate time series . . . 74
4.1.1 ML use of common information present in spectra: AMARES$_{ts}$ . . . 74
4.1.2 Suboptimal use of common information present in spectra: AMARES$_{sts}$ . . . 78
4.2 Extensions of HTLS to quantitate time series . . . 79
4.2.1 HTLSsum . . . 79
4.2.2 HTLSstack . . . 80
4.3 Numerical validation . . . 82
4.3.1 Example A . . . 82
4.3.2 Example B . . . 90
4.4 Experimental time series . . . 94
4.5 Conclusions . . . 95

5 Parameter estimation in non-ideal circumstances **99**
5.1 Corrections for model imperfections . . . 100
5.2 Visual SNR and/or resolution improvements . . . 102
5.3 Removal of unwanted components . . . 103
5.3.1 Removal of components separated in frequency . . . 104
5.4 Conclusions . . . 108

6 Frequency-selective parameter estimation 109

6.1 From FIR filter theory to accurate parameter estimation 110

6.1.1 FIR filter theory 110

6.1.2 Parameter estimation 117

6.1.3 Visualization 119

6.1.4 Design and use of the maximum-phase FIR filter 119

6.2 Numerical validation 122

6.2.1 Two-peak example 123

6.2.2 Frequency-selective quantification of an *in-vivo* ^{13}C spectrum 131

6.2.3 Frequency-selective quantification of an *in-vivo* ^{31}P spectrum 134

6.3 Conclusions 136

7 Fast removal of residual water in proton spectra 141

7.1 Derivation of the new algorithm 142

7.1.1 HSVD - Toeplitz version 142

7.1.2 HLR 143

7.2 Simulations 147

7.3 Application to *in-vivo* MRS image 149

7.4 Conclusions 151

8 Application of FS estimation in proton spectra 153

8.1 Optimized filter design scheme 154

8.2 Interpretation of other FIR filters appeared in literature 156

8.3 Numerical validation 158

8.3.1 Influence of FIR filter parameters on estimation precision 160

8.3.2 Influence of HSVD parameters on estimation precision . 161

8.3.3 Comparison of FIR filter method and HSVD method w.r.t. estimation precision 162

8.3.4 Comparison of FIR filter, HSVD and HLR methods for water removal w.r.t. computational complexity 164

8.4 Conclusions 165

9 Conclusions and new challenges **167**

9.1 Conclusions 167

9.2 New challenges 169

A Equivalence between time- and frequency domain model fitting **171**

B Publication list **173**

B.1 Articles in international journals 173

B.2 Articles in conference proceedings 174

B.3 Book chapters 174

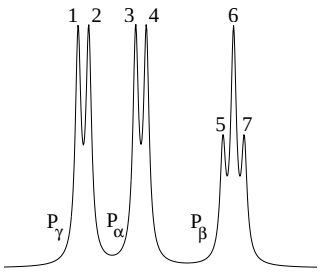
B.4 Abstracts international conferences 174

Geavanceerde methoden voor de analyse van nucleaire magnetische resonantie spectroscopie signalen in het tijdsdomein

Nederlandse samenvatting

Hoofdstuk 1: Inleiding

Nucleaire magnetische resonantie (NMR) is een techniek die de laatste jaren veelvuldig gebruikt wordt om beelden te maken van levende organismen (MRI of Magnetic Resonace Imaging) of om concentraties van chemische verbindingen, metabolieten genaamd, in diezelfde organismen, op een niet-invasieve manier te bepalen (MRS of Magnetische Resonantie Spectroscopie). Naargelang de ruimtelijke lokalisatietechniek die gebruikt wordt om MRS signalen van een welbepaald gebied te bekomen, kan men twee groepen onderscheiden: de single-voxel techniek die het signaal van één welbepaald volume (voxel) opneemt en de multi-voxel techniek die het signaal opneemt van een reeks voxels, gewoonlijk een matrix van voxels in een gegeven vlak (MRS imaging (MRSI) of beeldvorming). De grote troef van NMR is het gebruik als niet-invasieve analytische meettechniek. De parameters van het NMR signaal leveren rechtstreekse informatie over de moleculen van het onderzochte organisme: de frequenties van de spectrale componenten van het NMR signaal kenmerken de identiteit van de moleculen, de demping kenmerkt de mobiliteit van de overeenstemmende molecule en de intensiteit (amplitude) is rechtstreeks evenredig met de con-



Figuur 0.1: Theoretisch frequentiespectrum van ATP.

centratie van de moleculen. De nauwkeurige bepaling van deze parameters, in het bijzonder van de intensiteit, is dus primordiaal om MRS als een volwaardige analytische techniek te kunnen hanteren. Het doel van dit werk is derhalve technieken uit te werken en te evalueren voor de nauwkeurige analyse van MRS signalen voor biomedische toepassingen. De verwerking van MRS signalen van levende organismen wordt enorm bemoeilijkt door allerlei degradaties van deze signalen, inherent aan de moleculaire omgeving, de lage moleculaire concentraties, de technische beperkingen eigen aan meetprocessen op levende wezens en tenslotte aan de beperkte beschikbare meettijd. In het algemeen worden de signalen dus gekenmerkt door een lage signaal-ruis verhouding, overlappende spectrale componenten, afwijkingen van de modelfunctie en in sommige gevallen de aanwezigheid van storende componenten (zoals resterend water in proton spectra). De signalen worden gemodelleerd door een gepaste modelfunctie en vervolgens worden wiskundige technieken gebruikt om de modelparameters te bepalen.
De meest gebruikte modelfunctie is het volgend model bestaande uit een som van Lorentzianen

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-d_k + j2\pi f_k)t_n} + e_n \quad n = 0, 1, \ldots, N-1, \qquad (0.1)$$

K stelt het aantal verschillende componenten (ook pieken genoemd) in het signaal voor, $j = \sqrt{-1}$. Alternatieve modelfuncties die soms gebruikt worden zijn de Gaussiaanse en de Voigt modelfunctie.
Voorkennis is een erg belangrijk concept in de MRS gegevensverwerking. Ter illustratie nemen we de energiedrager adenosine trifosfaat (ATP). Het spectrum van ATP is weergegeven in Fig. (0.1). De aanwezige biochemische voorkennis kan in dit geval uitgedrukt worden als volgt

- De frequentieverschillen tussen de individuele resonanties binnen een multiplet zijn gelijk en gekend: $f_1 - f_2 = f_3 - f_4 = f_5 - f_6 = f_6 - f_7 = \Delta f$.
- De dempingen van alle pieken zijn dezelfde: $d_1 = d_2 = \ldots = d_7$.

- De amplitudeverhoudingen tussen de multipletten zijn $1:1$ voor P_α, $1:2:1$ voor P_β en $1:1$ voor P_γ. De totale amplitudes van P_α, P_β en P_γ verhouden zich als $1:1:1$. Samengevat: $a_1 = a_2 = a_3 = a_4 = 2a_5 = a_6 = 2a_7$.
- De fases van alle pieken zijn gelijk: $\phi_1 = \phi_2 = \ldots = \phi_7$.

De analyse van de MRS signalen moet niet alleen nauwkeurig zijn, ze moet ook efficiënt en liefst zo automatisch mogelijk gebeuren. Nauwkeurige resultaten liggen aan de basis van een betrouwbare diagnose in een klinische omgeving of van een juiste interpretatie van experimentele resultaten in het algemeen. Automatisatie speelt niet alleen een rol bij de verwerking van grote hoeveelheden data maar garandeert bovendien dat de resultaten gebruikersonafhankelijk en reproduceerbaar zijn. De resultaten van een analyse moeten zo vlug mogelijk beschikbaar zijn in een medische omgeving en efficiëntie is dan ook van uitermate groot belang. Het in deze thesis voorgestelde werk heeft als doel nauwkeurige, efficiënte en automatische MRS gegevensanalyse een stap dichter bij de dagdagelijkse realiteit te brengen.

Hoofdstuk 2: Overzicht van parameterschattingsmethoden

Hoofdstuk 2 geeft een overzicht van de meest gebruikte parameterschattingsmethoden. De methoden worden opgedeeld in tijdsdomein en frequentiedomein methoden. De onderliggende principes van de belangrijkste methoden worden uitgelegd en er worden verbanden tussen de methoden gelegd.

Tijdsdomein methoden worden meestal ingedeeld in de zogenaamde interactieve en zwarte-doos methoden. De interactieve methoden minimaliseren het verschil tussen de data en de modelfunctie,

$$\sum_{n=0}^{N-1} |y_n - \bar{y}_n|^2. \tag{0.2}$$

De modelfuncties voorgesteld in hoofdstuk 1 kunnen algemeen geschreven worden als

$$\bar{y}_n = \sum_{k=1}^{K} c_k \gamma_k(\boldsymbol{\alpha}_k, n), \quad n = 0, \ldots, N-1 \tag{0.3}$$

hierin zijn c_k de complexe amplitudes $a_k e^{j\phi_k}$ en $\gamma_k(\boldsymbol{\alpha}_k, n)$ zijn onafhankelijke functies van de niet-lineaire parameter vector $\boldsymbol{\alpha}_k$. In vector notatie wordt Vgl. (0.3)

$$\bar{\mathbf{y}} = \boldsymbol{\Gamma}\mathbf{c}, \tag{0.4}$$

met $\mathbf{c} = [c_1, \ldots, c_K]^T$ en

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1(\boldsymbol{\alpha}_1, 0) & \ldots & \gamma_K(\boldsymbol{\alpha}_K, 0) \\ \vdots & \ddots & \vdots \\ \gamma_1(\boldsymbol{\alpha}_1, N-1) & \ldots & \gamma_K(\boldsymbol{\alpha}_K, N-1) \end{bmatrix},$$

een $N \times K$ matrix van volle rang. Vgl. (0.2) wordt dan

$$\|\mathbf{y} - \mathbf{\Gamma c}\|^2, \tag{0.5}$$

met $\mathbf{y} = [y_0, \ldots, y_{N-1}]^T$.

De kostfunctie in Vgl. (0.5) kan herschreven worden als volgt

$$\|\mathbf{y} - \mathbf{\Gamma \Gamma}^\dagger \mathbf{y}\|^2, \tag{0.6}$$

met $\mathbf{\Gamma}^\dagger$ de pseudo-inverse van $\mathbf{\Gamma}$, d.w.z. $\mathbf{\Gamma}^\dagger = (\mathbf{\Gamma}^H \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^H$. Vgl. (0.6) definieert de variabele projectie functionaal. De lineaire parameters zijn geëlimineerd uit de kostfunctie en men bekomt aldus een optimalisatieprobleem in minder variabelen. De schattingen $\hat{\boldsymbol{\alpha}}_k$ worden bekomen als de parameters die Vgl. (0.6) minimaliseren. Gebruik makende van $\hat{\boldsymbol{\alpha}}_k$, wordt de matrix $\widehat{\mathbf{\Gamma}}$ berekend en de schattingen $\hat{c}_k$ van de complexe amplitudes worden bekomen als de kleinste kwadraten oplossing van

$$\widehat{\mathbf{\Gamma}} \mathbf{c} \approx \mathbf{y},$$

d.w.z. $\hat{\mathbf{c}} = \widehat{\mathbf{\Gamma}}^\dagger \mathbf{y}$. De parameters die op deze manier worden berekend zijn dezelfde als die bekomen door Vgl. (0.5) rechtstreeks te minimaliseren.

Deze methoden zijn zeer flexibel. Alle types van modelfuncties kunnen gebruikt worden en de aanwezige voorkennis kan heel gemakkelijk in rekening gebracht worden. Daarenboven worden grootste waarschijnlijkheidsschattingen bekomen indien de veronderstellingen aangaande de modelfunctie en de ruisverdeling vervuld zijn. De belangrijkste nadelen van deze methoden zijn de vereiste interactie met de gebruiker en het feit dat er geen methoden zijn die convergentie naar een globaal minimum garanderen in een redelijke tijd. VARPRO is een veel gebruikte interactieve methode. Zoals de naam suggereert wordt in VARPRO de variabele projectie functionaal, die gedefinieerd is in Vgl. (0.6), geminimaliseerd.
Zwarte-doos methoden vragen minimale gebruikersinteractie en expertise. De modelfunctie is echter beperkt tot een uniform bemonsterde, som van complex gedempte exponentiëlen. Een ander nadeel is het feit dat voorkennis slechts in beperkte mate kan ingebouwd worden. Lineaire prediktie en toestandsruimte gebaseerde methoden zijn voorbeelden van zwarte-doos methoden die de parameters schatten zonder een kostfunctie te minimaliseren. HSVD en HTLS zijn gekende voorbeelden van zwarte-doos methoden. Deze algoritmen lossen dan

ook het minimalisatieprobleem benaderend op en bijgevolg hebben de parameterschattingen geen grootste waarschijnlijkheidseigenschappen. In dit hoofstuk wordt aangetoond hoe de lineaire prediktie en toestandsruimte gebaseerde probleemstellingen kunnen opgelost worden zonder benaderingen in te voeren. Op die manier herleidt het bepalen van de parameters zich opnieuw tot het oplossen van een minimalisatieprobleem.

Verder wordt ook aangetoond dat modelfitting in het tijds- en frequentiedomein equivalent zijn. Bovendien wordt ook gewezen op het feit dat frequentiedomein technieken uit de controle- en identificatie literatuur mits minimale aanpassingen gebruikt kunnen worden bij de analyse van MRS signalen.

Hoofdstuk 3: VARPRO en AMARES

Het hoofddoel van dit hoofdstuk bestaat erin uit te leggen hoe, uitgaande van VARPRO, een nieuwe, verbeterde variant AMARES is ontwikkeld.

Eerst wordt een overzicht gegeven van de achterliggende theorie betreffende de minimalisatie van een kostfunctie gebruik makend van optimalisatiealgoritmen ontwikkeld voor het oplossen van niet-lineaire kleinste kwadraten problemen zonder beperkingen. Vervolgens wordt in detail beschreven hoe men, gebruik makend van deze algoritmen de kostfuncties in Vgl. (0.5) en Vgl. (0.6) kan minimaliseren. VARPRO minimaliseert Vgl. (0.6) gebruik makend van een eenvoudig Levenberg-Marquardt algoritme.

In dit hoofdstuk wordt een uitgebreide vergelijkende studie uitgevoerd. Er wordt nagegaan of er een verschil is tussen minimalisatie van Vgl. (0.5) en Vgl. (0.6) qua efficiëntie en robuustheid. Verder wordt onderzocht of er een verschil is in performantie naargelang het optimalisatiealgoritme dat gebruikt wordt. Uit deze vergelijkende studie blijkt dat minimalisatie van Vgl. (0.5) gebruik makende van het gesofisticeerd hybride algoritme NL2SOL leidt tot de beste performantie. In AMARES wordt dan ook voor deze aanpak gekozen.

De mogelijkheden tot het inbouwen van voorkennis, het gebruik van modelfuncties (Lorentz, Gauss) en het type signaal (FID of echo) in AMARES werden ook uitgebreid ten opzichte van VARPRO. De uitgebreide mogelijkheden om voorkennis in rekening te brengen worden geïllustreerd aan de hand van een *in-vivo* ^{1}H signaal van de prostaat.

Hoofdstuk 4: Kwantificeren van reeksen MRS signalen

Het kwantificeren van individuele signalen in het tijdsdomein is mogelijk door gebruik te maken van interactieve of zwarte-doos methoden. In biochemische studies worden echter dikwijls MRS signalen opeenvolgend in de tijd opgemeten om veranderingen in concentraties van metabolieten op te volgen. Dikwijls is er biochemische informatie aanwezig over de tijdsevolutie van de parameters. Een tijdsreeks kan als volgt mathematisch geformuleerd worden. Stel dat er S signalen zijn in de reeks, waarbij elk signaal afzonderlijk gemodelleerd kan worden door Vgl. (0.1). Het $n$de monster van signaal s wordt dan

$$y_{ns} = \sum_{k=1}^{K_s} a_{ks} e^{j\phi_{ks}} e^{(-d_{ks}+j2\pi f_{ks})t_n} + e_{ns}. \tag{0.7}$$

Het bekomen van grootste waarschijnlijkheidsschattingen komt vervolgens neer op het minimaliseren van de kostfunctie

$$\sum_{s=1}^{S} \sum_{n=0}^{N-1} |y_{ns} - \sum_{k=1}^{K_s} a_{ks} e^{j\phi_{ks}} e^{(-d_{ks}+j2\pi f_{ks})t_n}|^2. \tag{0.8}$$

Voorkennis tussen parameters behorend tot eenzelfde of verschillende signalen kan eenvoudig ingebracht worden. In dit hoofdstuk wordt aangetoond dat het in rekening brengen van deze tijdsinformatie leidt tot meer nauwkeurige resultaten. Zowel de interactieve als de zwarte-doos methoden kunnen uitgebreid worden tot de gelijktijdige verwerking van meerdere signalen. AMARES en HTLS worden gekozen als representatieve voorbeelden van de interactieve en zwarte-doos methoden en er wordt aangetoond hoe ze uitgebreid kunnen worden tot de analyse van meerdere signalen.

AMARES$_{ts}$ is een extensie van AMARES en minimaliseert Vgl. (0.8) rekening houdend met de beschikbare voorkennis. Een suboptimaal alternatief is AMARES$_{sts}$ dat kan gebruikt worden indien de voorkennis louter bestaat uit parameters die *constant* blijven in de tijd. Het gebruik van AMARES$_{ts}$ leidt tot betere en robuustere parameterschattingen dan de schattingen bekomen door de analyse van ieder signaal afzonderlijk met AMARES zonder gebruik te maken van de gemeenschappelijke informatie in de spectra. In de gevallen waar AMARES$_{sts}$ kan gebruikt worden, is de performantie van de methode beter dan die van AMARES en slechts lichtjes slechter dan die van AMARES$_{ts}$. HTLSsum en HTLSstack zijn uitbreidingen van HTLS om meerdere signalen tegelijk te analyseren. Alhoewel HTLSsum en HTLSstack slechter presteren qua nauwkeurigheid dan AMARES$_{ts}$, zijn het veelbelovende technieken om de analyse van MRS signalen verder te automatiseren.

Hoofdstuk 5: Parameterschatting in niet-ideale omstandigheden

In vele gevallen voldoen de biomedische MRS data niet aan de theoretische modelfuncties en/of bevatten ze storende componenten waarvan weinig bekend is. Dit hoofdstuk start met een overzicht van de methoden die bestaan om modelimperfecties te compenseren. De meeste van deze methoden bepalen de distorties aan de hand van een vergelijking tussen een referentiepiek en een piek met een ideale lijnvorm, waarna deze schatting van de distorties gebruikt wordt om het opgemeten signaal te corrigeren. Verder wordt een overzicht gegeven van bestaande technieken om de parameters van geselecteerde pieken in de aanwezigheid van ongewenste of ongekende componenten te bepalen. De meeste aandacht gaat naar de schatting van parameters in de aanwezigheid van ongewenste componenten met een frequentie verschillend van die van de gewenste metabolieten. Naar dit probleem wordt verwezen als frequentie-selectieve (FS) parameterschatting. AMARES$_w$, een methode die gebruik maakt van een gewichtsfunctie in het tijdsdomein in combinatie met AMARES, wordt kort uitgelegd. Ook het gebruik van het fitten van een welbepaald frequentiegebied met toevoeging van een polynoom ter modellering van de basislijn wordt besproken. Naar deze methode wordt verwezen als AMARES$_{FREQ}$. Met behulp van zwarte-doos methoden zoals HSVD kan een goede reconstructie van de oninteressante pieken verkregen worden. Deze kunnen vervolgens afgetrokken worden van het oorspronkelijk signaal waarna het resterend signaal kan gekwantificeerd worden met behulp van AMARES. Deze totale procedure krijgt de naam AMARES$_H$.

Hoofdstuk 6: Frequentie-selectieve parameterschatting

Een nieuwe techniek gebaseerd op het gebruik van "Finite Impulse Response" (FIR) filters met maximale fase wordt geïntroduceerd als alternatief om FS parameterschatting uit te voeren. Een FIR filter wordt gedefinieerd door de volgende modelfunctie

$$y_{f_n} = \sum_{m=0}^{M-1} h_m y_{n-m} \qquad n = 0, \ldots, N-1, \tag{0.9}$$

met $h_m, m = 0, \ldots, M-1$, de filtercoëfficiënten. Een probleem met de definitie van Vgl. (0.9) is het feit dat de monsters y_n voor $n < 0$ niet beschikbaar zijn. Dit wordt meestal opgelost door te veronderstellen dat het signaal cyclisch is of nul buiten het tijdsvenster. Beide veronderstellingen leiden tot distorties in de eerste $M-1$ monsters en bijgevolg moeten deze weggelaten worden. Wanneer

de FIR filter gedefinieerd in Vgl. (0.9) toegepast wordt op de modelfunctie van Vgl. (0.1) en de eerste $M-1$ monsters van het aldus bekomen signaal worden weggelaten verkrijgt men

$$\begin{aligned}\bar{y}_{f_n} &= \sum_{m=0}^{M-1} h_m \bar{y}_{n-m+M-1} \\ &= \sum_{k=1}^{K} \mathbf{h}^T \mathbf{b}_k \left(a_k e^{j\phi_k} e^{(-d_k + j2\pi f_k)t_n} \right) \qquad n = 0, \ldots, N-M,\end{aligned} \tag{0.10}$$

met

$$\mathbf{h} = \left[\begin{array}{ccc} h_{M-1} & \cdots & h_0 \end{array} \right]^T$$

en

$$\mathbf{b}_k = \left[\begin{array}{cccc} 1 & e^{(-d_k + j2\pi f_k)\Delta t} & \ldots & e^{(-d_k + j2\pi f_k)(M-1)\Delta t} \end{array} \right]^T .$$

Uit Vgl. (0.10) blijkt duidelijk dat de ongewenste componenten kunnen verwijderd worden door $|\mathbf{h}^T \mathbf{b}_k| \approx 0$ te kiezen. Men zorgt er verder voor dat voor de metabolieten waarin men geïnteresseerd is geldt dat $|\mathbf{h}^T \mathbf{b}_k| \approx 1$. Door het gebruik van een FIR filter met maximale fase wordt ervoor gezorgd dat de informatie-inhoud van de $M-1$ monsters die weggelaten worden zo klein mogelijk is. Verder wordt er een gedetailleerde uitleg gegeven over het ontwerp van deze FIR filter en het gebruik van de filter in combinatie met de parameterschattingsmethoden. Het gebruik van de FIR filter in combinatie met de parameterschattingsmethode AMARES wordt aangeduid met AMARES$_f$. Een schema voor het automatisch ontwerpen van de gewenste FIR filter wordt ontwikkeld.

De performantie van de FIR filter methode wordt vergeleken met reeds bestaande FS methoden. Daartoe worden de eigenschappen van AMARES$_f$, AMARES$_H$, AMARES$_w$, AMARES$_{FREQ}$ nauwkeurig onderzocht en met elkaar vergeleken aan de hand van simulaties.

AMARES$_w$, de methode gebaseerd op weging in het tijdsdomein, kan gebruikt worden om de statistische vertekening te verkleinen indien de pieken waarvan men de parameters wil bepalen behoorlijk gescheiden zijn van de ongewenste pieken. De prijs die men betaalt voor het gebruik van de methode is een verlaagde signaal-ruis verhouding waardoor de statistische variantie stijgt. Bovendien werkt de methode niet wanneer de gewenste pieken te dicht bij de ongewenste pieken liggen.

AMARES$_H$ is een uitstekende methode wat betreft nauwkeurigheid indien de gepaste keuze van de modelorde wordt gemaakt. Een nadeel echter is de gevoeligheid van de methode aan de gebruikte modelorde in moeilijke omstandigheden.

De analyse van een deel van het spectrum is zeer eenvoudig te implementeren in fittingsmethoden zoals AMARES$_{FREQ}$ die werken in het frequentiedomein. De situatie wordt echter bemoeilijkt door de invloed in het geselecteerd gebied van oninteressante pieken gelegen buiten dit gebied. Deze invloed moet in rekening gebracht worden, b.v. door het modelleren van de staarten van deze pieken door een polynoom van een zekere orde. Indien de polynoom de invloed van de oninteressante pieken slecht modelleert, wordt er een vertekening in de parameters geïntroduceerd. De keuze van het gebied waarin men de kostfunctie minimaliseert en de orde van de polynoom is in moeilijke gevallen zeer belangrijk.

Indien de onder- en bovengrens van het frequentiegebied dat onderdrukt moet worden goed gekozen zijn, levert het gebruik van AMARES$_f$ zeer goede resultaten op; de resulterende vertekening en variantie zijn laag. Het gebruik van het automatisch filterontwerp leidt tot een hoge gebruiksvriendelijkheid en reproduceerbaarheid van de resultaten. Het gebruik van AMARES$_f$ brengt bovendien een lage rekenkost met zich mee.

Hoofdstuk 7: Efficiënte verwijdering van resterend water in proton spectra

In proton spectra is steeds resterend water aanwezig, dat niet door een analytische functie kan worden beschreven. In hoofdstuk 5 werd uitgelegd hoe een zwarte-doos methode zoals HSVD kan gebruikt worden om het water uit de spectra te verwijderen. Het grote nadeel van HSVD is echter de hoge rekenkost die gepaard gaat met de berekening van de singuliere waarden ontbinding (SWO). In dit hoofdstuk wordt de SWO vervangen door een efficiëntere lage rang revelerende ontbinding om de signaalruimte te bepalen. De nieuwe methode krijgt de naam HLR. De winst in rekentijd door gebruik te maken van HLR in plaats van HSVD om het resterend water te verwijderen is aanzienlijk. De nauwkeurigheid van de uiteindelijke parameterschattingen bekomen na toepassing van HSVD en HLR gevolgd door fitting van het resterend signaal met AMARES is dezelfde. De winst in efficiëntie hangt af van het aantal datapunten in het signaal en de opgelegde modelorde en is groter naargelang het aantal datapunten groter is en de modelorde kleiner. Voor een signaal bestaande uit 512 datapunten en een modelorde van 10 vraagt HLR 35 tot 40 keer minder CPU tijd dan HSVD.

Hoofdstuk 8: Toepassing van FS parameterschatting in proton spectra

Aangezien de verwijdering van resterend water een erg belangrijk toepassingsgebied is, wordt in dit hoofdstuk een verdere studie gemaakt van de HSVD en de FIR filter methoden voor deze specifieke toepassing. Het automatisch FIR filterontwerp van hoofdstuk 6 wordt verfijnd om de procedure nog rekenefficiënter te maken. De performantie van beide methoden wordt vergeleken als functie van de frequenties van de metabolieten. De resultaten tonen aan dat de FIR filter methode in combinatie met het automatisch filterontwerp, de HSVD methode in de meeste gevallen overtreft qua performantie. Het gebruik van het automatisch filterontwerp waarborgt nauwkeurige parameterschattingen en een hoge reproduceerbaarheid van de resultaten. De HSVD methode is gevoelig aan de keuze van de modelorde. De FIR filter methode is bovendien minstens een grootteorde efficiënter dan HLR, de snelle versie van HSVD. De goede performantie van de FIR filter methode gecombineerd met de lage rekenkost motiveert het gebruik van AMARES$_f$ als alternatief voor AMARES$_H$.

Hoofdstuk 9: Besluit en nieuwe uitdagingen

In deze thesis worden verschillende aspecten van de kwantificatie van MRS signalen behandeld. De belangrijkste besluiten en eigen bijdragen worden hieronder samengevat.

- De verschillende methoden die bestaan om de parameters van een MRS signaal te bepalen werden uitgelegd en verbanden tussen methoden werden gelegd.
- AMARES (*a*dvanced *m*ethod for *a*ccurate, *r*obust and *e*fficient *s*pectral fitting) werd ontwikkeld. Er werd aangetoond dat AMARES het populaire VARPRO op twee manieren overtreft. Eerst en vooral werd een gemotiveerde keuze van de te minimaliseren functionaal en het te gebruiken niet-lineaire kleinste kwadraten algoritme gemaakt op basis van een uitgebreide simulatiestudie. Ten tweede laat AMARES het gebruik van meer vormen van voorkennis, modelfuncties en types van signaal (FID of echo) toe. Dit resulteert in een algoritme met verhoogde nauwkeurigheid, robuustheid en flexibiliteit.
- In biochemische studies worden MRS signalen vaak opeenvolgend in de tijd opgemeten om veranderingen in metabolietconcentraties in de tijd op te volgen. Dikwijls is er informatie aanwezig over de tijdsevolutie van sommige van de parameters in de tijdsreeks. In deze thesis werd

AMARES$_{ts}$, een uitbreiding van AMARES, voorgesteld. AMARES$_{ts}$ laat toe om de gemeenschappelijke informatie aanwezig in de spectra op een statistisch optimale manier in te bouwen. De methode doet het erg goed in praktische omstandigheden. De methode werd vergeleken met een suboptimale aanpak AMARES$_{sts}$. Verder werd AMARES$_{ts}$ vergeleken met het gebruik van HTLSsum en HTLSstack, uitbreidingen van HTLS om meerdere signalen tegelijk te verwerken.

- In deze thesis werden verschillende technieken vergeleken die in het verleden gebruikt werden om de parameters van een aantal pieken te bepalen in aanwezigheid van ongekende of oninteressante spectrale componenten, die in frequentie verschillen van de gewenste metabolieten. Dit probleem wordt aangeduid met de term frequentie-selectieve (FS) parameterschatting. Daar er tot op heden geen systematische studie werd gemaakt van de uiteindelijke nauwkeurigheid van de parameters werden de eigenschappen van weging in het tijdsdomein AMARES$_w$, HSVD filtering AMARES$_H$ en fitting in het frequentiedomein gebruik makend van een polynoom als basislijn AMARES$_{FREQ}$ onderzocht. De problemen verbonden met de hierboven vermelde methoden werden afgelijnd. AMARES$_H$ vertoont de beste performantie onder deze FS methoden. De rekenkost van de methode is echter vrij hoog en dit is een ernstig nadeel in toepassingen zoals MRSI.

- In de context van FS parameterschatting werd een nieuwe techniek, AMARES$_f$ voorgesteld die gebaseerd is op het gebruik van "finite impulse response" (FIR) filters met maximale fase in combinatie met AMARES. AMARES$_f$ werd vergeleken met de andere FS technieken en uit de vergelijkende studie kwam AMARES$_f$ naar voren als de methode met de hoogste nauwkeurigheid en gebruiksvriendelijkheid.

- HLR, een snelle variant van HSVD, gebaseerd op het vervangen van de singuliere waarden ontbinding door een lage rang revelerende ontbinding, werd ontwikkeld voor het verwijderen van resterend water in proton spectra. Deze snelle methode werd eveneens vergeleken met AMARES$_f$ in deze specifieke toepassing. AMARES$_f$ doet het nog steeds beter dan deze (snelle) HSVD-gebaseerde methode qua efficiëntie, nauwkeurigheid en gebruiksvriendelijkheid.

Niettegenstaande het feit dat er al een lange weg afgelegd is, dienen er zich nog steeds nieuwe uitdagingen aan op het gebied van de analyse van MRS signalen.

- De foutenschattingen op de parameters worden nog steeds berekend door de Cramér-Rao grenzen te evalueren in de uiteindelijk bekomen parameterschattingen. Het is niet duidelijk hoe deze foutenschattingen de echte fouten weerspiegelen. Een gecontrolleerde experimentele studie is nodig om dat te onderzoeken.

- In hoofdstuk 4 werd aangetoond dat het gebruik van AMARES$_{ts}$ kan leiden tot een aanzienlijke winst in nauwkeurigheid indien de signalen van de tijdsreeks samen verwerkt worden en de gemeenschappelijke informatie in rekening wordt gebracht. De benodigde rekentijd en het computergeheugen neemt echter drastisch toe wanneer gebruik gemaakt wordt van klassieke optimalisatietechnieken. Alternatieve optimalisatietechnieken moeten in deze context onderzocht worden.

- Tot nu toe is AMARES$_{ts}$ enkel toegepast op tijdsreeksen. De mogelijke voordelen van de methode in MRSI moeten onderzocht worden.

- In hoofdstuk 5 wordt een overzicht gegeven van methoden die in de loop der jaren werden ontwikkeld om modelimperfecties op te vangen. In de meeste gevallen werd de evaluatie van de methoden louter visueel uitgevoerd. Een systematische studie waarin gekeken wordt naar de uiteindelijke nauwkeurigheid van de parameters is wenselijk.

- Deze thesis richtte zich vooral op FS parameterschatting. Een moeilijker probleem bestaat echter uit die gevallen waarbij de ongewenste spectrale componenten noch in frequentie noch in demping gescheiden zijn van de gewenste parameters. Dit is vooral van belang bij de analyse van korte echotijd signalen in protonspectroscopie.

- VARPRO en AMARES zijn methoden die relatief veel interactie met de gebruiker vragen. Aangezien de hoeveelheid gegevens almaar toeneemt is er een nood aan verdere automatisering.

- Deze thesis behandelt uitsluitend signalen afkomstig uit 1D experimenten. Dit komt neer op het feit dat de signalen gemodelleerd worden als een som van Lorentzianen, Gaussianen, ... In 2D experimenten bestaat de modelfunctie echter uit producten van sinusoïden. Vele van de methoden die ontwikkeld werden in deze thesis kunnen uitgebreid worden tot de analyse van zulke signalen.

Chapter 1

Introduction

The purpose of this chapter is twofold:

1. *To provide the necessary background information on different aspects of magnetic resonance spectroscopy (MRS), including the basic vocabulary of the MRS jargon used further in this text.*

2. *To introduce and define the problems dealt with in this thesis.*

Section 1.1 situates in- and ex-vivo MRS, on which the emphasis of this work lies, among other nuclear magnetic resonance (NMR) techniques. In section 1.2 the basic physical principles of a MRS experiment are outlined. The mathematical model functions that are used to describe the MRS signal are given in section 1.3. In section 1.4 it is explained how the signals are visualized in the frequency domain and which conventions are used in MRS. Section 1.5 deals with the important concept of prior knowledge that is very often used in the remainder of this text. Some examples of signals often encountered in in- and ex-vivo MRS are given in section 1.6. The goals of this thesis are defined in section 1.7. The Monte-Carlo procedure used to evaluate the performance of different mathematical tools is described in section 1.8. Finally, a chapter-by-chapter overview is given in section 1.9.

More detailed background information can be found in [54, 35] and references therein.

1.1 Situation

Nuclear magnetic resonance (NMR) has become a routinely used method to obtain spatial images in living organisms or to study metabolic changes in a non-invasive way. When applied to living systems NMR can be divided in two classes, magnetic resonance imaging (MRI) and magnetic resonance spectroscopy (MRS).

MRI visualizes the spatial distribution of (mobile) protons. MRS can be used to obtain signals related to the presence of other chemical substances (metabolites). Numerous spatial localization techniques have been developed to measure signals from locally well-defined regions. Two broad classes of methods can be distinguished. On the one hand, there is single-voxel localization in which a signal is acquired from one spatial location, a voxel. Using multi-voxel techniques, signals from multiple spatial regions are acquired simultaneously in one measurement sequence. This application of MRS is denoted by MRS imaging (MRSI) or chemical shift imaging (CSI). Using MRSI a signal containing information of an array of voxels (typically 32×32) is obtained in one measurement, but the spatial definition of the voxel from which the signal was acquired is worse compared to single-voxel techniques. MRS signals measured in animals or human beings are called *in-vivo* signals. They are characterized by a low signal-to-noise ratio (SNR) due to the low concentrations of the metabolites and the limited measurement time and by overlapping spectral components due to the low magnetic field strength used. *Ex-vivo* signals denote signals typically obtained from an organ taken out of its natural context but kept alive artificially using perfusion. These signals can be obtained using higher field strengths and are usually of better quality. MRS signals obtained from tissue extracts, body fluids, ... , are called *in-vitro* signals. In this particular application, there are no limitations on the measurement time and higher magnetic field strengths can be used to increase the SNR and the spectral resolution. Since these extracts, body fluids, cells and so on are dissolved in water and measured in small sample tubes, the resolution is at least an order of magnitude better since the linewidth is usually only limited by the high resolution shim and no longer by local field inhomogeneities arising from differences in susceptibility between neighboring tissues, structures and interfaces in the in-vivo and in-situ environment. As a result these signals have a high SNR and are typically composed of a lot of spectral components.

In this study techniques for the accurate analysis of MRS signals are presented. The emphasis lies on the treatment of *in-vivo* and *ex-vivo* signals characterized by low SNR and overlapping peaks. To that end, the signals are first modeled by an appropriate model function and mathematical techniques are subsequently applied to determine the model parameters.

1.2 Basic principles of MRS

Certain atomic nuclei, like the ones of ^{1}H and ^{31}P, possess an inherent angular momentum or spin. A magnetic property is associated with this spin and when these nuclei are subject to a static magnetic field B_0, the laws of quantum mechanics come into play. ^{1}H and ^{31}P both have a spin quantum number of 1/2, corresponding to two different spin or energy states and the energy difference between both states is proportional to the applied magnetic field

$$\Delta E = \gamma \hbar B_0, \quad \hbar = h/(2\pi), \tag{1.1}$$

with γ the gyromagnetic ratio, characteristic of the isotope and h the Planck constant. To induce transitions between these energy states, an oscillating magnetic field is applied in a plane perpendicular to the static field B_0. This field has a frequency equal to the precession frequency of the nuclei

$$f_0 = \gamma B_0/(2\pi).$$

The amount of energy absorbed by the nuclei (and emitted in a later stage) depends on the population difference between the two energy states. The population density of the nuclei in the different energy states is determined by their Boltzmann distribution. In thermal equilibrium at absolute temperature T, the relative amounts n^- of the nuclei in the highest and n^+ of nuclei in the lowest energy state are given by

$$n^-/n^+ = e^{(-\Delta E/kT)}, \tag{1.2}$$

where k is the Boltzmann constant. The ratio n^-/n^+ is very close to unity in normal circumstances, resulting in low absorption (and as a consequence emission) of energy. This explains the inherently low SNR of the signal emitted when the spin system returns to equilibrium after excitation. This emitted signal is called a free induction decay (FID) signal.

The intensity of the signal is proportional to the number of nuclei that contributes to it. In order to obtain a reasonable SNR, a number of data acquisitions is performed consecutively and the final signal is the sum of all measured signals. Eqs. (1.1) and (1.2) show that increasing the field strength B_0 or lowering the temperature T results in a higher energy difference between both states, resulting in an increased amount of emitted energy and as a result higher SNR signals. At the same field strength and temperature, the sensitivity of the ^{1}H nucleus is higher than that of the ^{31}P one (larger γ).

Nuclei of different atoms can easily be distinguished from one another since they have a different γ and consequently resonate at other frequencies. Nuclei of the same isotope, but belonging to different molecules are differentiable due to the effect of shielding. The effective magnetic field experienced by a nucleus

is

$$B_{eff} = B_0(1 - \sigma'),$$

where σ' is the shielding constant, a dimensionless unit that depends on the electrical environment of a nucleus. Therefore, nuclei in a different chemical environment emit signals with different frequencies. This phenomenon makes MRS a very attractive tool since it allows to differentiate between molecular structures. Spin-spin (or J-) coupling, the interactions between neighboring spins in the same molecule, induces a further differentiation in resonance frequencies.

The FID signal approaches exponential decay if the molecules are sufficiently mobile such that interactions are averaged out. However, the local field inhomogeneities introduced by variations in tissue homogeneity and operator shim introduce departure from this exponential decay.

In its simplest form a MRS experiment thus consists of applying a radio frequency (RF) pulse with a bandwidth chosen to excite all the nuclei within the required frequency range and acquiring the response, the FID signal. The process is repeated until the desired SNR is obtained. Many special pulse sequences with or without the combined use of field gradients have been developed in order to enhance the quality and information content of spectra. The use of these specialized pulse sequences often leads to the formation of a so-called echo signal. In most modern systems, a RF coil is used to detect the signals. In the detector the frequencies of the signal are transformed to the audiofrequency range by subtracting a reference frequency.

1.3 Mathematical description of the MRS signal

1.3.1 FID signal

The model function most often used to represent the digitized FID signal is the following

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-d_k + j2\pi f_k)t_n} + e_n \quad n = 0, 1, \ldots, N-1, \qquad (1.3)$$

K represents the number of different resonances, $j = \sqrt{-1}$. The amplitude a_k is related to the concentration of the metabolite and the frequency f_k characterizes it. The damping d_k provides, among others, information about its mobility and molecular environment. Often the acquisition of the signal is only

started after a time delay t_0, the receiver dead time. Therefore, the sampling time instances t_n can be written as $t_n = n\Delta t + t_0$ with Δt the sampling interval. The quantity $2\pi f_k t_0$ is also called the first-order phase. The phase ϕ_k consists of a so-called zero-order phase ϕ_0 and an additional phase factor ϕ'_k which represents extra degrees of freedom that may be required under certain experimental conditions (but usually all ϕ'_k are zero). The noise term e_n is assumed to be circular[1] complex white Gaussian noise. It mainly consists of thermal noise from the sample and noise from the electronic components.

An alternative model function often used is the Gaussian model

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-g_k t_n + j2\pi f_k)t_n} + e_n \quad n = 0, 1, \ldots, N-1. \quad (1.4)$$

Recently, the Voigt model [84, 114] which combines the Lorentz and Gauss model, has received some attention

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-d_k - g_k t_n + j2\pi f_k)t_n} + e_n \quad n = 0, 1, \ldots, N-1. \quad (1.5)$$

1.3.2 Echo signal

An echo signal consists of a sum of sinusoids whose amplitudes grow at first, then reach a maximum at a certain point of time and subsequently decay into the noise. Assuming exponential growth, the model function of the left part of the echo consisting of N_l data points is given by (see also Fig. 1.1)

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-dl_k - j2\pi f_k)t_n} + e_n \quad n = 0, 1, \ldots, N_l - 1, \quad (1.6)$$

with $t_n = (N_l - n)\Delta - \delta$, $0 \leq \delta \leq \Delta$ and dl_k represents the damping of the kth sinusoid of the left part. The variable δ is here used since the top of an echo does not necessarily coincide with a data point. The corresponding model for the right part consisting of N_r data points is given by (see also Fig. 1.1)

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-dr_k + j2\pi f_k)t_n} + e_n \quad n = 0, 1, \ldots, N_r - 1, \quad (1.7)$$

with $t_n = n\Delta + \delta$, $0 \leq \delta \leq \Delta$ and dr_k represents the damping of the kth sinusoid of the right part. As is evident from a comparison between Eq. (1.3)

[1] The term circular means that the real and imaginary parts of the noise sequence are not correlated and have equal variance.

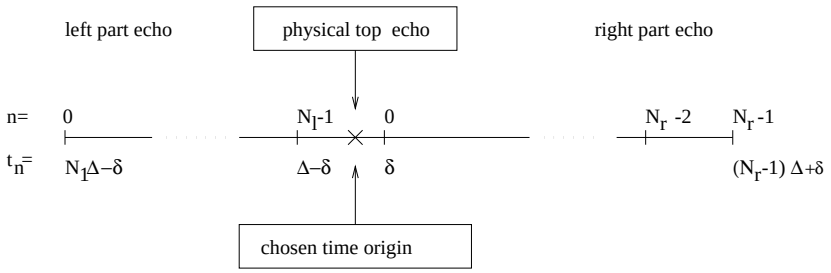


Figure 1.1: Time axis used in the representation of the model function of an echo signal. The time origin $t = 0$ coincides with the physical top of the echo. The time axis belonging to the left and the right part of the echo is taken positive.

and Eq. (1.7), the right part of the echo has exactly the same model function as the FID (the variable δ of Eq. (1.7) being equivalent to t_0 of Eq. (1.3)). In many cases, the left part of the echo is discarded and only the right part is retained for further processing since most software packages do not offer the possibility to process entire echo signals.

1.4 Visualizing the NMR signals

To visually interpret the measured MRS signal, the data are transformed to the frequency domain by the discrete fourier transform (DFT).

DFT transformation of $\bar{y}_n,\ n = 0, \ldots, N-1$, results in the sequence $\overline{y}_l,\ l = 0, \ldots, N-1$, which has the following analytical expression provided Eq. (1.3) is used to model the data,

$$\begin{aligned} \overline{y}_l &= \sum_{k=1}^{K} a_k e^{-d_k t_0 + j(2\pi f_k t_0 + \phi_k)} \frac{1 - e^{(-d_k + j2\pi(f_k - \nu_l))N\Delta t}}{1 - e^{(-d_k + j2\pi(f_k - \nu_l))\Delta t}} \qquad (1.8) \\ \nu_l &= l/(N\Delta t), \quad l = 0, \ldots, N-1. \end{aligned}$$

Since the DFT is periodic, one can as well evaluate Eq. (1.8) for other values of l. The convention in MRS is to evaluate Eq. (1.8) for $l = -\frac{N}{2}, \ldots, \frac{N}{2} - 1$ and to display the spectra from positive to negative frequency. After DFT each exponentially damped sinusoid gives rise to a peak, called a Lorentzian, in the obtained frequency spectrum.

Consider a signal consisting of two exponentially damped sinusoids ($K = 2$) with $\phi_1 = \phi_2 = \phi_0$.

If t_0 and ϕ_0 equal to zero, Eq. (1.8) can then be rewritten as

$$\overline{\mathcal{Y}}_l = AB_{1_l} + AB_{2_l} + jDI_{1_l} + jDI_{2_l} \quad (t_0 = 0, \phi_0 = 0),$$

where AB_{i_l} and DI_{i_l}, $i = 1, 2$, are called the absorption and dispersion components of the ith Lorentzian.

If $\phi_0 \neq 0$ and $t_0 = 0$, a mixture of the absorption and dispersion components is observed in the real and imaginary components of the DFT transformed signal

$$\begin{aligned} Re(\overline{\mathcal{Y}}_l) &= AB_{1_l}\cos\phi_0 + AB_{2_l}\cos\phi_0 - DI_{1_l}\sin\phi_0 - DI_{2_l}\sin\phi_0 \\ Im(\overline{\mathcal{Y}}_l) &= AB_{1_l}\sin\phi_0 + AB_{2_l}\sin\phi_0 + DI_{1_l}\cos\phi_0 + DI_{2_l}\cos\phi_0. \end{aligned}$$

Pure absorption and dispersion components can be obtained by so-called phasing of the spectrum, i.e. multiplying the DFT spectrum point-wise by $e^{-j\phi_0'}$, or equivalently

$$\begin{aligned} Re(\overline{\mathcal{Y}}_l) &= AB_{1_l}\cos(\phi_0 - \phi_0') + AB_{2_l}\cos(\phi_0 - \phi_0') \\ &\quad - DI_{1_l}\sin(\phi_0 - \phi_0') - DI_{2_l}\sin(\phi_0 - \phi_0') \\ Im(\overline{\mathcal{Y}}_l) &= AB_{1_l}\sin(\phi_0 - \phi_0') + AB_{2_l}\sin(\phi_0 - \phi_0') \\ &\quad + DI_{1_l}\cos(\phi_0 - \phi_0') + DI_{2_l}\cos(\phi_0 - \phi_0') \end{aligned}$$

and taking $\phi_0' = \phi_0$. The above procedure is also called zero-order phase correction.

Consider now the case where $\phi_0 \neq 0, t_0 \neq 0$. The following mixture of the absorption and dispersion components is observed in the real and imaginary components of the DFT transformed signal

$$\begin{aligned} Re(\overline{\mathcal{Y}}_l) &= AB_{1_l}\cos(\phi_0 + 2\pi f_1 t_0) + AB_{2_l}\cos(\phi_0 + 2\pi f_2 t_0) \\ &\quad - DI_{1_l}\sin(\phi_0 + 2\pi f_1 t_0) - DI_{2_l}\sin(\phi_0 + 2\pi f_2 t_0) \\ Im(\overline{\mathcal{Y}}_l) &= AB_{1_l}\sin(\phi_0 + 2\pi f_1 t_0) + AB_{2_l}\sin(\phi_0 + 2\pi f_2 t_0) \\ &\quad + DI_{1_l}\cos(\phi_0 + 2\pi f_1 t_0) + DI_{2_l}\cos(\phi_0 + 2\pi f_2 t_0). \end{aligned}$$

Besides the zero-order phase correction, a first-order phase correction needs to be performed. Therefore the DFT transformed spectrum is multiplied point-wise by $e^{-j\phi_0' - j2\pi t_0'\nu_l}$, or equivalently

$$\begin{aligned} Re(\overline{\mathcal{Y}}_l) &= AB_{1_l}\cos(\phi_0 + 2\pi f_1 t_0 - \phi_0' - 2\pi t_0'\nu_l) \\ &\quad + AB_{2_l}\cos(\phi_0 + 2\pi f_2 t_0 - \phi_0' - 2\pi t_0'\nu_l) \\ &\quad - DI_{1_l}\sin(\phi_0 + 2\pi f_1 t_0 - \phi_0' - 2\pi t_0'\nu_l) \\ &\quad - DI_{2_l}\sin(\phi_0 + 2\pi f_2 t_0 - \phi_0' - 2\pi t_0'\nu_l) \\ Im(\overline{\mathcal{Y}}_l) &= AB_{1_l}\sin(\phi_0 + 2\pi f_1 t_0' - \phi_0' - 2\pi t_0'\nu_l) \end{aligned}$$

$$
\begin{aligned}
&+AB_{2_l}\sin\left(\phi_0+2\pi f_2 t_0'-\phi_0'-2\pi t_0'\nu_l\right)\\
&+DI_{1_l}\cos\left(\phi_0+2\pi f_1 t_0'-\phi_0'-2\pi t_0'\nu_l\right)\\
&+DI_{2_l}\cos\left(\phi_0+2\pi f_2 t_0'-\phi_0'-2\pi t_0'\nu_l\right)\\
\nu_l \;&=\; l/(N\Delta t), \qquad l=-\frac{N}{2},\ldots,\frac{N}{2}-1,
\end{aligned}
$$

and taking $\phi_0' = \phi_0, t_0' = t_0$.

Note that this first-order phase correction is approximate in the sense that the correction for the t_0 influence need only be applied at the center frequency of the sinusoids and not over the entire frequency range. The different cases treated above are illustrated graphically in Fig. 1.2 for a signal consisting of two sinusoids with $f_1 = 200$ Hz, $d_1 = 50$ Hz, $f_2 = -100$ Hz, $d_2 = 50$ Hz.

The proper phasing of a spectrum is most often achieved by manually and interactively adjusting the zero- and first-order phase parameters until the real part of the DFT transformed spectrum coincides with the absorption spectrum. Different ways to automate this procedure have been proposed (see e.g. [73] and references therein).

Another less commonly used way of representing MRS spectra is in absolute value or magnitude mode: $\sqrt{(Re(\overline{y}_l) + Im(\overline{y}_l)}$.

1.5 The concept of prior knowledge

In MRS applications prior knowledge concerning the spectral components is often present. This prior knowledge can in some cases be derived from quantum mechanics as illustrated hereafter using the adenosine triphophate (ATP) example. In Fig. 1.3 the molecular structure of the ATP compound is depicted. The three phosphorus atoms, designated P_α, P_β and P_γ, experience a different chemical/electronical environment and as a result of shielding, they will resonate at slightly different frequencies, giving rise to three groups of spectral lines in the spectrum. As a result of the spin-spin coupling between the neighboring phosphorus nuclei, the groups corresponding to the α and γ phosphates of the ATP are split into two peaks (a doublet) and the one of the β phosphate is decomposed into three peaks (a triplet), as shown in Fig. 1.4. From quantum mechanics, the following relations between the parameters of ATP can be derived:

- The frequency differences between the individual resonances within a multiplet are equal and known: $f_1 - f_2 = f_3 - f_4 = f_5 - f_6 = f_6 - f_7 = \Delta f$.
- The dampings of all peaks are equal: $d_1 = d_2 = \ldots = d_7$.

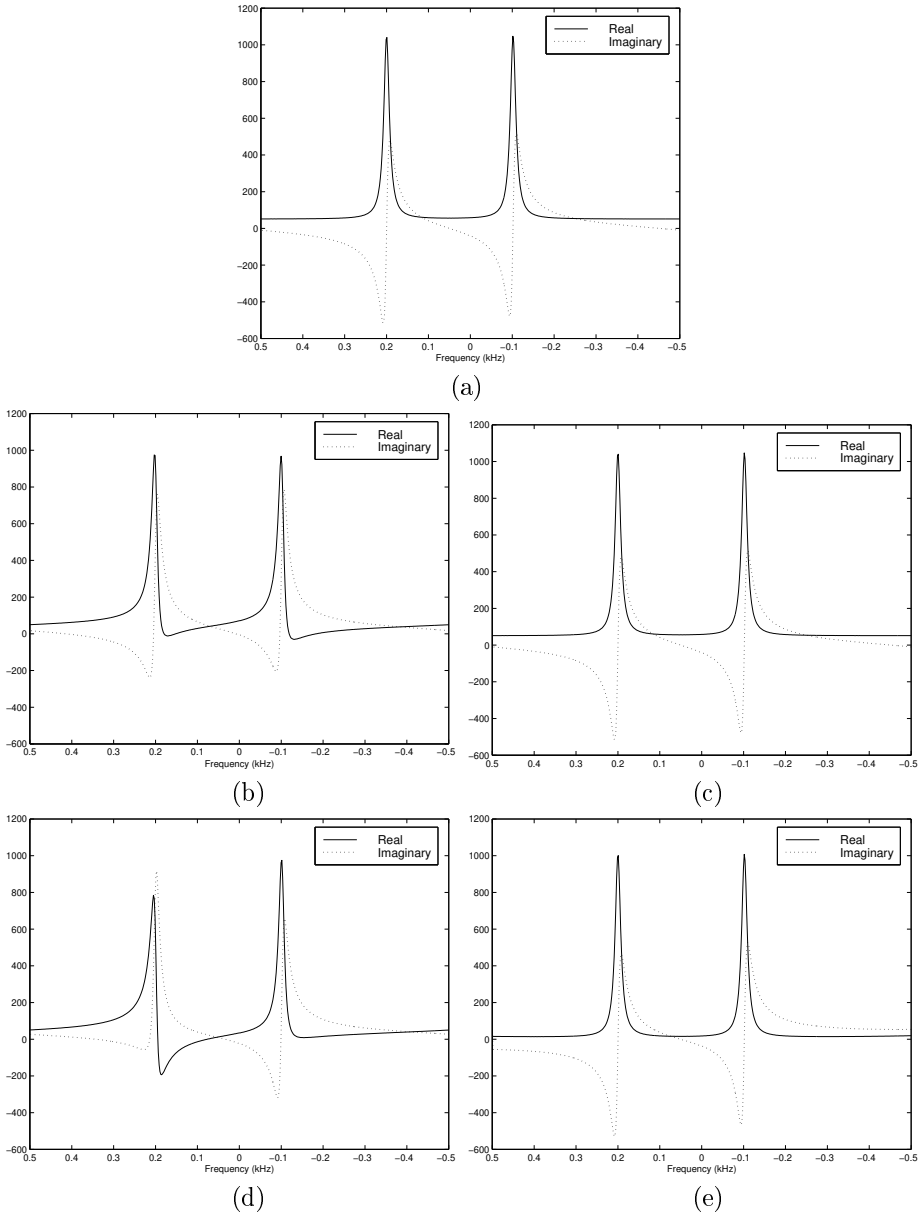


Figure 1.2: $\Delta t = 1$ ms, $d_1 = d_2 = 0.05$ kHz, $f_1 = 0.2$ kHz, $f_2 = -0.1$ kHz (a) $\phi_0 = 0, t_0 = 0$, no phase correction is necessary. (b) $\phi_0 = 30/180 * \pi$ radians, $t_0 = 0$, no phase correction was applied. (c) Same parameters as (b), but displayed after zero-order phase correction. (d) $\phi_0 = 30/180 * \pi$ radians, $t_0 = 0.4$ ms, no phase correction was applied. (e) Same parameters as (d), but displayed after zero- and first-order phase correction.

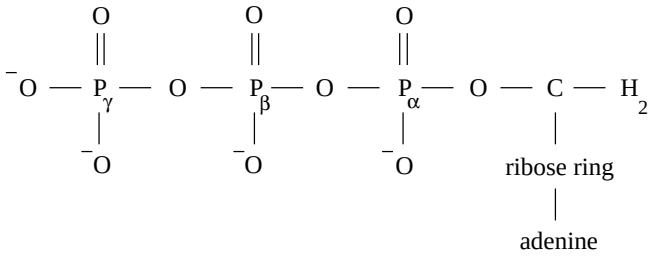


Figure 1.3: Structure of the chemical compound ATP. All three phosphorus atoms experience a different chemical environment.

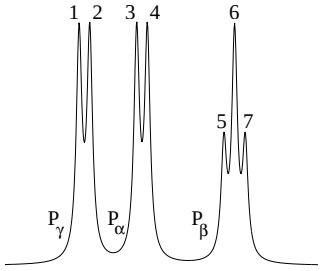


Figure 1.4: Theoretical frequency spectrum of ATP. Three groups of spectral lines can be observed corresponding to the P_α, P_β and P_γ atoms. In addition, each group is split into a multiplet of lines as a result of spin-spin coupling.

- The amplitude ratios between the multiplets are $1:1$ for P_α, $1:2:1$ for P_β and $1:1$ for P_γ. The total amplitudes of P_α, P_β and P_γ relate as $1:1:1$. In summary: $a_1 = a_2 = a_3 = a_4 = 2a_5 = a_6 = 2a_7$.

- The phases of all peaks are equal: $\phi_1 = \phi_2 = \ldots = \phi_7$.

The above prior knowledge is valid under some assumptions which must be evaluated for individual applications (for more information see [35]). The prior knowledge can also be derived from phantom spectra of pure metabolite solutions measured under the same experimental conditions as *in-vivo* spectra [90, 9]. Alternatively, the complete phantom metabolite spectra can be used as a model, circumventing the need for a mathematical model function [105]. In order to avoid the need to acquire phantom spectra for each field strength and pulse sequence used, quantum mechanical spectral simulation has been used to deduce a priori information [153, 123, 124, 125].

1.6 ^{31}P and ^{1}H MRS signals

In biological systems, the concentration of water is high compared to the metabolites of interest. The concentrations of e.g. N-acetyl-aspartate (NAA) and phosphocreatine (PCr) in the human brain are of the order of 1.5×10^{-2} and 5×10^{-3} mM per gram of tissue, respectively, compared to 55.5 mM per gram for water. In order to adequately digitize the signals from the metabolites of interest, it is common to suppress the water resonance by applying special RF pulse sequences. It is however not possible to completely suppress the water instrumentally without affecting the metabolites of interest over a relatively wide frequency range. Therefore a water resonance always remains present in the measured spectra. Compounds such as myo-inositol (mI), choline (Cho), creatine (Cr), PCr, glutamate (Glu), glutamine (Gln), NAA, lactate (Lac), citrate, glucose (Glc) are examples of important metabolites which can be observed in *in-vivo* MRS depending on the tissue under investigation and the parameters of the pulse sequence. Indeed, spectra obtained with different pulse sequence parameters can exhibit significant differences. An important parameter in that respect is the so-called echo time TE. In long echo proton spectra only resonances from NAA, Cho, Cr, PCr and residual water remain visible in the healthy brain. Lac can be observed when elevated in pathological disorders. When short TE values are used, resonances from e.g. Glu, Glc, Gln, mI can be observed. The problem with short echo proton spectra however is the fact that they exhibit extensive overlap of the peaks of these compounds, including the underlying signals from macromolecules [10]. Most clinical ^{1}H studies are performed in the human brain. To select the volume of which the signal is acquired, gradients are used. The switching of these gradients may introduce eddy currents. These create time-dependent shifts of the resonance frequency during acquisition, which result in phase modulation of the FID, leading to deviations from the theoretical model of Eq. (1.3). In section 5.1 it is explained how signal processing methods deal with this non-ideality. In Fig. 1.5 an example of a Siemens (Vision) localized long echo (TE=135 ms) signal is displayed. Resonances from Cr, Pcr, Cho and NAA can be seen.

^{31}P is the nucleus most often used for metabolic studies. The ^{31}P nucleus is less sensitive than the proton nucleus, but there is no need to suppress peaks during acquisition. Resonances from a wide variety of metabolites can be detected, such as phosphomonoesters (PME), inorganic phosphate (P_i), phosphodiesters (PDE), PCr and ATP. The phosphorus nuclei of adenosine diphosphate (ADP) have frequencies very similar to those of the α and γ phosphorus atoms of ATP. The spectral pattern will change for different organs. For example, the PCr concentration in muscle is higher than in the brain. PCr is absent in the liver. In Fig. 1.6 an ^{31}P-signal of an *ex-vivo* perfused rat liver is shown. Peaks from an external standard (ES), PME, P_i, PDE and α-, β-, and γ-ATP are observed. A feature often present in ^{31}P spectra is a broad resonance underlying

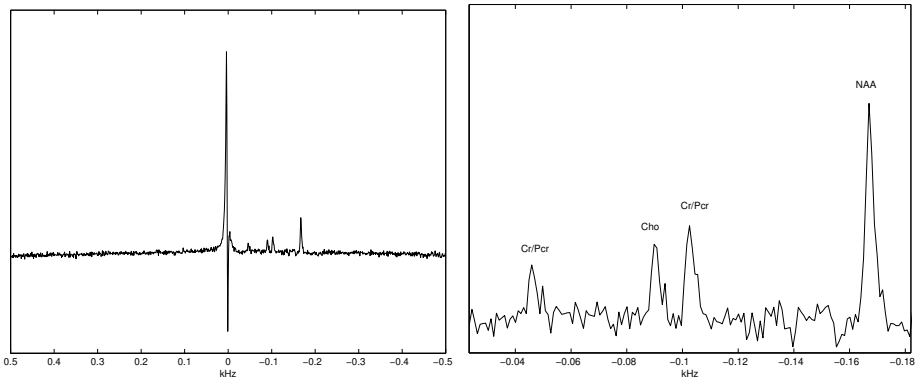


Figure 1.5: Long echo (TE=135 ms) proton signal from a $2 \times 2 \times 2$ cm^3 volume in the white matter of the human brain acquired in a 1.5 T magnet ($f = 63.6$ MHz). Left: the residual water peak dominates the spectrum. Right: zoomed into the region of interest. Resonances from creatine (Cr), phosphocreatine (PCr), choline containing compounds (Cho) and N-acetyl-aspartate (NAA) are observed.

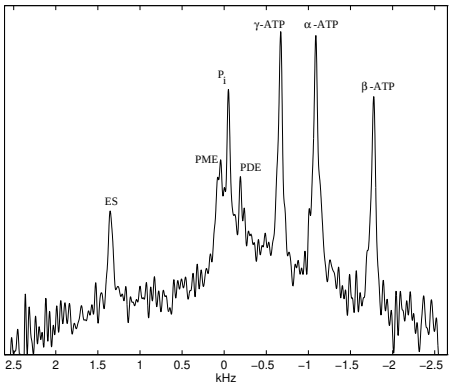


Figure 1.6: Ex-vivo ^{31}P signal from the perfused rat liver. Peaks from an external standard (ES), phosphomonoesters (PME), inorganic phosphate (P_i), phosphodiesters (PDE) and adenosine triphosphate (α-, β- and γ-ATP) can be observed. The broad resonance underlying the mentioned metabolites originates from phosphorus nuclei in less mobile molecules.

the metabolites of interest, originating from phosphorus nuclei in less mobile molecules.

Besides the ^{1}H and ^{31}P nuclei, other nuclei such as carbon-13 (^{13}C) and fluor (^{19}F) are also used for biomedical *in-vivo* MRS studies. ^{13}C has a very low

natural abundance and a low intrinsic sensitivity, making ^{13}C relatively insensitive. However, one often administers ^{13}C labeled substrates to study metabolic pathways (see e.g. [130]). ^{19}F has a high natural abundance and a sensitivity almost equal to that of protons. There are no endogenous ^{19}F containing compounds in biological tissue. ^{19}F is used *in vivo* e.g. to monitor the uptake and metabolism of drugs, to determine cerebral blood flow and to study the metabolism of anaesthetics [35].

1.7 Goals of the thesis

In this thesis, signal processing algorithms are developed to quantify *in-vivo* and *ex-vivo* MRS signals. As seen from the discussion and examples above, these signals are usually characterized by poor SNRs, overlapping peaks, deviations from the model function and in some cases the presence of disturbing components (e.g. the residual water in proton spectra). Ideally, the analysis of these signals should be accurate, automatic and efficient. Accurate results are the basis for a reliable diagnosis in a clinical setting or a sound interpretation of experimental results in general. Automation is not only important for the processing of large amounts of data but also to ensure that the obtained results are user-independent and as such reproducible. In a medical environment where the results of a MRS examination should be available as soon as possible, it goes without saying that efficiency is of utmost importance. Also in other domains, with ever increasing amounts of data to be processed, efficiency can not be considered a mere luxury. The work presented in this thesis aims at bringing accurate, efficient and automatic MRS data analysis one step closer to reality.

1.8 Monte-Carlo simulation procedure

In the next chapters, the performance of different methods is evaluated and compared. Visual comparison of the results of the analysis of a few signals is insufficient to draw any general conclusions. Bias, variance and root mean square error (RMSE) are commonly used to quantify the statistical performance. Ideally, the bias should be negligibly small and the variance as close as possible to the Cramér-Rao lower bound (CRB) [119], which indicates the best possible accuracy of an estimate for *any* unbiased estimator (see also section 2.1.1). Since experimental signals contain errors introduced by factors such as unknown lineshape, data acquisition errors and eddy currents, all inevitably present in *in-* and *ex-vivo* experiments, simulated signals are used to estimate the statistical parameters of interest. Extensive Monte-Carlo simulation studies are performed in which the statistics are calculated using the

same noise-free data set but different realizations of the noise. Unless otherwise specified the model function of Eq. (1.3) is used and the added complex noise is circular, white and Gaussian distributed. The relative root mean squared error (RRMSE), the relative bias (Rbias) and the relative standard deviation (RSTD) of e.g. the amplitude of peak k (in percent) are estimated as follows

$$\text{RRMSE peak } k{=}100\sqrt{\frac{1}{L}\sum_{l=1}^{L}\frac{(a_k-\hat{a}_k^l)^2}{a_k^2}}$$

$$\text{RBias peak } k{=}100\frac{a_k-\frac{1}{L}\sum_{l=1}^{L}\hat{a}_k^l}{a_k}$$

$$\text{RSTD peak } k{=}100\sqrt{\frac{1}{L-1}\sum_{l=1}^{L}\frac{(\hat{a}_k^l-\frac{1}{L}\sum_{l=1}^{L}\hat{a}_k^l)^2}{a_k^2}},$$

where L is the number of simulation runs and $\hat{a}_k^l$ denotes the estimate of a_k obtained in simulation run l.

The noise level is indicated either by its standard deviation σ both on the real and imaginary parts of the noise or expressed by the SNR of a peak in decibel (dB) and defined as

$$\text{SNR peak } k{=}20\log(\frac{a_k}{\sigma}).$$

1.9 Chapter-by-chapter overview

This thesis gives an overview of modern analysis methods in MRS. It is written to be as self-contained as possible and intended both for the engineer not familiar with the application and the scientist who uses MRS as a tool in his/her research.

This thesis mainly consists of two parts. Chapters 2 to 4 concern the analysis of signals in "ideal" circumstances, i.e. the model function of the entire signal is known. Chapters 5 to 8 deal with parameter estimation in "non-ideal" circumstances brought about by deviations from the model function and the presence of unknown or uninteresting spectral features. A schematic overview of the structure of this thesis is given in Fig. (1.7).

Chapter 2 gives an overview of methods currently used to quantitate individual MRS signals and also indicates the connections and similarities that exist between them. It allows to situate the methods that are developed in this thesis in the following chapters. Methods that operate in the time domain and the frequency domain are treated separately. The overview of time-domain methods includes:

1. Interactive methods that require a lot of expertise from the user, but are very flexible with respect to (w.r.t.) the choice of model function and prior knowledge that can be imposed. The first widely available and used interactive method is VARPRO.

2. Black-box methods that are easy to use, but have severe limitations concerning the model function and the prior knowledge that can be imposed, hence limiting their application in demanding *in-vivo* situations. HSVD and HTLS are examples of black-box methods that occur frequently in this thesis.

Chapter 3 deals with AMARES, an interactive time-domain method that was developed in the framework of this thesis. The principles underlying the method are outlined and the new features and advantages of the method over VARPRO are pointed out.

Chapter 4 concerns the analysis of so-called time series of MRS signals. These time series contain a number of individual MRS signals of which it is known that there exist relations between parameters of the different signals. It is explained how the black-box and interactive time-domain methods can be extended to the simultaneous processing of all spectra in the time series using the common information present in the spectra. It is shown that the simultaneous processing results in much more accurate and reliable parameter estimates.

Chapter 5 discusses parameter estimation in non-ideal circumstances. In many cases the biomedical MRS data do not obey the theoretical model functions of section 1.3.1 and contain disturbing resonances of which little is known. Examples are e.g. eddy currents which distort the lineshape and residual water in proton spectra. A short overview of techniques that can be used to improve the quantitation results in case model distortions are present is given. Methods that can be used to deal with unknown and/or unwanted components are also introduced. Especially the possibility to obtain accurate estimates of parameters of selected peaks in the presence of unknown or uninteresting spectral features separated in frequency from the metabolites of interest are treated in detail. This is denoted by frequency-selective (FS) quantification.

In chapter 6 a new method based on the use of maximum-phase finite impulse response (FIR) filters is presented. The principles underlying the method are explained in detail. The FIR filter method is compared to time-domain weighting, HSVD-based filtering and fitting in the frequency domain using a polynomial baseline. It is shown that the new method has considerable advantages over the other FS parameter estimation methods in terms of parameter accuracy, ease of use and computational complexity.

In chapter 7, HLR, a fast variant of HSVD is developed to be used in the context of removal of residual water in proton spectra and shown to give similar

parameter accuracy as the original method in this type of application.

Chapter 8 deals in some more detail with the application of FS estimation in the context of proton spectra. The FIR filter method is shown to outperform the HSVD/HLR-based filtering in terms of accuracy and computational efficiency.

Finally, chapter 9 summarizes the conclusions and outlines suggestions for further research.

1.10 Conclusions

In this chapter *in-* and *ex-vivo* MRS were situated in the more general context of NMR. The basic physical principles underlying the acquisition of a MRS signal were outlined. Thereafter, the mathematical model functions most often used to describe the measured signals were introduced. It was also explained how a MRS signal is normally visualized in the frequency domain. The concept of prior knowledge was illustrated and some examples of typical ^{31}P and ^{1}H signals were given. Based on this background material, the actual goals of this thesis were defined. The statistical procedure used to compare different mathematical algorithms was outlined and finally a chapter-by-chapter overview was provided.

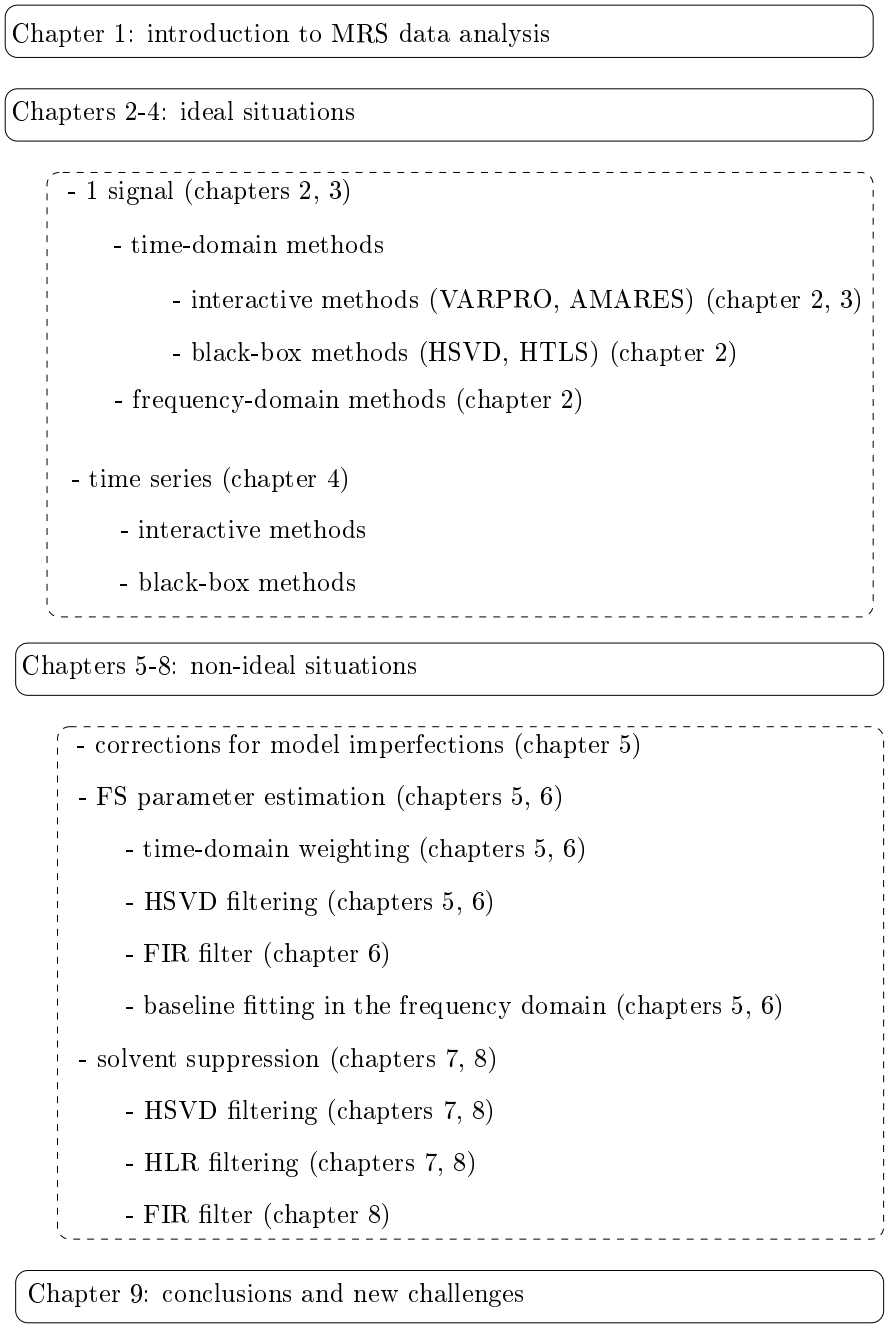


Figure 1.7: *Organization of the thesis.*

Chapter 2

Overview of parameter estimation methods

The goals of this chapter are:

1. *To provide a basic understanding of the underlying principles of existing MRS quantification methods.*
2. *To point out connections between methods.*

The practice of dividing the MRS quantification methods in time- and frequency domain methods is followed here. Section 2.1 and section 2.2 give an overview of time- and frequency-domain methods respectively. Section 2.3 discusses some other methods appeared in literature and section 2.4 summarizes the main conclusions of this chapter.

2.1 Time-domain methods

Time-domain methods are usually divided in interactive (section 2.1.1) and black-box (section 2.1.2) methods. The interactive methods are very flexible. All types of model functions can be used and prior knowledge can easily be incorporated. Above that, maximum likelihood (ML) parameter estimates are obtained if the underlying assumptions concerning the model function and noise distribution are satisfied. The most important drawbacks are the needed user interaction and the fact that no methods exist that guarantee convergence to the global minimum in a reasonable amount of time. Black-box methods require

minimal user involvement or expertise. One of the drawbacks however is the fact that the model function used is restricted to a sum of uniformly sampled, damped complex exponentials as specified in Eq. (1.3). Another disadvantage is that only limited forms of prior knowledge can be used.

2.1.1 Interactive methods

An approach often used in all kinds of parameter estimation problems is to maximize $P(\boldsymbol{\theta})$, the probability density function of the parameters $\boldsymbol{\theta}$, given the data vector $\mathbf{y} = [y_0, \ldots, y_{N-1}]^T$

$$\max_{\boldsymbol{\theta}} P(\boldsymbol{\theta}|\mathbf{y}) = \max_{\boldsymbol{\theta}} \frac{P(\mathbf{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{y})} \tag{2.1}$$

Since the denominator of Eq. (2.1) is independent of the parameters, it is sufficient to maximize the numerator. If the probability density function of the parameters is unavailable, the most reasonable choice for $P(\boldsymbol{\theta})$ is that of a uniform distribution. $P(\mathbf{y}|\boldsymbol{\theta})$ is known as the likelihood function and the estimates obtained by maximizing $P(\mathbf{y}|\boldsymbol{\theta})$ are called ML estimates. In the case of circular, white Gaussian noise with standard deviation σ for the real and imaginary parts, the following holds

$$\begin{aligned} P(\mathbf{y}|\boldsymbol{\theta}) &= \prod_{n=0}^{N-1} P(Re(e_n))P(Im(e_n)) \\ &= \frac{1}{(2\pi\sigma^2)^N} \prod_{n=0}^{N-1} e^{-\frac{(Re(e_n))^2}{2\sigma^2}} e^{-\frac{(Im(e_n))^2}{2\sigma^2}} \\ &= \frac{1}{(2\pi\sigma^2)^N} \prod_{n=0}^{N-1} e^{-\frac{(Re(y_n-\bar{y}_n))^2}{2\sigma^2}} e^{-\frac{(Im(y_n-\bar{y}_n))^2}{2\sigma^2}}, \end{aligned}$$

where $Re(.)$, $Im(.)$ denote the real, respectively imaginary part of a complex quantity. Maximizing $P(\mathbf{y}|\boldsymbol{\theta})$ is equivalent to maximizing the log-likelihood function L, i.e.

$$L = ln(P(\mathbf{y}|\boldsymbol{\theta})) = -Nln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} |y_n - \bar{y}_n|^2, \tag{2.2}$$

$|.|$ represents the modulus of a complex entity. The above equation shows clearly that in the case of circular, white Gaussian noise, obtaining the ML estimates boils down to minimizing the squared difference between the data and the model function or equivalently, solving a nonlinear least-squares (NLLS) problem. In vector notation the functional to be minimized is written as

$$\|\mathbf{y} - \bar{\mathbf{y}}\|^2, \tag{2.3}$$

$\|.\|$ denotes the Euclidean vector norm, $\bar{\mathbf{y}} = [\bar{y}_0, \ldots, \bar{y}_{N-1}]^T$.

It is possible to obtain the ML estimates by minimizing a so-called variable projection functional [59] since the model functions used in MRS (see section 1.3.1) can be written as

$$\bar{y}_n = \sum_{k=1}^{K} c_k \gamma_k(\boldsymbol{\alpha}_k, n), \quad n = 0, \ldots, N-1 \tag{2.4}$$

where c_k are the complex amplitudes $a_k e^{j\phi_k}$ and $\gamma_k(\boldsymbol{\alpha}_k, n)$ are independent functions of the nonlinear parameter vector $\boldsymbol{\alpha}_k$. E.g. in the case of the model function of Eq. (1.3), $\boldsymbol{\alpha}_k$ is equal to $[f_k \; d_k \; t_0]$. Using matrix notation, Eq. (2.4) becomes

$$\bar{\mathbf{y}} = \boldsymbol{\Gamma}\mathbf{c}, \tag{2.5}$$

with $\mathbf{c} = [c_1, \ldots, c_K]^T$ and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1(\boldsymbol{\alpha}_1, 0) & \ldots & \gamma_K(\boldsymbol{\alpha}_K, 0) \\ \vdots & \ddots & \vdots \\ \gamma_1(\boldsymbol{\alpha}_1, N-1) & \ldots & \gamma_K(\boldsymbol{\alpha}_K, N-1) \end{bmatrix},$$

an $N \times K$ matrix of full rank. Suppose that the nonlinear parameters $\boldsymbol{\alpha}_k$, $k = 1, \ldots, K$ are known; then the matrix $\boldsymbol{\Gamma}$ can be computed and an estimate for the linear parameters $\mathbf{c}$ is obtained as the solution of a linear least squares (LS) problem $\widehat{\mathbf{c}} = \boldsymbol{\Gamma}^{\dagger}\mathbf{y}$, where $\boldsymbol{\Gamma}^{\dagger}$ denotes the pseudo-inverse of $\boldsymbol{\Gamma}$, i.e. $\boldsymbol{\Gamma}^{\dagger} = (\boldsymbol{\Gamma}^H\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^H$ (the superscript ^{H} denotes the complex conjugate). The cost function of Eq. (2.3) then becomes

$$\|\mathbf{y} - \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\dagger}\mathbf{y}\|^2 = \|\mathbf{P}_{\boldsymbol{\Gamma}}^{\perp}\mathbf{y}\|^2. \tag{2.6}$$

Eq. (2.6) defines the variable projection functional and $\mathbf{P}_{\boldsymbol{\Gamma}}^{\perp}$ is the orthogonal projector on the column space of $\boldsymbol{\Gamma}$. The linear parameters have been eliminated from the cost function of Eq. (2.6). As a result a minimization problem in fewer variables is obtained. The estimates $\hat{\boldsymbol{\alpha}}_k$ are obtained as the parameters that minimize Eq. (2.6). Using $\hat{\boldsymbol{\alpha}}_k$, the matrix $\widehat{\boldsymbol{\Gamma}}$ is computed and the estimates $\hat{c}_k$ of the complex amplitudes are obtained as the LS solution of

$$\widehat{\boldsymbol{\Gamma}}\mathbf{c} \approx \mathbf{y},$$

i.e. $\hat{\mathbf{c}} = \widehat{\boldsymbol{\Gamma}}^{\dagger}\mathbf{y}$. In [59] it is proven that the parameters obtained in this way are the same as the ones obtained by minimizing Eq. (2.3) directly.

As already indicated in section 1.8, there exists a fundamental lower limit on the accuracy of the parameter estimates that can be obtained by any unbiased

estimator. Let E denote the expected value operator and assume that $\widehat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$

$$E(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta}.$$

The covariance matrix of $\widehat{\boldsymbol{\theta}}$ is defined as

$$\mathbf{P} = E\left((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right).$$

It can be shown [119] that under quite general conditions the following holds

$$\mathbf{P} \geq \mathbf{P}_{CR} = \mathbf{F}^{-1} = -\left(E\left(\frac{\delta^2 L}{\delta\boldsymbol{\theta}\ \delta\boldsymbol{\theta}^T}\right)\right)^{-1}, \tag{2.7}$$

where $\mathbf{F}$ is the Fisher information matrix and $\mathbf{P}_{CR}$ the Cramér-Rao matrix. In the case of circular, white Gaussian noise expressions for the elements of $\mathbf{F}$ can be worked out explicitly [33]. In case L is given by Eq. (2.2), the element of the ith row and jth column of $\mathbf{F}$ can be written as [33]

$$\frac{1}{\sigma^2}\sum_{n=0}^{N-1} Re\left(\frac{\delta\bar{y}_n^*}{\delta\theta_i}\frac{\delta\bar{y}_n}{\delta\theta_j}\right), \tag{2.8}$$

where * denotes the complex conjugate and θ_i is the ith element of $\boldsymbol{\theta}$.
If the parameter θ' is a function of the parameters $\boldsymbol{\theta}$, $\theta' = f(\boldsymbol{\theta})$, the CRB corresponding to θ' is given by

$$\left[\begin{array}{ccc} \frac{\delta f}{\delta\theta_1} & \cdots & \frac{\delta f}{\delta\theta_M} \end{array}\right] \mathbf{F}^{-1} \left[\begin{array}{ccc} \frac{\delta f}{\delta\theta_1} & \cdots & \frac{\delta f}{\delta\theta_M} \end{array}\right]^T, \tag{2.9}$$

where $\theta_1, \ldots, \theta_M$ are the individual elements of the parameter vector $\boldsymbol{\theta}$. Note that further on in this text, the term CRB is also used to refer to the square root of ((derived quantities of) elements of) the Cramér-Rao matrix. This is not mentioned explicitly but it is clear from the context.

The prior knowledge described in section 1.5 can easily be incorporated in these functionals. By imposing the prior knowledge, the CRBs decrease, resulting in more accurate parameter estimates [20]. The prior knowledge is expressed as linear equality constraints which are substituted in the original functional, resulting in an unconstrained NLLS problem. The latter can be solved using local or global optimization theory. Global optimization has already been used in MRS [111, 45, 88] but the main disadvantage of these methods is the poor computational efficiency. On the other hand it is possible to obtain good starting values for the frequencies and dampings by a procedure called peak picking for use in local optimization methods, resulting in an acceptable solution in a reasonable time in most circumstances. Peak picking can be explained as follows. The signal is Fourier transformed and the real part is phased properly

(see section 1.4). For each peak, the user indicates the top and a point belonging to the peak at half height. The top value provides a starting value for the frequency of the peak and the difference between the top value and the value at half height is directly proportional to the damping of that peak. This step in combination with the proper choice of prior knowledge to be imposed involves a lot of interaction with the user, therefore these model fitting methods are mostly called *interactive methods*.

The first widely available and used interactive method in MRS data quantification is VARPRO [137]. As the name suggests, the method actually minimizes Eq. (2.6). In this work, an improved method, called AMARES (*a*dvanced *m*ethod for *a*ccurate, *r*obust and *e*fficient *s*pectral fitting) is derived. AMARES actually minimizes the cost function of Eq. (2.3). Detailed implementations aspects and results of comparisons between VARPRO and AMARES are discussed in the next chapter.

Alternative ways to minimize the cost function have been developed. In [91, 28] the expectation-maximization (EM) algorithm is used to reduce the high dimensional search for the optimum. The EM algorithm divides the problem into K independent optimizations, each involving the parameters of a single peak. The methods presented in [82] and [129] are very similar in nature and also try to reduce the complexity of the minimization problem by performing multiple one-dimensional searches. The possibility of imposing prior knowledge was not addressed in those publications. In [11], the methods presented in [82] and [129] are extended such that the algorithms use the a priori knowledge of the possible frequency intervals of the damped exponentials.

2.1.2 Black-box methods

Linear prediction (LP) methods and state-space methods are treated separately in sections 2.1.2.1 and 2.1.2.2 respectively. For each of these two classes, the principles are first outlined for the ideal case of noiseless data. Subsequently, noisy data are considered and an overview of the methods which compute a statistically suboptimal solution is provided. Finally, it is indicated how ML estimates can be obtained for both black-box approaches. The basic underlying principles of the black-box methods are explained without going into all mathematical details. More details and additional references can e.g. be found in [118, 56, 72, 33].

2.1.2.1 LP methods

2.1.2.1.1 Basic theory - noiseless data

The model of Eq. (1.3) can be rewritten in terms of complex amplitudes c_k and signal poles z_k as follows

$$\bar{y}_n = \sum_{k=1}^{K} (a_k e^{j\phi_k})(e^{(-d_k + j2\pi f_k)n\Delta t}) = \sum_{k=1}^{K} c_k z_k^n, \; n = 0, \ldots, N-1. \tag{2.10}$$

A polynomial of degree K having the signal poles as roots can be formed

$$p(z) = \prod_{k=1}^{K} (z - z_k) = \sum_{i=0}^{K} p_i z^{K-i} \;\; (p_0 = 1), \tag{2.11}$$

p_k are the so-called forward prediction coefficients. Using the fact that $p(z_k) = 0, \;\; 1 \leq k \leq K$, the following holds for all k

$$z_k^m = -\sum_{i=1}^{K} p_i z_k^{m-i}, \;\; m \geq K. \tag{2.12}$$

Combining Eqs. (2.12) and (2.10) gives

$$\begin{aligned} \bar{y}_n &= -\sum_{k=1}^{K} c_k \left(\sum_{i=1}^{K} p_i z_k^{n-i} \right) \\ &= -\sum_{i=1}^{K} p_i \left(\sum_{k=1}^{K} c_k z_k^{n-i} \right) \\ &= -\sum_{i=1}^{K} p_i \bar{y}_{n-i}, \;\; n \geq K. \end{aligned} \tag{2.13}$$

The theoretical series $\bar{y}_n, n = 0, \ldots, N-1$ of Eq. (2.10) can be prolonged by an arbitrary number of points using Eq. (2.13) and this procedure is therefore called forward linear prediction. A similar reasoning as above can be repeated to define backward linear prediction.

To determine the prediction coefficients two different approaches have been proposed. The first is based on the autocorrelation coefficients of the data points (see [72] and references therein), while the second works on the data points directly. Here only the second option is explored further.

Writing down Eq. (2.13) for $n = K, \ldots, N-1$ results in the following set of equations

$$\begin{bmatrix} \bar{y}_K \\ \bar{y}_{K+1} \\ \vdots \\ \bar{y}_{N-1} \end{bmatrix} = - \begin{bmatrix} \bar{y}_0 & \cdots & \bar{y}_{K-1} \\ \bar{y}_1 & \cdots & \bar{y}_K \\ \vdots & \ddots & \vdots \\ \bar{y}_{N-K-1} & \cdots & \bar{y}_{N-2} \end{bmatrix} \begin{bmatrix} p_K \\ p_{K-1} \\ \vdots \\ p_1 \end{bmatrix}$$

$$\bar{\mathbf{h}}_{lp} = -\bar{\mathbf{H}}_{lp}\mathbf{p}. \tag{2.14}$$

Eq. (2.14) is exactly solvable and gives the exact linear prediction coefficients $\mathbf{p}$. By rooting the polynomial $p(z)$ specified in Eq. (2.11) the signal poles $z_k, \quad k = 1, \ldots, K$ are obtained, from which the frequencies and dampings can easily be computed. The signal poles are filled in in Eq. (2.10) and the exact amplitudes and phases can then be determined from this set of equations.

2.1.2.1.2 Noisy data - suboptimal solutions

If noise is introduced, Eq. (2.13) becomes

$$y_n = -\sum_{k=1}^{K} p_k(y_{n-k} - e_{n-k}) + e_n = -\sum_{k=1}^{K} p_k y_{n-k} + \sum_{k=0}^{K} p_k e_{n-k} \quad (p_0 = 1). \tag{2.15}$$

Eq. (2.15) represents a case of an autoregressive moving average (ARMA) model with identical autoregression and moving average parameters. Since an ARMA model can be arbitrarily well approximated by a high order AR model, Eq. (2.15) is replaced by

$$y_n \approx -\sum_{k=1}^{M} p_m y_{n-k} + v_n, \tag{2.16}$$

where $v_n, \quad n = 0, \ldots, N-1$ is considered to be circular, white Gaussian noise.

Eq. (2.16) represents a much easier to solve AR model. It is of course evident, that the parameters obtained by using the formalism of Eq. (2.16) are statistically suboptimal due to the simplified underlying noise model. The prediction coefficients are found by setting up Eq. (2.16) for $n = M, \ldots, N-1$, i.e.

$$\begin{bmatrix} y_M \\ y_{M+1} \\ \vdots \\ y_{N-1} \end{bmatrix} \approx - \begin{bmatrix} y_0 & \cdots & y_{M-1} \\ y_1 & \cdots & y_M \\ \vdots & \ddots & \vdots \\ y_{N-M-1} & \cdots & y_{N-2} \end{bmatrix} \begin{bmatrix} p_M \\ p_{M-1} \\ \vdots \\ p_1 \end{bmatrix}$$

$$\mathbf{h}_{lp} \approx -\mathbf{H}_{lp}\mathbf{p}. \tag{2.17}$$

From Eq. (2.17) the maximum value for M can be derived. Since $N - M$ has to be greater or equal to M in order not to have a underdetermined set of equations, the largest possible value for M is $N/2$. No exact solution exists for the above equations.

The popular Kumeresan and Tufts' LPSVD method [75, 6] starts by computing the singular value decomposition (SVD) of $\mathbf{H}_{lp}$. A rank K approximation

of $\mathbf{H}_{lp}$, the matrix $\mathbf{H}_{lp_K}$, is then obtained by retaining the largest K singular values and setting the others to zero. In this way, a significant noise contribution is removed. This latter procedure is also called discrete regularization. Solving the set of equations $\mathbf{h}_{lp} \approx \mathbf{H}_{lp_K}\mathbf{p}$ in a LS sense results in an estimate $\hat{\mathbf{p}}$ of the linear prediction coefficients. The estimates $\hat{z}_k$ of the signal poles z_k, $k = 1, \ldots, K$, are derived from the K largest (in absolute value) roots of the prediction polynomial $1 + \sum_{m=1}^{M} \hat{p}_m z^{-m} = 0$.

A lot of variants of the above outlined algorithm exist. Truncating e.g. the SVD of $[\mathbf{H}_{lp}\ h_{lp}]$ to rank K and computing the total least squares (TLS) solution, results in an improved algorithm, called LPTLS [126]. Instead of truncating the SVD to rank K, a continuous regularization technique, called LPSVD(CR) [74] can be applied to allow automatic determination of the number of components. In EPLPSVD [46, 47], the Hankel structure of the $\mathbf{H}_K$ matrix is restored by applying a SVD-based signal enhancement algorithm due to Cadzow. In [23] the truncation step in Cadzow's algorithm is followed by a correction of the retained singular values according to the principles of minimum variance. All mentioned algorithms are however statistically suboptimal due to the fact that the ARMA model is not taken properly into account.

2.1.2.1.3 Noisy data - ML solutions

It is now shown how the LP problem can be solved in a ML way. Two different approaches are considered. The first is based on the optimal solution of the LP equations directly, the other rewrites Eq. (2.6) in terms of the LP coefficients.

Consider again Eq. (2.15) and set up the linear prediction equations as in Eq. (2.17) with $M = K$. Basically, the LP problem solution comes down to computing a solution to the set of incompatible linear equations $\mathbf{h}_{lp} \approx -\mathbf{H}_{lp}\mathbf{p}$. To compute a solution to this set of equations, corrections $[\Delta\mathbf{H}_{lp}\ \Delta\mathbf{h}_{lp}]$ have to be made to $[\mathbf{H}_{lp}\ \mathbf{h}_{lp}]$ in order to ensure rank deficiency of $[\mathbf{H}_{lp} + \Delta\mathbf{H}_{lp}\ \mathbf{h}_{lp} + \Delta\mathbf{h}_{lp}]$. In the classical LS approach the following problem is solved

$$\begin{aligned} &\min_{\Delta\mathbf{h}_{lp},\mathbf{p}} \|\Delta\mathbf{h}_{lp}\| \\ &\text{such that } -\mathbf{H}_{lp}\mathbf{p} = \mathbf{h}_{lp} + \Delta\mathbf{h}_{lp}. \end{aligned} \tag{2.18}$$

The solution $\mathbf{p}$ to Eq. (2.18) is only statistically optimal in case the errors are identically, independently distributed (i.i.d.) and confined to $\mathbf{h}_{lp}$. The latter condition is obviously violated in this case. The TLS method is more general than the LS approach in the sense that it solves

$$\begin{aligned} &\min_{\Delta\mathbf{H}_{lp},\Delta\mathbf{h}_{lp},\mathbf{p}} \|[\Delta\mathbf{H}_{lp}\ \Delta\mathbf{h}_{lp}]\|_F \\ &\text{such that } -(\mathbf{H}_{lp} + \Delta\mathbf{H}_{lp})\mathbf{p} = \mathbf{h}_{lp} + \Delta\mathbf{h}_{lp}, \end{aligned} \tag{2.19}$$

yielding statistically optimal results in case of row-wise, i.i.d. errors on $[\mathbf{H}_{lp}\ \mathbf{h}_{lp}]$. Note that $\|.\|_F$ denotes the Frobenius norm of a matrix. Since $[\mathbf{H}_{lp}\ \mathbf{h}_{lp}]$ has a Hankel structure, the row-wise independency assumption is not valid either. The only correct way to obtain ML estimates for $\mathbf{p}$ is to use the structured TLS (STLS) formulation [36, 37, 109], which comes as a natural extension to the TLS problem when structured matrices are involved. The STLS approach can be formulated as follows

$$\begin{aligned}&\min_{\Delta\mathbf{H}_{lp},\Delta\mathbf{h}_{lp},\mathbf{p}} \mathbf{q}^H\mathbf{q}\\ &\text{such that } -(\mathbf{H}_{lp}+\Delta\mathbf{H}_{lp})\mathbf{p} = \mathbf{h}_{lp}+\Delta\mathbf{h}_{lp},\\ &\text{and } [\Delta\mathbf{H}_{lp}\ \Delta\mathbf{h}_{lp}] \text{ has the same structure as } [\mathbf{H}_{lp}\ \mathbf{h}_{lp}],\end{aligned} \tag{2.20}$$

where $\mathbf{q} \in \mathbb{C}^{N\times 1}$ contains the different entries of $[\Delta\mathbf{H}_{lp}\ \Delta\mathbf{h}_{lp}]$, i.e. the elements in the first column and last row of this matrix. In recent years many different formulations have been proposed for the STLS problem: the constrained TLS (CTLS) approach [1, 2], the structured total least norm (STLN) approach [109, 139] and the Riemannian SVD (RiSVD) approach [37]. All these formulations start more or less from a formulation similar to the one above, but the final formulation for which an algorithm is developed can be quite different. In any case, there exists no closed-form solution to the minimization problem of Eq. (2.20) and it has to be solved using an iterative algorithm.

Once the ML estimates of the linear prediction coefficients have been obtained by solving the above STLS problem formulation, ML estimates of frequencies, dampings, amplitudes and phases are determined as outlined before. Since the parameters obtained in this way are optimal in a ML sense, it comes as no surprise, that the model fitting methods as presented in section 2.1.1 and this STLS approach of the LP problem formulation are totally equivalent. To prove this, the following derivation is made. Let $\Delta y_k = y_k - \bar{y}_k$, the problem stated in Eq. (2.20) can then be written down more explicitly

$$\begin{aligned}&\min_{\substack{\Delta y_n,\ n=0,\ldots,N-1\\ p_k,\ k=1,\ldots,K}} \sum_{n=0}^{N-1} |\Delta y_n|^2\\ &\text{such that } -(\mathbf{H}_{lp}+\Delta\mathbf{H}_{lp})\mathbf{p} = \mathbf{h}_{lp}+\Delta\mathbf{h}_{lp}\end{aligned} \tag{2.21}$$

with

$$\Delta\mathbf{H}_{lp} = \begin{bmatrix} \Delta y_0 & \cdots & \Delta y_{K-1} \\ \Delta y_1 & \cdots & \Delta y_K \\ \vdots & \ddots & \vdots \\ \Delta y_{N-K-1} & \cdots & \Delta y_{N-2} \end{bmatrix}, \tag{2.22}$$

$$\Delta\mathbf{h}_{lp} = [\Delta y_K\ \Delta y_{K+1}\ \cdots\ \Delta y_{N-1}]^T. \tag{2.23}$$

The constraint in Eq. (2.21) basically expresses that $[\mathbf{H}_{lp}+\Delta\mathbf{H}_{lp}\ \mathbf{h}_{lp}+\Delta\mathbf{h}_{lp}]$ is a rank-deficient matrix. Each rank-deficient Hankel matrix $\mathbb{C}^{(N-K)\times(K+1)}$ can however be constructed from a signal representing a sum of damped complex exponentials $\sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-d_k+j2\pi f_k)n\Delta t}$, $n=0,\ldots,N-1$. This parameterization follows directly from the fact that if $[\mathbf{H}_{lp}+\Delta\mathbf{H}_{lp}\ \mathbf{h}_{lp}+\Delta\mathbf{h}_{lp}]$ is a rank-deficient Hankel matrix, there exists a vector $\mathbf{p}$ such that $-[\mathbf{H}_{lp}+\Delta\mathbf{H}_{lp}]\mathbf{p} = [\mathbf{h}_{lp}+\Delta\mathbf{h}_{lp}]$. The latter equation is nothing else than a set of LP equations, similar to Eq. (2.14). The complex amplitudes and signal poles can be retrieved using the procedure outlined in the beginning of this section. Using this parameterization of the elements y_n, $n=0,\ldots,N-1$ of the Hankel matrix, it is possible to recast Eq. (2.21) as the minimization of the cost function $\sum_{n=0}^{N-1} |y_n - \sum_{k=1}^{K} a_k e^{j\phi_k} e^{(-d_k+j2\pi f_k)n\Delta t}|^2$ w.r.t. a_k, d_k, f_k, ϕ_k, $k=1,\ldots,K$. The constraint present in Eq. (2.21) has been eliminated by using the particular choice of parameterization of the rank-deficient matrix $[\mathbf{H}_{lp}+\Delta\mathbf{H}_{lp}\ \mathbf{h}_{lp}+\Delta\mathbf{h}_{lp}]$. This cost function is totally equivalent to the one of Eq. (2.3) in case $\bar{\mathbf{y}}$ is modeled by a sum of uniformly sampled, damped complex exponentials. By solving the STLS formulation an optimal Hankel matrix of rank K is found as an approximation to $[\mathbf{H}_{lp}\ \mathbf{h}_{lp}]$. Note that Cadzow's algorithm computes only an approximate solution to the STLS problem in the sense that not the closest (to $[\mathbf{H}_{lp}\ \mathbf{h}_{lp}]$) rank-deficient matrix is found. Cadzow's algorithm uses an iterative two-step procedure. First a rank K matrix is computed using the truncated SVD. Since the latter approximation has no Hankel structure anymore, the second step proceeds by replacing each matrix entry by the value obtained by averaging over the corresponding anti-diagonal. These two steps are repeated until convergence.

It is now shown how Eq. (2.6) can be rewritten in terms of the LP coefficients. In case the model function of Eq. (2.10) is used, the matrix $\mathbf{\Gamma}$ defined in Eq. (2.5) becomes

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & \ldots & 1 \\ z_1 & \ldots & z_K \\ \vdots & \vdots & \vdots \\ z_1^{N-1} & \ldots & z_K^{N-1} \end{bmatrix}.$$

The cost function of Eq. (2.6) can be rewritten as

$$\|\mathbf{y} - \mathbf{\Gamma}\mathbf{\Gamma}^{\dagger}\mathbf{y}\|^2 = \mathbf{y}^H(\mathbf{I} - \mathbf{\Gamma}\mathbf{\Gamma}^{\dagger})\mathbf{y}, \tag{2.24}$$

$(\mathbf{I} - \mathbf{\Gamma}\mathbf{\Gamma}^{\dagger})$ projects on the orthogonal complement of the column space of $\mathbf{\Gamma}$. Defining the $((N-K)\times N)$ Toeplitz matrix $\mathbf{P}$ as

$$\mathbf{P} = \begin{bmatrix} p_K & p_{K-1} & \ldots & p_1 & 1 & 0 & \ldots & \ldots & 0 \\ 0 & p_K & p_{K-1} & \ldots & p_1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & p_K & p_{K-1} & \ldots & \ldots & \ldots & p_1 & 1 \end{bmatrix} \tag{2.25}$$

and taking into account that $p(z_k) = 0,\ k = 1, \ldots, K$, the relation $\mathbf{P}\boldsymbol{\Gamma} = \mathbf{0}$ holds. Combining this with the observation that the rank of $\boldsymbol{\Gamma}$ is K and the one of $\mathbf{P}$ is $N - K$, $\mathbf{P}^H$ is seen to be a basis for the orthogonal complement of the column space of $\mathbf{P}$. Therefore, the associated projector becomes

$$\mathbf{I} - \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\dagger} = \mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}\mathbf{P}. \tag{2.26}$$

When $\mathbf{H}$ is defined as the extended matrix $\mathbf{H} = [\mathbf{H}_{lp}\ \mathbf{h}_{lp}]$, the following holds

$$\mathbf{P}\mathbf{y} = \mathbf{H}\mathbf{p}. \tag{2.27}$$

By combining Eqs. (2.26) and (2.27) the cost function of Eq. (2.24) becomes

$$\mathbf{p}^H\mathbf{H}^H(\mathbf{P}\mathbf{P}^H)^{-1}\mathbf{H}\mathbf{p}. \tag{2.28}$$

Minimizing Eq. (2.28) w.r.t. $\mathbf{p}$ results in ML estimates for the linear prediction coefficients. The iterative quadratic maximum likelihood (IQML) algorithm, initially proposed in [13], is often used (see e.g. [155]) to minimize Eq. (2.28). However, it is not widely known that the solution obtained by IQML is suboptimal. In fact, in [155] it is still claimed that the IQML method is a ML estimation algorithm. For a formal proof of the suboptimality of IQML, the interested reader is referred to [80]. In [80], an alternative algorithm to determine the minimum of Eq. (2.28) is outlined.

2.1.2.2 State-space methods

2.1.2.2.1 Basic theory - noiseless data

The term state-space originates from the corresponding theory derived in the control and identification literature [77]. In the following, the basic principles of the method are explained using principles from linear algebra.

Arrange the noiseless data $\bar{y}_n,\ n = 0, \ldots .N-1$ in a Hankel matrix as follows

$$\bar{\mathbf{H}}_s = \begin{bmatrix} \bar{y}_0 & \bar{y}_1 & \cdots & \bar{y}_M \\ \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_{M+1} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{y}_{N-M-1} & \bar{y}_{N-M} & \cdots & \bar{y}_{N-1} \end{bmatrix}, \text{ with } M \geq K,\ N - M > K. \tag{2.29}$$

Using the model function of Eq. (2.10), the following Vandermonde decomposition of $\bar{\mathbf{H}}$ is easily obtained

$$\bar{\mathbf{H}}_s = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_K \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{N-M-1} & z_2^{N-M-1} & \cdots & z_K^{N-M-1} \end{bmatrix} \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c_K \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_K \\ \vdots & \vdots & \vdots & \vdots \\ z_1^M & z_2^M & \dots & z_K^M \end{bmatrix}^T \tag{2.30}$$

$$= \bar{\mathbf{S}}\bar{\mathbf{C}}\bar{\mathbf{T}}^T \tag{2.31}$$

From this Vandermonde decomposition the complex amplitudes and signal poles can immediately be derived. However, no standard algorithms exist that compute the Vandermonde decomposition directly and as a result the parameters have to be determined indirectly. To this end, the following reasoning is made. From Eq. (2.31) it is seen that $\bar{\mathbf{S}}$ (but also $\bar{\mathbf{T}}$) exhibit the shift-invariant property

$$\bar{\mathbf{S}}^{\uparrow} = \bar{\mathbf{S}}_{\downarrow}\bar{\mathbf{Z}}, \tag{2.32}$$

where the up (down) arrow stands for deleting the top (bottom) row, and $\bar{\mathbf{Z}} \in \mathbb{C}^{K\times K}$ is a diagonal matrix composed of the K signal poles $z_k,\ k = 1, \dots, K$. The rank of $\bar{\mathbf{S}}$, $\bar{\mathbf{C}}$ and $\bar{\mathbf{T}}$ is equal to K. From the LP relations which exist between the columns of $\bar{\mathbf{H}}_s$, it follows that $\bar{\mathbf{H}}_s$ itself has exactly rank K. That means that the SVD of $\bar{\mathbf{H}}_s$ has the following form

$$\begin{aligned} \bar{\mathbf{H}}_s &= \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^H = [\bar{\mathbf{U}}_K\ \bar{\mathbf{U}}_2] \begin{bmatrix} \bar{\mathbf{\Sigma}}_K & 0 \\ 0 & 0 \end{bmatrix} [\bar{\mathbf{V}}_K\ \bar{\mathbf{V}}_2]^H \\ &= \bar{\mathbf{U}}_K\bar{\mathbf{\Sigma}}_K\bar{\mathbf{V}}_K^H, \end{aligned} \tag{2.33}$$

$\bar{\mathbf{U}}_K \in \mathbb{C}^{(N-M)\times K}$, $\bar{\mathbf{U}}_2 \in \mathbb{C}^{(N-M)\times(N-M-K)}$, $\bar{\mathbf{\Sigma}}_K \in \mathbb{C}^{K\times K}$, $\bar{\mathbf{V}}_K \in \mathbb{C}^{(M+1)\times K}$, $\bar{\mathbf{V}}_2 \in \mathbb{C}^{(M+1)\times(M+1-K)}$. From the comparison of Eq. (2.31) and Eq. (2.33) it can be inferred that $\bar{\mathbf{U}}_K$ and $\bar{\mathbf{S}}$ span the same column space and hence are equal up to a multiplication by a non-singular matrix $\bar{\mathbf{Q}} \in \mathbb{C}^{K\times K}$

$$\bar{\mathbf{U}}_K = \bar{\mathbf{S}}\bar{\mathbf{Q}}. \tag{2.34}$$

Using Eq. (2.34), the shift-invariance property of Eq. (2.32) translates to

$$\bar{\mathbf{U}}_K^{\uparrow} = \bar{\mathbf{U}}_{K\downarrow}\bar{\mathbf{Q}}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{Q}}. \tag{2.35}$$

Computation of the SVD of $\bar{\mathbf{H}}$, subsequent truncation of the SVD and setting up Eq. (2.35), allows to determine $(\bar{\mathbf{Q}}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{Q}})$. Since the eigenvalues of $(\bar{\mathbf{Q}}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{Q}})$ and $\bar{\mathbf{Z}}$ are equal, the signal poles are easily derived

$$eig(\bar{\mathbf{Q}}^{-1}\bar{\mathbf{Z}}\bar{\mathbf{Q}}) = eig(\bar{\mathbf{Z}}) = \{z_k,\ k = 1, \dots, K\},$$

the function $eig(.)$ determines the eigenvalues of the matrix between brackets. From these signal poles, frequencies and dampings are easily calculated. Amplitudes and phases are determined by filling in the obtained signal poles in the model function (2.10) and solving the resulting equations. Note that using the shift-invariant property of $\bar{\mathbf{T}}$ a similar reasoning can be made to derive a relation between $\bar{\mathbf{V}}_{K\downarrow}$ and $\bar{\mathbf{V}}_K^{\uparrow}$ similar to Eq. (2.35).

2.1.2.2.2 Suboptimal solutions - noisy data

If noise is considered, relation (2.33) does not hold anymore and there is no exact solution ($\mathbf{Q}^{-1}\mathbf{Z}\mathbf{Q}$) of the shift-invariant property in $\mathbf{U}_K$ anymore. However $\mathbf{H}_s$, having the same structure as $\bar{\mathbf{H}}_s$, but constructed from the noisy data, can be approximated by the truncated SVD of $\mathbf{H}_s$

$$\mathbf{H}_s = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \approx \mathbf{U}_K\mathbf{\Sigma}_K\mathbf{V}_K^H = \mathbf{H}_{s_K}, \tag{2.36}$$

where $\mathbf{U}_K$, $\mathbf{V}_K$ are respectively the first K columns of $\mathbf{U}, \mathbf{V}$ and $\mathbf{\Sigma}_K$ is the leading $(K \times K)$ submatrix of $\mathbf{\Sigma}$. This step brings along the suboptimality of the method. $\mathbf{H}_K$ is a matrix of rank K but its Hankel structure has been destroyed by the truncation of the SVD. Therefore, there exists no exact solution to the set of equations

$$\mathbf{U}_K^{\uparrow} \approx \mathbf{U}_{K\downarrow}\mathbf{Q}^{-1}\mathbf{Z}\mathbf{Q}. \tag{2.37}$$

Instead, in the method called HSVD [77, 7], an estimate of the matrix ($\mathbf{Q}^{-1}\mathbf{Z}\mathbf{Q}$) is obtained by solving Eq. (2.37) in a LS sense. The eigenvalues of the latter matrix are the estimates $\hat{z}_k$, $k = 1, \ldots, K$, of the signal poles. These estimates are filled in in the model equations of Eq. (2.10) and fitted to the original data with the LS method.

A lot of variants of HSVD exist. An improved variant is the HTLS algorithm [138] which computes the TLS solution of Eq. (2.37). Cadzow's method or the minimum variance technique can evidently also be used as preprocessing techniques to improve quantitation by means of the basic HSVD and HTLS algorithms [23].

Since the HSVD en HTLS methods are used often in the remainder of this thesis, a short outline of the methods is given below.

Algorithm 2.1.1 (HSVD/HTLS)

Input: data samples $y_n, n = 0, \ldots, N-1$, model order K.
Output: $\hat{a}_k, \hat{d}_k, \hat{f}_k, \hat{\phi}_k, k = 1, \ldots, K$.

Step 1. Arrange the data points $y_n, n = 0, \ldots, N-1$ in a Hankel matrix H_s of dimensions $(N-M) \times (M+1)$, $M \geq K$, $N - M > K$ as follows

$$\mathbf{H}_s = \begin{bmatrix} y_0 & y_1 & \cdots & y_M \\ y_1 & y_2 & \cdots & y_{M+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N-M-1} & y_{N-M} & \cdots & y_{N-1} \end{bmatrix}. \tag{2.38}$$

In order to obtain the best parameter accuracy it is recommended to choose $\mathbf{H}_s$ as square as possible [138].

Step 2. Compute the SVD of $\mathbf{H}_s$

$$\mathbf{H}_s = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H.$$

Step 3. Truncate $\mathbf{H}_s$ to a matrix $\mathbf{H}_{s_K}$ of rank K

$$\mathbf{H}_{s_K} = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^H.$$

The model order K is chosen equal to the number of sinusoids that comprise the measured signal.

Step 4. Form the following overdetermined set of equations

$$\mathbf{U}_K^{\uparrow} \approx \mathbf{U}_{K\downarrow}\mathbf{E},$$

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

- HSVD: compute an estimate of $\mathbf{E}$ as the LS solution to the above set of equations.
- HTLS: compute an estimate of $\mathbf{E}$ as the TLS solution to the above set of equations.

Once $\mathbf{E}$ is estimated, its K eigenvalues give the signal pole estimates

$$\hat{z}_k = e^{(-\hat{d}_k + j2\pi\hat{f}_k)\Delta t}, \quad k = 1, \ldots, K.$$

From these signal poles it is easy to obtain estimates of the dampings d_k and frequencies f_k.

Step 5. Finally fill in the estimate $\hat{z}_k, k = 1, \ldots, K$, in the N model equations and compute the LS solution $\hat{c}_k = \hat{a}_k e^{j\hat{\phi}_k}$, $k = 1, \ldots, K$, of

$$y_n \approx \sum_{k=1}^{K} c_k \hat{z}_k^n \, , n = 0, \ldots, N-1.$$

In this way estimates for the amplitudes a_k and phases ϕ_k are obtained.

2.1.2.2.3 Noisy data - ML solutions

This state-space formulation can also be solved in a statistically optimal way. Consider again $\mathbf{H}_s$, with $M = K$

$$\mathbf{H}_s = \begin{bmatrix} y_0 & \cdots & y_K \\ y_1 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots \\ y_{N-1-K} & \cdots & y_{N-1} \end{bmatrix} = [\mathbf{H}_1 \ \mathbf{h}_s], \tag{2.39}$$

with $\mathbf{H}_1 \in \mathbb{C}^{(N-K)\times K}$, $\mathbf{h}_s \in \mathbb{C}^{(N-K)\times 1}$. When no noise is present $\mathbf{H}_s = \bar{\mathbf{H}}_s = [\bar{\mathbf{H}}_1 \ \bar{\mathbf{h}}_s]$ and the rank of $[\bar{\mathbf{H}}_1 \ \bar{\mathbf{h}}_s]$ is K implying the existence of a vector $\mathbf{x}$ such that $\bar{\mathbf{H}}_1\mathbf{x} = \bar{\mathbf{h}}_s$. By adding noise, the corresponding set of linear equations $\mathbf{H}_1\mathbf{x} \approx \mathbf{h}_s$ is no longer compatible, but corrections $[\Delta\mathbf{H}_1 \ \Delta\mathbf{h}_s]$ can be computed using a STLS algorithm such that $[\mathbf{H}_1 + \Delta\mathbf{H}_1 \ \mathbf{h}_s + \Delta\mathbf{h}_s]$ has exactly rank K and still has Hankel structure. Comparing $\mathbf{H}_1$, $\mathbf{h}_s$ with Eq. (2.17), learns that $\mathbf{h}_s = \mathbf{h}_{lp}$ and $\mathbf{H}_1 = \mathbf{H}_{lp}$ if M is taken to be equal to K in Eq. (2.17). The problem as stated above thus reduces to the one stated in Eq. (2.20) or Eq. (2.21). Once the optimal rank K approximation of the matrix $\mathbf{H}_s$ is found, the parameters are determined as outlined above for the HSVD algorithm. Note that Eq. (2.36) and Eq. (2.37) now hold exactly and that the such obtained parameters have ML properties. It is clear that the optimal solution of the LP and state-space problems give the same results since both start from the same optimal rank K matrix. Only the way in which the parameters are derived differs.

2.1.3 Remarks

The underlying principles of the matrix pencil (MP) method [67] applied to MRS data in [81], are very similar to the ideas used to derive the state-space formalism. Polynomial rooting and root selection are avoided since similar to the state-space methods, the method finds the estimates of the signal poles as eigenvalues of a matrix. In fact in the noiseless data case the MP and state-space method give the same result. In the noisy data case, there is almost no difference between estimates obtained by both methods [106].

Recently black-box methods have been proposed in which some form of prior knowledge can be incorporated, but the limitations to the imposition of prior knowledge about model function parameters are inherent to this type of methods. In the literature, the following types of prior knowledge have been incorporated:

- known signal poles [22, 26, 27].

- known (frequency and) phase [25].
- known frequency difference and equal dampings [140].
- known signal poles and phases [25].

2.1.4 Summary

It was shown in the previous sections that minimizing

$$\|\mathbf{y} - \bar{\mathbf{y}}\|^2$$

gives the same ML parameter estimates as:

- minimizing the variable projection functional $\|\mathbf{y} - \mathbf{\Gamma}\mathbf{\Gamma}^{\dagger}\mathbf{y}\|^2$ and obtaining the estimates for the linear parameters by solving $\widehat{\mathbf{\Gamma}}\mathbf{c} \approx \mathbf{y}$ in the LS sense.
- Solve

$$\begin{aligned} &\min_{\mathbf{\Delta H}, \mathbf{\Delta h}, \mathbf{p}} \mathbf{q}^H \mathbf{q} \\ &\text{such that } -(\mathbf{H} + \mathbf{\Delta H})\mathbf{p} = \mathbf{h}_{lp} + \mathbf{\Delta h}, \\ &\text{and } [\mathbf{\Delta H}\ \mathbf{\Delta h}] \text{ has the same structure as } [\mathbf{H}\ \mathbf{h}], \\ &\mathbf{q} \text{ contains the different elements of } [\mathbf{\Delta H}\ \mathbf{\Delta h}]. \end{aligned}$$

  - $\mathbf{H} = \mathbf{H}_{lp} \in \mathbb{C}^{(N-K)\times K}, \mathbf{h} = \mathbf{h}_{lp} \in \mathbb{C}^{(N-K)\times 1}, \mathbf{\Delta H} = \mathbf{\Delta H}_{lp}, \mathbf{\Delta h} = \mathbf{\Delta h}_{lp}$ in the LP problem setting. Once the K optimal LP coefficients $\hat{\mathbf{p}}$ are obtained, estimates of the frequencies and dampings can be derived from the roots of the polynomial $1 + \sum_{m=1}^{K} \hat{p}_m z^{-m} = 0$. The estimates for amplitudes and phases are obtained by solving $\widehat{\mathbf{\Gamma}}\mathbf{c} \approx \mathbf{y}$ in the LS sense.
  - $\mathbf{H} = \mathbf{H}_1 \in \mathbb{C}^{(N-K)\times K}, \mathbf{h} = \mathbf{h}_s \in \mathbb{C}^{(N-K)\times 1}, \mathbf{\Delta H} = \mathbf{\Delta H}_1, \mathbf{\Delta h} = \mathbf{\Delta h}_s$ in the state-space setting. Determine the eigenvalues from $\hat{\mathbf{E}}$ obtained by solving $\mathbf{U}_K^{\uparrow} = \mathbf{U}_{K\downarrow}\mathbf{E}$. Note that this equation holds *exactly*. $\mathbf{U}_K$ is here the matrix derived by the SVD of the optimal rank K approximation $[\mathbf{H}_1 + \mathbf{\Delta H}_1\ \ \mathbf{h}_s + \mathbf{\Delta h}_1]$. Estimates for frequencies and dampings are obtained from the eigenvalues of $\mathbf{E}$. The estimates for amplitudes and phases are obtained by solving $\widehat{\mathbf{\Gamma}}\mathbf{c} \approx \mathbf{y}$ in the LS sense.
- minimizing $\mathbf{p}^H\mathbf{H}^H(\mathbf{P}\mathbf{P}^H)^{-1}\mathbf{H}\mathbf{p}$ w.r.t. $\mathbf{p}$. Once the K optimal LP coefficients $\hat{\mathbf{p}}$ are obtained, estimates of the frequencies and dampings can be derived from the roots of the polynomial $1 + \sum_{m=1}^{K} \hat{p}_m z^{-m} = 0$. The estimates for amplitudes and phases are obtained by solving $\widehat{\mathbf{\Gamma}}\mathbf{c} \approx \mathbf{y}$ in the LS sense.

On the other hand, LPSVD, LPTLS, HSVD, HTLS possibly combined with signal enhancement e.g. the minimum variance technique or Cadzow's method provide statistically suboptimal solutions.

2.2 Frequency-domain methods

Frequency-domain methods can be divided into two classes. The *non-parametric* frequency-domain methods are based on integration of the peak area of the frequency-domain signal [89]. The advantage of these methods is that no assumptions have to be made concerning the signal. The major drawback is the low estimation accuracy. The accuracy of the integration based methods requires appropriate phasing (which is far from trivial) and is dependent on the definition of the integration area, especially in those cases where the peaks are not well separated or when missing time-domain data, acquisition artifacts or unidentified broad resonances distort the frequency-domain baseline.

More advanced *parametric* frequency-domain methods have been presented that rely on a model function for the metabolite peaks. The methods are often based on a time-domain model. The model is transformed into a frequency-domain model by use of the DFT as in done in Eq. (1.8) for a Lorentzian model. If another model than the Lorentzian one is used (e.g. Gauss or Voigt), a simple exact analytical expression for the DFT of the model function is not available [62]. It is however always possible to numerically compute the model function in the frequency domain by taking the DFT of the model function in the time domain [114]. To obtain the parameters, the difference between the DFT of the data and the frequency-domain model function is minimized

$$\sum_{l=0}^{N-1} |y_l - \overline{y}_l|^2. \tag{2.40}$$

The solution of this NLLS problem is the same as the one obtained by minimizing Eq. (2.2) directly in the time domain. See Appendix A for a formal proof of this statement. In order to get rid of distorted signal samples or to reduce the influence of broad resonances underlying metabolites of interest, the first data points are often discarded. It is often believed that this effect can only be accounted for correctly in the time domain. This is however not true at all since this truncation of data points is perfectly modeled by the parameter t_0. The formalism above also allows the imposition of prior knowledge into the function to be minimized in the same way as is done in the time domain. In summary, if carried out properly, time- and frequency-domain fitting are equivalent from a theoretical point of view. Note that in the early stages of model fitting in the frequency domain, often a sampled version of the theoretical spectrum obtained by *continuous* Fourier transformation of a continuous time-domain signal was

used to fit the measured data. Since there are several discrepancies between the DFT spectrum and the continuous one, the results obtained as such are not optimal (see also [3, 56, 85] for further details). Examples of methods in which Eq. (2.40) is either solved directly or using some approximations can be found in [96, 76, 64, 30].

It is now shown that if the MRS data are modeled as uniformly sampled, damped complex exponentials, methods developed in the frequency domain identification and control literature can be used to retrieve the parameters. Transforming the model of Eq. (2.10) by use of the DFT results in the following expression for $\overline{\mathcal{Y}}_l,\ l = 0, \ldots, N-1$

$$\begin{aligned}
\overline{\mathcal{Y}}_l &= \sum_{n=0}^{N-1} \bar{y}(n) e^{\frac{-j2\pi nl}{N}}, \\
&= \sum_{k=1}^{K} c_k \frac{1 - z_k^N}{1 - z_k e^{-j2\pi l/N}} \\
&= \frac{\sum_{k=0}^{K-1} \alpha_k s_l^{-k}}{1 + \sum_{k=1}^{K} \beta_k s_l^{-k}} \quad \text{with } s_l = e^{j2\pi l/N}.
\end{aligned}$$

The problem has been reformulated in terms of the parameter vector $\boldsymbol{\beta} = [\alpha_0 \ldots \alpha_{K-1}\ \beta_1 \ldots \beta_K]$. The DFT of a circular complex white noise sequence with standard deviation σ for the real and imaginary parts, is again circular complex white with standard deviation $N\sigma$ [56]. The Gaussian distribution is retained. So the DFT of the complete model, including noise is given by

$$\mathcal{Y}_l = \frac{\sum_{k=0}^{K-1} \alpha_k s_l^{-k}}{1 + \sum_{k=1}^{K} \beta_k s_l^{-k}} + N_l, \tag{2.41}$$

with N_l circular complex white Gaussian noise.

In the control and identification literature one is often concerned with estimating the real parameters $\boldsymbol{\beta}' = [\alpha'_0 \ldots \alpha'_n\ \beta'_0 \ldots \beta'_d]$ of the rational transfer function $H(s_l, \boldsymbol{\beta}')$ [102]

$$H(s_l, \boldsymbol{\beta}') = \frac{\sum_{k=0}^{n} \alpha'_k s_l^k}{1 + \sum_{k=1}^{d} \beta'_k s_l^k} \quad \text{with } s_l = e^{j2\pi l/N}, \tag{2.42}$$

using a discrete set of measured input and output spectra $\mathcal{X}'_l$, $\mathcal{Y}'_l$. The following relations exist between input and output spectra in a stochastic setting [102]

$$\begin{aligned}
\mathcal{Y}'_l &= H\overline{\mathcal{X}_l}' + N_{1_l} \\
\mathcal{X}'_l &= \overline{\mathcal{X}_l}' + N_{2_l},
\end{aligned} \tag{2.43}$$

where N_{1_l}, N_{2_l} denote noise sequences which are correlated with each other and the bar on the variables indicates that these are noise-free quantities. The

estimation of the parameters of Eq. (2.41) and $\boldsymbol{\beta}'$ of Eq. (2.42) is equivalent if:

- n and d of Eq. (2.42) are chosen equal to $K-1$ and K respectively
- the parameter vector $\boldsymbol{\beta}'$ of Eq. (2.42) is generalized to complex values
- $\overline{\mathcal{X}_l}' = 1$ for all values of l
- the input noise N_{2_l} is equal to zero

2.3 Other methods

In Bayesian estimation a prior probability function of each model parameters is required. As an outcome of the analysis a posterior probability function containing the wanted information is provided. For an application of Bayesian estimation in the time domain see [14, 15, 16, 17, 18, 4]. In [107] the Bayesian principle is applied in the frequency domain. Another conceptually different approach to *quantify* MRS signals and first reported in [65] is the use of an artificial neural network (ANN). The problem with the use of an ANN is the necessity of obtaining an acceptable experimental training data set of which the actual amplitudes of the metabolites of each spectrum are known accurately. In an attempt to overcome these problems, the training sets were *simulated* in [68]. A different approach to LP spectral analysis used e.g. in [8] is offered by determining the final parameter estimates based on a statistical analysis of the results obtained by performing a large number of LP runs using different prediction lengths and number of data points. Wavelets have also been used for quantitation purposes, see e.g. [113].

2.4 Conclusions

In this chapter an overview of algorithms used to estimate the parameters of MRS signals was given. Time- and frequency domain methods were discussed. The time-domain methods can be divided in two classes:

- Interactive methods that fit a model function to the data and minimize a cost function. The methods are very flexible in terms of possible model functions and prior knowledge that can be imposed. If the assumptions concerning the model function and the noise are satisfied, ML parameter estimates are obtained. The drawback is that a solution to a difficult minimization problem has to be obtained.

- Black-box methods which are restricted in the sense that they can only be applied to uniformly sampled, complex, damped exponentials and that only limited forms of prior knowledge can be imposed. They are however easy to use and require no user expertise. LP methods and state-space variants exist. Most algorithms only approximately solve the LP or state-space problem setting. It was shown here that the LP and state-space problems provide ML parameter estimates if they are solved without introducing approximations. However, then again the problems reduce to the solution of a minimization problem.

It was also shown that model fitting in the frequency and time domain are totally equivalent. The relation with frequency domain techniques used in the control and identification literature was pointed out.

Chapter 3

VARPRO and AMARES

The aim of the study presented in this chapter is to develop an improved variant of VARPRO, a MRS tuned fitting algorithm that minimizes the variable projection function of Eq. (2.3) using a simple NLLS algorithm. It was seen in the previous chapter that in the class of interactive methods one has the choice between minimizing Eq. (2.3) and Eq. (2.6). From a theoretical point of view, the parameter vector which minimizes those two equations is the same. Since computational efficiency is important and good starting values can be obtained, only traditional optimization methods are considered here. In this chapter, an experiment is set up to determine whether minimizing the general functional of Eq. (2.3) or the variable projection functional of Eq. (2.6) is better from a computational and robustness point of view. Furthermore, various NLLS algorithms are used to minimize Eq. (2.3) and Eq. (2.6) to investigate if they exhibit a significant difference in performance. Based on these findings, AMARES is derived.
The main objectives of this chapter are:

- *To explain the algorithmic principles underlying VARPRO and AMARES.*
- *To show how AMARES improves VARPRO in terms of*
  - *algorithmic aspects,*
  - *increased flexibility concerning the imposition of prior knowledge and choice of model function.*

Section 3.1 starts with a short overview of the most important unconstrained minimization algorithms and treats the solution to NLLS problems as a special case. The NLLS algorithms actually used in the comparative study which is

carried out in section 3.3 are explained in more detail. Section 3.2 discusses practical implementation aspects specific to the MRS model fitting problem. In section 3.3 a study is carried out in which different NLLS algorithms are used to minimize both types of functionals. Section 3.4 explains the types of prior knowledge that can be imposed in VARPRO and AMARES. In section 3.5 the features of AMARES are summarized. Section 3.6 presents the main conclusions of this chapter.

This chapter is based on work presented in [146, 147, 143].

3.1 Unconstrained minimization

This section contains background material on unconstrained minimization algorithms which form the basis for the remainder of this chapter. For a thorough treatment of the topic see e.g. [43, 52, 58, 97]. The aim of this section is to situate the NLLS algorithms that appear in the study of section 3.3 where they are used to minimize both the general and variable projection functional. Section 3.1.1 treats the most common general purpose unconstrained minimization algorithms. Section 3.1.2 discusses NLLS algorithms and section 3.1.3 explains the principles of the versions that are actually used in the comparative study of section 3.3.

3.1.1 General purpose unconstrained minimization

The unconstrained minimization problem seeks a local minimizer of a real-valued function $f(\mathbf{x})$ where $\mathbf{x}$ is a vector consisting of real variables $x_1, \ldots, x_n$. A point $\mathbf{x}^*$ is called a strong local minimizer if there is a neighborhood around $\mathbf{x}^*$, in which no other point $\mathbf{x}$ gives a value of f as small as $f(\mathbf{x}^*)$. A weak local minimizer is a point $\mathbf{x}^*$ such that there is a neighborhood around $\mathbf{x}^*$ in which no $\mathbf{x}$ gives a smaller value of $f(\mathbf{x}^*)$. In many cases an analytic expression for the minimum is not available. The minimum has to be obtained by an iterative procedure which generates a sequence of points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$, hopefully converging to the solution $\mathbf{x}^*$ of the problem. It will be assumed that f is a smooth, continuous and continuously differentiable function. This implies that at any point $\mathbf{x}$ there is a vector of first partial derivatives, the gradient vector $[\delta f/\delta x_1, \ldots, \delta f/\delta x_n]^T = \nabla f(\mathbf{x})$. If $f(\mathbf{x})$ is twice continuously differentiable, there exists a matrix of second partial derivatives, the Hessian matrix $\nabla^2 f(\mathbf{x})$ of which the element in row i and column j is given by $\delta^2 f/(\delta x_i \delta x_j)$.
The sufficient conditions for $\mathbf{x}^*$ to be a strong local minimum of the problem are

1. $\nabla f(\mathbf{x}^*) = \mathbf{0}$

2. $\nabla^2 f(\mathbf{x}^*)$ is positive definite.

A point $\mathbf{x}$ for which $\nabla f(\mathbf{x}) = 0$ is called a stationary point.

In many iterative procedures, the function around the current iterate is approximated by a simpler function obtained from a Taylor series expansion. The Taylor series expansion of $f(\mathbf{x})$ around $\mathbf{x}^{(k)}$ is

$$f(\mathbf{x}^{(k)} + \boldsymbol{\delta}) = f(\mathbf{x}^{(k)}) + \boldsymbol{\delta}^T \nabla f(\mathbf{x}^{(k)}) + \frac{1}{2}\boldsymbol{\delta}^T \nabla^2 f(\mathbf{x}^{(k)})\boldsymbol{\delta} + \ldots$$

In a sufficiently small neighborhood of $\mathbf{x}^{(k)}$, the function $f(\mathbf{x})$ can be closely approximated by the quadratic function $q^{(k)}(\boldsymbol{\delta})$

$$f(\mathbf{x}^{(k)} + \boldsymbol{\delta}) \approx q^{(k)}(\boldsymbol{\delta}) = f(\mathbf{x}^{(k)}) + \boldsymbol{\delta}^T \nabla f(\mathbf{x}^{(k)}) + \frac{1}{2}\boldsymbol{\delta}^T \nabla^2 f(\mathbf{x}^{(k)})\boldsymbol{\delta}.$$

In the basic *Newton* method, the next iterate $\mathbf{x}^{(k+1)}$ is derived from the result of the minimization of $q^{(k)}(\boldsymbol{\delta})$. Provided the Hessian matrix is positive definite in $\mathbf{x}^{(k)}$, the quadratic function has a unique minimizer given by the solution to

$$\begin{aligned} \nabla q^{(k)}(\boldsymbol{\delta}) &= 0 \\ \text{or} \quad \nabla^2 f(\mathbf{x}^{(k)})\boldsymbol{\delta} &= -\nabla f(\mathbf{x}^{(k)}). \end{aligned} \tag{3.1}$$

The solution of Eq. (3.1) provides the step $\boldsymbol{\delta} = \boldsymbol{\delta}^{(k)}$ in the basic Newton iteration scheme which proceeds as follows

$$\text{Solve} \quad \nabla^2 f(\mathbf{x}^{(k)})\boldsymbol{\delta} = -\nabla f(\mathbf{x}^{(k)}) \text{ for } \boldsymbol{\delta} = \boldsymbol{\delta}^{(k)} \tag{3.2}$$

$$\text{Set} \quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}. \tag{3.3}$$

Convergence is guaranteed if the starting point $\mathbf{x}^{(1)}$ is sufficiently close to a local minimum $\mathbf{x}^*$ at which the Hessian is positive definite. Problems with this basic scheme arise since the Hessian at $\mathbf{x}^{(k)}$ may not be positive definite when $\mathbf{x}^{(k)}$ is remote from the solution. If the Hessian is not positive definite the quadratic model need not have a minimum, nor even a stationary point. A unique stationary point exists only if the Hessian is non-singular. In case the Hessian is indefinite, the function is unbounded below. If the Hessian is singular there will be a stationary point only if the gradient vector lies in the column range of the Hessian. If the Hessian is positive semi-definite, a stationary point (if it exists) is a weak local minimum. Moreover, even if the Hessian is positive definite, the Newton iterates may not even decrease. To ensure convergence to a local minimum, the basic Newton method has to be extended by either a line-search or a trust-region approach.

Line-search methods generate the iterates by setting

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \nu^{(k)}\mathbf{p}^{(k)},$$

where $\mathbf{p}^{(k)}$ is the search direction and $\nu^{(k)} > 0$ is chosen such that $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$. Different variants exist to determine a suitable $\nu^{(k)}$. Most line-search versions of the basic Newton method generate the direction $\mathbf{p}^{(k)}$ by modifying the Hessian matrix computed at $\mathbf{x}^{(k)}$ to ensure that the quadratic model has a unique minimizer.

In the trust-region methods a different way of reasoning is used. If the Hessian is not positive definite, the region about $\mathbf{x}^{(k)}$ in which the Taylor series is adequate, does not include a point that minimizes $q^{(k)}(\boldsymbol{\delta})$. Therefore the method assumes that there is a neighborhood of $\mathbf{x}^{(k)}$ where $q^{(k)}(\boldsymbol{\delta})$ agrees with $f(\mathbf{x}^{(k)}+\boldsymbol{\delta})$ in some way and looks for the correction $\boldsymbol{\delta}^{(k)}$ that minimizes $q^{(k)}(\boldsymbol{\delta})$ for all the values in that neighborhood

$$\begin{aligned} \min_{\boldsymbol{\delta}} \quad & q^{(k)}(\boldsymbol{\delta}) \\ \text{subject to} \quad & \|\mathbf{D}^{(k)}\boldsymbol{\delta}\| \leq h^{(k)}, \end{aligned} \tag{3.4}$$

where $h^{(k)}$ can be seen as the region in which the quadratic model is valid and $\mathbf{D}^{(k)}$ is a scaling matrix. It can be shown [52] that if $\lambda^{(k)}$ is a scalar ≥ 0 such that the matrix $(\nabla^2 f(\mathbf{x}^{(k)}) + \lambda^{(k)} {\mathbf{D}^{(k)}}^T \mathbf{D}^{(k)})$ is positive semi-definite, the solution of the equations

$$(\nabla^2 f(\mathbf{x}^{(k)}) + \lambda^{(k)} {\mathbf{D}^{(k)}}^T \mathbf{D}^{(k)})\boldsymbol{\delta} = -\nabla f(\mathbf{x}^{(k)}), \tag{3.5}$$

solves problem (3.4) if either $\lambda^{(k)} = 0$ and $\|\mathbf{D}^{(k)}\boldsymbol{\delta}\| \leq h^{(k)}$ or $\lambda^{(k)} > 0$ and $\|\mathbf{D}^{(k)}\boldsymbol{\delta}\| = h^{(k)}$. If $h^{(k)}$ is large enough, the solution to problem (3.4) is simply the Newton direction. Again several versions exist which differ in the particular choice of parameters involved in this scheme.

If the Hessian matrix is not available, it can be replaced by an approximation. The most obvious way to do this is by replacing the Hessian by its finite difference approximation. A much more important class of methods in this respect however are the so-called *quasi-Newton* methods. An approximation of the Hessian is corrected or updated from iteration to iteration. Various conditions are imposed on the approximate Hessian, like e.g. in most cases positive definiteness. Different Hessian updating formulas exist.

3.1.2 NLLS minimization

The NLLS problem has the general form

$$f(\mathbf{x}) = \sum_{i=1}^{m} r_i(\mathbf{x})^2 = \mathbf{r}^T\mathbf{r}. \tag{3.6}$$

An alternative way of viewing such problems is that they arise from an attempt to solve the system of m equations $r_i(\mathbf{x}) \approx 0, i = 1, \ldots, m$. Therefore the functions $r_i(\mathbf{x})$ are also called the *residuals* of the equations. In MRS applications

m is always larger than the number of variables n. The derivatives of $f(\mathbf{x})$ are given by

$$\nabla f(\mathbf{x}) = 2\mathbf{J}^T\mathbf{r} \tag{3.7}$$

$$\nabla^2 f(\mathbf{x}) = 2\mathbf{J}^T\mathbf{J} + 2\sum_{i=1}^{m} r_i(\mathbf{x})\nabla^2 r_i(\mathbf{x}), \tag{3.8}$$

where

$$\mathbf{J}(\mathbf{x}) = [\nabla r_1, \ldots, \nabla r_m]^T \tag{3.9}$$

is the Jacobian matrix, the rows of which are the transpose of the first derivative vectors $\nabla r_i, i = 1, \ldots, m$ of the components of $\mathbf{r}$ [1]. It is possible to use these formulae in a conventional way, that is either Eq. (3.7) in combination with a quasi-Newton method, or both Eqs. (3.7) and (3.8) with a Newton method. However, in many cases the components r_i are small and this suggests that a good approximation of the Hessian can be obtained by using the following formula

$$\nabla^2 f(\mathbf{x}) \approx 2\mathbf{J}^T\mathbf{J}. \tag{3.10}$$

In most practical circumstances, the first term in Eq. (3.8) is indeed more important than the second term. These type of problems are called small-residual problems.

An algorithm that is particularly suited to solve the small-residual case and uses Eq. (3.10) to approximate the Hessian is the *Gauss-Newton* algorithm. The basic Gauss-Newton iteration scheme proceeds as follows

$$\begin{aligned} &\text{Solve} \quad \mathbf{J}^{(k)^T}\mathbf{J}^{(k)}\boldsymbol{\delta} = -\mathbf{J}^{(k)^T}\mathbf{r}^{(k)} \text{ for } \boldsymbol{\delta} = \boldsymbol{\delta}^{(k)} \\ &\text{Set} \quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}. \end{aligned} \tag{3.11}$$

Convergence is only guaranteed if $\mathbf{x}^{(1)}$ is sufficiently close to $\mathbf{x}^*$ and $\sum_{i=1}^{m} r_i(\mathbf{x}^*)\nabla^2 r_i(\mathbf{x}^*)$ is small. In view of these observations, Gauss-Newton codes in practice perform a line search along the direction $\boldsymbol{\delta}^{(k)}$ to obtain the new iterate. This direction is always a descent direction in case the Jacobian is non-singular. When the Jacobian is of rank less than n, Eq. (3.11) has a multiplicity of solutions. In these circumstances, Gauss-Newton algorithms select a particular solution (e.g. the one of least norm).

The *Levenberg-Marquardt* algorithm can be seen as a trust-region modification of the basic Gauss-Newton algorithm. The correction $\boldsymbol{\delta}^{(k)}$ is obtained by solving the system

$$(\mathbf{J}^{(k)^T}\mathbf{J}^{(k)} + \lambda^{(k)}\mathbf{D}^{(k)^T}\mathbf{D}^{(k)})\,\boldsymbol{\delta} = -\mathbf{J}^{(k)^T}\mathbf{r}^{(k)} \text{ for } \boldsymbol{\delta} = \boldsymbol{\delta}^{(k)}. \tag{3.12}$$

[1] Note that in some textbooks, like e.g. in [52], $\mathbf{J}$ is defined as the transpose of the matrix defined in Eq. (3.9).

Many different algorithms of this type have been suggested. In some (early) versions the restriction of $\|\mathbf{D}^{(k)}\boldsymbol{\delta}\| \leq h^{(k)}$ on the length of $\boldsymbol{\delta}$ is not imposed. Instead $\lambda^{(k)}$ is used as the controlling parameter in the iteration and the length of $\boldsymbol{\delta}$ is determined by whatever value $\lambda^{(k)}$ happens to take (see also section 3.1.3.1 for an example of a Levenberg-Marquardt implementation where this approach is adopted).

The Gauss-Newton and Levenberg-Marquardt methods usually exhibit quadratic convergence for zero-residual ($r_i(\mathbf{x}^*) = 0, i = 1, \ldots, m$) problems. Otherwise, the convergence is only linear. To increase the convergence rate, the so-called *hybrid* methods try to increase the convergence speed by using some of the ideas applied in quasi-Newton methods which are superlinearly convergent (see section 3.1.3.3).

3.1.3 Examples of practical NLLS algorithms

3.1.3.1 Levenberg-Marquardt -version 1

Below a simple version of the Levenberg-Marquardt algorithm [100] as it is actually implemented in VARPRO is discussed in detail to give a basic understanding of the computational complexity associated with the use of a NLLS algorithm.

At every iteration step, the solution to Eq. (3.12) has to be computed. The solution to Eq. (3.12) can be obtained by solving the equations directly, but they can also be solved by recognizing that Eq. (3.12) are the normal equations for the LS problem

$$\begin{bmatrix} \mathbf{J}^{(k)} \\ \sqrt{\lambda^{(k)}}\mathbf{D}^{(k)} \end{bmatrix} \boldsymbol{\delta}^{(k)} = \begin{bmatrix} -\mathbf{r}^{(k)} \\ \mathbf{0} \end{bmatrix}. \tag{3.13}$$

The main advantage of solving the normal equations is speed, Eq. (3.12) can be solved twice as fast as Eq. (3.13). On the other hand, solving Eq. (3.13) leads in general to more accurate results [59]. In the following it is assumed that $\mathbf{J}^{(k)}$ is of full column rank n. To solve Eq. (3.13) the following procedure can be used. In the first step a QR factorization of $\mathbf{J}^{(k)}$ is computed [59]

$$\mathbf{Q}_1^{(k)}\mathbf{J}^{(k)} = \begin{bmatrix} \mathbf{R}_1^{(k)} \\ \mathbf{0} \end{bmatrix},$$

where $\mathbf{Q}_1^{(k)} \in \mathbb{R}^{m\times m}$ is an orthogonal matrix and $\mathbf{R}_1^{(k)} \in \mathbb{R}^{n\times n}$ is an upper triangular matrix. This factorization can e.g. be accomplished using Householder transformations.

A Householder transformation is defined by an orthogonal matrix $\mathbf{P} \in \mathbb{R}^{m \times m}$ of the form

$$\mathbf{P} = \mathbf{I}_m - 2\mathbf{v}\mathbf{v}^T/(\mathbf{v}^T\mathbf{v}),$$

in which $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ is the identity matrix and $\mathbf{v} \in \mathbb{R}^m$. Given a non-zero vector $\mathbf{x}$, $\mathbf{P}\mathbf{x}$ can be made a multiple of the unit vector $\mathbf{u}_1 = [1 \quad 0 \ldots 0]^T$ by choosing the Householder vector $\mathbf{v}$ equal to

$$\mathbf{v} = \mathbf{x} + \text{sign}(x_1)\|\mathbf{x}\|\mathbf{u}_1,$$

where x_1 is the first element of $\mathbf{x}$ and $\text{sign}(a) = 1$ if $a \geq 0$ and $\text{sign}(a) = -1$ if $a < 0$. Multiplication of a matrix $\mathbf{A}$ with the Householder matrix $\mathbf{P}$ can be done efficiently by first performing the matrix-vector multiplication $\mathbf{v}^T\mathbf{A}$, followed by multiplication with the scalar $-2/(\mathbf{v}^T\mathbf{v})$ and the multiplication of $\mathbf{v}$ and the thus obtained row vector

$$\mathbf{P}\mathbf{A} = \mathbf{A} + \mathbf{v}\left(\left(\frac{-2}{\mathbf{v}^T\mathbf{v}}\right)(\mathbf{v}^T\mathbf{A})\right).$$

The matrix $\mathbf{Q}_1^{(k)}$ is obtained by multiplication of n Householder matrices $\mathbf{P}_n, \ldots, \mathbf{P}_1$. The ith Householder matrix $\mathbf{P}_i$ has the form

$$\mathbf{P}_i = \mathbf{I} - 2\mathbf{v}_i\mathbf{v}_i^T/(\mathbf{v}_i^T\mathbf{v}_i)$$

and $\mathbf{v}_i = \begin{bmatrix} \mathbf{0}_{i-1}^T & \mathbf{v}_i' \end{bmatrix}$ where $\mathbf{v}_i'$ is designed to zero the last $(m-i)$ elements of the ith column of $\mathbf{P}_{i-1}\mathbf{P}_{i-2}\ldots\mathbf{P}_1\mathbf{J}^{(k)}$, $\mathbf{0}_{i-1}$ is the zero vector of length $(i-1)$. The matrix $\mathbf{Q}_1^{(k)}$ is not formed explicitly. The computation of the QR decomposition of an $m \times n$ matrix using Householder transformations requires around $2n^2(m - n/3)$ floating point operations (flops) [59].

In the second stage the matrix $\begin{bmatrix} \mathbf{R}_1^{(k)^T} & \mathbf{0} & \sqrt{\lambda^{(k)}}\mathbf{D}^{(k)^T} \end{bmatrix}^T$ is further reduced to triangular form $\mathbf{R}_2^{(k)}$ by applying consecutive Householder transformations represented by $\mathbf{Q}_2^{(k)}$, thereby exploiting the fact that the matrix contains a band of zeros as is done in [59]. The transformations $\mathbf{Q}_1^{(k)}$ and $\mathbf{Q}_2^{(k)}$ are also applied to the right hand side of Eq. (3.13). Finally, the solution to the LS problem is obtained by back substitution. The total number of flops required to solve this LS problem is still of the order of $2n^2(m - n/3)$, the flops associated with the reduction of $\begin{bmatrix} \mathbf{R}_1^{(k)^T} & \mathbf{0} & \sqrt{\lambda^{(k)}}\mathbf{D}^{(k)^T} \end{bmatrix}^T$ to upper triangular form, the application of the Householder transformations to the right hand side of Eq. (3.13) and the back substitution are not significant compared to the work required to factor $\mathbf{J}^{(k)}$. An outline of the complete Levenberg-Marquardt algorithm as implemented in VARPRO is given below. This Levenberg-Marquardt code is part of the varpro routines obtainable from the opt library of Netlib [2] [53].

[2] Netlib is a repository of public-domain numerical software.

Algorithm 3.1.1 (Levenberg-Marquardt)

Input: starting values for the unknown parameters $\mathbf{x}^{(1)}$.
Output: parameter vector $\mathbf{x}^{(it)}$, obtained after it iterations.

Step 1. $\lambda^{(1)} = 1$.

Step 2. Calculate $\mathbf{J}^{(k)}$.

Step 3. Compute QR factorization of $\mathbf{J}^{(k)}$ using Householder transformations, set IC=0.

Step 4. Reduce $\begin{bmatrix} \mathbf{R}_1^{(k)} \\ \mathbf{0} \\ \sqrt{\lambda^{(k)}}\mathbf{D}^{(k)} \end{bmatrix}$ to an upper triangular matrix using Householder transformations which preserve the band of zeros. $\mathbf{D}^{(k)}$ is a diagonal matrix consisting of the norms of the columns of $\mathbf{J}^{(k)}$.

Step 5. Calculate $\mathbf{r}^{(k)}$.

Step 6. Apply transformations in *Step 3* and *Step 4* to $\begin{bmatrix} -\mathbf{r}^{(k)} \\ \mathbf{0} \end{bmatrix}$ and calculate $\boldsymbol{\delta}^{(k)}$ by backward substitution.

Step 7. Compute $f(\mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)})$ and set IC=IC+1.

Step 8.
if $\|f(\mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)})\| < \|f(\mathbf{x}^{(k)})\|$
$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$
if IC=1 then $\sqrt{\lambda^{(k+1)}} = 0.5\sqrt{\lambda^{(k)}}$
go to *Step 9*
else $\sqrt{\lambda^{(k)}} = 1.5\sqrt{\lambda^{(k)}}$
go to *Step 4*
end

Step 9. Test convergence. If no convergence go to *Step 2*.

It is important to note that if in *Step 8* no satisfactory next iterate is obtained and *Step 4* has to be performed with a new value of $\lambda^{(k)}$, the factorization of $\mathbf{J}^{(k)}$ must not be redone. $\mathbf{J}^{(k)}$ and $\mathbf{r}^{(k)}$ in *Step 2* and *Step 5* respectively are calculated analytically and this is done based on subroutines provided by the user.

3.1.3.2 Levenberg-Marquardt - MINPACK version

The *lmder* routine from the MINPACK library [93] of Netlib is an implementation of the Levenberg-Marquardt algorithm in which the trust region approach is taken. The implementation is based on the ideas described in [92]. The *lmder* routine requests the user to provide a subroutine to calculate the residuals and the Jacobian.

3.1.3.3 Hybrid algorithm - NL2SOL

The basic principles underlying the most recent NL2SOL [41, 42] version from the PORT library of netlib are outlined. The algorithm combines approaches used in Gauss-Newton and quasi-Newton methods. A trust-region approach is used with an approximate Hessian matrix of the form $(\mathbf{J}^{(k)^T}\mathbf{J}^{(k)} + \mathbf{S}^{(k)})$, where $\mathbf{S}^{(k)}$ is a quasi-Newton approximation to the second order term in the true Hessian of Eq. (3.8). Low-rank corrections are applied to $\mathbf{S}^{(k)}$ at every iteration together with a scaling strategy that ensures that this matrix stays small when the residuals are small. At each iteration, a decision is made whether to take the Gauss-Newton step or the step that is computed by including the $\mathbf{S}^{(k)}$ term. An interesting feature of this code is the fact that it is possible to place bound constraints $\mathbf{low} \leq \mathbf{x} \leq \mathbf{upp}$ on the variables $\mathbf{x}$. The *dn2gb* routine also relies on subroutines provided by the user to calculate the residuals and the Jacobian.

3.2 Specific implementation issues

Some practical implementation details concerning the actual minimization of Eqs. (2.3) and (2.6) using either one of the algorithms in section 3.1.3 are given. The algorithms outlined in section 3.1.3 all request the user to supply subroutines to calculate the residual $\mathbf{r}$ and Jacobian $\mathbf{J}$. In this section it is shown how this can be done for both the functionals. To give an idea of the difference in computational complexity between minimizing V and minimizing G, an approximate complexity analysis is provided in case algorithm 3.1.1 is used to minimize the functionals. For the ease of notation, the data are assumed to be modeled by Eq. (1.3) and t_0 is supposed to be fixed in the remainder of this chapter unless otherwise specified. Extensions to other model functions are straightforward.

3.2.1 General functional

To perform the actual minimization of Eq. (2.3) the data are split into a real and imaginary part as follows

$$G = \|\mathbf{y} - \bar{\mathbf{y}}\|^2 = \|\mathbf{y}^G - \mathbf{\Gamma}^G \mathbf{l}^G\|^2 \tag{3.14}$$

where

$$\mathbf{\Gamma}^G = \begin{bmatrix} e^{-d_1 t_0}\cos(2\pi f_1 t_0 + \phi_1) & \dots \\ e^{-d_1 t_0}\sin(2\pi f_1 t_0 + \phi_1) & \dots \\ \vdots & \ddots \\ e^{-d_1 t_{N-1}}\cos(2\pi f_1 t_{N-1} + \phi_1) & \dots \\ e^{-d_1 t_{N-1}}\sin(2\pi f_1 t_{N-1} + \phi_1) & \dots \end{bmatrix}$$

$$\begin{matrix} \dots & e^{-d_K t_0}\cos(2\pi f_K t_0 + \phi_K) \\ \dots & e^{-d_K t_0}\sin(2\pi f_K t_0 + \phi_K) \\ \ddots & \vdots \\ \dots & e^{-d_K t_{N-1}}\cos(2\pi f_K t_{N-1} + \phi_K) \\ \dots & e^{-d_K t_{N-1}}\sin(2\pi f_K t_{N-1} + \phi_K) \end{matrix} \Bigg], \tag{3.15}$$

$$\mathbf{y}^G = \begin{bmatrix} Re(y_0) \\ Im(y_0) \\ \vdots \\ Re(y_{N-1}) \\ Im(y_{N-1}) \end{bmatrix} \quad \text{and} \quad \mathbf{l}^G = \begin{bmatrix} a_1 \\ \vdots \\ a_K \end{bmatrix}. \tag{3.16}$$

The starting values for the amplitudes and phases are derived from the LS solution $(\mathbf{\Gamma}^V)^\dagger \mathbf{y}^G$, with the starting values for the frequencies and dampings inserted in $\mathbf{\Gamma}^V$ (for the definition of $\mathbf{\Gamma}^V$ see Eq. 3.20).

The residual vector $\mathbf{r}$ of Eq. (3.6) in this specific case is given by

$$\mathbf{r}^G = \begin{bmatrix} r_1^G \\ r_2^G \\ \vdots \\ r_{2N-1}^G \\ r_{2N}^G \end{bmatrix} = \begin{bmatrix} Re(y_1) - \sum_{k=1}^K a_k e^{-d_k t_0}\cos(2\pi f_k t_0 + \phi_k) \\ Im(y_1) - \sum_{k=1}^K a_k e^{-d_k t_0}\sin(2\pi f_k t_0 + \phi_k) \\ \vdots \\ Re(y_{N-1}) - \sum_{k=1}^K a_k e^{-d_k t_{N-1}}\cos(2\pi f_k t_{N-1} + \phi_k) \\ Im(y_{N-1}) - \sum_{k=1}^K a_k e^{-d_k t_{N-1}}\cos(2\pi f_k t_{N-1} + \phi_k) \end{bmatrix}. \tag{3.17}$$

The iteration superscripts are dropped to simplify the notation involved. Minimization is performed w.r.t. $a_k, d_k, f_k, \phi_k, k = 1, \dots, K$. The Jacobian matrix

$\mathbf{J}^G$ can be straightforwardly evaluated at every iteration point by computing

$$\mathbf{J}^G = \begin{bmatrix} \frac{\delta r_1^G}{\delta a_1} & \cdots & \frac{\delta r_1^G}{\delta a_K} & \frac{\delta r_1^G}{\delta d_1} & \cdots & \frac{\delta r_1^G}{\delta d_K} & \frac{\delta r_1^G}{\delta f_1} & \cdots & \frac{\delta r_1^G}{\delta f_K} & \frac{\delta r_1^G}{\delta \phi_1} & \cdots & \frac{\delta r_1^G}{\delta \phi_K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\delta r_{2N}^G}{\delta a_1} & \cdots & \frac{\delta r_{2N}^G}{\delta a_K} & \frac{\delta r_{2N}^G}{\delta d_1} & \cdots & \frac{\delta r_{2N}^G}{\delta d_K} & \frac{\delta r_{2N}^G}{\delta f_1} & \cdots & \frac{\delta r_{2N}^G}{\delta f_K} & \frac{\delta r_{2N}^G}{\delta \phi_1} & \cdots & \frac{\delta r_{2N}^G}{\delta \phi_K} \end{bmatrix}. \quad (3.18)$$

The dimension of the matrix is $2N \times 4K$.

The computational complexity associated with the solution of the Levenberg-Marquardt equations using the scheme presented in section 3.1.3.1 is about $2(4K)^2(2N - 4K/3) = 64K^2N - (128/3)K^3$ flops in this specific case.

Until now, no prior knowledge was imposed. Introduction of prior knowledge typically leads to smaller dimensions of the Jacobian since the number of variables is reduced. As a result the computational cost per iteration decreases. The way prior knowledge is introduced into the minimization process is illustrated using an artificial example of a signal consisting of the 7 peaks of ATP. The prior knowledge as described in section 1.5 is imposed. The frequency difference between the individual resonances within a multiplet Δf is equal to 16 Hz. If the peaks are numbered as in Fig. 1.4, the prior knowledge imposed in terms of the parameters $a_1, \ldots, a_7, d_1, \ldots, d_7, f_1, \ldots f_7, \phi_1, \ldots, \phi_7$ e.g. becomes

- a_1 is a free parameter and is estimated. The amplitudes of the other peaks are not free, but linked to a_1

  $$a_2 = a_1,\ a_3 = a_1,\ a_4 = a_1,\ a_5 = 0.5a_1,\ a_6 = a_1,\ a_7 = 0.5a_1.$$

- d_1 is a free parameter and is estimated. The dampings of the other peaks are not free, but linked to d_1

  $$d_2 = d_1,\ d_3 = d_1,\ d_4 = d_1,\ d_5 = d_1,\ d_6 = d_1,\ d_7 = d_1.$$

- f_1, f_3, f_5 are free parameters and are estimated. The frequencies of the other peaks are linked to them

  $$f_2 = f_1 - 16,\ f_4 = f_3 - 16,\ f_6 = f_5 - 16,\ f_7 = f_5 - 32.$$

- ϕ_1 is a free parameter and is estimated. The phases of the other peaks are linked to ϕ_1

  $$\phi_2 = \phi_1,\ \phi_3 = \phi_1,\ \phi_4 = \phi_1,\ \phi_5 = \phi_1,\ \phi_6 = \phi_1,\ \phi_7 = \phi_1.$$

As a result of the imposition of prior knowledge the number of unknowns reduces to 6: $a_1, d_1, f_1, f_3, f_5, \phi_1$. The expression for r_1^G of Eq. (3.17) simplifies to

$$r_1^G = Re(y_1) - a_1 e^{-d_1 t_0}\left(\cos(2\pi f_1 t_0 + \phi_1) + \cos(2\pi(f_1 - 16)t_0 + \phi_1) + \cos(2\pi f_3 t_0 + \phi_1)\right.$$

$$+\cos(2\pi(f_3-16)t_0+\phi_1)+0.5\cos(2\pi f_5 t_0+\phi_1)+$$
$$\cos(2\pi(f_5-16)t_0+\phi_1)+0.5\cos(2\pi(f_5-32)t_0+\phi_1))\,.$$

The Jacobian $\mathbf{J}^{\mathrm{G}}$ becomes the $2N\times 6$ matrix

$$\mathbf{J}^{\mathrm{G}}=\begin{bmatrix} \frac{\delta r_1^G}{\delta a_1} & \frac{\delta r_1^G}{\delta d_1} & \frac{\delta r_1^G}{\delta f_1} & \frac{\delta r_1^G}{\delta f_3} & \frac{\delta r_1^G}{\delta f_5} & \frac{\delta r_1^G}{\delta \phi_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\delta r_{2N}^G}{\delta a_1} & \frac{\delta r_{2N}^G}{\delta d_1} & \frac{\delta r_{2N}^G}{\delta f_1} & \frac{\delta r_{2N}^G}{\delta f_3} & \frac{\delta r_{2N}^G}{\delta f_5} & \frac{\delta r_{2N}^G}{\delta \phi_1} \end{bmatrix}.$$

3.2.2 Variable projection functional

To minimize the variable projection functional the data are also partitioned into a real and imaginary part

$$V=\|\mathbf{y}^V-\mathbf{\Gamma}^V\mathbf{l}^V\|^2=\|\mathbf{y}^V-\mathbf{\Gamma}^V{\mathbf{\Gamma}^V}^\dagger\mathbf{y}^V\|^2=\|\mathbf{P}^\perp_{\mathbf{\Gamma}^V}\mathbf{y}^V\|^2 \tag{3.19}$$

where

$$\mathbf{\Gamma}^V=\begin{bmatrix} e^{-d_1t_0}\cos(2\pi f_1t_0) & -e^{-d_1t_0}\sin(2\pi f_1t_0) & \dots \\ e^{-d_1t_0}\sin(2\pi f_1t_0) & e^{-d_1t_0}\cos(2\pi f_1t_0) & \dots \\ \vdots & \vdots & \ddots \\ e^{-d_1t_{N-1}}\cos(2\pi f_1t_{N-1}) & -e^{-d_1t_{N-1}}\sin(2\pi f_1t_{N-1}) & \dots \\ e^{-d_1t_{N-1}}\sin(2\pi f_1t_{N-1}) & e^{-d_1t_{N-1}}cos(2\pi f_1t_{N-1}) & \dots \end{bmatrix}$$

$$\begin{bmatrix} \dots & e^{-d_Kt_0}\cos(2\pi f_Kt_0) & -e^{-d_Kt_0}\sin(2\pi f_Kt_0) \\ \dots & e^{-d_Kt_0}\sin(2\pi f_Kt_0) & e^{-d_Kt_0}\cos(2\pi f_Kt_0) \\ \ddots & \vdots & \vdots \\ \dots & e^{-d_Kt_{N-1}}\cos(2\pi f_Kt_{N-1}) & -e^{-d_Kt_{N-1}}\sin(2\pi f_Kt_{N-1}) \\ \dots & e^{-d_Kt_{N-1}}\sin(2\pi f_Kt_{N-1}) & e^{-d_Kt_{N-1}}\cos(2\pi f_Kt_{N-1}) \end{bmatrix}, \tag{3.20}$$

$$\mathbf{y}^V=\begin{bmatrix} Re(y_0) \\ Im(y_0) \\ \vdots \\ Re(y_{N-1}) \\ Im(y_{N-1}) \end{bmatrix},\quad \mathbf{l}^{\mathrm{V}}=\begin{bmatrix} a_1\cos(\phi_1) \\ a_1\sin(\phi_1) \\ \vdots \\ a_K\cos(\phi_K) \\ a_K\sin(\phi_K) \end{bmatrix}=\begin{bmatrix} Re(c_1) \\ Im(c_1) \\ \vdots \\ Re(c_K) \\ Im(c_K) \end{bmatrix}, \tag{3.21}$$

where $c_k=a_ke^{j\phi_k},\ k=1,\dots K$ are the previously defined complex amplitudes.

If $\mathbf{Q}$ is an orthogonal matrix, then for every vector $\mathbf{z}$,

$$\|\mathbf{Qz}\|=\|\mathbf{z}\|. \tag{3.22}$$

This property is used to simplify the evaluation of the residual $\|\mathbf{P}_{\mathbf{\Gamma}^V}^{\perp}\mathbf{y}^{\mathrm{V}}\|$. Consider the orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{2N\times 2N}$ such that

$$\mathbf{Q}\mathbf{\Gamma}^{\mathrm{V}} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}, \tag{3.23}$$

with $\mathbf{R} \in \mathbb{R}^{2K\times 2K}$ upper-triangular. The matrix $\mathbf{Q}$ is the product of $2K$ Householder matrices. Using Eq. (3.23) $\mathbf{P}_{\mathbf{\Gamma}^V}$ can be expressed as

$$\mathbf{P}_{\mathbf{\Gamma}^V} = \mathbf{Q}^T \begin{bmatrix} \mathbf{I}_{2K\times 2K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}, \tag{3.24}$$

where $\mathbf{I}_{2K\times 2K} \in \mathbb{R}^{2K\times 2K}$ is the identity matrix. If $\mathbf{Q}$ is partioned into

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix},$$

with $\mathbf{Q}_1 \in \mathbb{R}^{2K\times 2N}, \mathbf{Q}_2 \in \mathbb{R}^{(2N-2K)\times 2N}$ and using Eq. (3.24), the following property results

$$\mathbf{Q}\mathbf{P}_{\mathbf{\Gamma}^V}^{\perp} = \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}_2 \end{bmatrix}. \tag{3.25}$$

Define

$$\begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \mathbf{y}^{\mathrm{V}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \tag{3.26}$$

then using Eq. (3.22) and Eq. (3.25) the following is derived

$$\|\mathbf{P}_{\mathbf{\Gamma}^V}^{\perp}\ \mathbf{y}^{\mathrm{V}}\| = \|\mathbf{Q}\mathbf{P}_{\mathbf{\Gamma}^V}^{\perp}\ \mathbf{y}^{\mathrm{V}}\| = \|\mathbf{Q}_2\mathbf{y}^{\mathrm{V}}\| = \|\mathbf{y}_2\|.$$

Below a schematic outline of the evaluation of the residual as implemented in VARPRO is given.

Algorithm 3.2.1 (Residual evaluation (as implemented in VARPRO))

Input: current iteration point.
Output: residual vector.

Step 1. Calculate $\mathbf{\Gamma}^{\mathrm{V}}$ in the current iteration point using a user-supplied subroutine.

Step 2. Compute the QR factorization of $\mathbf{\Gamma}^{\mathrm{V}}$ using Householder transformations

$$\mathbf{Q}\mathbf{\Gamma}^{\mathrm{V}} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}.$$

The matrix $\mathbf{Q}$ is not formed explicitly. Only the information required to generate the transformation, i.e. the Householder vectors, are stored.

Step 3. Set $\mathbf{Q}\mathbf{y}^{\mathrm{V}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$. The residual is given by $\mathbf{y}_2$.

In [59], an expression for the Jacobian is derived

$$\mathbf{J}^{\mathrm{V}} = -\mathbf{P}^{\perp}_{\mathbf{\Gamma}^V}\ \mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})\mathbf{\Gamma}^B\mathbf{y}^{\mathrm{V}} - (\mathbf{\Gamma}^B)^T\left(\mathbf{P}^{\perp}_{\mathbf{\Gamma}^V}\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})\right)^T\mathbf{y}^{\mathrm{V}} \tag{3.27}$$

where

$$\mathbf{\Gamma}^B = \begin{bmatrix} \mathbf{R}^{-1} \\ \mathbf{0} \end{bmatrix} Q$$

and $\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})$ denotes a tridimensional tensor, consisting of $2K(2N \times 2K)$ matrices, each one containing the partial derivatives of the elements of $\mathbf{\Gamma}^{\mathrm{V}}$ w.r.t. one of the $2K$ variables. Since in general this tensor has many zero columns, only the non-zero columns are stored. If γ_{ij} denotes the element in row i and column j of $\mathbf{\Gamma}^{\mathrm{V}}$, a compact representation of $\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})$ is e.g. the $2N \times 4K$ matrix

$$\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}}) = \begin{bmatrix} \frac{\delta\gamma_{11}}{\delta d_1} & \frac{\delta\gamma_{12}}{\delta d_1} & \cdots & \frac{\delta\gamma_{1(2K-1)}}{\delta d_K} & \frac{\delta\gamma_{1(2K)}}{\delta d_K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\delta\gamma_{(2N)1}}{\delta d_1} & \frac{\delta\gamma_{(2N)2}}{\delta d_1} & \cdots & \frac{\delta\gamma_{(2N)(2K-1)}}{\delta d_K} & \frac{\delta\gamma_{(2N)(2K)}}{\delta d_K} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\delta\gamma_{11}}{\delta f_1} & \frac{\delta\gamma_{12}}{\delta f_1} & \cdots & \frac{\delta\gamma_{1(2K-1)}}{\delta f_K} & \frac{\delta\gamma_{1(2K)}}{\delta f_K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\delta\gamma_{(2N)1}}{\delta f_1} & \frac{\delta\gamma_{(2N)2}}{\delta f_1} & \cdots & \frac{\delta\gamma_{(2N)(2K-1)}}{\delta f_K} & \frac{\delta\gamma_{(2N)(2K)}}{\delta f_K} \end{bmatrix}.$$

In [69] Kaufman proposes to ignore the second term of Eq. (3.27) in order to reduce the computational complexity associated with the minimization of V. See [110] and [55] for a discussion concerning the convergence of the algorithm when this second term is ignored. Eq. (3.13) can then be rewritten in terms of the explicit expression for the residual and the approximate Jacobian

$$\begin{bmatrix} \mathbf{P}^{\perp}_{\mathbf{\Gamma}^V}\ \mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})\mathbf{\Gamma}^B\mathbf{y}^{\mathrm{V}} \\ \sqrt{\lambda}\mathbf{D} \end{bmatrix}\boldsymbol{\delta} = \begin{bmatrix} \mathbf{P}^{\perp}_{\mathbf{\Gamma}^V}\mathbf{y}^{\mathrm{V}} \\ \mathbf{0} \end{bmatrix}. \tag{3.28}$$

After multiplication of both sides of Eq. (3.28) from the left with $\begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$ and using Eqs. (3.25) and (3.26), Eq. (3.28) reduces to

$$\begin{bmatrix} \mathbf{Q}_2\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})\mathbf{\Gamma}^B\mathbf{y}^{\mathrm{V}} \\ \sqrt{\lambda}\mathbf{D} \end{bmatrix}\boldsymbol{\delta} = \begin{bmatrix} \mathbf{y}_2 \\ \mathbf{0} \end{bmatrix}. \tag{3.29}$$

The factor $\mathbf{\Gamma}^B\mathbf{y}^{\mathrm{V}}$ in Eq. (3.29) is nothing else than the LS solution to $\mathbf{\Gamma}^{\mathrm{V}}\mathbf{l}^{\mathrm{V}} \approx \mathbf{y}^{\mathrm{V}}$, i.e. an estimate of the complex amplitudes. Consider further the expression $\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})\mathbf{l}^{\mathrm{V}}$. The result of this tensor-vector multiplication is a matrix of size $(2N \times 2K)$ and is obtained as follows. Each column of $\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})$ corresponding to the derivative w.r.t. the same nonlinear parameter is multiplied with its corresponding element in $\mathbf{l}^{\mathrm{V}}$ and the columns are added. As a result the part of the matrix $\mathbf{J}^{\mathrm{G}}$ of Eq. (3.18) corresponding to the derivatives of the nonlinear parameters, d_k and $f_k, k = 1, \ldots, K$ is obtained. E.g. for d_1 it can easily be verified that the following holds

$$\begin{bmatrix} \frac{\delta\gamma_{11}}{\delta d_1} \\ \vdots \\ \frac{\delta\gamma_{(2N)1}}{\delta d_1} \end{bmatrix} l_1^{\mathrm{V}} + \begin{bmatrix} \frac{\delta\gamma_{12}}{\delta d_1} \\ \vdots \\ \frac{\delta\gamma_{(2N)2}}{\delta d_1} \end{bmatrix} l_2^{\mathrm{V}} = \begin{bmatrix} \frac{\delta r_1^{\mathrm{G}}}{\delta d_1} \\ \vdots \\ \frac{\delta r_{2N}^{\mathrm{G}}}{\delta d_1} \end{bmatrix}.$$

A schematic outline of the approximate Jacobian evaluation as implemented in VARPRO is given below.

Algorithm 3.2.2 (Jacobian evaluation (as implemented in VARPRO))

Input: current iteration point.
Output: approximate Jacobian.

Step 1. Calculate $\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})$ in the current iteration point using a user-supplied subroutine. Only the non-zero columns are stored.

Step 2. Solve $\mathbf{R}\mathbf{l}^{\mathrm{V}} = \mathbf{y}_1$ by back substitution.

Step 3. Apply Householder transformations, represented by $\mathbf{Q}$, to $\mathcal{D}(\mathbf{\Gamma}^{\mathrm{V}})$: $\mathbf{W} = -\mathbf{Q}\mathcal{D}(\Gamma^{\mathrm{V}})$. Partition $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix}$, with $\mathbf{W}_1 \in \mathbb{R}^{2K \times 2K}, \mathbf{W}_2 \in \mathbb{R}^{(2N-2K) \times 2K}$.

Step 4. The approximate Jacobian is given by $\mathbf{W}_2\mathbf{l}^{\mathrm{V}}$.

The computational complexity involved in the QR factorization of the approximate Jacobian is around $2(2K)^2(2N - 2K/3) = 16K^2N - (16/3)K^3$ flops which is considerably less than the amount of computations needed to factor the corresponding Jacobian matrix $\mathbf{J}^{\mathrm{G}}$. However, it is clear from the description above that the calculation complexity involved in evaluating the residual and approximate Jacobian is significant in case of minimization of the variable projection functional. From analysis of the two computational schemes above, the following can be derived:

- residual evaluation
  - *Step 2*: QR factorization of a $(2N \times 2K)$ matrix
    $$2(2K)^2(2N - (2K)/3) = 16K^2N - (16/3)K^3 \text{ flops.}$$
- Jacobian evaluation
  - *Step 3*: application of $\mathbf{Q}$ to $\mathcal{D}(\mathbf{\Gamma}^{\text{V}}) \in \mathbb{R}^{2N \times 4K}$. $\mathbf{Q}$ is stored by its $2K$ Householder vectors and if the operations are performed economically, the process takes [60]
    $$2(4K)(2K)(2(2N) - 2K) = 64K^2N - 32K^3 \text{ flops.}$$
  - *Step 4*: a tensor-vector multiplication. In this particular case the amount of computations is equal to
    $$\begin{aligned}&(2N - 2K)4K \text{ multiplications } + (2N - 2K)2K \text{ additions}\\ &= 10NK - 12K^2 \text{ flops.}\end{aligned}$$

The work in *Step 2* of the residual calculation and *Step 3* of the Jacobian calculation are extra work compared to the evaluation of the residual and the Jacobian in the case of minimization of G. Recall that the work involved in the tensor-vector multiplication in *Step 4* of the Jacobian evaluation of V needs also to be done when $\mathbf{J}^{\text{G}}$ is formed. The work involved in evaluating $\mathcal{D}(\mathbf{\Gamma}^{\text{V}})$, $\mathbf{\Gamma}^{\text{V}}$ and performing *Step 4* of the approximate Jacobian evaluation is of the same order of magnitude as evaluating $\mathbf{J}^{\text{G}}$. For the particular matrix sizes involved here the extra work required to compute a next iteration point using minimization of V compared to minimization of G is of the order of $32K^2N$ flops (assuming only 1 function evaluation is needed and ignoring the computational cost involved in computing $\mathbf{r}^{\text{G}}$).

An important remark can be made here. The computation of the factor $\mathbf{Q}_2\mathcal{D}(\mathbf{\Gamma}^{\text{V}})\mathbf{l}^{\text{V}}$ can be performed slightly different than in the scheme above. Instead of applying $\mathbf{Q}_2$ to the $2N \times 4K$ tensor $\mathcal{D}(\mathbf{\Gamma}^{\text{V}})$ and subsequently multiplying the obtained $(2N - 2K) \times 4K$ matrix with $\mathbf{l}^{\text{V}}$, $\mathbf{Q}_2$ can be applied directly to the $2N \times 2K$ matrix $\mathcal{D}(\mathbf{\Gamma}^{\text{V}})\mathbf{l}^{\text{V}}$. The latter is more efficient for the particular matrix sizes involved in the example above. In case the factor $\mathbf{Q}_2\mathcal{D}(\mathbf{\Gamma}^{\text{V}})\mathbf{l}^{\text{V}}$ is obtained in the latter way, the amount of work involved in the minimization of V and G is approximately equal for the matrix sizes involved here. This is however not always the case. E.g. in case every nonlinear variable is associated with only one element in the $\mathbf{l}^{\text{V}}$ vector, the implementation used in VARPRO is more efficient.

The introduction of prior knowledge in this variable projection framework also reduces the sizes of the matrices involved. If all ATP prior knowledge is imposed as in section 3.2.1 the approximate Jacobian reduces to a $(2N - 2) \times 2$ matrix. For every specific case, a complexity analysis as above can be carried out.

3.3 Numerical evaluation

A Monte-Carlo study is performed to determine which of the functionals G or V is best minimized w.r.t. efficiency and robustness. The three different NLLS algorithms described in section 3.1.3 are used to minimize the two different functionals. The Levenberg-Marquardt version of section 3.1.3.1 will be referred to as LM. The procedures outlined in section 3.2 are used to calculate the Jacobians and residuals involved. In total 6 different methods are compared

- LM to minimize V.
- LM to minimize G.
- NL2SOL (*dn2gb*) to minimize V.
- NL2SOL (*dn2gb*) to minimize G.
- MINPACK (*lmder*) to minimize V.
- MINPACK (*lmder*) to minimize G.

VARPRO uses the LM algorithm to minimize V. An important fact is also that in VARPRO minimization is done w.r.t. the square roots of the damping factors b_k, i.e., $b_k^2 = d_k$ $(k = 1, \ldots, K)$, to make sure the damping is positive as required by physical considerations. This variable transformation works in most practical situations, but is not without danger as explained by Gill et al. [58]. It will be demonstrated with an example that minimizing w.r.t. d_k $(k = 1, \ldots, K)$, and imposing the positivity constraint on the dampings leads to better results.

The simulation signal used is derived from an *in-vivo* ^{31}P spectrum measured in the human brain and consists of 256 complex data points and 11 exponentials (see Fig. 3.1 and Table 3.1). The ^{31}P peaks from brain tissue, from left to right PME, P_i, PDE, PCr, γ-ATP, α-ATP and β-ATP can all be observed. From the noiseless signal 300 noisy realizations were generated with standard deviation σ (both on the real and imaginary parts). A low, intermediate and high noise level ($\sigma = 5, 15, 25$ - this corresponds to a SNR expressed in dB of 29.5, 20 and 15.6 for the middle peak of β-ATP) were used. A method is considered to fail if the minimization algorithm claims not to have found the solution (according to its corresponding stopping criterion), if the damping of one of the peaks is negative or if not all 11 peaks are resolved within specific intervals lying symmetrically around the exact frequencies.

For all the presented results it was verified that these simple criteria guarantee that all methods find the same minimum if no failure occurs. As a result, RMSE, bias and standard deviation of the parameter estimates computed by

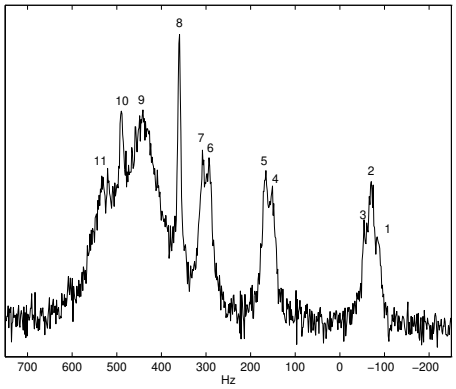


Figure 3.1: Real, phased DFT spectrum of a simulated ^{31}P MRS signal, the standard deviation of the added noise is 15. See Table 3.1 for the actual parameter values.

peak k	f_k (Hz)	d_k(Hz)	a_k (a.u.)	θ_k (°)
1	-86	50	75	135
2	-70	50	150	135
3	-54	50	75	135
4	152	50	150	135
5	168	50	150	135
6	292	50	150	135
7	308	50	150	135
8	360	25	150	135
9	440	28.6	1400	135
10	490	25	60	135
11	530	200	500	135

Table 3.1: Exact parameters values of the simulated ^{31}P MRS signal, modeled by Eq. (1.3). $t_0 = 0, \Delta t = 0.333$ ms. The values are based on the fit of an *in-vivo* ^{31}P MRS signal, $\theta_k = \phi_k * 180/\pi$ expresses the phase in degrees, a.u. denotes arbitrary units .

any of these methods as a function of the noise level are the same and will not be shown.

Starting values for the dampings and the frequencies can be obtained by peak picking or by using a black-box method like HSVD. In a previous study [146] it was found that for signals with a high SNR, HSVD is the best way to provide starting values. However for low SNR signals peak picking is preferred. Here

only peak picking is considered. For every noise level used in the Monte Carlo simulation one signal out of 300 was chosen randomly and damping and frequency starting values were determined by peak picking. These starting values were then used to process all the signals affected by the same noise level.

In section 3.3.1 it is investigated what the best approach is to ensure that the computed dampings are positive. The effect of using starting values of different quality and of imposing different forms of prior knowledge on the performance of the methods is also investigated in sections 3.3.2 and 3.3.3 respectively.

The methods are compared in terms of robustness and efficiency. The robustness can be assessed by the number of times a method fails. To compare the efficiency, the average number of functional and Jacobian evaluations and the average overall CPU time are measured. All experiments were performed on a SUN ULTRA2 (200 MHz).

3.3.1 Influence of different problem formulations to ensure that the dampings are positive

To ensure the damping values are positive there are several possibilities. The straightforward approach is to minimize w.r.t. d_k $(k = 1, \ldots, K)$ and to impose the constraint that all dampings must be positive. This can easily be done in the *dn2gb* version of NL2SOL since it handles simple bounds on the variables. In the LM and the MINPACK algorithm it can be done by imposing a penalty on the function if the algorithm computes damping estimates that are negative. In this way the penalty forces all the dampings to be positive. Another well-known approach is to represent the dampings d_k by b_k^2 and to minimize w.r.t. b_k $(k = 1, \ldots, K)$. This transformed problem may however be more difficult to solve.

The two different approaches were tested for a set of starting values "peakpick1". No prior knowledge is imposed, hence 4 parameters for each of the 11 peaks are to be estimated. In the case of minimization of G a problem in 44 variables is obtained, compared to a problem in 22 variables when minimizing V. The results are displayed in Tables 3.2 and 3.3.

In Table 3.2 the LM algorithm used to minimize V is the VARPRO program. In the case of minimization w.r.t. b_k $(k = 1, \ldots, K)$ VARPRO performs very bad in terms of failures. The problem disappears when minimization is done w.r.t. d_k $(k = 1, \ldots, K)$ as shown in Table 3.3 or when a more sophisticated algorithm like MINPACK or NL2SOL is used. Also if MINPACK is used to minimize V there is a significant reduction in the number of evaluations of the functional and the Jacobian (and corresponding CPU time) when minimization is done w.r.t. d_k $(k = 1, \ldots, K)$. This example shows that in order to make

σ	crite-rium	*LM*		*NL2SOL*		*MINPACK*	
		G	V	G	V	G	V
5	% fail	0	0	0	0	0	0
	jev ±std	9.9±0.3	10.4±1.0	11.0±1.6	8.5±0.5	8.0±0.2	13.5±1.8
	fev ±std	10.9±0.3	13.5±1.3	14.3±2.6	11.5±0.5	11.0±0.2	17.1±2.2
	cpu ±std	1.7±0.1	2.6±0.3	2.0±0.3	2.1±0.1	1.5±0.06	3.2±0.4
15	% fail	0.33	12.3	0.33	0	0.33	0
	jev ±std	10.6±1.3	11.5±3.8	13.5±2.6	11.8±1.5	10.2±2.0	10.2±2.6
	fev ±std	11.7±1.7	15.8±6.9	17.9±3.8	14.8±2.0	13.7±2.9	13.3±3.0
	cpu ±std	1.8±0.2	2.9±1.0	2.5±0.5	2.7±0.3	1.9±0.4	2.5±0.6
25	% fail	7.0	13.0	7.3	6.7	7.0	7.0
	jev ±std	12.7±4.5	11.3±4.2	15.4±5.4	12.3±3.7	12.3±6.0	12.6±4.2
	fev ±std	14.5±5.9	14.1±5.8	20.8±8.5	15.1±4.1	15.9±7.5	15.4±4.7
	cpu ±std	2.2±0.8	2.8±1.0	2.9±1.0	2.8±0.8	2.2±1.0	3.1±1.0

Table 3.2: Sample Mean and standard deviation values of Monte Carlo simulation (300 runs) using "peakpick1" starting values and no prior knowledge. Minimization is done w.r.t. b_k, $b_k^2 = d_k$ $(k = 1, \ldots, K)$. LM, NL2SOL and MINPACK are used to minimize G and V. The computed quantities are: % fail, the percentage of failures, jev, the average number of Jacobian evaluations, fev, the average number of functional evaluations, cpu, the average CPU time in seconds. std denotes the standard deviation.

a fair comparison between all six methods, minimization has to be done w.r.t. d_k $(k = 1, \ldots, K)$. Otherwise differences in performance of the methods might be due to the influence of minimization w.r.t. b_k $(k = 1, \ldots, K)$. In the experiments of sections 3.3.2 and 3.3.3 minimization is carried out w.r.t. d_k.

3.3.2 Influence of starting values on the minimization of the functionals

In this section the influence of using starting values of different quality is examined. As a first set of starting values the "peakpick1" starting values used in the studies of Tables 3.2 and 3.3 are taken. The procedure outlined before is repeated to obtain a second set of starting values "peakpick2". The "peakpick2" set of starting values was derived more carefully than the "peakpick1" set and is hence of better quality. No prior knowledge was imposed.

σ	crite-rium	*LM*		*NL2SOL*		*MINPACK*	
		G	V	G	V	G	V
5	% fail	0	0	0	0	0	0
	jev ±std	9.9±0.4	9.0±0.3	11.5±1.7	8.3±0.5	8.1±0.3	7.0±0.2
	fev ±std	10.9±0.4	12.1±0.5	15.6±2.4	11.3±0.5	11.1±0.3	9.0±0.2
	cpu ±std	1.7±0.06	2.2±0.07	2.1±0.3	2.0±0.1	1.4±0.05	1.7±0.04
15	% fail	0.33	0	0.33	0	0.33	0
	jev ±std	10.6±1.3	10.2±0.9	14.7±3.2	10.5±1.3	10.1±2.3	8.6±1.4
	fev ±std	11.7±1.7	14.4±1.1	20.9±5.8	13.2±1.6	13.7±3.7	10.8±1.7
	cpu ±std	1.9±0.2	2.8±0.2	2.6±0.6	2.5±0.3	1.8±0.4	2.1±0.3
25	% fail	7.0	6.7	7.0	6.7	7.0	6.7
	jev ±std	12.7±4.4	11.4±3.5	15.0±6.1	12.1±3.7	12.1±5.6	9.8±3.6
	fev ±std	14.4±5.7	14.9±3.6	20.4±9.7	14.6±3.9	15.7±7.0	11.8±3.6
	cpu ±std	2.2±0.8	2.9±0.8	2.6±1.1	2.8±0.8	2.1±0.9	2.4±0.8

Table 3.3: Sample mean and standard deviation values of Monte Carlo simulation using "peakpick1" starting values. Minimization is done w.r.t. d_k, $(k = 1, \ldots K)$ and the constraint that the dampings must be positive is imposed. Further as in Table 3.2.

The results of this Monte Carlo study are displayed in Tables 3.3 and 3.4.

In terms of failures all methods perform equally well for both sets of starting values. Comparing the three different algorithms for the minimization of G, learns that NL2SOL needs the most functional and Jacobian evaluations and has the highest overall CPU time for the two sets of starting values. MINPACK is the most efficient method. When the algorithms for the minimization of V are compared, it can be concluded that MINPACK is the most efficient method. For the "peakpick1" starting values LM is the slowest method while for the "peakpick2" starting values NL2SOL is the worst method in terms of efficiency.

For the "peakpick1" starting values, minimization of G is more efficient than minimization of V for LM and MINPACK. In the case of NL2SOL there is no significant difference. However in the case of the "peakpick2" starting values there is no real difference in terms of efficiency between minimizing G and V for LM and NL2SOL. For MINPACK it is still more efficient to minimize G. These results show that in this case minimization of G is comparable in efficiency as minimization of V.

σ	crite-rium	*LM*		*NL2SOL*		*MINPACK*	
		G	V	G	V	G	V
5	% fail	0	0	0	0	0	0
	jev ±std	9.6 ± 0.5	7.0 ± 0.2	9.8 ± 0.8	8.4 ± 0.5	7.8 ± 2.7	6.6 ± 0.6
	fev ±std	10.6 ± 0.5	8.0 ± 0.2	12.8 ± 0.9	10.6 ± 0.7	10.0 ± 3.7	7.6 ± 0.6
	cpu ±std	1.7 ± 0.1	1.7 ± 0.04	1.8 ± 0.1	2.0 ± 0.1	1.4 ± 0.4	1.6 ± 0.1
15	% fail	0	0	0	0	0	0
	jev ±std	11.0 ± 2.2	8.0 ± 0.9	13.9 ± 3.5	10.1 ± 1.1	9.7 ± 3.7	8.4 ± 1.7
	jev ±fev	12.3 ± 3.4	9.9 ± 0.9	17.8 ± 5.7	12.1 ± 1.1	13.2 ± 5.4	9.9 ± 2.2
	cpu ±std	1.9 ± 0.4	2.0 ± 0.2	2.5 ± 0.6	2.4 ± 0.2	1.7 ± 0.6	2.0 ± 0.4
25	% fail	6.7	6.7	6.7	6.7	6.7	7.3
	jev ±std	12.5 ± 4.4	9.4 ± 3.4	15.8 ± 6.4	11.9 ± 2.9	11.2 ± 6.5	10.3 ± 4.1
	fev ±std	14.0 ± 5.7	10.4 ± 3.4	21.0 ± 9.4	14.4 ± 3.4	14.6 ± 8.1	12.2 ± 4.6
	cpu ±std	2.2 ± 0.7	2.3 ± 0.8	2.8 ± 1.1	2.7 ± 0.6	2.0 ± 1.1	2.5 ± 0.9

Table 3.4: Sample mean and standard deviation values of Monte Carlo simulation using "peakpick2" starting values. Further as in Table 3.3.

For the LM algorithm, the time needed to compute the residual vector, the (approximate) Jacobian and the solution to Eq. (3.13) were measured separately. In case of minimization of V the following execution times were measured:

- calculation of approximate Jacobian: $1.5\ 10^{-1}$ s.
- computation residual: $5.5\ 10^{-2}$ s.
- solution of Eq. (3.13): $1.9\ 10^{-2}$ s.

Leading to an overall execution time of approximately:

$$1.7\ 10^{-1}\ \text{jev}\ +\ 5.5\ 10^{-2}\ \text{fev}.$$

In case of minimization of G the following execution times were measured:

- calculation of Jacobian: $7.6\ 10^{-2}$ s.
- computation residual: $1.8\ 10^{-2}$ s.
- solution of Eq. (3.13): $7.1\ 10^{-2}$ s.

peak k	f_k (Hz)	d_k	a_k	θ_k (°)
1	f_1	d_1	a_1	135
2	$f_1 + 16$	d_1	$2 \times a_1$	135
3	$f_1 + 32$	d_1	a_1	135
4	f_2	d_1	$2 \times a_1$	135
5	$f_2 + 16$	d_1	$2 \times a_1$	135
6	f_3	d_1	$2 \times a_1$	135
7	$f_3 + 16$	d_1	$2 \times a_1$	135
8	f_4	d_2	a_2	135
9	f_5	d_3	a_3	135
10	f_6	d_4	a_4	135
11	f_7	d_5	a_5	135

Table 3.5: Imposed relations between the parameters of the simulation signal ("prior2"), $\theta_k = \phi_k * 180/\pi$ expresses the phase in degrees.

Leading to an overall execution time of approximately:

$$1.5\,10^{-1}\text{ jev } +\ 1.8\,10^{-2}\text{ fev.}$$

The above analysis explains the fact that although the number of Jacobian and function evaluations necessary to minimize V is in general lower than the number needed to minimize G, the overall CPU time associated with the minimization of V is larger than for minimization of G for most cases in this example. Note that by computing the factor $\mathbf{Q}_2\mathcal{D}(\mathbf{\Gamma}^{\mathrm{v}})\mathbf{l}^{\mathrm{v}}$ in a smart way the calculation time for the computation of the Jacobian associated with the minimization of V can be significantly reduced.

3.3.3 Influence of prior knowledge on the minimization of the functionals

Until now all signals were processed using no prior knowledge at all. In Table 3.5 all available prior knowledge about the simulation signal is given.

In the first comparison the prior knowledge that all peaks have a phase of 135° ("prior1") is imposed. This reduces the number of linear variables to 11. The number of nonlinear parameters is equal to 22. In the second comparison all prior knowledge of Table 3.5 ("prior2") is imposed. The number of nonlinear parameters then drops to 12 and the number of linear parameters to 5. The results are displayed in Tables 3.6 and 3.7.

From the results of the "prior1" set of prior knowledge, it is clear that NL2SOL has the lowest overall failure rate in case of minimization of V and G. MINPACK performs particularly bad when minimizing V. This suggests that it

σ	crite-	*LM*		*NL2SOL*		*MINPACK*	
	rium	G	V	G	V	G	V
5	% fail	0	0	0	0	0	0
	jev $\pm$std	8.6 ± 0.6	7.3 ± 0.5	9.3 ± 0.6	8.8 ± 0.6	7.1 ± 1.4	6.3 ± 0.7
	fev $\pm$std	9.6 ± 0.6	9.3 ± 0.5	12.4 ± 0.7	11.0 ± 0.7	8.5 ± 2.1	7.3 ± 0.8
	cpu $\pm$std	1.0 ± 0.08	0.8 ± 0.05	1.1 ± 0.06	1.0 ± 0.06	0.8 ± 0.2	0.7 ± 0.07
15	% fail	0.67	1	0	0	0.33	1.7
	jev $\pm$std	9.7 ± 3.7	8.2 ± 1.6	10.1 ± 1.0	10.3 ± 1.3	9.8 ± 2.9	9.1 ± 2.6
	fev $\pm$std	10.7 ± 3.7	10.2 ± 1.6	13.1 ± 1.0	12.7 ± 1.2	12.0 ± 3.5	10.6 ± 3.2
	jev $\pm$cpu	1.1 ± 0.4	0.9 ± 0.2	1.2 ± 0.1	1.2 ± 0.1	1.1 ± 0.3	1.0 ± 0.3
25	% fail	3.0	2.7	0.33	0.67	1.0	7.0
	jev $\pm$std	11.4 ± 5.2	9.7 ± 6.9	10.6 ± 1.6	11.6 ± 1.9	11.3 ± 4.1	10.6 ± 4.1
	fev $\pm$std	12.5 ± 5.2	11.7 ± 6.9	13.7 ± 1.8	14.3 ± 2.2	13.3 ± 4.6	12.4 ± 4.5
	cpu $\pm$std	1.3 ± 0.6	1.1 ± 0.7	1.3 ± 0.2	1.3 ± 0.2	1.3 ± 0.4	1.2 ± 0.4

Table 3.6: Results of Monte Carlo simulation (300 runs), using the prior knowledge that all phases are equal ("prior1").

is better to minimize G from the point of view of robustness. Indeed, in all performed experiments, it was seen that if an algorithm performed badly, it was on the minimization of V. The overall CPU times of all methods are very comparable. On the "prior2" set of prior knowledge all methods perform almost identical. The imposition of more prior knowledge decreases the overall computational time significantly. This means that imposition of prior knowledge not only increases the accuracy of the results but also the computational efficiency because of the reduction in the number of variables.

Also here the approximate timings for different parts of the LM algorithm are reported. In case of minimization of V and the prior knowledge that all phases are equal the following execution times were measured:

- calculation of approximate Jacobian: $6.6\ 10^{-2}$ s.
- computation residual: $2.6\ 10^{-2}$ s.
- solution of Eq. (3.13): $2.0\ 10^{-2}$ s.

Leading to an overall execution time of approximately:

$$8.6\ 10^{-2}\ \text{jev}\ +\ 2.6\ 10^{-2}\ \text{fev}.$$

σ	crite-rium	*LM*		*NL2SOL*		*MINPACK*	
		G	V	G	V	G	V
5	% fail	0	0	0	0	0	0
	jev ±std	8±0	6.1±0.2	7.7±0.5	7.6±0.5	5.0±0.1	5.0±0.06
	fev ±std	9±0	7.1±0.2	10.6±0.5	10.7±0.5	6.0±0.1	6.0±0.06
	cpu ±std	0.6±0.02	0.4±0.01	0.6±0.03	0.6±0.03	0.4±0.007	0.4±0.003
15	% fail	0	0	0	0	0	0
	jev ±std	7.3±0.5	6.2±0.4	8.3±0.4	8.4±0.5	5.1±0.2	5.0±0.2
	fev ±std	8.3±0.4	7.2±0.4	10.3±0.5	10.8±0.8	6.1±0.2	6.0±0.2
	cpu ±std	0.6±0.03	0.4±0.02	0.7±0.03	0.6±0.03	0.4±0.02	0.4±0.01
25	% fail	0	0	0	0	0	0
	jev ±std	7.7±0.6	6.7±0.5	8.6±0.6	9.3±0.7	7.5±0.9	7.2±0.8
	fev ±std	8.7±0.6	7.7±0.5	11.6±0.7	12.0±0.9	8.8±1.0	8.5±1.0
	cpu ±std	0.6±0.04	0.5±0.03	0.7±0.04	0.7±0.04	0.6±0.06	0.5±0.005

Table 3.7: Results of Monte Carlo simulation (300 runs) using the prior knowledge as specified in Table 3.5 ("prior2").

In case of minimization of G and the prior knowledge that all phases are equal the following execution times were measured:

- calculation of Jacobian: $5.6\,10^{-2}$ s.
- computation residual: $1.8\,10^{-2}$ s.
- solution of Eq. (3.13): $4.2\,10^{-2}$ s.

Leading to an overall execution time of approximately:

$$9.8\,10^{-2}\,\text{jev} \;+\; 1.8\,10^{-2}\,\text{fev}.$$

In case of minimization of V and the "prior2" prior knowledge the following times were reported:

- calculation of approximate Jacobian: $4.6\,10^{-2}$ s.
- computation residual: $2.1\,10^{-2}$ s.
- solution of Eq. (3.13): $7.6\,10^{-3}$ s.

Leading to an overall execution time of approximately:

$$5.4\ 10^{-2}\ \text{jev}\ +\ 2.1\ 10^{-2}\ \text{fev.}$$

In case of minimization of G and the "prior2" prior knowledge the following execution times were measured:

- calculation of approximate Jacobian: $5.5\ 10^{-2}$ s.
- computation residual: $1.8\ 10^{-2}$ s.
- solution of Eq. (3.13): $1.3\ 10^{-2}$ s.

Leading to an overall execution time of approximately:

$$6.8\ 10^{-2}\ \text{jev}\ +\ 1.8\ 10^{-2}\ \text{fev.}$$

Since *dn2gb* is able to impose upper and lower bounds on the variables, this feature can be used to impose the natural bounds on the variables. Amplitudes (like the dampings) can not become negative (negative amplitudes correspond to a 180° phase shift), whereas the upper and lower bounds on the frequencies are determined by the spectral width. Phases can be constrained to lie between $-180°$ and $+180°$. Since NL2SOL already performed well in the previous experiments without imposing these bounds, it is expected that imposing these natural bounds explicitly on the variables has little effect on the overall behavior. The average number of functional and Jacobian evaluations and resulting CPU time decreases only slightly in the cases examined here. The results are therefore not shown. In general however it is recommended to add those extra bounds on the variables [58] in order to ensure maximal accuracy and minimize the failure rate.

3.3.4 General findings

It was demonstrated that minimization has to be done w.r.t. the damping value with imposition of a positivity constraint instead of minimization w.r.t. the square root of the damping value as is done in VARPRO. In the results shown there is one case in which an algorithm has problems with the minimization of V: MINPACK with as prior knowledge the set "prior1". This indicates that minimization of G is preferable to get a more robust method. The minimization of V and G is comparable in efficiency. It is certainly not true to state that minimization of V is always more efficient than minimization of G, as was formerly believed. NL2SOL exhibits excellent behavior in terms of robustness in all examined cases and this compensates for the somewhat lower efficiency of the method. The results shown here indicate that in order to get a robust

method, one should minimize the general functional using NL2SOL. In addition NL2SOL enables imposition of simple bounds (so-called "soft constraints") on the variables, making it possible to impose the natural constraints of the problem in a simple way. It was also shown that the use of prior knowledge leads to a significant reduction in computation time.

3.4 Possibilities of imposing prior knowledge in VARPRO and AMARES

In section 3.4.1 the possibilities of imposing prior knowledge in VARPRO are discussed. Section 3.4.2 describes the possible types of prior knowledge that can be imposed using AMARES. In section 3.4.3 the increased possibilities to impose prior knowledge using AMARES are illustrated.

3.4.1 VARPRO: multiplet approach

VARPRO puts peaks belonging to the same multiplet into one group. Amplitudes within one group can be left unconstrained or they can be linked to each other when the intensity ratios of the peaks are known. Dampings are treated in the same way, but with the extra possibility of fixing the damping to a value which has to be the same for all peaks in that group. Frequencies within a group can be left unconstrained or they can be linked in case the frequency splittings between neighboring peaks are known exactly. If the frequency splittings between neighboring peaks are unknown, but equal for all the peaks in the group, a new variable Δ is introduced. One then has f_1, $f_2 = f_1 + \Delta$, $f_3 = f_1 + 2\Delta$, $\ldots$, $f_k = f_1 + (k-1)\Delta$, for all k peaks belonging to the same group, f_k is the arbitrarily chosen reference peak, neighboring peaks in the spectrum are denoted by consecutive numbers. In addition, the frequency of the first component (chosen arbitrarily) can be fixed together with the frequency splitting of the multiplet. In that case all the frequencies of a group are fixed. Note that it is impossible to fix the frequencies of the peaks of a certain group independently: the frequency differences between neighboring peaks are the same. The overall phase ϕ_k of a peak is considered to consist of an individual phase ϕ'_k and a zero-order phase ϕ_0 which is the same for all peaks: $\phi_k = \phi'_k + \phi_0$. The individual phases can be left unconstrained in which case the zero-order phase must be fixed. The phases of all peaks within a group can be fixed to a value (the same for all the peaks within the group) relative to the zero-order phase, which can be estimated or kept fixed. Of all the groups with constraints on amplitudes one group can be taken as a reference and the other groups can be linked to the reference group. This means that the relative amplitude of the first component (chosen arbitrarily) of the reference group w.r.t. the first

component of a linked group can be imposed. Groups with constrained dampings are treated in the same way. Groups with variable frequency splittings can also be linked in the same way, the only difference being that the frequency splitting ratio between the linked groups must be one.

Some of the above-mentioned possibilities are illustrated using the ^{31}P simulation example (see Fig. 3.1 and Table 3.1). The peaks belonging to the β-ATP triplet are put into one group. The amplitude ratios are $a_1 : a_2 : a_3 = 1 : 2 : 1$. If this prior knowledge is imposed in VARPRO and the corresponding phases are left unconstrained, the program will automatically assume these phases to be equal (this is not the case when amplitudes are left unconstrained). This is equivalent to linking the complex amplitudes $a_k e^{j\phi_k}$ or in this example $Re(c_1) : Re(c_2) : Re(c_3) = 1 : 2 : 1$ and $Im(c_1) : Im(c_2) : Im(c_3) = 1 : 2 : 1$. Peaks 4 and 5 belong to the α-ATP doublet and are put in a second group. The following is known: $a_4 : a_5 = 1 : 1$ and $a_4 : a_1 = 2 : 1$. VARPRO offers the possibility of imposing this prior knowledge between the two groups as long as the constraints on the phases are equal for both groups.

In the case of the β-ATP triplet a fixed frequency splitting of 16 Hz between the individual peaks within the triplet can be imposed, leading to: $f_2 = f_1 + 16$ and $f_3 = f_1 + 32$. Suppose however that the exact frequency splittings between the peaks are not known but known to be equal. This can be expressed by introducing a new variable Δ: $f_2 = f_1 + \Delta$ and $f_3 = f_1 + 2\Delta$. The fact that the frequency splitting in the α-ATP doublet is the same as in the β-ATP triplet can also be expressed. Using VARPRO this is imposed as $f_5 = f_4 + \Delta$.

This implementation of prior knowledge in VARPRO already offers many possibilities of imposing various linear relations between parameters and has proven crucial in earlier studies [134, 133, 120]. However, in recent years, more prior knowledge has become available and the present implementation of VARPRO can no longer satisfy all the needs [38]. One example is formed by the six glycogen resonances in ^{13}C MRS, for which the relative frequency shifts are known but different. This can therefore not be implemented as a regular multiplet in the VARPRO algorithm. For the quantification of a ^{13}C MRS signal using AMARES, see [149]. AMARES has greatly increased possibilities concerning the prior knowledge that can be invoked as demonstrated in the next section.

3.4.2 AMARES: singlet approach

In AMARES a singlet approach is chosen making an identical treatment of all parameters possible. Each of the parameters can be left unconstrained or kept fixed. One can impose a fixed shift or ratio w.r.t. any unconstrained parameter of the same type. Finally, it is possible to impose a *variable* shift or ratio w.r.t. any unconstrained or fixed parameter of the same type. These variable shifts or

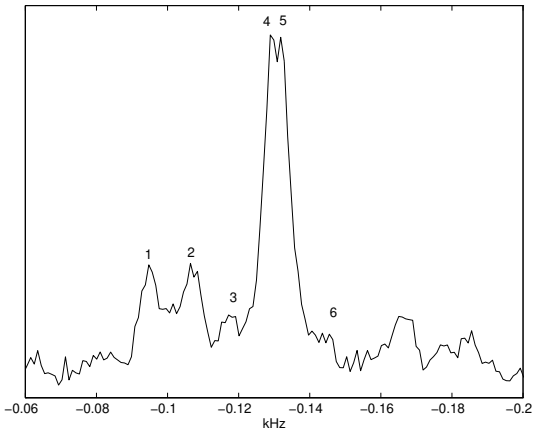


Figure 3.2: *In-vivo* spectrum of the prostate gland of a BPH patient. Peak 1 is choline, peak 2 creatine and peaks 3 to 6 constitute the citrate lines.

ratios can then be linked between different groups of peaks. Imposing a fixed shift or ratio and a variable shift or ratio deserves some more explanation; the other forms of prior knowledge are trivial to implement.

As an arbitrary example, take a signal consisting of three peaks. Imposing a fixed ratio is done in the following way; e.g. for the amplitudes: $a_2 = a_1 x, a_3 = a_1 y, x, y \in \mathbb{R}$. x and y are values provided by the user. Imposing a variable ratio leads to the introduction of a new variable Δ, $a_2 = a_1(x\Delta), a_3 = a_1(y\Delta), x, y \in \mathbb{R}$. x and y are values provided by the user. Imposing a variable shift or ratio is only useful if the new variable is shared by at least two peaks, otherwise the total number of unknowns is not changed.

3.4.3 Example

To illustrate the features of AMARES it was tested on a challenging data set provided by Dr. A. Heerschap of the University Hospital Nijmegen (see acknowledgments). An *in-vivo* spectrum was acquired from the prostate gland of a benign prostatic hyperplasia (BPH) patient, at 1.5 T using a PRESS localization sequence (τ_1=11 ms, τ_2=67.5 ms) and an endorectal coil for signal reception. 256 scans of 1024 complex echo data points were acquired from a volume of 8 cc, with a TR of 1.6 s (see Fig. 3.2). The goal of this measurement was the quantification of the citrate content, which is an important tool in the discrimination between prostate adenocarcinoma and benign prostatic hyperplasia (BPH). The citrate protons form a complicated AB-type multiplet (peaks 3 to 6 in Fig. 3.2), consisting of four Lorentzian lines with different

peak k	$f_k \pm$ std (kHz)	$d_k \pm$ std (kHz)	$a_k \pm$ std (a.u.)	θ'_k (°)
1	-0.0949 ± 0.0004	-0.0222 ± 0.0032	9.70 ± 1.30	0
2	-0.1069 ± 0.0003	-0.0251 ± 0.0040	10.42± 1.97	0
3	-0.1140 ± 0.0005	-0.0111 ± 0.0023	1.55 ± 0.05	-98.74
4	-0.1296 ± 0.0001	-0.0126 ± 0.0006	19.51 ± 0.64	35.60
5	-0.1317 ± 0.0001	-0.0129 ± 0.0006	19.51 ± 0.64	-35.60
6	-0.1472 ± 0.0005	-0.0111 ± 0.0023	1.55 ± 0.05	98.74

fitted zero-order phase (°)
225.6297 ± 25.8974
fitted begin time (ms)
2.4339 ± 0.5528

Table 3.8: Results of analysis of the citrate signal with VARPRO, $\theta'_k = \phi'_k * 180/\pi$ expresses the phase in degrees.

phase. Relative intensities of the citrate lines and their phase deviations from the rest of the spectrum can be calculated [95]. It is also known that the dampings of the main peaks should be equal as well as those of the side peaks.

To analyze the signal with VARPRO all peaks are put in separate groups. The peaks of choline and creatine could also be put in one group but that would not have made any difference. The four peaks of citrate have to be put in separate groups in order to impose the fixed (but different for each peak) phase shifts w.r.t. the zero-order phase. This implies that the known frequency shifts between the citrate peaks can not be imposed since these can only be imposed between peaks belonging to the same group. Amplitudes were linked between the different groups and the individual phases of the peaks fixed relative to the zero-order phase. Since VARPRO allows only one overall linking for dampings between groups, all four peaks could be linked to each other (theoretically incorrect), or either the side peaks, or the main peaks. In this example with the VARPRO method the linewidths (dampings) of the small side peaks were linked, as these are more difficult to fit. Zero-order phase and begin time are estimated. The results are displayed in Table 3.8.

Only with AMARES it is possible to impose all the available prior knowledge. The results are displayed in Table 3.9.

The most important difference between the VARPRO and AMARES results is that the latter can be obtained in a much more consistent fashion. This is entirely due to the increased use of important prior knowledge on the AB-type multiplet of citrate, in particular the chemical shift differences between its respective components. This is especially crucial at echo times where the side peaks reach their minimum intensities and sink into the noisy baseline

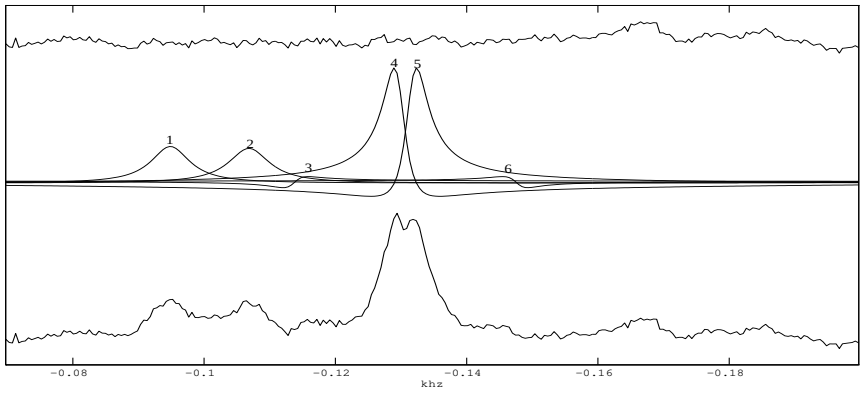


Figure 3.3: Graphical result of the analysis of the citrate signal with AMARES. From bottom to top: the DFT spectrum of the original signal, the individual Lorentzians (DFT of fitted sinusoids) and the residual, which is the difference between the original signal and the reconstructed signal.

[136]. With the combined prior knowledge of the chemical shift differences and amplitude (peak area) ratios with respect to the prominent main peaks of the quartet, the small side peaks can still be estimated reliably. Including the small side peaks in the fit is not only necessary for complete quantitation of the citrate multiplet, but also enhances the accuracy of the creatine fit, as their resonances overlap closely (see Fig. 3.3).

Another advantage of AMARES is that, in combination with all other prior knowledge, the linewidths of the side peaks can be linked to each other, and those of the main peaks can be linked to each other, see the results in Table 3.9.

The main peaks were found to be a little broader than the side peaks, which corresponds to the expectations for the citrate multiplet [12]. With the VARPRO method where only the side peaks could be linked to each other, slightly different (although not significant, due to the high level of the noise) linewidths of the main peaks were obtained, which is undesired. Despite the slight differences in linewidths and chemical shifts between the two methods, the estimated amplitudes are identical for the citrate multiplet. This is mainly due to the available prior knowledge for the amplitude ratios with respect to the prominent main peaks, and to the precise phase constraints. As an aside, it can be noted that, perhaps because of the improved fitting of the left side peak of citrate, the fitting of creatine and choline is more according to expectations

peak k	$f_k \pm$ std (kHz)	$d_k \pm$ std (kHz)	$a_k \pm$ std (a.u.)	θ'_k (°)
1	-0.0949 ± 0.0004	-0.0227 ± 0.0032	9.93 ± 1.27	0
2	-0.1068 ± 0.0003	-0.0240 ± 0.0037	9.95 ± 1.83	0
3	-0.1137 ± 0.0001	-0.0117 ± 0.0025	1.55 ± 0.05	-98.74
4	-0.1296 ± 0.0001	-0.0127 ± 0.0005	19.50 ± 0.63	35.60
5	-0.1317 ± 0.0001	-0.0127 ± 0.0005	19.50 ± 0.63	-35.60
6	-0.1477 ± 0.0001	-0.0117 ± 0.0025	1.55 ± 0.05	98.74

fitted zero-order phase (°)
221.1242 ± 25.2350
fitted begin time (ms)
2.3313 ± 0.5380

Table 3.9: Results of analysis of the citrate signal with AMARES, $\theta'_k = \phi'_k * 180/\pi$ expresses the phase in degrees.

for AMARES. Their linewidths should not differ too much, due to the leveling effect on the effective linewidths of the magnetic field inhomogeneities and magnetic susceptibility effects. Creatine may indeed be a little broader than choline, as its T2 is shorter than that of choline [131].

3.5 AMARES: an overview

AMARES uses *dn2gb*, the most recent version of NL2SOL to minimize the general functional. Imposition of simple bounds, so-called soft constraints, is possible. The possibilities of imposing the prior knowledge were extended by using a singlet approach as explained in section 3.4.2. The software is written in Fortran77. It can be used as a stand-alone program, but it is also implemented within MRUI [132]. The MRUI (magnetic resonance user interface) software package provides a user-friendly graphical user interface to a number of sophisticated analysis routines for MRS data quantification. In this way methods usually developed by mathematics- or physics-related research groups find their way to users in biomedical/biochemical laboratories and the clinical environment. The current implementation of the GUI requires the basic Matlab™ toolbox.

For every peak one has the choice between a pure Lorentzian line form or a pure Gaussian line form. In VARPRO the line form used must be the same for all peaks.

AMARES also offers the possibility to work with echo signals. The left and the

right part of the echo are considered to have the same amplitudes, frequencies and phases but different dampings. The dampings of the right and left part can however be linked to each other. In VARPRO the first part of the echo is truncated in order to be able to work with the model function of a FID signal. As a consequence, part of the signal and thus part of the information is not used. Like in VARPRO the user has the possibility to perform frequency-selective quantification in the time domain [71].

A measure of precision is also computed. Therefore an approximation to the CRBs is computed. The Cramér-Rao matrix in Eq. (2.7) can be rewritten in terms of the Jacobian matrix $\mathbf{J}^{\mathrm{G}}$, evaluated in the *true* parameters

$$\left(\frac{1}{\sigma^2}\mathbf{J}^{\mathrm{G}^T}\mathbf{J}^{\mathrm{G}}\right)^{-1}. \tag{3.30}$$

The square roots of the diagonal elements of Eq. (3.30) give a measure of the standard deviation of the parameters. In practical circumstances the true value of the parameters and the noise standard deviation are not known. As an estimate of the true parameter values, the parameters determined by AMARES are used. The noise standard deviation can be estimated from the last samples of the original data sequence

$$\hat{\sigma} = \sqrt{\frac{1}{P-1}\sum_{p=1}^{P}(y_{N-p} - m_y)^H (y_{N-p} - m_y)},$$

where m_y is the estimated mean

$$m_y = \frac{1}{P}\sum_{p=1}^{P} y_{N-p}$$

and P is chosen equal to a (small) number of samples containing mainly noise. Alternatively, the value of the cost function G computed in the final parameter estimates can be used as an estimate of the noise standard deviation. Note that the CRBs do not change by imposing simple bounds on the variables as proven in [61]. A schematic outline of AMARES is given below.

Algorithm 3.5.1 (AMARES)

Input: data samples $y_n, n = 0, \ldots, N-1$, starting values for the unknowns, available prior knowledge, line form per peak, FID or echo, upper and lower bounds on the variables.
Output: estimates of the parameters $\hat{a}_k, \hat{d}_k, \hat{f}_k, \hat{\phi}_k, k = 1, \ldots, K$ and their respective error bounds.

Step 1. Split complex data in real and imaginary part and form $\mathbf{y}^{\mathrm{G}}$.

Step 2. Call *dn2gb*. *dn2gb* uses user-supplied routines to evaluate $\mathbf{r}^G$ and $\mathbf{J}^G$ as outlined in section 3.2.1.

Step 3. *dn2gb* returns parameter values.

Step 4. Calculate error estimates of the parameters.

3.6 Conclusions

AMARES was presented which outperforms the currently used reference time-domain method VARPRO in two ways. First of all an appropriate choice of the functional and the NLLS algorithm was made. When estimating the parameters of a MRS signal in the time domain using optimization methods a choice between two functionals has to be made. On the one hand, a general functional, consisting of the sum of squared differences between the data and the model function can be minimized. On the other hand, a so-called variable projection functional derived from the general functional by eliminating all linear parameters can be optimized. It was shown in this chapter that minimizing the general functional leads to a more robust determination of parameter estimates. In addition it was demonstrated that the functional should be minimized with respect to the damping with imposition of a positivity constraint instead of minimizing the functional with respect to the square root of the damping as is done in VARPRO. In this way numerical problems are avoided. Different NLLS algorithms were compared and it was found that NL2SOL is the most robust method at the expense of a somewhat lower efficiency. In AMARES NL2SOL is used to minimize the general functional. As a result AMARES has a higher numerical accuracy and robustness than VARPRO.

Second, AMARES allows to include more prior knowledge about the signal parameters, the model function (Lorentz, Gauss and Voigt) and the type of signal (FID or echo). This results in an increased accuracy and flexibility. In addition, imposing all prior knowledge leads to a problem in a smaller number of variables which can be solved in less time. Overall an algorithm is obtained which outperforms VARPRO in terms of accuracy, robustness and flexibility.

Chapter 4

Quantification of series of spectra

As seen in chapters 2 and 3, quantification of individual MRS signals is possible in the time domain using interactive NLLS fitting methods which provide maximum likelihood parameter estimates under certain assumptions or using fully automatic, but statistically suboptimal black-box methods. In biochemical studies MRS signals are often acquired consecutively to monitor metabolic changes over time. Time series are e.g. measured to follow-up reactions occurring after administration of a certain substance, to observe changes in metabolites after pinching off an artery in perfusion experiments where an organ such as the heart is isolated and to study the metabolic changes during and after muscle exercise. Often information concerning the time evolution of some of the parameters is present. It is shown in this chapter that incorporation of this additional information leads to more accurate parameter estimates.
The main objectives of this chapter are:

- *To show how AMARES and HTLS two representative examples of the interactive and black-box methods respectively, can be extended to the simultaneous processing of all spectra in the time series using the common information present in the spectra.*
- *To point out the advantages and disadvantages of the presented methods.*
- *To analyze the statistical properties of the presented methods.*

In section 4.1 possible ways of extending AMARES to the processing of time series are explained. In section 4.2 ways to extend HTLS to time series processing are reviewed. The methods presented are illustrated and validated using

a Monte-Carlo study on two different time series in section 4.3. In section 4.4 the methods are compared based on the analysis of an experimental time series. In section 4.5 the main results obtained in this chapter are summarized.

This work is based on material presented in [148, 150].

4.1 Extensions of AMARES to quantitate time series

Two possible ways of extending AMARES to the processing of time series using additional information on the time evolution of the parameters are presented. In section 4.1.1 it is explained how AMARES can be extended to the simultaneous processing of all spectra in the time series with imposition of prior knowledge on relations between different spectra in a ML setting. In the literature a suboptimal approach consisting of applying AMARES twice is sometimes used to analyze time series in case the common information among spectra consists of parameters which remain constant in time [38]. This approach is outlined in section 4.1.2. Note that in [142, 115], the optimal use of common information in time series has been presented for some special cases.

4.1.1 ML use of common information present in spectra: AMARES$_{ts}$

The quantification of time series can be formulated in a mathematical way as follows.
Suppose there are S signals, each of which can be modeled by Eq. (1.3). The nth sample of signal s, y_{ns}, is then given by

$$y_{ns} = \sum_{k=1}^{K_s} a_{ks} e^{j\phi_{ks}} e^{(-d_{ks}+j2\pi f_{ks})t_n} + e_{ns}. \tag{4.1}$$

The additional parameter index s indicates that the parameter refers to the sth signal of the time series and K_s denotes the number of peaks present in the sth signal. ML fitting of time series with relations between spectra boils down to minimizing the following cost function

$$\sum_{s=1}^{S}\sum_{n=0}^{N-1} |y_{ns} - \sum_{k=1}^{K_s} a_{ks} e^{j\phi_{ks}} e^{(-d_{ks}+j2\pi f_{ks})t_n}|^2 = \|\mathbf{y}_{series} - \Gamma_{series}\, \mathbf{l}_{series}\|^2, \tag{4.2}$$

where

$$\mathbf{y}_{series} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_S^T]^T = [y_{01}, \ldots, y_{(N-1)1}, \ldots, y_{0S}, \ldots, y_{(N-1)S}]^T,$$

$$
\begin{aligned}
\mathbf{l}_{series} &= [\mathbf{l}_1^T, \ldots, \mathbf{l}_S^T]^T \\
&= [a_{11}e^{j\phi_{11}}, \ldots, a_{K_1 1}e^{j\phi_{K_1 1}}, \ldots, a_{1S}e^{j\phi_{1S}}, \ldots, a_{K_S S}e^{j\phi_{K_S S}}]^T, \\
\mathbf{\Gamma}_{series} &= \begin{bmatrix} \mathbf{\Gamma}_1 & 0 & \ldots & \ldots & 0 \\ 0 & \ddots & & & 0 \\ \vdots & & \mathbf{\Gamma}_s & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \ldots & \ldots & & \mathbf{\Gamma}_S \end{bmatrix} \qquad (4.3)
\end{aligned}
$$

with (4.4)

$$
\mathbf{\Gamma}_s = \begin{bmatrix} e^{(-d_{1s}+j2\pi f_{1s})t_0} & \ldots & e^{(-d_{K_s s}+j2\pi f_{K_s s})t_0} \\ \vdots & \ddots & \vdots \\ e^{(-d_{1s}+j2\pi f_{1s})t_{N-1}} & \ldots & e^{(-d_{K_s s}+j2\pi f_{K_s s})t_{N-1}} \end{bmatrix}.
$$

The relations present between spectra are expressed as relations between the corresponding parameters. In principle, any relation between parameters of the same type can be expressed.

The possibilities currently implemented in this extension of AMARES, denoted by AMARES$_{ts}$ from now on, are:

- Each of the parameters can be left unconstrained or kept fixed.
- A parameter can be expressed as the sum of an unconstrained and a fixed parameter of the same type.
- A parameter can be written as an unconstrained parameter of the same type multiplied with a fixed number.
- A variable can be expressed as the sum of an unconstrained or a fixed parameter (of the same type), an unknown (and to be estimated) shift and a fixed shift.
- An unknown (and to be estimated) ratio can also be used to express relations between parameters.

The use of these unknown shifts or ratios leads to the introduction of a new variable and as a result this new variable has to be shared by at least two peaks in order to reduce the total number of unknowns. Some of the prior knowledge that can be imposed is illustrated using an arbitrary example of a time series consisting of two ^{31}P signals of the ATP molecule. Using the peak numbering of Fig. 1.4, the prior knowledge available between the parameters of these 7 peaks within one signal can e.g. be expressed as follows for every signal in the time series, $s = 1, 2$:

1. $a_{2s} = a_{1s}$, $a_{4s} = a_{3s}$, $a_{6s} = 2a_{5s}, a_{7s} = a_{5s}$; a_{1s}, a_{3s} and a_{5s} are unconstrained, the other amplitudes are linked to them by a fixed ratio (1 or 2).

2. $d_{2s} = d_{1s}$, ..., $d_{7s} = d_{1s}$; d_{1s} is unconstrained and the other dampings are linked to it by a fixed ratio of 1.

3. $f_{2s} = f_{1s} - \Delta_{fixed}$, $f_{4s} = f_{3s} - \Delta_{fixed}$, $f_{6s} = f_{5s} - \Delta_{fixed}$, $f_{7s} = f_{5s} - 2\Delta_{fixed}$; $\Delta_{fixed} = 16$ Hz; f_{1s}, f_{3s} and f_{5s} are unconstrained and the other frequencies are linked to them by the fixed shift Δ_{fixed}.

4. $\phi_{2s} = \phi_{1s}$, ..., $\phi_{7s} = \phi_{1s}$; ϕ_{1s} is unconstrained and the other phases are linked to it using a fixed ratio of 1.

Suppose now, that due to a change in the field homogeneity after the measurement of the first signal (see also section 4.3, Example B), an unknown increase in damping Δ_{var1} and an unknown shift in frequency Δ_{var2} occurs for all peaks. The prior knowledge for the dampings and frequencies of the peaks of the second signal has to be changed to the following in order to express this extra information:

1. $d_{12} = d_{11} + \Delta_{var1}$, ..., $d_{72} = d_{11} + \Delta_{var1}$; d_{12}, ..., d_{72} are linked to the unconstrained parameter d_{11} by introduction of a new variable Δ_{var1}, the unknown shift in damping.

2. $f_{12} = f_{11} + \Delta_{var2}$, $f_{22} = f_{11} - \Delta_{fixed} + \Delta_{var2}$, ..., $f_{72} = f_{51} - 2\Delta_{fixed} + \Delta_{var2}$; f_{12}, f_{22} are linked to f_{11}; f_{32}, f_{42} are linked to f_{31} and f_{52}, f_{62}, f_{72} are linked to f_{51} by a fixed shift Δ_{fixed} and by the unknown shift in frequency Δ_{var2}.

Similar as was done in the processing of one signal, the prior knowledge imposed is a set of linear relations between parameters and as a consequence a minimization problem with linear equality constraints is obtained which are substituted in the original functional resulting in an unconstrained NLLS problem.

Analyzing the signals in one round is only useful when information on relations between different signals is present. This can be seen as follows. Consider each signal of the time series separately. The CRB on the jth parameter of the sth signal is given by $(1/\sigma^2 \mathbf{J}_s^T \mathbf{J}_s)^{-1}_{jj}$, the subscript jj denotes the element on the jth row and the jth column of the matrix and $\mathbf{J}_s$ is the Jacobian matrix. Consider now the processing of all signals simultaneously. If no prior knowledge

between the signals is present the associated Jacobian $\mathbf{J}_{series}$ becomes

$$\mathbf{J}_{series} = \begin{bmatrix} \mathbf{J}_1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & & & 0 \\ \vdots & & \mathbf{J}_s & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & & \mathbf{J}_S \end{bmatrix}.$$

The CR matrix corresponding to the simultaneous estimation of all the parameters of the time series then becomes

$$(\frac{1}{\sigma^2}\mathbf{J}_{series}^T\mathbf{J}_{series})^{-1} = \begin{bmatrix} (\frac{1}{\sigma^2}\mathbf{J}_1^T\mathbf{J}_1)^{-1} & 0 & \dots & \dots & 0 \\ 0 & \ddots & & & 0 \\ \vdots & & (\frac{1}{\sigma^2}\mathbf{J}_s^T\mathbf{J}_s)^{-1} & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & & (\frac{1}{\sigma^2}\mathbf{J}_S^T\mathbf{J}_S)^{-1} \end{bmatrix},$$

indicating clearly that the CRBs on the individual parameters do not decrease by processing the signals simultaneously and not using any information between spectra.

The simultaneous processing of multiple signals of the time series increases the complexity of the optimization problem since the number of unknowns becomes larger. The needed storage for the Jacobian increases dramatically. Alternative optimization routines to alleviate these problems do exist. Large scale methods (e.g. LANCELOT [29], VE10 [128]) for solving large-scale NLLS problems, exploit the so-called group partially separable structure of the NLLS problem. The exploitation of the latter structure allows for efficient storage and calculation of gradients and Hessians [127]. Since the main interest here is the analysis of the statistical properties of the proposed method, *dn2gb* is still used to minimize this cost function. This means that also here the Jacobian is calculated analytically and that upper and lower bounds on the variables can be imposed (see also the previous chapter). To perform the actual minimization, $\mathbf{\Gamma_s}$, $\mathbf{l}_s$ and $\mathbf{y}_s$, $s = 1, \dots, S$ are split into a real and imaginary part as done in Eq. (3.14). The starting values for the amplitudes and phases are obtained by solving $\mathbf{y}_s = \mathbf{\Gamma}_s\mathbf{l}_s,\ s = 1, \dots, S$, with the starting values for the frequencies and dampings inserted in $\mathbf{\Gamma}_s,\ s = 1, \dots, S$.

An outline of AMARES$_{ts}$ as used in the remainder of this chapter is given below.

Algorithm 4.1.1 (AMARES$_{ts}$)

Input: data samples $y_{ns}, n = 0, \ldots, N-1, s = 0, \ldots, S$, starting values for the unknowns, available prior knowledge within and between signals, upper and lower bounds on the variables.
Output: $\hat{a}_{ks}, \hat{d}_{ks}, \hat{f}_{ks}, \hat{\phi}_{ks}, k = 1, \ldots, K_s, s = 1, \ldots, S$ and error estimates.

Step 1. For every signal s, $s = 1, \ldots, S$: form $\mathbf{y}_s^{\mathrm{G}}$ as in Eq. (3.16) and compose $[\mathbf{y}_1^{\mathrm{G}^T}, \ldots, \mathbf{y}_S^{\mathrm{G}^T}]^T$.

Step 2. Call *dn2gb*. *dn2gb* uses user-supplied routines to evaluate the residual and the Jacobian.

Step 3. *dn2gb* returns parameter values.

Step 4. Calculate error estimates of the parameters.

4.1.2 Suboptimal use of common information present in spectra: AMARES$_{sts}$

An alternative to AMARES$_{ts}$, denoted here by AMARES$_{sts}$, that can be used to analyze time series in those cases where some of the parameters remain *constant* in time, has been described in literature [38]. The method proceeds as follows. In a first round every signal of the time series is analyzed separately using AMARES. Mean values of all the parameters known to remain constant in time are computed. Then an additional AMARES run is performed on every signal with these parameters fixed to the mean value found, in order to obtain more accurate parameter estimates for the time-varying parameters. This procedure leads to ML estimates as $S \to \infty$. Since only a limited number of signals are used to calculate the mean values, this approach may lead to biased estimates.

One could think of ways to extend AMARES$_{sts}$ to processing series of signals in which other types of time information are present, but the application of AMARES$_{sts}$ is less straightforward and more ad hoc in those cases and will not be pursued here.

4.2 Extensions of HTLS to quantitate time series

In [24] it is shown how HTLS can be extended to the simultaneous processing of spectra in analogy with the so-called multichannel representation in other applications [48]. In [24] the multichannel approach is illustrated on some simple signals comprising two sinusoids. Below, the principles of two HTLS-extensions, HTLSsum and HTLSstack are reviewed. Attention is drawn to the difficulties associated with these approaches in more realistic situations encountered in MRS and some possible remedies are presented.

4.2.1 HTLSsum

By summing the S signals of the time series a new signal is obtained

$$\begin{aligned} y'_n &= \sum_{s=1}^{S}(\sum_{k=1}^{K_s} a_{ks} e^{j\phi_{ks}} e^{(-d_{ks}+j2\pi f_{ks})n\Delta t}) = \sum_{s=1}^{S}(\sum_{k=1}^{K_s} c_{ks} z_{ks}^n) \\ &= \sum_{k=1}^{K_{series}} c'_k {z'_k}^n \quad n = 0, \ldots, N-1, \end{aligned} \tag{4.5}$$

where $z'_k, k = 1, \ldots, K_{series}$ are the different signal poles present in y'_n and $c'_k, k = 1, \ldots, K_{series}$ are the corresponding complex amplitudes. The signal is arranged in a Hankel matrix and steps 2 to 4 of the HTLS algorithm (Algorithm 2.1.1) are performed. The main drawback of this approach is the fact that the obtained signal poles still have to be assigned to a particular signal. This turns out to be a serious problem in practical situations as illustrated in section 4.3. Two approaches are investigated here.

The first one is based on the estimates of the amplitudes: for every signal the estimates of the K_{series} signal poles are substituted in the N model equations

$$y_{ns} \approx \sum_{k=1}^{K_{series}} c''_k {\hat{z}'_k}^n \, , n = 0, \ldots, N-1. \tag{4.6}$$

The K_s signal poles belonging to the largest amplitude estimates are assigned to the sth signal. Similarly to the processing of one signal, this approach returns the exact results when no noise is present but is statistically suboptimal when applied to noisy signals.

A second approach is based on the following principle. The K_s signal poles in each signal s are composed of K_{fixed} signal poles corresponding to time-invariant frequencies and K_{shift} signal poles corresponding to the time-varying

frequencies. Using biochemical information and/or peak picking on one of the signals, the approximate frequencies of the K_{fixed} peaks which do not shift over time are determined. In a first step, the K_{fixed} poles out of the set of K_{series} estimated poles, which frequency lie closest to these approximate frequencies, are selected. Moreover, usually the frequency ranges in which each of the K_{shift} shifting poles lie, are known. Therefore, a second step consists of selecting from the remaining set of $K_{series} - K_{fixed}$ poles, the K_d poles whose frequencies lie in the indicated frequency ranges. If at least one pole can be detected within each frequency range, the algorithm does not fail. If more than one pole per interval is detected, a further selection has to be made in the following way. For each signal of the time series, the previously determined $K_{fixed} + K_d$ poles are filled in. The corresponding amplitudes are determined and the pole having the largest amplitude is selected within each frequency interval. In a final step amplitudes and phases belonging to the finally retained signal poles are calculated.

4.2.2 HTLSstack

Consider noiseless data and arrange every signal of the time series in a Hankel matrix $\bar{\mathbf{H}}_{s_s}, s = 1, \ldots, S \in \mathbb{C}^{(N-M)\times(M+1)}$ as in Eq. (2.29). The S Hankel matrices are subsequently stacked horizontally resulting in a block Hankel matrix $\bar{\mathbf{H}}_{series} = [\bar{\mathbf{H}}_{s_1}, \ldots, \bar{\mathbf{H}}_{s_S}]$. $\bar{\mathbf{H}}_{series}$ can be decomposed as

$$\bar{\mathbf{H}}_{series} = \bar{\mathbf{S}}_{series} \left[\begin{array}{ccc} \bar{\mathbf{C}}_1 \bar{\mathbf{T}}^T_{series} & \ldots & \bar{\mathbf{C}}_S \bar{\mathbf{T}}^T_{series} \end{array} \right] \tag{4.7}$$

with

$$\bar{\mathbf{S}}_{series} = \left[\begin{array}{cccc} 1 & 1 & \ldots & 1 \\ z'_1 & z'_2 & \ldots & z'_{K_{series}} \\ \vdots & \vdots & \vdots & \vdots \\ z_1'^{N-M-1} & z_2'^{N-M-1} & \ldots & z_{K_{series}}'^{N-M-1} \end{array} \right]$$

and

$$\bar{\mathbf{T}}_{series} = \left[\begin{array}{cccc} 1 & 1 & \ldots & 1 \\ z'_1 & z'_2 & \ldots & z'_{K_{series}} \\ \vdots & \vdots & \vdots & \vdots \\ z_1'^{M} & z_2'^{M} & \ldots & z_{K_{series}}'^{M} \end{array} \right],$$

$\mathbf{C}_s \in \mathbb{C}^{K_{series}\times K_{series}}, s = 1, \ldots, S$ is a diagonal matrix where the kth diagonal element is zero if signal pole z'_k does not enter in signal s. There is again a shift-invariant property in the column space of $\bar{\mathbf{H}}_{series} = \bar{\mathbf{U}}_{series_K} \bar{\mathbf{\Sigma}}_{series_K} \bar{\mathbf{V}}^H_{series_K}$, i.e.

$$\bar{\mathbf{S}}^{\uparrow}_{series} = \bar{\mathbf{S}}_{series_{K\downarrow}} \bar{\mathbf{Z}}_{series} \Rightarrow \bar{\mathbf{U}}^{\uparrow}_{series_K} = \bar{\mathbf{U}}_{series_{K\downarrow}} \bar{\mathbf{Q}}^{-1}_{series} \bar{\mathbf{Z}}_{series} \bar{\mathbf{Q}}_{series},$$

$\bar{\mathbf{Z}}_{series} \in \mathbb{C}^{K_{series} \times K_{series}}$ is a diagonal matrix composed of the K_{series} different signal poles, $\bar{\mathbf{Q}}_{series} \in \mathbb{C}^{K_{series} \times K_{series}}$ is a non-singular matrix. Then steps 2 to 4 of the HTLS algorithm (Algorithm 2.1.1) are performed with the matrix $\mathbf{H}$ replaced by $\mathbf{H}_{series}$ and with $K = K_{series}$. In this way estimates for the dampings d_k and frequencies f_k, $k = 1, \ldots, K_{series}$ are obtained. The problem of assigning signal poles to a particular signal remains the same as with HTLSsum. In a final step amplitudes and phases belonging to the finally retained signal poles are calculated. When the Hankel matrices are stacked vertically instead of horizontally, the shift-invariant property holds in the row space. Below a schematic outline of the HTLSstack algorithm is given.

Algorithm 4.2.1 (HTLSstack)

Input: data samples $y_{ns}, n = 0, \ldots, N-1, s = 1, \ldots, S$, model order K_{series}, individual model order K_s for each signal s.
Output: $\hat{a}_{ks}, \hat{d}_{ks}, \hat{f}_{ks}, \hat{\phi}_{ks}, k = 1, \ldots, K_s, s = 1, \ldots, S$.

Step 1. For every signal s, $s = 1, \ldots, S$: arrange the data points $y_{ns}, n = 0, \ldots, N-1$ in a Hankel matrix $\mathbf{H}_{s_s}$ of dimensions $(N-M) \times (M+1)$. Form a block Hankel matrix $\mathbf{H}_{series} = [\mathbf{H}_{s_1}, \ldots, \mathbf{H}_{s_S}]$.

Step 2. Compute the SVD of $\mathbf{H}_{series}$

$$\mathbf{H}_{series} = \mathbf{U}_{series} \mathbf{\Sigma}_{series} \mathbf{V}^H_{series}.$$

Step 3. Truncate $\mathbf{H}_{series}$ to a matrix $\mathbf{H}_{series_K}$ of rank K_{series}

$$\mathbf{H}_{series_K} = \mathbf{U}_{series_K} \mathbf{\Sigma}_{series_K} \mathbf{V}^H_{series_K}$$

Step 4. Form the following overdetermined set of equations

$$\mathbf{U}^{\uparrow}_{series_K} \approx \mathbf{U}_{series_K \downarrow} \mathbf{E}_{series}$$

and compute the TLS solution $\hat{\mathbf{E}}_{series}$. Once $\mathbf{E}_{series}$ is estimated, its K_{series} eigenvalues give the signal pole estimates

$$\hat{z}'_k = e^{(-\hat{d}'_k + j2\pi \hat{f}'_k)\Delta t}, \quad k = 1, \ldots, K_{series}.$$

Step 5. For every signal s, K_s signal poles are selected out of the K_{series} signal poles. The corresponding estimates for amplitudes a_{ks} and phases ϕ_{ks}, $k = 1, \ldots, K_s$ are derived from the LS solution of

$$y_{ns} \approx \sum_{k=1}^{K_s} c_{ks} \hat{z}^n_{ks}, \quad n = 0, \ldots, N-1.$$

4.3 Numerical validation

In this section the methods presented are evaluated using Monte-Carlo studies on two typical time series derived from experimental data.

In all examples each of the signals is modeled exactly by Eq. (1.3). Statistical parameters like the (R)RMSE are computed from the estimated parameters (using 400 runs), excluding parameters belonging to signals on which the method fails. A failure or bad run occurs if not all peaks are resolved within specified intervals lying symmetrically around the exact frequencies.

4.3.1 Example A

Monte-Carlo studies were performed on a series of 40 simulated signals, derived from 40 ^{31}P-signals of an *ex-vivo* perfused rat liver, acquired consecutively during administration of the fructose analogue 2,5-anhydro-D-mannitol (aHMol) (signals 4-21), subsequent administration of potassium cyanide (KCN) (signals 22 to 30) and removal of aHMol (signals 31 to 40) [19]. The corresponding changes in the amplitude of the PME (in casu, aHMolP esters), P_i and ATP peaks and in the frequency of the PME and Pi peaks are shown in Fig. 4.1.

Peak linewidths remain constant over the experiment. The experimental signals were recorded at 81.1 MHz (4.7 T Bruker Biospec) with a temporal resolution of 100 s. Each signal consists of 128 complex data points. The sampling interval is 0.2 ms, t_0 and $\phi_k, k = 1, \ldots, K$, are zero. In Fig. 4.2 peaks from ES, PME, P_i), PDE and α-, β-, and γ-ATP are observed. The top figure of Fig. 4.2 shows a noiseless version of the first signal in the time series.

Due to the line broadening, the frequency splitting between the ATP multiplets is not visible from Fig. 4.2 but it is present. Noise from a Gaussian distribution with standard deviation σ is added to the signals. Seven noise levels varying form $\sigma = 500$ to 3500 in steps of 500 were used. This implies that the SNR for e.g. the reference peak (ES), expressed in dB ranges from 57.6 to 18.7 dB. One noisy realization ($\sigma = 3500$) is shown in Fig. 4.2. Noise levels of $\sigma = 2000$ or 2500 are typical for *in-vivo* experiments.

4.3.1.1 AMARES-based methods

In all results based on AMARES the following prior knowledge is used:

1. Within a signal: ATP prior knowledge as explained in section 4.1.1, $\phi_k = 0, k = 1, \ldots, K$, $t_0 = 0$; damping of PME and P_i equal.
2. Between signals: the frequencies of ATP, PDE and ES, the dampings of

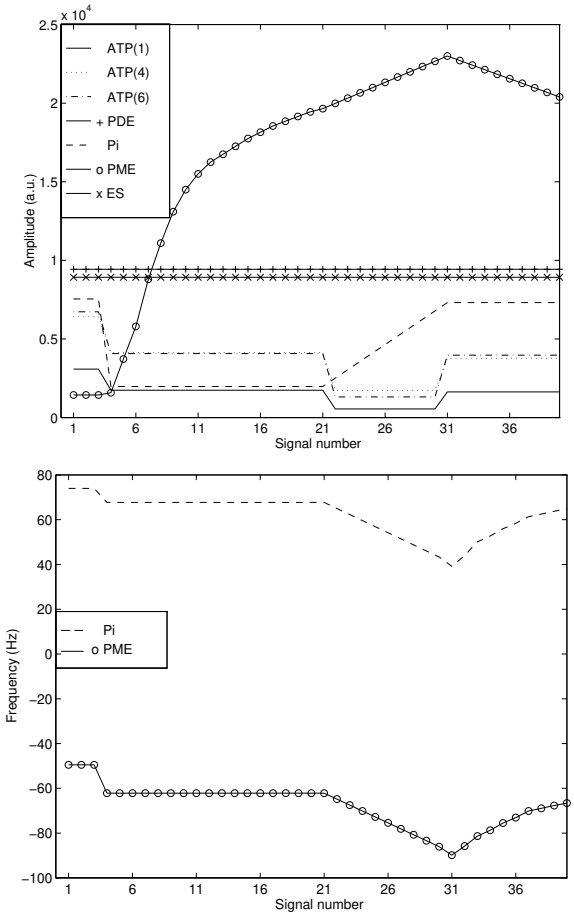


Figure 4.1: Time variation of the parameters used in Example A. Top: amplitude variation of ES, PME, P_i, PDE and α-, β- and γ-ATP as a function of time (signal number). The number in brackets in the ATP multiplets refers to the peak number as defined in Fig. 1.4. Bottom: frequency variation of P_i and PME as a function of time (signal number).

all peaks and the amplitude of ES remain constant.

This leads to the following number of parameters to be estimated: $(6nrsignals+1)$ unknown amplitudes, 4 unknown dampings, $(2nrsignals+5)$ frequencies, $nrsignals$ denotes the number of signals that is processed simultaneously. One set of starting values for the frequencies and dampings was obtained by peak-picking the first signal of the time series. These starting values were used for all the signals in the time series.

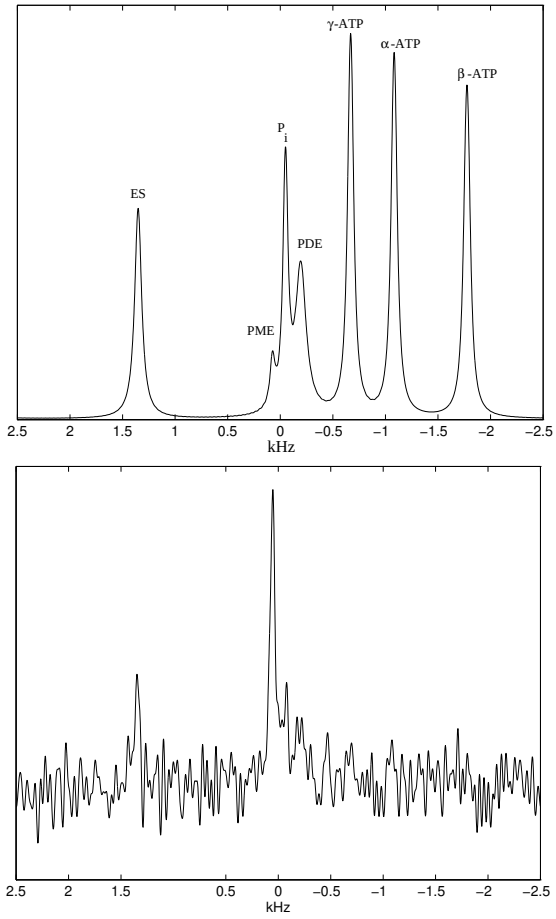


Figure 4.2: Top: first signal of time series used in Example A without noise. Bottom: noisy realization ($\sigma = 3500$) of signal number 22.

The number of signals analyzed simultaneously has to be chosen. There is a trade-off between accuracy and speed. On the one hand, results get more accurate as more signals are analyzed together. On the other hand, computer time and storage requirements increase. To gain some insight in this matter, the possible improvements that can be made using the AMARES$_{ts}$ approach were analyzed by looking at the theoretical CRBs. A timing experiment was also performed. In Fig. 4.3 the (relative) CRBs of the amplitude, damping and frequency of the peaks of signal 1 are plotted versus the total number of signals that is processed simultaneously. The accuracy of parameters which are constrained to remain constant over time (i.e. the dampings of all peaks,

the frequencies of all peaks except the ones of PME and P_i and the amplitude of ES) increases substantially when more signals are processed simultaneously. The accuracy of the amplitudes of PDE and P_i is also increased significantly. Worth noting here is the fact that the accuracy of these peaks does not improve much further by processing more than about 10 signals at the same time.

Fig. 4.4 shows the required CPU time (measured on a SUN ULTRA2 (200 MHz)). To process 2 signals individually with AMARES approximately 1.9 s are needed, compared to 2 s with AMARES$_{ts}$. When every signal is analyzed separately using AMARES the required CPU time varies linearly with the number of signals processed. For AMARES$_{ts}$ the needed CPU time increases approximately cubically. For processing 20 signals at once 11 minutes are needed compared to 75 s when 10 signals are processed at once. Based on these observations concerning accuracy and required CPU time the time series are divided into 4 consecutive time frames of each 10 signals (frame 1 consists of signals 1 to 10 of the time series, frame 2 is made up of signals 11 to 20 and so on).

On each of these frames three methods in the Monte-Carlo study are compared: AMARES on each signal separately, AMARES$_{ts}$ and AMARES$_{sts}$.

In Table 4.1 the number of times these 3 algorithms fail on the 4 different time frames as a function of the noise standard deviation is shown. Both AMARES$_{ts}$ and AMARES$_{sts}$ have a lower failure rate than AMARES applied to every signal individually.

In the top of Fig. 4.5 the three methods are compared in terms of the accuracy of the amplitude of PDE for the different signals in the time series ($\sigma = 3500$). The methods are compared to the theoretical CRBs. In Fig. 4.5 three different CRBs are plotted. One corresponding to the analysis of each signal separately, one corresponding to analyzing the signals simultaneously in sets of 10 signals and one which shows the accuracy that can be obtained when the 40 signals are processed together. Again it can be observed that not much accuracy is lost by splitting up the time series in 4 frames of 10 signals. AMARES and AMARES$_{ts}$ lie close to the corresponding CRBs. Note that the erratic behavior of AMARES in time frame 1 and 3 is due to the fact that the failure rate of the method is very high (78% and 84.5% respectively) in these frames and that the RRMSE shown is computed over a limited amount of successful runs. The erratic behavior disappears for lower noise levels. When AMARES$_{ts}$ and AMARES$_{sts}$ are compared for all the parameters it can be observed that the first follows the CRBs more closely. As an example of this the RRMSE of the damping of P_i as a function of time is shown in the bottom of Fig. 4.5. In time frame one, three and four, the estimates for the damping of P_i are further away from the CRB than the estimates obtained by AMARES$_{ts}$ which closely follow the CRB. The gain in accuracy obtained by using AMARES$_{ts}$ is evident from these figures.

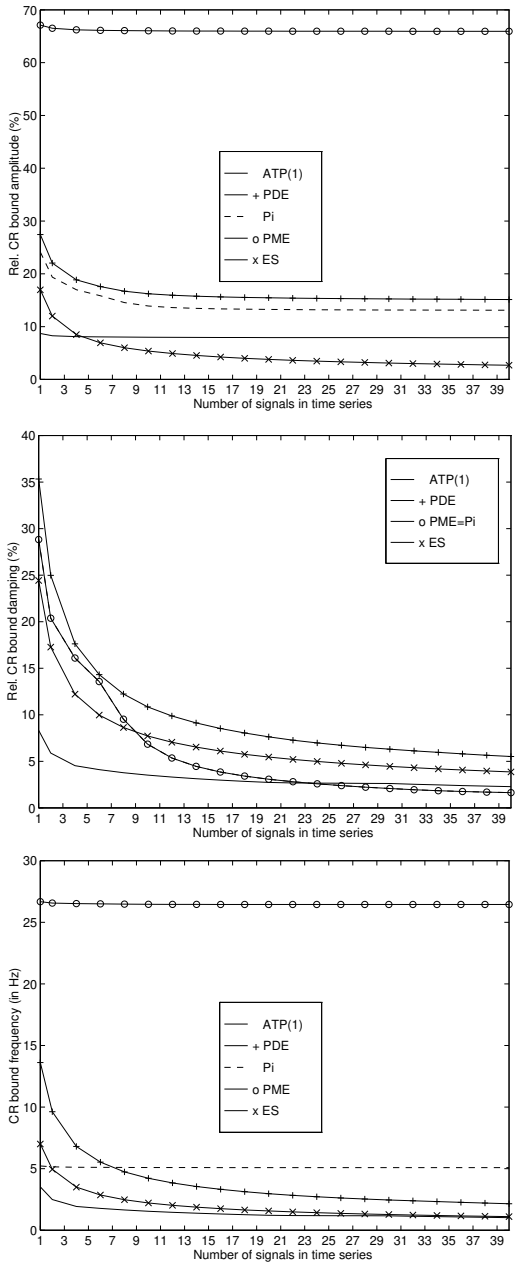


Figure 4.3: (Rel.) CRBs (calculated at $\sigma = 3500$) of amplitudes, dampings and frequencies of ATP(1), PDE, P_i, PME and ES of the first signal in the time series of Example A are shown as a function of the number of consecutive signals of the time series that are processed simultaneously.

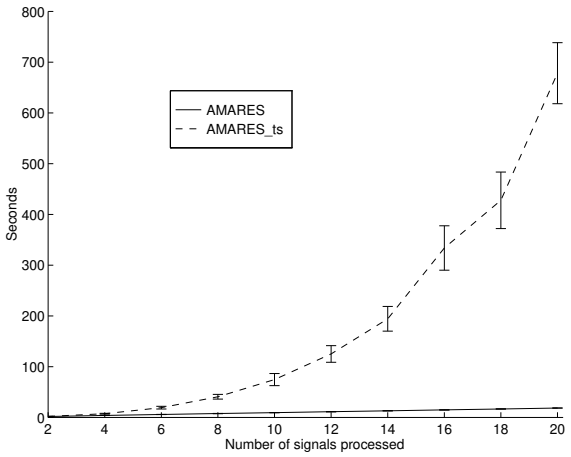


Figure 4.4: Mean value and standard deviation (denoted by the error bars) of the CPU time in seconds needed by AMARES and AMARES$_{ts}$ as a function of the number of signals processed simultaneously (Example A).

4.3.1.2 HTLS-based methods

First it is very important to note that the HTLS-based methods can not be expected to resolve the multiplet structure of ATP. Even with AMARES it is impossible to do so without imposing prior knowledge. This is also found in the CRBs: if no prior knowledge about the multiplet structure is imposed, extremely high CRBs are found on the parameters. Since no prior knowledge about frequency splittings, equal dampings and relations between amplitudes is imposed HTLS is only expected to find 1 peak for each of the α-, β- and γ-ATP peaks. For this study only frame 1 of the time series is considered. In this time frame, as can be seen in Fig. 4.1, the frequencies of PME and P_i are constant for signals 1 to 3, then drop to a lower value for signal 4 after which they remain constant. This means that the total number of different frequencies is equal to 13. The theoretical model order is thus 13. However, the course of the frequency variations is normally not known. So from a practical point of view the most logical order to impose is equal to 25, with 3 poles for ATP, 1 for PDE and ES and 2 *nrsignals* poles for the shifting PME and P_i. Simulations using HTLSstack and HTLSsum were performed on the 10 signals of frame 1 using the two different pole selection procedures as outlined in section 4.2.1. When using the procedure based on amplitude selection, no reasonable results could be obtained. It was almost never possible to retrieve the 7 peaks of interest, irrespective of the model order used. Using the approach based on the selection of frequencies, much better results are obtained. The methods HTLS, HTLSstack and HTLSsum are compared using model orders 13 and 25.

σ	method	frame 1	frame 2	frame 3	frame 4
1000	AMARES	0	0	0	0
	AMARES$_{sts}$	0	0	0	0
	AMARES$_{ts}$	0	0.25	0	0
1500	AMARES	0.75	0	0.25	0
	AMARES$_{sts}$	1	0	0	0
	AMARES$_{ts}$	0	0	0	0
2000	AMARES	7.75	0.5	6.25	1.25
	AMARES$_{sts}$	5.75	0.5	0	0
	AMARES$_{ts}$	4.75	0	1	0
2500	AMARES	27.5	12.25	30.25	9.75
	AMARES$_{sts}$	18.75	8	4.5	0
	AMARES$_{ts}$	19.5	6.25	3.75	0
3000	AMARES	57	26.5	55.25	21.5
	AMARES$_{sts}$	30.5	21.75	22.75	0.25
	AMARES$_{ts}$	41	25.25	21.25	0.5
3500	AMARES	78	59	84.25	43.75
	AMARES$_{sts}$	49.75	42.25	41	2.5
	AMARES$_{ts}$	60.75	41	39.5	2

Table 4.1: Failure rates in % for different methods used on 4 consecutive time frames consisting of 10 signals each as a function of the noise level σ. Used methods: AMARES applied to every signal separately, AMARES$_{sts}$ and AMARES$_{ts}$.

The corresponding failure rates are depicted in Table 4.2. All multichannel approaches have drastically lower failure rates than HTLS. The failure rates are higher when 25 is taken as the model order. HTLSsum and HTLSstack have almost the same performance.

In Fig. 4.6 the multichannel HTLS methods and AMARES$_{ts}$ are compared in terms of the relative RMSE of the amplitudes of ATP(1) and PDE as a function of the signal number. As expected all multichannel HTLS approaches perform worse than AMARES$_{ts}$, but they still manage to track the amplitude variations over time. Using a higher model order results in a higher RMSE. For the theoretical model order of 13 HTLSsum and HTLSstack perform equally well. For model order 25 there is no clear winner either since for some peaks HTLSsum performs better than HTLSstack and for other peaks it is the other way around. The results are however promising enough to conduct further

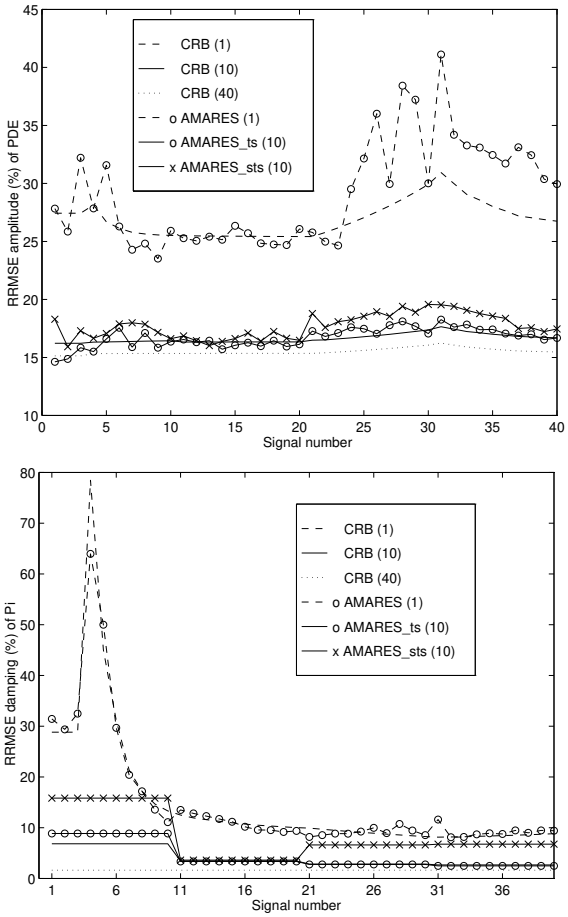


Figure 4.5: Comparison of methods on the time series of Example A. Top: RRMSE ($\sigma = 3500$) of the amplitude of PDE as a function of time (signal number) obtained by AMARES applied separately to each signal, AMARES$_{sts}$ (4 frames of 10 signals) and AMARES$_{ts}$ (4 frames of 10 signals). Bottom: RRMSE ($\sigma = 3500$) of the damping of P_i as a function of time (signal number) obtained by AMARES applied separately to each signal, AMARES$_{sts}$ (4 frames of 10 signals) and AMARES$_{ts}$ (4 frames of 10 signals). Theoretical CRBs are also depicted and denoted by CRB in the figure. The number in brackets indicates the number of signals that is processed simultaneously.

studies by using other pole selection techniques and recently developed techniques to incorporate some types of prior knowledge [26, 25, 140].

σ	HTLS(11)	HTLS stack(13)	HTLS sum(13)	HTLS stack(25)	HTLS sum(25)
500	6	1	0	4	1
1000	99	2	1	6	3
1500	100	3	0	7	7
2000	100	1	4	6	9
2500	100	2	7	9	11
3000	100	7	10	16	24
3500	100	11	16	23	26

Table 4.2: Failure rates (in %) for HTLS, HTLSsum, HTLSstack used on time frame 1 as a function of the noise level. The number in brackets denotes the model order used.

4.3.2 Example B

A Monte-Carlo study is performed on a series of 40 simulated signals, derived from 40 *in-vivo* ^{31}P signals of the calf muscle of a healthy human, acquired consecutively during an initial rest period (signals 1-3), an isometric contraction (signals 4-15) and the subsequent recovery period (signal 16-40). The experimental signals were acquired at 81.1 MHz (4.7 T Bruker Biospec) with a temporal resolution of 10 s, using a 5 cm diameter surface coil positioned against the calf muscle. The corresponding changes in amplitude of the P_i and PCr peaks and of the frequency of the P_i peak are shown in Fig. 4.7. Also simulated is the often observed additional line broadening and frequency shift of all peaks during contraction, arising e.g. from the altered B_0 homogeneity profile.

Each signal consists of 128 data points. The sampling interval is 0.3 ms, t_0 and $\phi_k, k = 1, \ldots, K$, are zero. In Fig. 4.8 a noisy realization of the first signal of the simulated time series is displayed. Peaks from P_i, PCr and α-, β-, and γ-ATP can be observed. The standard deviation σ used for the study is derived from the *in-vivo* time series and is equal to 5.2. This corresponds to a SNR of 7.6 dB for the peaks in the α- and γ-ATP doublets.

For the ATP peaks the prior knowledge as described in section 4.1.1 is used. The time dependencies imposed are the following:

1. The amplitudes of all the ATP peaks remain constant.

2. Dampings and frequencies of ATP and PCr remain constant for signals 1 to 3. In signal 4 the frequencies of PCr and ATP are shifted with a fixed

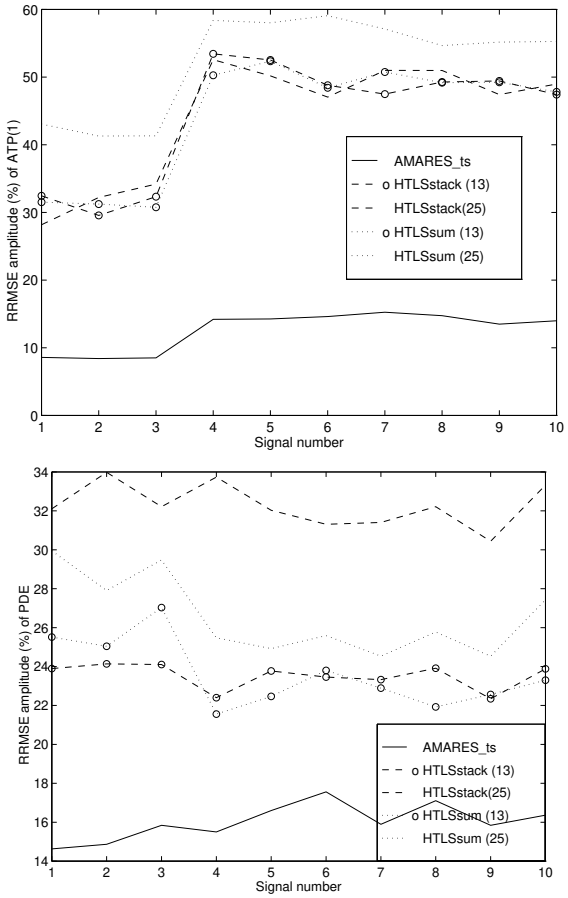


Figure 4.6: RRMSE ($\sigma = 3500$) of the amplitudes of ATP(1) and PDE as a function of time (signal number) obtained by applying HTLSstack and HTLSsum. The used model orders are shown in brackets. For the sake of comparison the result obtained by AMARES$_{ts}$ is also shown.

unknown amount which remains constant throughout signal 15. Similarly, the dampings of PCr and ATP are increased from signal 4 throughout 15 with a fixed, unknown amount which remains constant. In signal 16 they return to their original values and remain constant throughout signal 40.

3. No temporal information is available for the P_i peak.

The prior knowledge can be taken into account by AMARES$_{ts}$ as explained in section 4.1.1. The time series is again analyzed in 4 frames, consisting of 10

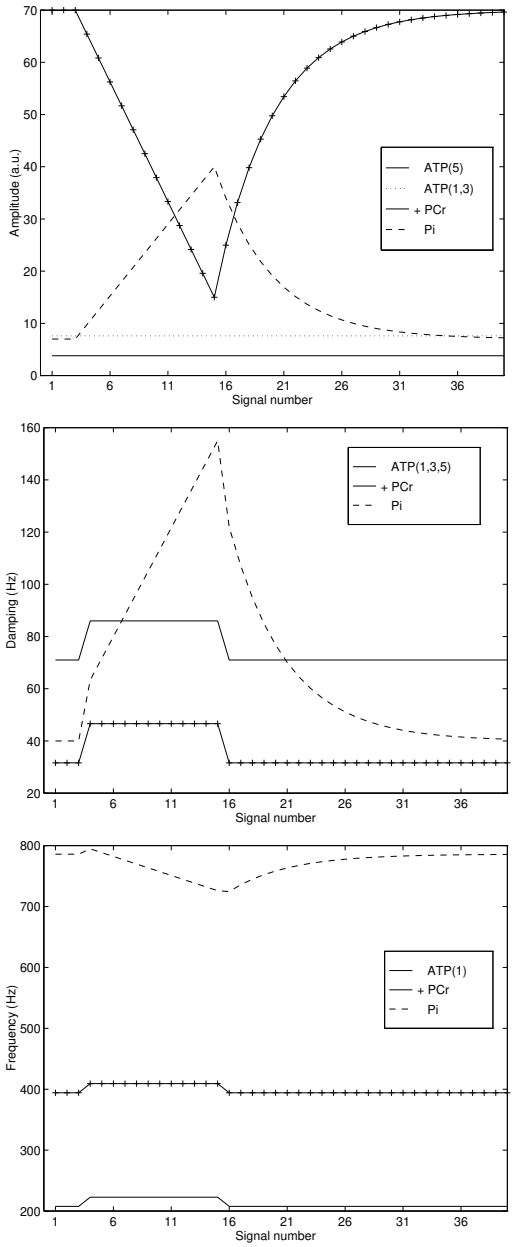


Figure 4.7: Time variation of the parameters used in Example B. Top: amplitude variation of P_i, PCr and α-, β-, and γ-ATP as a function of time (signal number). Middle: damping variation of P_i, PCr and ATP (α-, β-, and γ-ATP) as a function of time (signal number). Bottom: frequency variation of P_i, PCr and γ-ATP (peak number 1 in Fig. 1.4; the variation of the frequencies of α- and β-ATP (not shown) is similar to that of γ-ATP) as a function of time (signal number). The number in brackets in the ATP multiplets refers to the peak number as defined in Fig. 1.4.

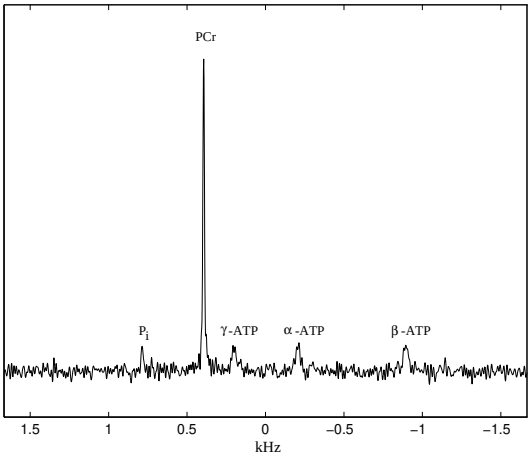


Figure 4.8: Noisy realization ($\sigma = 5.2$) of the first signal of the time series used in Example B.

signals each. In this study only AMARES and AMARES$_{ts}$ were compared since there is no straightforward way to apply AMARES$_{sts}$ in this case. In Fig. 4.9 the two methods are compared in terms of the accuracy of the amplitude of α-ATP and PCr for the different signals in the time series. The methods are compared to the theoretical CRBs. Three different CRBs are plotted. One corresponding to the analysis of each signal separately, one corresponding to analyzing the signals simultaneously in sets of 10 signals and one which shows the accuracy that can be obtained when the 40 signals are processed together. Again it can be seen that not much accuracy is lost by splitting up the time series in 4 frames of 10 signals. AMARES and AMARES$_{ts}$ lie close to the corresponding CRBs. The gains in accuracy obtained by applying AMARES$_{ts}$ are considerable for those peaks. For the P_i peak, for which no time information is present, there is no significant gain in amplitude accuracy (not shown). The gain in amplitude accuracy of the β- and γ-ATP peaks (not shown) is the same as for the α-ATP peaks.

In general, the gain in accuracy highly depends on the values and the time evolution of the parameters in the time series. For every specific case however, it is possible to calculate the corresponding CRBs, which gives an excellent indication of the gain in accuracy that can be obtained by using AMARES$_{ts}$.

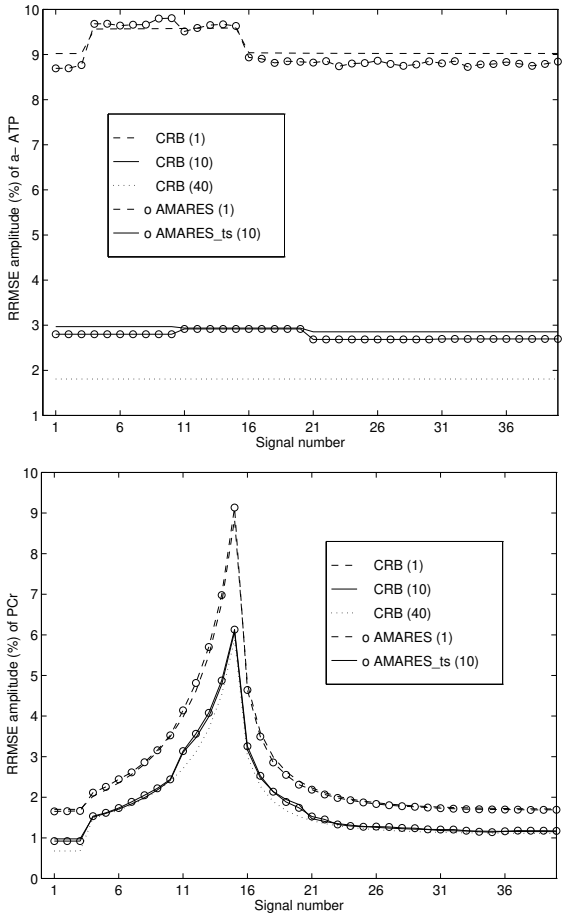


Figure 4.9: Comparison of methods on the time series of Example B. Top: RRMSE ($\sigma = 5.2$) of the amplitude of the α-ATP peaks as a function of time (signal number) obtained by AMARES applied separately to each signal and AMARES$_{ts}$ (4 frames of 10 signals). Bottom: RRMSE ($\sigma = 5.2$) of the amplitude of PCr as a function of time (signal number) obtained by AMARES applied separately to each signal and AMARES$_{ts}$ (4 frames of 10 signals). Theoretical CRBs are also depicted and denoted by CRB in the figure. The number in brackets indicates the number of signals that is processed simultaneously.

4.4 Experimental time series

The experimental time series already described in section 4.3.1 is analyzed. An example of a signal of this time series is shown in Fig. 1.6. In all the applied

methods the first 5 data points are omitted to reduce the influence of the broad resonance underlying the metabolites of interest.

The forty signals were analyzed using AMARES applied to each signal individually. The same prior knowledge as in section 4.3.1 was imposed; the phases and the begin time were fixed, the ATP prior knowledge was imposed and the dampings of PME and P_i were constrained to be equal. In two cases out of forty the algorithm failed to converge. In three other cases the PDE peak is obviously badly fitted. An example of such a bad fit is shown in the top of Fig. 4.10. The time series were also fitted with AMARES$_{ts}$ and in addition to the prior knowledge present within a signal it was imposed that the frequencies of ATP, PDE and ES, the dampings of all peaks and the amplitude of ES, remain constant over time. When the signals were analyzed using AMARES$_{ts}$ no convergence problems were encountered and all the signals were nicely fitted. In the top of Fig. 4.10 the corresponding successful AMARES$_{ts}$ fit of the signal badly fitted with AMARES is shown. When AMARES$_{sts}$ is used there are no convergence problems either. However, since this is an experimental signal there is no real objective way of drawing conclusions out of the small differences in fit obtained by AMARES$_{ts}$ and AMARES$_{sts}$.

The time series was also analyzed using the multichannel HTLS extensions. The second signal pole assignment procedure as described in section 4.2.1 was used. When a model order of 11 is used, it is possible to retrieve the 7 peaks in 11 cases, but only in 4 cases the fit is more or less acceptable. HTLSstack is able to give a reasonable fit of 29 of the spectra. In the badly fitted spectra the problems mainly occur in the PME, P_i and PDE region. The results obtained with HTLSsum are very similar. In the top of Fig. 4.11 a signal reasonably fitted with HTLSstack is shown, in the bottom a signal is shown where HTLSstack has problems fitting the PME, P_i and PDE region. It is clear however that the extensions of AMARES and HTLS to process signals simultaneously perform better than their respective counterparts.

4.5 Conclusions

Quantification of individual MRS spectra in the time domain is possible using interactive methods like AMARES or black-box methods like HTLS. Four methods (AMARES$_{ts}$, AMARES$_{sts}$, HTLSsum and HTLSstack) which are able to process time series taking into account information concerning the time evolution of the parameters were presented and evaluated. AMARES$_{ts}$, an extension of AMARES, allows taking into account in a statistically optimal way the common information present in spectra of a time series. The method performs very well in practical situations. AMARES$_{ts}$ leads to improved and more robust estimates than the ones obtained by processing the signals individually

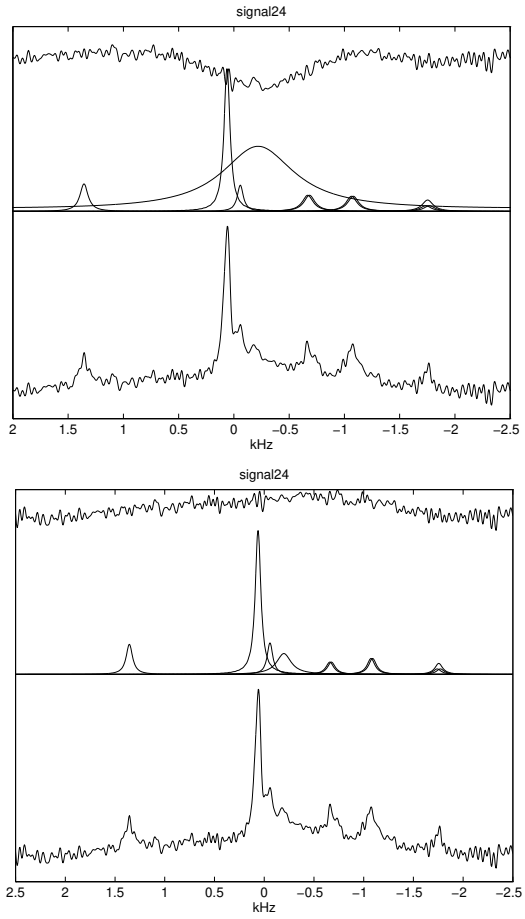


Figure 4.10: Top: signal 24 of the time series fitted with AMARES. Bottom: signal 24 of the time series fitted with AMARES$_{ts}$.

with AMARES since in the latter case the information present between the spectra can not be taken into account. AMARES$_{sts}$ makes suboptimal use of the common information present in spectra and is restricted in the sense that the types of prior knowledge that can be imposed in a straightforward way are rather limited. In those situations where the method can be used, it performs better than AMARES applied on every signal and slightly worse than AMARES$_{ts}$. Further studies using the HTLS multichannel extensions HTLSstack and HTLSsum were conducted. Although these algorithms do not perform as well in terms of precision as AMARES$_{ts}$ and need some more fine tuning, they perform better than HTLS applied on every signal individually and are promising black-box techniques to further automate the MRS data

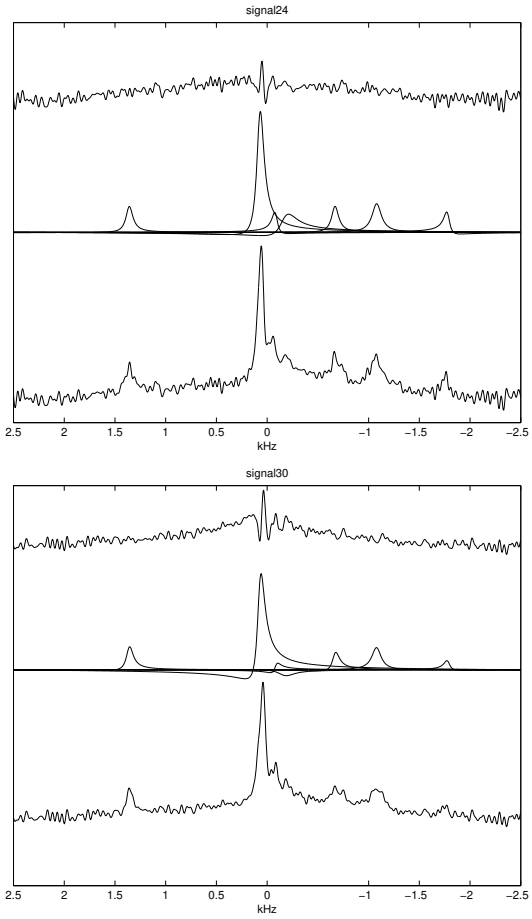


Figure 4.11: Top: reasonable fit of signal 24 obtained with HTLSstack (used model order is 25). Bottom: HTLSstack has problems quantifying the PME, P_i and PDE region of signal 30.

processing. It is also straightforward to extend the ideas used in the derivation of the AMARES$_{ts}$ algorithm to frequency domain fitting methods. The same concepts as explained here can also be applied to MRSI.

Chapter 5

Parameter estimation in non-ideal circumstances

In many cases the biomedical MRS data do not obey the theoretical model functions of section 1.3.1 and contain disturbing resonances of which little is known. Examples are e.g. eddy currents which distort the lineshape and residual water in proton spectra. The quantitation results can often be improved using techniques to adjust for the data imperfections. Many of the methods are actually preprocessing methods which can be carried out before the actual quantitation is performed in the time or frequency domain.

The goals of this chapter are:

- *To give an idea of the methods that can be used to compensate for model imperfections.*
- *To provide some remarks on the use of techniques used to improve the visual SNR and/or resolution in relation to parameter accuracy.*
- *To give an overview of methods that can be used to reduce the influence of unwanted spectral features on the parameter estimates of the metabolites of interest.*

Section 5.1 discusses methods that deal with model imperfections. Section 5.2 discusses the use of techniques to improve the visual SNR and/or resolution and their relation w.r.t. parameter accuracy. Different techniques to reduce the influence of unwanted spectral features in the spectrum are discussed in section 5.3. The focus of section 5.3 lies on the discussion of methods used to remove components which are separated in frequency from the metabolites of interest. Section 5.3 serves as an introduction to chapters 6, 7 and 8.

5.1 Corrections for model imperfections

Severe signal distortions most often appear in biomedical MRS due to restriction of molecular motion, instrumental imperfections and sample inhomogeneities. These shortcomings might limit the spectral resolution and result in systematic quantitation errors. These errors occur in particular for model-based quantitation algorithms since in general the resonance peaks do not have the assumed ideal lineshapes (such as Lorentzian, Gaussian or Voigt). Different techniques to handle such problems have been proposed in the literature.

The common idea for time-domain preprocessing techniques used in this context is to determine the signal distortions by comparing a reference peak to a peak having an ideal lineshape. The required reference peak can be acquired as a separate measurement or extracted from the spectrum itself. If a reference peak is obtained from a separate measurement, this results in a high quality estimate of the deviations. Time is however spent on performing the additional measurements and furthermore it is questionable if the distortions are actually equal through different experiments even though the same measurement sequences and hardware are used. Extracting a reference peak from the spectrum itself offers as obvious advantage that no separate measurement needs to be performed. Unfortunately in many cases no suitable reference peak can be found in the spectrum that is sufficiently separated from other peaks and has a high enough SNR. Some of the work in this area based on both these approaches is described in more detail in the following.

Initial efforts dealt with the correction of eddy current induced distortions. The eddy currents give rise to a time-varying phase-shift in the acquired data. In [99] it was proposed to determine the phase-shift from a one-peak reference signal $yref_n, n = 0, \ldots, N-1$ and subsequently correct the data according to

$$ycorr_n = y_n e^{-j\theta ref_n}, \tag{5.1}$$

where

$$\theta ref_n = \tan^{-1}\left(\frac{Im(yref_n)}{Re(yref_n)}\right) \tag{5.2}$$

denotes the instantaneous phase of the reference signal. The method was revisited in [70] for the special case of ^{1}H MRS. There the instantaneous phase is obtained from separate data acquired without using any hardware water suppression. The high signal energy of the unsuppressed water peak compared to the metabolite peaks is expected to give a reliable measure of the instantaneous phase of the water peak only. Thereafter the corrections are applied to the acquired water suppressed data. This method is standardly used for eddy current corrections in ^{1}H MRS.

More general methods that are able to adjust for more general data imperfections have also been proposed. In [94] a reference peak is chosen as one of

the peaks in the experimental data. The time-domain data for the isolated reference peak, $yref_n, n = 0, \ldots, N-1$, are obtained by setting the spectral values outside the frequency region of the reference peak to zero and using the inverse Fourier transform (FT). An ideal reference peak signal *yideal* is constructed via the inverse FT of the predicted reference peak spectrum. Finally, the corrected data $ycorr_n, n = 0, \ldots, N$ are obtained by

$$ycorr_n = y_n \frac{yideal_n}{yref_n} \qquad n = 0, \ldots, N-1. \tag{5.3}$$

A potential drawback is that the reference signal might be equal or close to zero what can lead to numerical problems. This can be avoided if the values of $yref_n$ are examined and replaced when necessary. The effect of setting some spectral values to zero before transforming the data to the time domain is equivalent to multiplying the frequency domain data with a rectangular window. This operation gives rise to a ringing effect in the time-domain signal. These distortions were identified and discussed in [57]. Furthermore it was found that the effect can be reduced by smoothing the spectrum before transforming the data to the time domain.

An algorithm based on the same principle as [94] is called QUALITY (Quantitation by converting lineshapes to the Lorentzian type). It was proposed in [34] and addresses the removal of arbitrary distortions by element-wise division by an estimated lineshape deviation using either separate data or an isolated peak in the data to be quantitated as a reference peak. A useful theoretical discussion of the limitations and advantages of this kind of methods is provided. Further development for automating the above procedures for various specific spectroscopic experiments has been proposed in e.g., [87] (^{1}H spectroscopic imaging), [152] (^{1}H spectroscopy).

In [86] a technique called self-deconvolution is used avoiding the need for a reference peak. The method is iterative and based on an initial estimate of the parameters for the spectral components. The estimates are obtained assuming some user-defined ideal lineshape (e.g. Lorentz or Gauss). A time-domain signal is thereafter reconstructed and an estimate of the lineshape distortions can be found from dividing the original signal with the reconstructed signal. The estimated distortions are corrupted by noise and the final corrections are improved if the estimates are smoothed using some window function. As an alternative to smoothing it is proposed to fit a model to the estimated distortions using either a polynomial or a sum of damped sinusoids (using some black-box method e.g., HSVD). The modeled distortions are less corrupted by noise and were seen to improve the corrections in particular for low SNR data. The corrected data are finally used to re-estimate the parameters of interest. The method only performs well if all resonances have the same lineshape and if all peaks present in the spectrum can be sufficiently well quantitated in the initial step.

In [5] wavelets are used to remove the phase artifacts arising from eddy currents. In [117, 151] an experimental spatial map of the variations in the magnetic field is set up. Those variations are included in the quantitation thereby eliminating distortions due to field inhomogeneities. More general approaches can be found in e.g., [108, 114, 105] where lineshape distortions are parameterized in various ways and included in the fitting procedures. By including the distortions into the model function directly -either as additional parameters to be estimated or by multiplying the model function with the derived model distortions - these methods also avoid the problems associated with division by a reference signal that is equal or close to zero.

No final conclusion can be drawn about what method should be used for a specific scenario since no systematic study of the accuracy of the methods has been performed. Most evaluations have been done only by visual inspection of the corrected spectrum. A systematic study of the parameter accuracy achieved by the different techniques using both simulations and an extensive experimental study is desirable.

5.2 Visual SNR and/or resolution improvements

To *visually* improve the SNR and/or the resolution, the acquired spectrum is often multiplied point-wise by a function in the time domain. By multiplication with a decreasing mono-exponential, the noisy data points at the end of the FID are attenuated and the width of the peaks in the DFT spectrum increases, resulting in a spectrum which looks less noisy but with less resolution. To increase the spectral resolution, at the expense of a more noisy looking spectrum, the signal can be multiplied with an increasing mono-exponential. Other functions like e.g. the Voigt function have been used in this respect. If however, the goal is optimal parameter accuracy, these weighting functions should be used with care. The model fitting methods as described in section 2.1.1 assume that every data point included in the fit is affected by noise with standard deviation σ. After multiplication of the acquired data with a non-constant function, this assumption is no longer valid. It is possible to take this into account in the model fitting procedures, but the effect is a mere reversion of the operation. E.g. in the case of multiplication with $e^{-dn\Delta t}$, this is seen as follows

$$\sum_{n=1}^{N} \frac{1}{\sigma^2 e^{-2dn\Delta t}} |e^{-dn\Delta t} y_n - \bar{y}_n|^2.$$

If instead $\sum_{n=1}^{N} |e^{-dn\Delta t} y_n - \bar{y}_n|^2$ is minimized, the parameter estimates will not have ML properties. In [85] a decreased parameter accuracy due to line-broadening was reported based on results on simulated data. In [9] the same conclusion was reached based on the analysis of *in-vivo* short-echo proton spectra.

Cadzow's method or the principle of minimum variance can also be applied as a preprocessing step prior to DFT to visually enhance the signal [22]. Black-box quantitation methods can be used in combination with such a preprocessed signal, resulting in increased parameter accuracy [22]. Interactive methods such as VARPRO or AMARES should be applied on the original and not on the preprocessed data in order to preserve ML properties.

5.3 Removal of unwanted components

In biomedical MRS the acquired signal is often composed of many resonance frequencies out of which only a few are of clinical interest. Preprocessing the MRS signal to eliminate the unwanted features prior to quantitation can be favorable in some cases. Removal of the features of no interest (called *nuisance peaks* in the following) can reduce the computational burden in the final estimation phase, as well as improve the accuracy of the parameter estimates (remove bias and/or reduce variance). Depending on the characteristics of the nuisance peaks different approaches can be taken. Three scenarios for which the removal of nuisance peaks have to be handled in different ways can be distinguished.

The most straightforward case is when the nuisance peaks and the peaks of interest are separated in frequency. Estimation of parameters of selected peaks in the presence of unknown or uninteresting spectral features separated in frequency from the peaks of interest is denoted by frequency selective (FS) parameter estimation in the following. A large number of methods that handle such cases have been presented in the literature. An important example in that respect is solvent suppression in ^{1}H spectroscopy which has received much attention. The water resonance which is inevitably present in those spectra can not be described by an analytical function, mainly because of magnetic field inhomogeneity and measurement suppression techniques. An overview of techniques that are used to remove the influence from nuisance peaks separated in frequency from the peaks of interest is given in section 5.3.1.

Another possibility is that the nuisance peaks have a much larger linewidth than the peaks of interest. Such features originate from large less-mobile molecules or from sequence or hardware artifacts. These peaks can be efficiently removed even though they severely distort the spectrum and apparently overlap with peaks of interest by discarding the first samples of the acquired data. The loss of SNR for the peaks of interest when discarding a few samples is most often negligible.

Finally, the most difficult scenario occurs when the nuisance peaks present are not separated in frequency nor in damping from the peaks of interest. This is for example encountered in short-echo proton spectra where macromolecular baseline resonances pose problems. A number of different solutions have been

proposed. Most methods use some kind of model for the baseline. In [66] the baseline is modeled by a sum of Voigt lines and in [9] by a set of Gaussian peaks. In the frequency domain, the baseline has also been approximated by a spline function [105] and by wavelet coefficients [154]. No method can be seen to solve this problem in a satisfactory manner and up till now no comparisons between methods have been made.

5.3.1 Removal of components separated in frequency

In this section some often used methods which have appeared in the literature to deal with nuisance peaks separated in frequency from the peaks of interest are described. For each of the methods the basic principles are outlined and attention is paid to possible problems associated with these approaches.

5.3.1.1 Time-domain weighting

The influence of nuisance peaks in NLLS parameter estimation techniques such as VARPRO and AMARES was studied in [71]. The bias term of the amplitude estimates, assuming correct estimates of frequencies and dampings, was derived and it was seen that this term could be reduced by introducing an appropriate weighting vector w_n, in the NLLS cost function

$$\sum_{n=0}^{N-1} |w_n \left(y_n - \bar{y}_n\right)|^2. \tag{5.4}$$

The choice of the weighting vector is a trade-off between reducing the nuisance peaks influence and the loss of SNR. The weighting vector should be matched with the frequency distance, amplitude and damping of the nuisance peaks to find the optimal trade-off. This is however not feasible in practice and instead a generic weighting vector is applied that is expected to work reasonably well in most cases. In [71] a weighting consisting of a quarter-wave sinusoid for the first (and last) 20 samples was recommended

$$w_n = \begin{cases} \sin\left(\frac{n\pi}{40}\right) & n \in [0, 19] \\ 1 & n \in [20, N-21] \\ \sin\left(\frac{(N-n-1)\pi}{40}\right) & n \in [N-20, N-1]. \end{cases} \tag{5.5}$$

The method is expected to give good results for relatively well-separated peaks. However, the technique always leads to a loss of SNR resulting in an increased variance of the parameter estimates. If the nuisance peaks have a large amplitude or are close in frequency to the peaks of interest, the method breaks down. These properties are illustrated in section 6.2.

5.3.1.2 Use of black-box methods

Another approach to solve the problem of FS estimation is to model the nuisance peaks by a black-box method and subtract the reconstructed time-domain signal from the original signal prior to parameter estimation. In general black-box methods provide a very good mathematical fit of the data by a sum of exponentially damped complex-valued sinusoids. A black-box method can therefore be used to approximate complicated features of the nuisance peaks by modeling them by a superposition of Lorentzian peaks. After subtraction of the nuisance peaks, the residual signal can be quantified using any parameter estimation method. This technique was first used in the context of water suppression in proton MRS [135, 79]. A black-box method that can be used in this context is the HSVD method [135]. The following scheme can e.g. be used to process spectra with HSVD for FS estimation:

1. The user specifies the total number of Lorentzians in the spectrum, i.e. the model order K to be used in HSVD and the R frequency regions $[fl_r, fh_r]$, $r = 1, \cdots, R$ in which the nuisance peaks lie.

2. HSVD is used to model the original signal by a sum of K exponentially damped complex-valued sinusoids.

3. The modeled peaks lying within the user-defined regions are used to reconstruct the nuisance peaks. Afterwards the reconstructed nuisance signal is subtracted from the original signal.

4. The residual signal is quantified with AMARES (or any other parameter estimation method).

A drawback of using HSVD to model the nuisance peaks is the high computational complexity involved with computing an entire SVD. This is particularly important in proton MRSI. A single metabolite image typically requires $32 \times 32 = 1024$ times the removal of the water resonance. In [101] a fast version called HLSVD is developed which is based on the Lanczos procedure [60] and gives computational savings in most cases. However, HLSVD has the disadvantage that it can slow down in case of repeated or close singular values [101]. The Lanczos method also suffers from the loss of orthogonality of the Lanczos vectors in finite precision. Chapter 7 presents a method that offers nearly the same computational savings as the Lanczos procedure but does not suffer from these drawbacks [50] in the context of water removal.

In addition to the frequency regions to be suppressed, the user has to choose the model order K. The choice of model order K in HSVD is not trivial in case the nearby nuisance peaks have an unknown or other than Lorentzian model function. Up till now no study of the influence of the model order had

been made. In chapters 6 and 8 it is shown that in some cases the choice of the model order is important. Therefore, the use of methods that make an automatic choice of model order based on various information criteria and on the dominant singular values (see e.g. [81] and reference therein) was considered here. In this thesis the Minimum Description Length (MDL) criterion used in [81] is examined in the context of FS parameter estimation. The optimal model order M is determined by minimizing a discrete function of the singular values $\sigma_k, k = 1, \ldots, p$ of the $(N - M) \times (M + 1)$ Hankel matrix consisting of the noisy data

$$MDL(k) = -\ln\left[\frac{(\prod_{i=k+1}^{M+1}\sigma_i)^N}{\left(\frac{1}{M+1-k}\sum_{i=k+1}^{M+1}\sigma_k\right)^{(M+1-k)N}}\right] + \frac{lnN}{2}k(2(M+1)-k).$$

The MDL criterion however requires the knowledge of all the singular values. Consequently, the entire SVD needs to be calculated when the MDL criterion is used to estimate the model order and no fast methods such as presented in [101] and chapter 7 can be used.

5.3.1.3 Filtering

In the context of removal of residual water in proton spectra, the earliest preprocessing techniques were based on filtering. In [78] a simple high-pass finite impulse response (FIR) filter was used to suppress the water peak. From a filtering point of view, these are very basic filters which will strongly influence the metabolite peaks making an accurate quantitation using the filtered signal impossible. A technique that is more elaborate in that sense was proposed in [83] where the main idea consists of using a *low-pass* FIR filter of relatively high order to suppress all metabolite resonances at higher frequencies. The filtered signal ideally consists of the pure water signal which can be subtracted from the original signal. Recently, a simple preprocessing method, ER-filter [21], has been proposed in the more general context of suppressing any wanted region in the spectrum. Although this technique inherently distorts the signal it can be used in case the wanted spectral region is small compared to the full spectral width and the number of data points is large. A general comment on these filter related papers is the lack of a satisfactory discussion of the design of the proposed filters. The estimation results are strongly influenced by the choice of filter type and filter order and only limited guidelines have been provided for making that choice. Moreover, none of these papers discusses the influence of the used techniques on the parameter estimates of the metabolites of interest. The methods presented in the literature are mostly only evaluated based on a visual inspection of the filtered signal. Recently, in [49] a particular narrow pass-band FIR filter was proposed for suppressing neighboring peaks in the

spectrum. The peak of interest was thereafter quantitated using the HSVD method and the parameters adjusted for the filter influence. This method was proposed for reduction of the computational burden involved with a total quantitation and to achieve robust estimates with high accuracy. In chapter 6 a method based on the use of maximum-phase FIR filters is presented and a detailed study of the use of FIR filters for suppressing (a) region(s) in the spectrum is made. It is shown how the influence of the filter on the parameters of interests can be taken into account. Furthermore, an automatic filter design is proposed to ensure the reproducibility and ease of use of the method. In chapter 8 the FIR filter method presented in chapter 6 is applied to the specific case of removal of residual water in proton spectra. It is also shown how the filters briefly mentioned above are related to this maximum-phase FIR filter.

5.3.1.4 Frequency-domain fitting

FS parameter estimation is usually considered to be very easy in the frequency domain. FS estimation using frequency-domain model fitting is done by minimizing the difference between the DFT-transformed data and a model function for the frequency region(s) of interest only. The influence from the other peaks in the spectrum is however not zero in the frequency region(s) of interest and has to be taken into account. This is usually handled by modeling the nuisance peaks in the frequency region of interest as an additional baseline by some choice of basis function (e.g. polynomials, sinusoids, damped sinusoids, splines etc.) [114, 40, 64, 39]. E.g. if a polynomial basis functions is chosen and the peaks of interest are represented by Lorentzians, the complete function used to model the spectrum in the frequency region(s) of interest is given by

$$\begin{aligned} \overline{y}_l &= \sum_{k=1}^{K} a_k e^{-d_k t_0 + j(2\pi f_k t_0 + \phi_k)} \frac{1 - e^{(-d_k + j2\pi(f_k - \nu_l))N\Delta t}}{1 - e^{(-d_k + j2\pi(f_k - \nu_l))\Delta t}} + \sum_{p=0}^{P} d_p \nu_l^p \\ \nu_l &= l/(N\Delta t),\ l = 0, \ldots, N-1 \\ \nu_l &\in [fl_r, fh_r],\ r = 1, \cdots, R. \end{aligned} \tag{5.6}$$

where P is the model order of the polynomial, d_p are the complex polynomial coefficients and $[fl_r, fh_r], \quad r = 1, \ldots, R$ is (are) the frequency region(s) of interest.

To obtain the parameters of interest, the minimization is only carried out over the frequency region(s) of interest instead of over the entire frequency range as done in Eq. (2.40)

$$\sum_{\substack{\nu_l = l/(N\Delta t) \in [fl_r, fh_r] \\ r = 1, \ldots, R}} |\mathcal{Y}_l - \overline{y}_l|^2. \tag{5.7}$$

Minimization has to be carried out w.r.t. the parameters of the peaks of interest and the polynomial coefficients. The user has to specify the frequency range of the regions of interest and as a function thereof the polynomial order has to be chosen. In difficult cases where the peaks of interest are not lying on a flat baseline the choice of the polynomial order is not trivial. If the polynomial model order is chosen appropriately, i.e. if the polynomial is a good model for the underlying baseline in the frequency region included, the estimation bias is expected to be low. The introduction of the extra polynomial coefficients to be estimated inevitably increases the variance of the parameter estimates. If the polynomial is not a good model for the baseline, bias is introduced in the estimated parameters. The influence of the choice of the frequency region and the corresponding polynomial order are further investigated in section 6.2. Note that other choices of basis functions (e.g. (damped) sinusoids) might in some cases lead to a better modeling of the baseline.

5.3.1.5 Other methods

In [44] the baseline caused by the presence of the water peak is approximated and removed from the spectrum while the frequency region around the center of the water peak is replaced by random noise. In [156] a technique based on the idea that the largest singular value usually corresponds to the predominant component in the signal is presented for removal of the solvent. The data matrix is formed, the SVD is calculated and the first singular value is set to zero and the signal is reconstructed. It can however be noted that such a procedure can lead to serious distortions since there is no reason to expect that the first singular value refers to the water resonance only [72]. In [122] a method (LP-ZOOM) was suggested that uses a linear-prediction based method that focuses on a small part of the frequency axis. The fit is performed for data points in the frequency domain after z-transform of the time-domain data. In [5] it was visually illustrated how wavelets can be used to subtract a large spectral component from the spectrum.

5.4 Conclusions

In this chapter an overview of methods to deal with parameter estimation in non-ideal circumstances was given. Methods applicable to situations in which non-idealities result in deviations from the ideal lineshapes (Lorentz, Gauss, Voigt) were revisited. An overview of methods to deal with disturbing resonances was given. The methods used to deal with components separated in frequency from the peaks of interest were treated in detail. This FS parameter estimation problem is further dealt with in chapters 6, 7 and 8.

Chapter 6

Frequency-selective parameter estimation

In this chapter the possibility to obtain accurate estimates of parameters of selected peaks in the presence of unknown or uninteresting spectral features separated in frequency from the metabolites of interest is investigated. This is denoted by frequency-selective (FS) parameter estimation.

The main goals of this chapter are:

- *To explain how maximum-phase FIR filters can be used in the context of FS parameter estimation by providing a detailed explanation of the choice of filter, the filter design parameters and the use of the filters in combination with parameter estimation methods.*

- *To investigate the performance of this FIR filter based method compared to earlier proposed FS selection techniques, i.e. time-domain weighting, black-box filtering, frequency-domain fitting using a polynomial baseline as presented in section 5.3.1. The sensitivity w.r.t. the choice of the user parameters in these methods is also examined.*

Section 6.1 contains all the theory behind the use of maximum-phase FIR filters in the context of FS parameter estimation. The FIR filter method is compared to other FS techniques in section 6.2. This is partly done by a simple simulation example but also by applying the methods to two relevant biomedical MRS scenarios. The influence of the choice of user parameters involved in each of the FS selection techniques which are studied here is investigated. In section 6.3 the main conclusions are formulated.

The work in this chapter is also presented in [145, 121].

6.1 From FIR filter theory to accurate parameter estimation

In section 6.1.1 the FIR filter theory is presented and the use of maximum-phase FIR filters is motivated. In section 6.1.2 it is pointed out how parameter estimation methods can be used in combination with the FIR filters. Section 6.1.3 explains how FIR filtered signals can be displayed in the frequency domain. The filter design aspects are covered and an automatic filter design scheme is proposed in section 6.1.4. A detailed discussion is given on how the user has to choose the parameters in the automatic filter design scheme.

6.1.1 FIR filter theory

A FIR filter is defined by the convolution sum

$$y_{f_n} = \sum_{m=0}^{M-1} h_m y_{n-m} \qquad n = 0, \ldots, N-1, \tag{6.1}$$

where $h_m, m = 0, \ldots, M-1$, are the constant (possibly complex) filter coefficients (see e.g. [98]). A problem with the definition in Eq. (6.1) is that the samples y_n for $n < 0$ are not available for filtering. This is normally solved by assuming that the signal is zero outside the time window (i.e. $y_n = 0$ for $n < 0$) or assuming a cyclic signal (i.e. $y_n = y_{N+n}$ for $n < 0$). Either of these assumptions leads to distortions in the first $(M-1)$ samples of y_{f_n}, which therefore have to be discarded resulting in the following signal

$$y_{f_n} = \sum_{m=0}^{M-1} h_m y_{n-m+M-1}, \quad n = 0, \ldots, N-M. \tag{6.2}$$

Assuming the data has been collected long enough such that the $(M-1)$ last signal samples contain mainly noise, there is practically no loss of signal energy due to the fact that these samples are shifted out of the time window.

Such a FIR filter can be used as follows to eliminate peaks in certain frequency regions. The influence of applying a filter on the exponentially damped sinu-

soids can be examined by applying Eq. (6.2) to the signal model of Eq. (1.3)

$$\begin{aligned}
\bar{y}_{f_n} &= \sum_{m=0}^{M-1} h_m \bar{y}_{n-m+M-1} \\
&= \sum_{m=0}^{M-1} h_m \left(a_1 e^{j\phi_1} e^{(-d_1+j2\pi f_1)t_{n-m+M-1}} + a_2 e^{j\phi_2} e^{(-d_2+j2\pi f_2)t_{n-m+M-1}} + \dots \right. \\
&\qquad \left. + a_K e^{j\phi_K} e^{(-d_K+j2\pi f_K)t_{n-m+M-1}} \right) \\
&= \mathbf{h}^T \mathbf{b}_1 \left(a_1 e^{j\phi_1} e^{(-d_1+j2\pi f_1)t_n} \right) + \mathbf{h}^T \mathbf{b}_2 \left(a_2 e^{j\phi_2} e^{(-d_2+j2\pi f_2)t_n} \right) + \dots \\
&\qquad + \mathbf{h}^T \mathbf{b}_K \left(a_K e^{j\phi_K} e^{(-d_K+j2\pi f_K)t_n} \right) \\
&= \sum_{k=1}^{K} \mathbf{h}^T \mathbf{b}_k \left(a_k e^{j\phi_k} e^{(-d_k+j2\pi f_k)t_n} \right) \qquad\qquad n = 0, \dots, N-M,
\end{aligned} \tag{6.3}$$

where

$$\mathbf{h} = \begin{bmatrix} h_{M-1} & \cdots & h_0 \end{bmatrix}^T$$

and

$$\mathbf{b}_k = \begin{bmatrix} 1 & e^{(-d_k+j2\pi f_k)\Delta t} & \dots & e^{(-d_k+j2\pi f_k)(M-1)\Delta t} \end{bmatrix}^T .$$

From Eq. (6.3) it is seen that the filtered NMR signal consists of the original damped sinusoids (same frequency and damping) altered by a complex scalar $\mathbf{h}^T\mathbf{b}_k$. The filter can thereby be designed to suppress the nuisance peaks (i.e. make $|\mathbf{h}^T\mathbf{b}_k| \approx 0$) including their frequency-domain "tails" while leaving the peaks of interest practically undistorted (i.e. make $|\mathbf{h}^T\mathbf{b}_k| \approx 1$). Signals not exactly obeying the Lorentzian model are also efficiently removed if they can be approximated by a sum of (damped) complex sinusoids with a frequency located in the stop band of the filter.

The filter (or convolution) methods are often believed to have difficulties in removing the frequency-domain "tail" of peaks without affecting peaks lying on this tail. This misunderstanding comes from the well-established fact that time-domain convolution is equal to frequency-domain multiplication. It is however usually forgotten that this is not exactly true unless the signal is cyclic or identically zero outside the actual time window (see e.g. [31]). The actual effect of the filtering procedure in the frequency domain can be examined by taking the FT of the filtered signal model in Eq. (6.3)

$$\mathrm{FT}(\bar{y}_{f_n}) = \sum_{k=1}^{K} \mathbf{h}^T \mathbf{b}_k \mathrm{FT}\left(a_k e^{j\phi_k} e^{(-d_k t_n + j2\pi f_k)t_n} \right). \tag{6.4}$$

It is thereby clear that the broad peaks representing the exponentially damped sinusoids can be suppressed individually including their broad "tails". In [31] this same feature was pointed out using relations in the frequency domain. Note that suppressing entire peaks including their large frequency-domain "tails" is something that can not be obtained by simple operations directly in the frequency domain.

A suitable filter can be proposed from analyzing properties of the filter frequency response $H(f)$ defined by

$$H(f) = \mathbf{h}^T \mathbf{g}(f) \qquad f \in [-0.5, 0.5],$$

where f denotes the normalized frequency (sampling frequency = 1), and $\mathbf{g}(f)$ is the Fourier vector

$$\mathbf{g}(f) = \left[\begin{array}{cccc} 1 & e^{j2\pi f} & \cdots & e^{j2\pi f(M-1)} \end{array} \right]^T. \tag{6.5}$$

The frequency dependent amplification of the filter is given by the magnitude response $|H(f)|$. Secondly, the frequency dependent phase delay is equal to the phase response $\phi(f)$ defined as

$$\phi(f) = \tan^{-1}\left(\frac{Im(H(f))}{Re(H(f))}\right).$$

Finally, the frequency dependent time delay (in number of samples) is equivalent to the filter group delay $Gd(f)$

$$Gd(f) = \frac{d\phi(f)}{df}.$$

A suitable filter type is found by first studying the properties of the so-called *linear-phase* filter (see e.g. [98]). This type of filter is characterized by its phase response which is a linear function of the frequency. The linear dependency implies that the filter has a group delay which is equal to a constant (i.e. independent of the frequency). The constant group delay equals a pure time delay of the filtered signal. The time delay is equal to *half* the filter length (i.e. $(M-1)/2$ samples). This means that $(M-1)/2$ information carrying samples are lost when the first $M-1$ samples of the filtered signal are discarded. It is advantageous to use a filter with a time delay equal to the filter length (i.e. $M-1$ samples). Such a filter is in general not exactly realizable. Here it is proposed to use a so called *maximum-phase* filter which has the largest possible group delay for a given magnitude response (see e.g. [104]). This kind of filter is characterized by a phase response which is a nonlinear function of the frequency, leading to a non-constant group delay. A maximum-phase filter has by definition all its zeroes of the filter coefficient polynomial outside the unit circle. A linear-phase filter can therefore easily be transformed into a maximum-phase filter by reflecting the zeroes of the filter coefficient polynomial

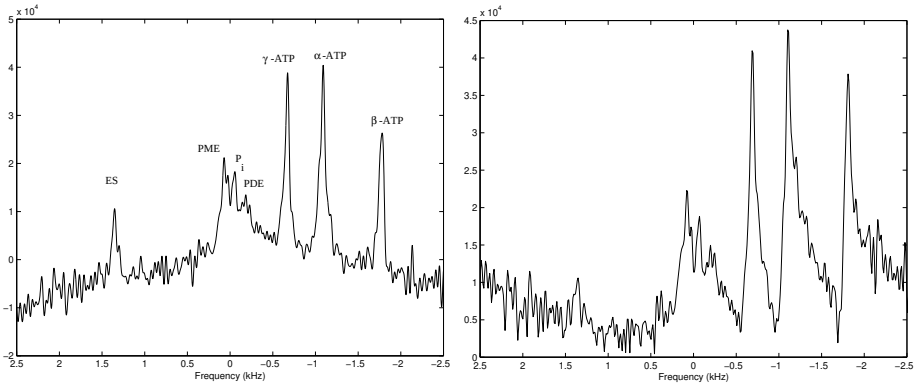


Figure 6.1: Ex-vivo ^{31}P signal of the perfused rat liver, acquired at 4.7 T (81 MHz). Peaks of the external standard (ES), PME, P_i, PDE and α-, β- and γ-ATP can be observed. Left: zero- and first-order phase-corrected real part of the spectrum. Right: magnitude spectrum.

outside the unit circle. Note that this operation does not change the magnitude response of the filter.

The properties of the described filter are illustrated using an *ex-vivo* ^{31}P signal of the perfused rat liver, acquired at 4.7 T (81 MHz). The signal consists of 128 complex data points and the sampling interval is 0.2 ms. In the left hand side of Fig. 6.1 the original signal is displayed after zero- and first-order correction. The magnitude spectrum of the signal is displayed in the right hand side of the figure.

The properties of a linear-phase filter of length 31 and the corresponding maximum-phase filter used to extract the β-ATP signal are studied. The magnitude response of the filters (both magnitude responses are equal) is displayed in Fig. 6.2.

The group delay of the filters is displayed in Fig. 6.3. It can be seen that the linear-phase filter delays the signal exactly 15 samples for all frequencies. For the maximum-phase filter the time delay is around 28 in the pass-band and varying slightly as a function of the frequency. In Fig. 6.4 the filter coefficients (i.e. the impulse responses) are displayed. The figure clearly shows that the linear-phase filter coefficients are symmetric while the maximum-phase filter has most of its energy in the last coefficients. These are typical properties of the impulse responses of the respective filters (see e.g. [98]).

The filters described above are applied to the ^{31}P signal. The top left figure of Fig. 6.5 displays the magnitude spectrum of the linear-phase filtered signal applied as in Eq. (6.1) *without* discarding the initial $M-1$ distorted

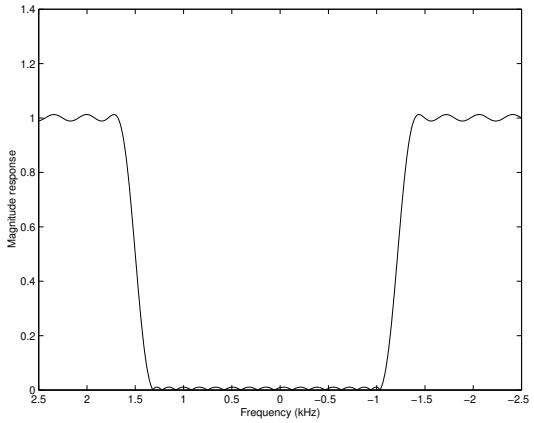


Figure 6.2: Magnitude response of the FIR filter designed to extract the β-ATP peaks of Fig. 6.1.

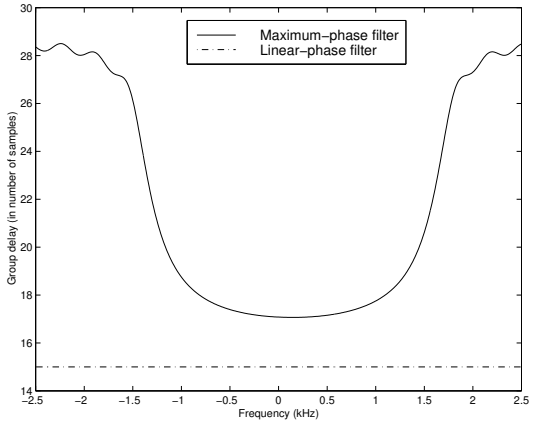


Figure 6.3: Group delay of the maximum- and linear-phase filter designed to extract the β-ATP peaks of Fig. 6.1.

samples. This example is interesting since it shows that the spectrum of the filtered signal is nicely explained by the convolution theorem (i.e. the resulting spectrum is exactly given by the multiplication of the signal spectrum and the magnitude response of the filter). In the top right figure the magnitude spectrum of the filtered signal is displayed after applying the linear-phase filter and discarding 15 of the distorted samples in the beginning of the signal. The remaining distorted samples still give rise to distortions. Note that the filtered spectrum can no longer be explained using the convolution theorem. In the bottom left figure the magnitude spectrum of the linear-phase filtered signal with all distorted samples discarded is displayed. The spectrum shows that

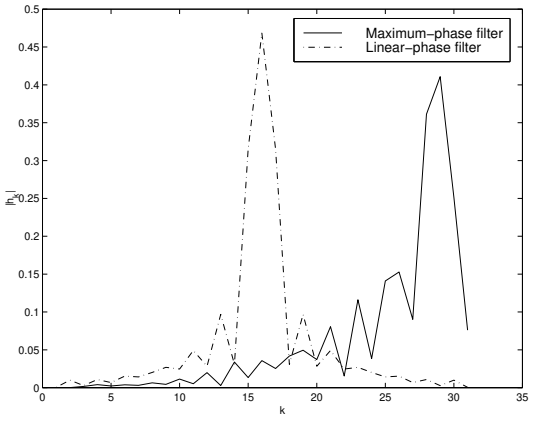


Figure 6.4: Filter coefficients (impulse response) of linear-phase and maximum-phase FIR filter.

there is a perfect suppression of the nuisance peaks including their tails as was predicted by the above discussion and the result in Eq. (6.3). In the bottom right figure the result from applying the corresponding maximum-phase filter and discarding all the distorted samples is displayed. Also here the nuisance peaks are perfectly removed. The difference between the two last spectra shows the influence of the different group delays of the filters. The 15 sample group delay of the linear-phase filter leads to a loss of 15 signal samples which contain an important part of the signal. The maximum-phase filter has a group delay around 28 for frequencies in the pass-band and therefore there is practically no loss of the initial high-amplitude samples of the signal. This can be seen from the higher metabolite peak amplitudes in the maximum-phase filtered spectrum compared to the linear-phase filtered spectrum.

The importance of the large group delay of the maximum-phase filter can also be studied using the result derived in Eq. (6.3). It is seen that the $\mathbf{h}^T\mathbf{b}_k$ term gives a measure of the effect of the filter time delay when filtering damped sinusoids and discarding all the distorted samples. In Fig. 6.6 the loss of SNR as a function of the damping of the peaks is displayed for the linear-phase filter and the maximum-phase filter used above. The magnitude response of the linear-phase FIR filter is displayed in the left hand side together with the values of the $|\mathbf{h}^T\mathbf{b}_k|$ term as a function of the frequency for two different damping coefficients ($d = 0.2$ kHz and $d = 0.4$ kHz). The corresponding figures for the maximum-phase filter are displayed on the right hand side. It is clear that for the same value of d, the metabolite peaks of interest are much more suppressed in case the linear-phase filter is applied.

The advantage of the use of maximum-phase filters can also be understood in the following way. Note that the $\mathbf{b}_k$ vector is equal to the Fourier vector $\mathbf{g}(f)$

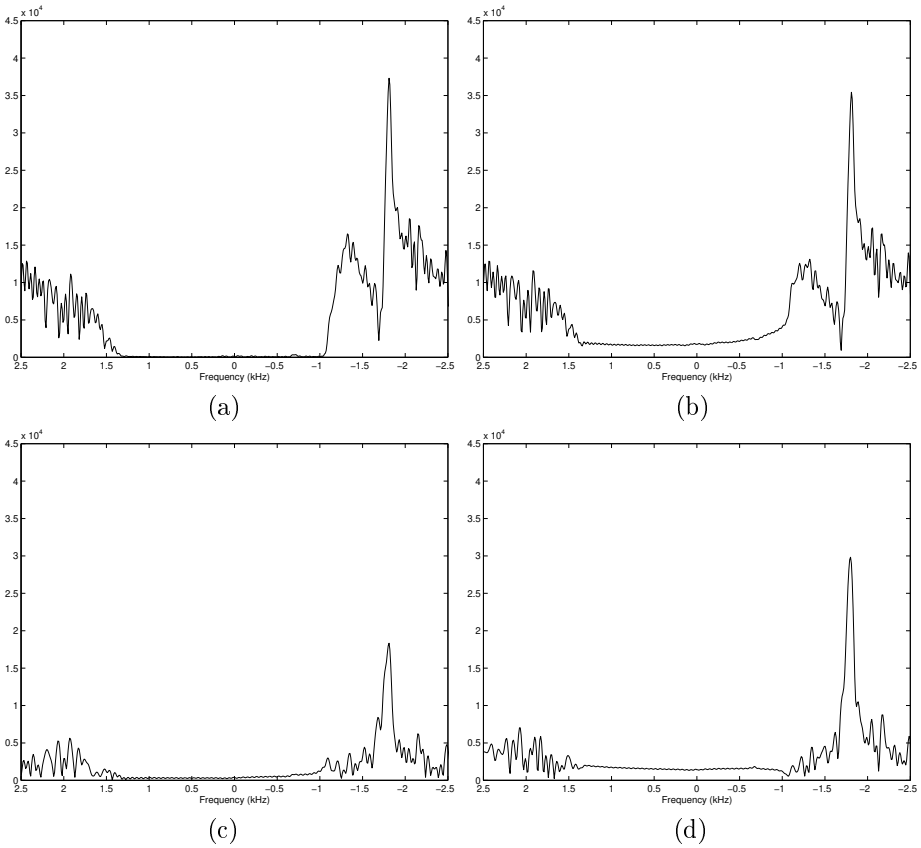


Figure 6.5: Illustration of distortions introduced by applying a FIR filter. (a) Magnitude spectrum of linear-phase filtered signal without discarding distorted samples. (b) Magnitude spectrum of linear-phase filtered signal discarding 15 distorted samples. (c) Magnitude spectrum of linear-phase filtered signal with all distorted samples (30) discarded. (d) Magnitude spectrum of maximum-phase filtered signal with all distorted samples (30) discarded.

for sinusoids without damping (i.e. $d = 0$). The influence of the damping is that the magnitude response is weighted element-wise by the damped exponentials in $\mathbf{b}_k$: $\begin{bmatrix} 1 & e^{-d_k \Delta t} & \dots & e^{-d_k (M-1) \Delta t} \end{bmatrix}^T$. The influence of the weighting will be minimized if the filter has most of its power in the last filter coefficients. This desirable property is, as already stated, typical for maximum-phase FIR filters.

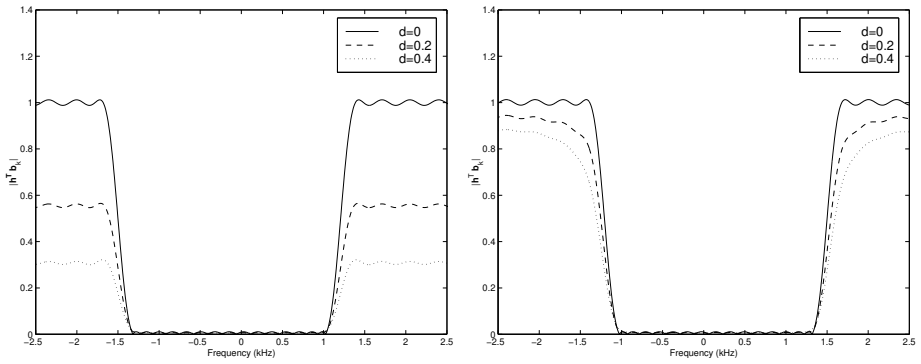


Figure 6.6: $|\mathbf{h}^T\mathbf{b}_k|$ term as a function of frequency for different damping coefficients d. Left: linear-phase FIR filter of length 31. Right: maximum-phase FIR filter of length 31.

6.1.2 Parameter estimation

From the filtered signal, $y_{f_n}, n = 0, \ldots, N-M$, parameter estimates can be obtained using any of the standard estimation methods. This is possible since the filtered damped sinusoids are still damped sinusoids with altered amplitude and phase as can be seen by Eq. (6.3). Applying an estimation method to the filtered signal y_{f_n} results in estimates (called $\hat{a}'_k, \hat{d}'_k, \hat{f}'_k$ and $\hat{\phi}'_k$ respectively) that are related to the estimates of the parameters of interest in the following way

$$\begin{aligned} \hat{a}_k &= \frac{\hat{a}'_k}{|\mathbf{h}^T\mathbf{b}_k|} \\ \hat{d}_k &= \hat{d}'_k \\ \hat{f}_k &= \hat{f}'_k \\ \hat{\phi}_k &= \hat{\phi}'_k - \tan^{-1}\left(\frac{Im(\mathbf{h}^T\mathbf{b}_k)}{Re(\mathbf{h}^T\mathbf{b}_k)}\right). \end{aligned} \tag{6.6}$$

However, the estimation can not be performed with standard methods if prior knowledge concerning the amplitudes or phases is to be taken into account. The filter influence on these parameters has to be considered directly in the estimation procedure to yield correct estimates. Here it is proposed to minimize the squared difference between the filtered signal y_{f_n} and the filtered model function derived in Eq. (6.3)

$$\sum_{n=0}^{N-M} |y_{f_n} - \sum_{k=1}^{K} \mathbf{h}^T\mathbf{b}_k (a_k e^{j\phi_k} e^{(-d_k + j2\pi f_k)t_n})|^2. \tag{6.7}$$

It is straightforward to correctly incorporate any kind of prior knowledge in the above criterion. The minimization in Eq. (6.7) has been numerically implemented for evaluation by modifying the AMARES algorithm. The new algorithm is referred to as AMARES$_f$ in the following.

The use of the NLLS fit in Eq. (6.7) can be motivated by studying the ML-estimator for the model in Eq. (1.3) based on a more general noise assumption. Let

$$\mathbf{R} = E(\mathbf{e}\mathbf{e}^H) \tag{6.8}$$

denote the covariance matrix of the circular Gaussian distributed noise, where $\mathbf{e} = [e_0, \ldots, e_{N-1}]^T$. The ML estimator is then given by minimizing the following weighted NLLS (WNLLS) function

$$[\mathbf{y} - \bar{\mathbf{y}}]^H \mathbf{R}^{-1} [\mathbf{y} - \bar{\mathbf{y}}]. \tag{6.9}$$

It is easy to check that Eq. (6.9) simplifies to

$$\sum_{n=0}^{N-1} |y_n - \bar{y}_n|^2. \tag{6.10}$$

for white noise, with covariance matrix $R = \sigma^2 I$, where I denotes the identity matrix and σ^2 is the noise variance. The cost function in Eq. (6.10) is the one that is minimized in algorithms such as VARPRO and AMARES.

The relation between minimizing the WNLLS function in Eq. (6.9) and the proposed function in Eq. (6.7) is easily seen if the latter is slightly reformulated. Let

$$\mathbf{H}_f = \begin{bmatrix} h_{M-1} & h_{M-2} & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_{M-1} & h_{M-2} & \cdots & h_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h_{M-1} & h_{M-2} & \cdots & h_0 \end{bmatrix} \tag{6.11}$$

denote the $(N - M + 1) \times N$ FIR filter matrix. Using Eq. (6.11), the following matrix multiplication replaces the convolution sum in Eq. (6.2)

$$\mathbf{y}_f = \mathbf{H}_f \mathbf{y}, \tag{6.12}$$

with $\mathbf{y}_f = [y_{f_0}, \ldots, y_{f_{N-M}}]^T$. The NLLS cost function in Eq. (6.7) can then be written in the following way

$$[\mathbf{y} - \bar{\mathbf{y}}]^H \mathbf{H}_f^H \mathbf{H}_f [\mathbf{y} - \bar{\mathbf{y}}]. \tag{6.13}$$

When the NLLS function in Eq. (6.13) is compared with Eq. (6.9) it is seen that the modified NLLS function is equal to the ML estimator if the FIR filter is an ideal prewhitening filter

$$\mathbf{H}_f^H \mathbf{H}_f = \mathbf{R}^{-1}. \tag{6.14}$$

The basic idea of the FIR filter technique is to regard the nuisance signals as part of the noise term and use the filter to whiten the total noise term prior to the estimation phase. It is however important to note that the ML interpretation of the filtered NLLS fit in Eq. (6.7) is only valid if the noise term is Gaussian distributed. This is not the case when a non-Gaussian signal (such as the nuisance peaks) is included in the noise term. Furthermore, it is impossible to completely decorrelate the noise term and the metabolite signals since Eq. (6.14) can not be solved in general. The use of an approximate prewhitening FIR filter and the NLLS fit in Eq. (6.7) for quantification of the spectra is still expected to perform well and its good performance is illustrated in section 6.2.

6.1.3 Visualization

The phase changes introduced by applying the FIR filter also influence the visualization of the filtered signals. One solution is to display the magnitude of the spectra. To display the phased, real spectrum, the DFT transformed signal has to be corrected not only for the zero- and first-order phase but also for the frequency-dependent phase response of the filter. Therefore the DFT transformed spectrum is multiplied point-wise by $e^{-j\phi(f)}$.

This is illustrated using the ^{31}P signal of the perfused rat liver of Fig. 6.1. In Fig. 6.1 the real part of the original signal is displayed after zero- and first-order phase correction. In the left hand side of Fig. 6.7 the β-ATP signal has been extracted from the signal using the filter, and phase corrected to zero and first order. In the right hand side the filtered signal is shown with an additional phase correction to compensate for the influence of the filter.

6.1.4 Design and use of the maximum-phase FIR filter

The objective of this section is to give guidelines on how to find a well-matched prewhitening filter that minimizes the filter influence on the final parameter estimates. An automatic scheme for finding the appropriate values of the filter design parameters is proposed. This scheme reduces the required user interaction and ensures the reproducibility of the proposed method.

The design of maximum-phase filters is closely related to the design of minimum-phase filters. The design of the latter type of filter is often based on the approach proposed in [63] where primarily a linear-phase FIR filter is designed and transformed to a minimum-phase filter by spectral factorization (i.e. reflecting all zeroes of the filter coefficient polynomial inside the unit circle). The maximum-phase filter is then found by reversing the order of the filter coefficients. The design of the initial linear-phase FIR filter is here done with a

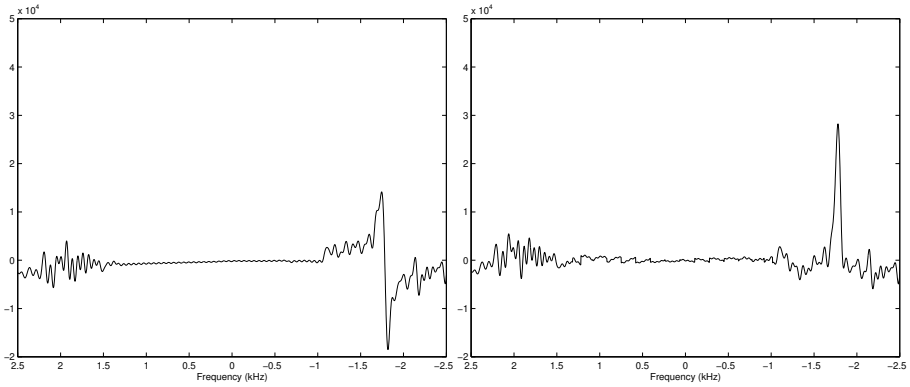


Figure 6.7: Real part of the maximum phase FIR filtered spectrum. Left: zero- and first-order phase-corrected as in Fig. 6.1. Right: corrected for zero- and first-order phase as in Fig. 6.1 and for the frequency-dependent phase of the filter.

constrained LS fit (using the algorithms proposed (and provided) by [112]) in which the filter order M, normalized cut-off frequency f_c, stop-band suppression sup, and pass-band ripple r, have to be specified (see e.g. [98]). The use of the above design algorithm transforms the problem of finding a suitable prewhitening filter into choosing appropriate values of the design parameters M, f_c, sup and r. The automatic filter design scheme proposed here is based on the specification of the regions to be suppressed by the user. Furthermore, it is assumed that the sinusoidal components have been sufficiently damped out such that an estimate of the noise standard deviation can be obtained from the last samples of the original data sequence. Below, an outline of the automatic filter design scheme is given.

Algorithm 6.1.1 (Filter design scheme)

Input: data samples $y_n, n = 0, \ldots, N-1$, the filter cut-off frequencies fl_r and fh_r corresponding to the user-selected frequency region(s) containing the nuisance peaks $[fl_r, fh_r]$, $r = 1, \cdots, R$, where R is the number of frequency regions.
Output: the filter coefficients $\mathbf{h}$.

Step 1. An estimate $\hat{\sigma}$ of the noise standard deviation is calculated as the standard deviation of the last samples of the signal. The peak value, s_0, is computed as the value of the strongest nuisance peak

$$s_0 = \max_l |\frac{\mathcal{Y}_l}{\sqrt{N}}|,$$

where $l = \nu_l(N\Delta t)$ and $\nu_l \in [fl_r, fh_r]$, $r = 1, \cdots, R$.

Step 2. The initial filter suppression sup is taken equal to

$$sup = \frac{2\hat{\sigma}}{s_0}$$

and starting values for the filter order (e.g. $M = 30$) and passband ripple (e.g. $r = 5\%$) are chosen.

Step 3. The linear-phase filter is designed by the LS algorithm of [112] and transformed into a maximum-phase filter.

Step 4. The signal is filtered and the maximum value s_m in the frequency regions of the nuisance peaks is calculated

$$s_m = \max_l |\frac{\mathcal{Y}_{f_l}}{\sqrt{N}}|,$$

where $l = \nu_l(N\Delta t)$ and $\nu_l \in [fl_r, fh_r]$, $r = 1, \cdots, R$.
If

$$s_m > 2\hat{\sigma}$$

the filter suppression is increased (e.g. $sup = sup/5$) and the process is restarted from *Step 3*. If s_m does not decrease in the next iteration, the filter order is increased (e.g. $M = M + 10$).

If the nuisance peaks are relatively narrow (slowly decaying), the above scheme results in a suitable filter in the first iteration. A higher filter suppression is required if some of the nuisance peaks have a fast decay. In this case the iterative procedure will find a filter with a sufficient suppression within a few iterations (< 10). The computational complexity is thereby increased compared to the original filter design scheme. The advantage is however that a suitable filter is found independently of the shape and signal energy of the nuisance peaks. See section 8.1 for a detailed analysis concerning the computational complexity associated with the use of the presented FIR filters in the context of water removal in proton spectra.

Some guidelines on how to choose the regions to be suppressed are given below. A narrow stop band is advisable since in general it leads to a higher group delay in the pass band than a filter with the same order, stop-band suppression and pass-band ripple. This principle is illustrated using the same ^{31}P example described above. First the automatic design scheme is applied with -1.2 kHz as a lower bound and 1.5 kHz as an upper bound for the region to be suppressed. The group delay of the resulting maximum phase filter of length 31 is shown in

the left hand side of Fig. 6.8. The group delay of the filter with -1.5 kHz and 2.5 kHz as a lower and upper bound respectively is shown in the right hand side of the figure. Two effects can be observed. First, the filter with the larger stop band has an overall smaller group delay in the pass band and secondly, the peak closest to the filter lower bound is located in the transition band of the filter, leading to an even smaller group delay for this specific peak. The smaller group delay for the frequencies of interest will lead to an overall loss in SNR and a decrease in resulting parameter accuracy. This shows that care has to be taken on how to specify the lower and upper bounds of the region(s) to be suppressed and it is therefore advisable to inspect the group delay of the filter in the pass band in difficult scenarios (large stop band and/or bound close to frequency of interest). When choosing the bound closest to the region to be suppressed, it is important to make sure that no signal parts of the nuisance region are located within the transition band of the filter. Otherwise, the nuisance peaks would be insufficiently suppressed, introducing bias in the parameters of interest.

In summary, the main reasons for using maximum-phase FIR filters to suppress peaks in certain frequency regions are the following. A FIR filter (in general) is powerful since the filter influences the different peaks individually. Thereby it is possible to suppress entire peaks - including the frequency-domain "tails" - while not influencing other peaks provided all distorted samples are discarded. If the upper and lower bounds of the region(s) to be suppressed are chosen with some care, the nuisance peaks are efficiently removed and the bias of the estimates is eliminated. The resulting signal is undistorted in the sense that the filtered signal is exactly described by the relation in Eq. (6.3). Note that this property is not dependent on the type of filter used. The actual choice of filter type becomes important if one studies the signal samples lost when the $M-1$ distorted samples of the filtered signal are discarded. The high group delay of the maximum-phase filter was seen to minimize this loss. The SNR gain compared to using a linear-phase filter can be substantial.

6.2 Numerical validation

In this section different Monte Carlo studies are performed to compare and evaluate different FS methods. The methods studied are:

- AMARES$_f$ as presented in section 6.1. The influence of the choice of the upper and lower frequency bound in the automatic filter design scheme on the parameters of interest is investigated.
- Time-domain weighting (see section 5.3.1.1): AMARES$_w$. Inserting the weighting function of Eq. (5.5) into Eq. (5.4) and solving it using the NLLS solver used in AMARES is referred to as AMARES$_w$ in the following.

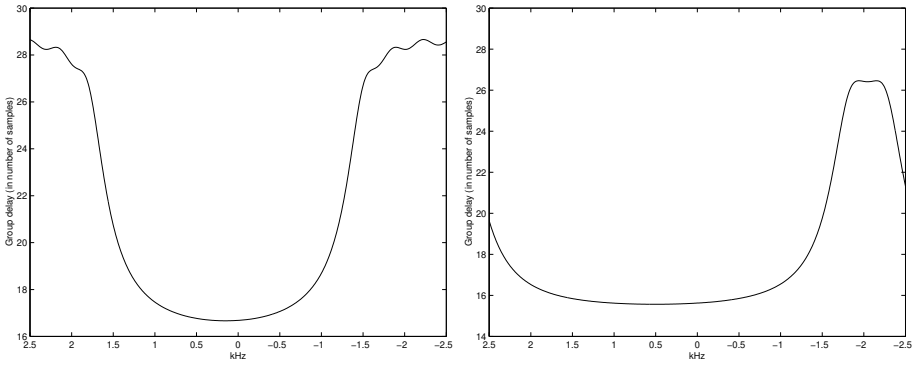


Figure 6.8: Group delay of the maximum-phase FIR filter designed to extract the β-ATP peaks of Fig. 6.1. Left: lower cut-off frequency=-1.2 kHz, upper cut-off frequency=1.5 kHz. Right: lower cut-off frequency=-1.5 kHz, upper cut-off frequency=2.5 kHz.

- AMARES$_H$: HSVD filtering followed by fitting the residual using AMARES as explained in section 5.3.1.2. The influence of the choice of the order in HSVD is investigated. Also the performance of the automatic order selection criterion as discussed in section 5.3.1.2 is evaluated.
- Minimizing Eq. (5.7) using the NLLS solver of AMARES is denoted by AMARES$_{\text{FREQ}}$ in the following. The influence of the choice of the polynomial order and the size of the frequency region included in Eq. (5.7) are investigated.

The different methods are also compared with each other on 3 different examples. First an artificial two-peak example is constructed to examine the estimation properties of the different methods as a function of the amplitude of the nuisance peaks and the frequency distance between the nuisance peaks and the peaks of interest. The performance of AMARES$_f$ in case of a non-Lorentzian lineshape is also examined. Thereafter the methods are evaluated for two relevant biomedical MRS examples. Unless otherwise specified the model function of Eq. (1.3) is used and the added complex noise is circular, white and Gaussian distributed. The methods are compared based on the calculated RRMSE, RBias and RSTD. The RRMSE is compared to the CRB. The CRB is calculated from a model consisting of the peaks of interest only.

6.2.1 Two-peak example

In this section an example consisting of two exponentially damped sinusoids in which only one is of interest is considered. The peak of interest (peak1) has

the following parameter values:

$$f_1 = 20 \text{ Hz}$$
$$d_1 = 10 \text{ Hz}$$
$$\phi_1 = 0 \text{ radians}$$
$$a_1 = 20 \text{ (a.u.)}.$$

The damping and phase of the nuisance peak (peak2) are fixed to:

$$d_2 = 10 \text{ Hz}$$
$$\phi_2 = 0 \text{ radians}$$

whereas the amplitude and frequency of peak2 are varied. The sampling frequency is 1 kHz and the number of data points is 512. Unless otherwise specified the SNR used in the simulations for peak1 is equal to 10 dB.

6.2.1.1 Frequency-domain baseline fitting using AMARES$_{FREQ}$

The purpose is to assess the influence of the polynomial baseline fit on the accuracy of the parameters of peak1 estimated with AMARES$_{FREQ}$. The choice of the polynomial order and the size of the frequency region included in Eq. (5.7) are investigated.

In the first experiment simulations were run with and without inclusion of a polynomial baseline. The amplitude of peak2 is varied between 10-320 a.u. while its frequency is fixed to 0.1 kHz. Polynomial orders varied from $P = 0$ to $P = 5$ and the frequency region included in Eq. (5.7) was -0.2 kHz to 0.04 kHz. The RRMSE, RSTD and RBias of the amplitude estimates of peak1 are shown as a function of the amplitude of peak2 for different polynomial orders in Fig. 6.9. The figure shows that a higher polynomial order is required to achieve good estimates as the amplitude of peak2 increases. Furthermore, it is seen that the bias almost vanishes if the polynomial order is chosen to be sufficiently high. The price for including additional polynomial parameters, is an increased standard deviation on the estimates.

To get an idea of the influence of the size of the frequency region included in Eq. (5.7), in a second experiment, the amplitude of peak2 is fixed and the lower and upper bounds of the frequency interval are varied between -0.5 and 0.01 kHz and between 0.03 and 0.09 kHz respectively. The corresponding parameter accuracy of peak1 (RRMSE, RSTD, RBias) with the amplitude of peak2 fixed to 320 a.u. is shown in Fig. 6.10 (for $P = 2$) and Fig. 6.11 (for $P = 4$). The following points can be derived from these figures. The standard deviation is overall lower for $P = 2$, but the bias is in general much larger than the bias for $P = 4$, thereby confirming the result of Fig. 6.9. It can also be observed that if the region of interest is small, the standard deviation is higher since the number of points included in the optimization problem is smaller. The

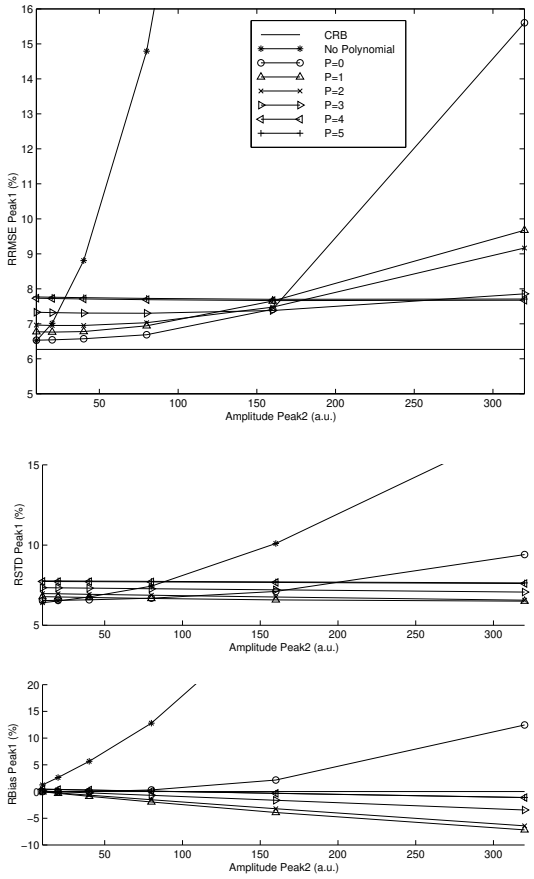


Figure 6.9: RRMSE, RSTD and RBias of amplitude estimates of peak1 as a function of the amplitude of the nuisance peak (peak2) for the frequency-domain fitting procedure AMARES$_{FREQ}$ using a polynomial baseline fit with different polynomial orders P, SNR = 10 dB. Top: RRMSE. Middle: RSTD. Bottom: RBias. CRB denotes the theoretical lower bound.

polynomial of order $P = 2$ is seen to be a bad model for the baseline for almost any choice of the frequency region. The polynomial model of order $P = 4$ is a good approximation for most choices of the frequency region but also breaks down when the frequency region becomes too large.
Additional simulations show that as the amplitude of peak2 decreases, a lower polynomial order suffices to model the baseline and that the results becomes less sensitive to the frequency range included in Eq. (5.7).

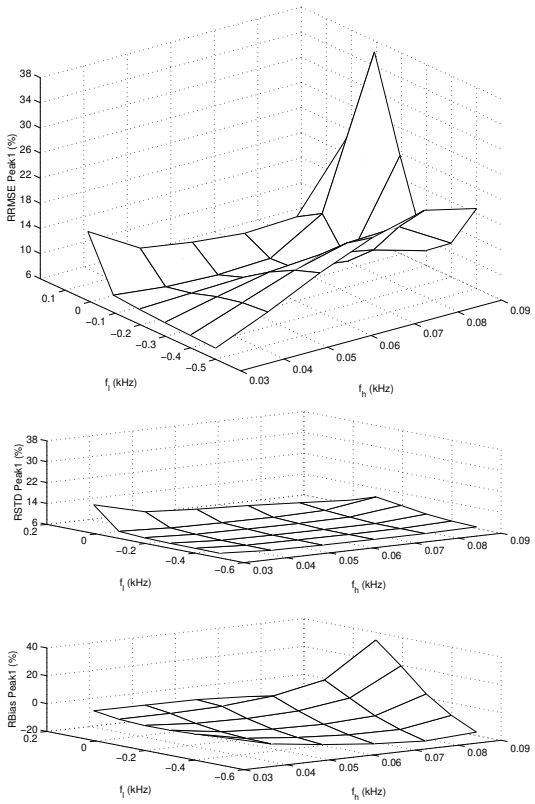


Figure 6.10: RMSE, RSTD and RBias of the AMARES$_{FREQ}$ amplitude estimates of peak1 ($a_1 = 20, f_1 = 20$ Hz) as a function of the width of the frequency region included in Eq. (5.7). The lowest frequency varies between -0.5 and 0.01 kHz and the highest frequency varies between 0.03 and 0.09 kHz. The polynomial order is $P = 2$. The parameters of peak2 are: $a_2 = 320, f_2 = 0.1$ kHz. Top: RRMSE. Middle: RSTD. Bottom: RBias.

6.2.1.2 Time-domain FIR filtering using AMARES$_f$

The purpose is twofold, first to assess the influence of the value of the upper and lower bound of the region to be suppressed in the AMARES$_f$ filter design scheme and second to assess the performance of the method in case the nuisance peak does not have a Lorentzian lineshape.

The amplitude of peak2 is fixed to 320 a.u. and its frequency is fixed to 0.1 kHz. The upper and lower bound used in the filter design scheme are varied between 0.11 and 0.5 kHz and 0.03 and 0.09 kHz respectively. The RMSE, RSTD and RBias of the amplitude estimates of peak1 are shown in Fig. 6.12

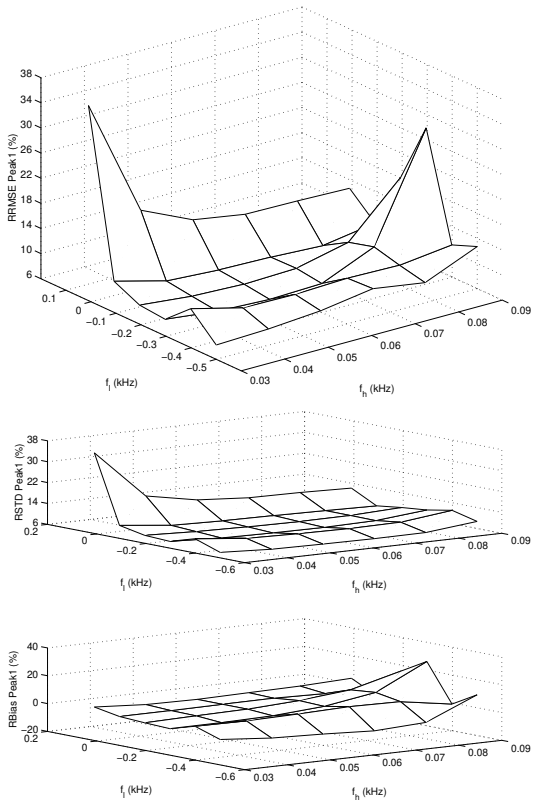


Figure 6.11: RMSE, RSTD and RBias of the amplitude estimates of peak1 ($a_1 = 20, f_1 = 20$ Hz) obtained by AMARES$_{FREQ}$ as a function of the size of the frequency region included in Eq. (5.7). The lowest frequency varies between -0.5 and 0.01 kHz and the highest frequency varies between 0.03 and 0.09 kHz. The polynomial order is $P = 4$. The parameters of peak2 are: $a_2 = 320, f_2 = 0.1$ kHz. Top: RRMSE. Middle: RSTD. Bottom: RBias.

as a function of the bounds. It is seen that the bias is effectively removed for almost all combinations of upper and lower bounds used in the filter design scheme. The method breaks down when the upper and/or lower bound is chosen too close to the frequency of peak2, resulting in insufficient suppression of the nuisance peak. A slight increase in standard deviation can be observed when the lower cut-off frequency is chosen closer to the peak of interest. This is explained by the resulting lower group delay for peak1 leading to a loss of SNR and subsequently a higher standard deviation.

To test the performance of AMARES$_f$ in case the lineshape is non-Lorentzian, an example is used consisting of two Gaussian peaks:

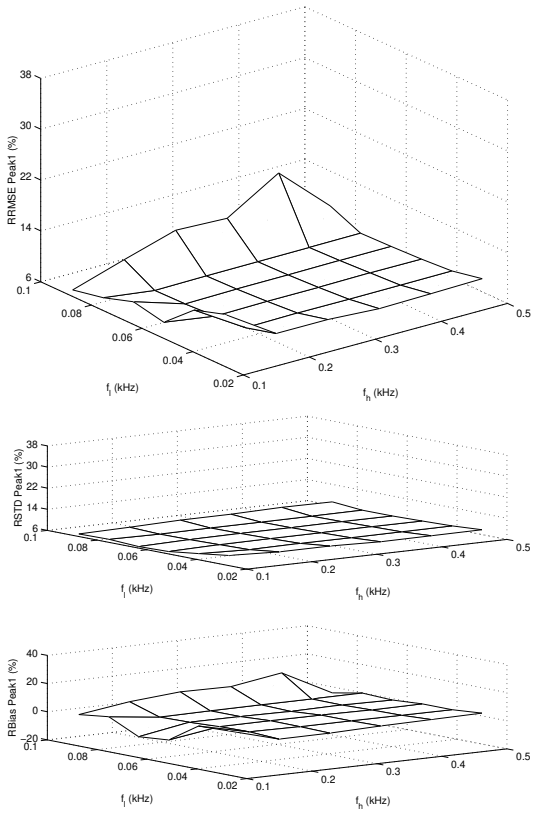


Figure 6.12: RMSE, RSTD and RBias of the AMARES$_f$ amplitude estimates of peak1 ($a_1 = 20, f_1 = 20$ Hz) as a function of the lower and upper bounds in the filter design scheme. The parameters of peak2 are: $a_2 = 320, f_2 = 100$ Hz. SNR of peak1= 10 dB. Top: RRMSE. Middle: RSTD. Bottom: RBias.

$\sum_{k=1}^{2} a_k e^{-g_k (n\Delta t)^2} e^{j(2\pi f_k n\Delta t + \phi_k)}$, with

$$
\begin{aligned}
f_1 &= 20 \text{ Hz} \\
a_1 &= 20/(\sqrt{\pi ln2}) \text{ a.u.} \\
a_2 &= 230/(\sqrt{\pi ln2}) \text{ a.u.} \\
g_1 &= g_2 = 10^2/(4ln2) \text{ Hz}^2 \\
\phi_1 &= \phi_2 = 0 \text{ radians}
\end{aligned}
$$

and f_2 varying between 0.06 and 0.22 kHz. Note that the Gaussian peaks have the same full width at half height and the same maximum value of the real part of the DFT spectrum as Lorentzian peaks with amplitudes of 20 and 320 a.u. and dampings of 10 Hz. The SNR of peak1 is equal to 6.6 dB. The filter design scheme is used with upper and lower bounds of the region to be suppressed

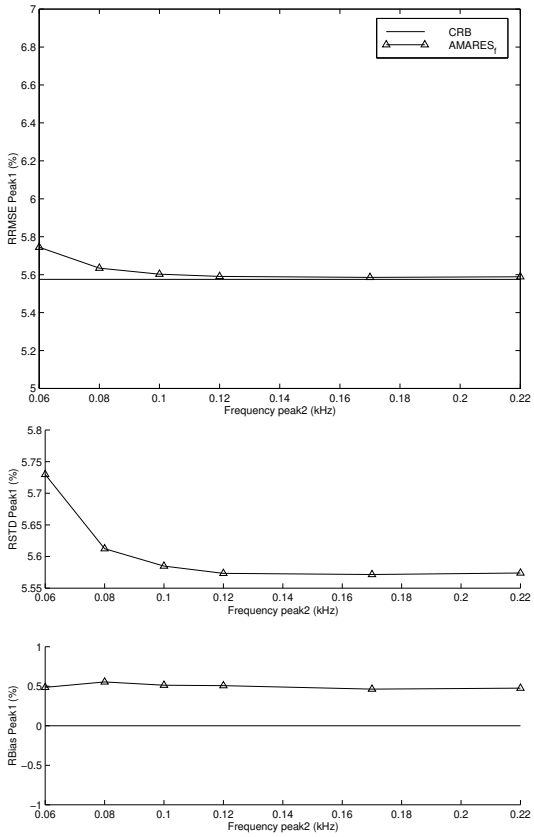


Figure 6.13: RRMSE, RSTD and RBias of the amplitude estimates of peak1 ($a_1 = 13.6$, $f_1 = 20$Hz) as a function of the frequency of the nuisance peak (peak2, $a_2 = 216.9$) for AMARES$_f$. The two peaks have a Gaussian lineshape. SNR of peak1= 6.6 dB. Top: RRMSE. Middle: RSTD. Bottom: RBias. CRB denotes the theoretical lower bound.

equal to the frequency of peak2 plus and minus 0.02 kHz, thereby fulfilling the basic principles as explained in section 6.1.4. Eq. (6.7) was adapted to compensate for the filter influence in case of a Gaussian lineshape. The results are displayed in Fig. 6.13. As can be seen from the figure, AMARES$_f$ has no problems in dealing with the Gaussian nuisance peak. The bias is removed and the RRMSE follows the CRB very closely.

6.2.1.3 Method comparison

The accuracy of the four FS estimation methods is compared.

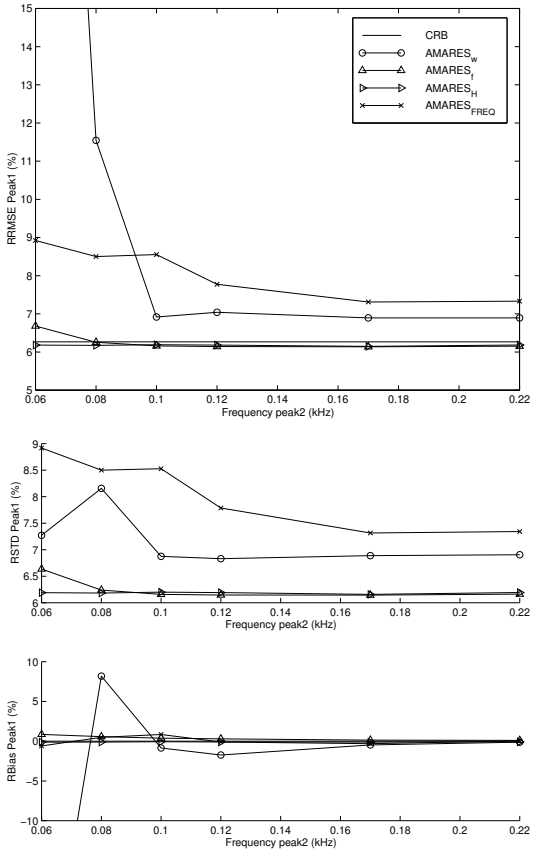


Figure 6.14: RRMSE, RSTD and RBias of the amplitude estimates of peak1 ($a_1 = 20$, $f_1 = 20$Hz) as a function of the frequency of the nuisance peak (peak2, $a_2 = 320$) for different FS estimation procedures. SNR of peak1= 10 dB. Top: RRMSE. Middle: RSTD. Bottom: RBias. CRB denotes the theoretical lower bound.

The frequency of peak2 is varied between 0.06 and 0.22 kHz while its amplitude is kept fixed at 320 a.u.. The results are displayed in Fig. 6.14.

For AMARES$_{FREQ}$, the polynomial order that eliminates the bias and has the lowest standard deviation is shown for every frequency. The frequency region from -0.15 kHz to 0.04 kHz is included in Eq. (5.7) for all frequencies of peak2. It was verified that for the polynomial orders shown, the width of this frequency interval was not critical. For AMARES$_f$, the upper and lower bound in the filter design scheme are chosen as the frequency of peak2 plus

and minus 0.02 kHz, thereby fulfilling the basic principles explained in section 6.1.4. In AMARES$_H$ the model order is taken equal to 2, the theoretical one.

For AMARES$_w$ a "break-down" frequency is found below which the assumptions of a "large" frequency separation no longer hold. The method is successful in eliminating the estimator bias outside this region, but the loss of SNR associated with this approach increases the standard deviation of the estimates. AMARES$_{FREQ}$ is capable of handling close nuisance peaks if the polynomial order is chosen high enough. This is confirmed by the low estimator bias. The standard deviation is however increasing as a function of the polynomial order thereby degrading the estimates. The time-domain methods based on FIR filtering and HSVD filtering are clearly outperforming the other methods and their RRMSEs are very close to the theoretical lower bounds even for a small frequency separation.

6.2.2 Frequency-selective quantification of an *in-vivo* ^{13}C spectrum

A simulation study is performed using signals derived from an *in-vivo* proton decoupled ^{13}C spectrum of subcutaneous adipose tissue in the human forearm. The spectrum was acquired at 4.7 T (50.3 MHz) using a 5 cm diameter double-tuned surface coil. In some applications one is only interested in obtaining estimates of the two peaks around 1.57 kHz and 1.66 kHz respectively. The peak at 1.66 kHz corresponds to carbons in double bonds of mono- and polyunsaturated fatty acid chains while the peak at 1.57 kHz corresponds exclusively to carbons in double bonds of polyunsaturated chains. The amplitudes of these peaks can be used to study non-invasively the relative amounts of poly- and monosaturation of fatty acids in human tissue. The real, phased spectrum of the original *in-vivo* signal is displayed in the left hand side of Fig. 6.15 while the real, phased spectrum of the FIR filtered signal is shown in the right hand side.

The simulation signal was derived as follows. The *in-vivo* signal was quantified using AMARES. The model fitted to the signal consisted of 15 Lorentzian peaks and no prior knowledge was imposed. The estimated parameters were then used as parameters for the noiseless simulation signal. The sampling frequency is 10 kHz and the number of data points is 512. It is important to note that the region around -3.5 kHz was difficult to model correctly. In that region some signal features are still present in the residual, while outside that frequency region the residual looks like white noise. In this case it is certainly advantageous to use FS estimation, since no satisfactory model is present for the entire signal.

The accuracy of the four FS estimation methods are compared for the peak at

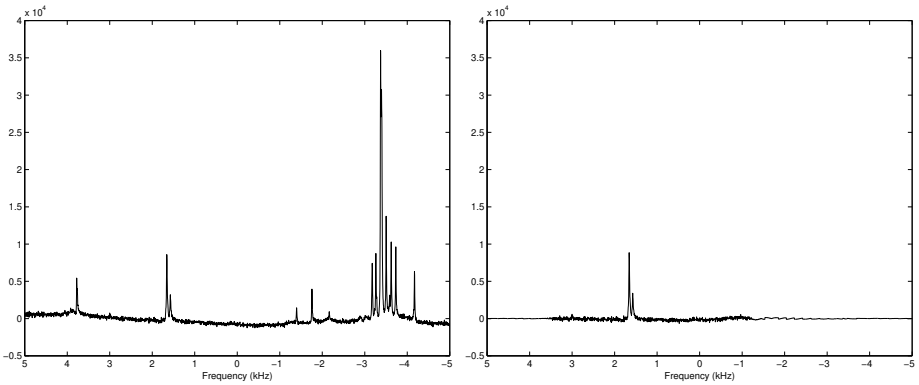


Figure 6.15: In-vivo proton decoupled ^{13}C spectrum of subcutaneous adipose tissue in the human forearm (acquired at 4.7 T (50.3 MHz)). The peaks to be quantified are at 1.57 and 1.66 kHz. Left: real, phased part of ^{13}C spectrum. Right: real, phased part of the same signal after maximum phase FIR filtering.

1.66 kHz for different noise levels.

For AMARES$_{FREQ}$, different polynomial orders and widths of frequency regions were tested. In AMARES$_{FREQ}$ quantification of the ^{13}C spectrum turned out to be impossible without inclusion of a polynomial. A polynomial order of $P = 0$ is only valid in a small region around the two peaks of interest. A polynomial order of $P = 1$ provides a bias close to zero in a large frequency region: the lower bound can vary between -0.5 to 1 kHz and the upper bound between 2 and 3.2 kHz. There is no significant change in standard deviation and RMSE in this region. There is only a slight increase in RRMSE by using a higher polynomial order.

The results obtained for different model orders in AMARES$_H$ were compared and the performance of the automatic order estimation scheme was examined. For AMARES$_H$, there are no significant differences between the results obtained for the theoretical model order and those determined by the automatic model order selection criterion.

In AMARES$_f$ the influence of the choice of the upper and lower bounds in the filter design scheme was studied. The results are very insensitive to the choice of the lower and upper bounds in the filter design scheme. Only if the bounds are chosen close to the peaks of interest, there is a degeneration in the results.

In Fig. 6.16 the 4 methods are compared for different noise levels. The results displayed for AMARES$_{FREQ}$ were generated with $P = 1$, lower bound equal to 1 kHz, upper bound equal to 2.5 kHz. The results displayed for AMARES$_f$ were generated with suppression of the region below -1.2 kHz and above 3.5

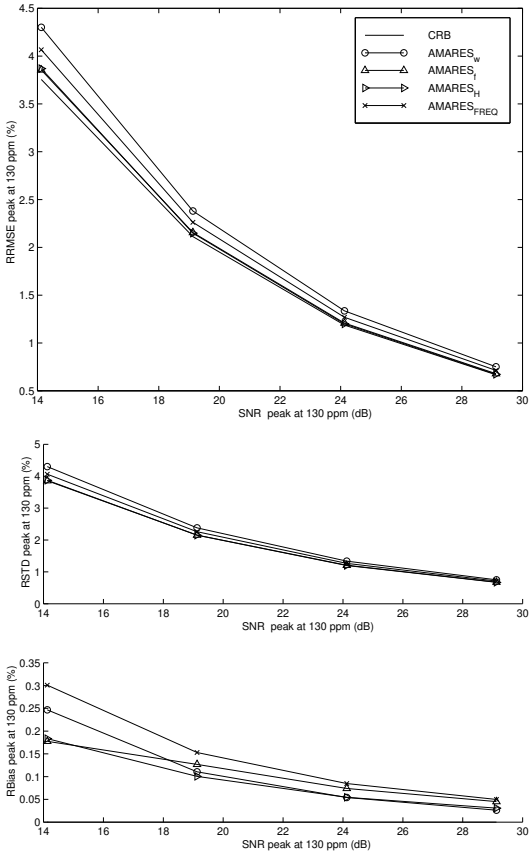


Figure 6.16: RRMSE, RSTD and RBias of amplitude estimates of the peak located at 1.66 kHz in the simulated ^{13}C spectrum derived from Fig. 6.15 for different SNRs using different FS estimation methods. The polynomial order P used in AMARES$_{FREQ}$ is 1 for all SNR levels. Top: RRMSE. Middle: RSTD. Bottom: RBias. CRB denotes the theoretical lower bound.

kHz. For AMARES$_H$ the results were obtained for a model order of 15. As seen in Fig. 6.16, the results for both AMARES$_f$ and AMARES$_H$ are very close to the theoretical CR bound. Since the peaks of interest are well separated from the other peaks in the spectrum, the other methods perform only slightly worse, as expected. Note also that all methods are able to remove the bias and that the RSTD contributes most to the RRMSE.

FS estimation in this case also leads to a reduction in calculation time. Quantifying all 15 peaks of the spectrum with AMARES takes about 7.6 s, while it takes 0.6 s to analyze the two peaks of interest with AMARES$_f$. The filter

design itself takes less than 1 second in this case. Overall, FS estimation in this example reduces the computation time with a factor 5.

6.2.3 Frequency-selective quantification of an *in-vivo* ^{31}P spectrum

The simulation study in this section is based on a signal derived from an *in-vivo* ^{31}P signal of the perfused rat liver, acquired at 4.7 T (81 MHz), displayed in the left hand side of Fig. 6.1. The β-ATP signal is used for the quantification of the ATP concentration, since unlike the other ATP signals it is essentially free from underlying contributions from other low concentration molecules. Therefore, in some applications, it is interesting to extract only the parameters of the β-ATP triplet. In Fig. 6.7, it is illustrated how the triplet can be extracted using the FIR filter method.

To derive the parameters of the simulation signal, the *in-vivo* signal was quantified using AMARES. A Lorentzian line shape was used and the first data points were excluded from the fit to reduce the influence of the broad contribution. The following prior knowledge was imposed:

1. Amplitude ratios: 1/1 in doublets of α- and γ-ATP, 1/2/1 in β-ATP triplet.
2. Frequency splittings of 16 Hz within the multiplets.
3. Dampings of all ATP peaks equal.
4. Phases of all peaks equal.

The noise level in the simulations is similar to the one in the *in-vivo* signal. In all 4 compared methods the dampings and phases of the three β-ATP peaks were constrained to be equal, an amplitude ratio of 1/2/1 and frequency splittings of 16 Hz were imposed. The sampling frequency is 5 kHz and the number of data points is 128. The amplitude of the ATP peaks was varied from 20 to 100 % of the originally derived value. These variations correspond to changes typically encountered in *in-vivo* ^{31}P signals.

Different polynomial orders ($P = 0 - 5$) and widths of frequency regions were tested for AMARES$_{FREQ}$. The upper bound of the frequency region included in Eq. (5.7) was varied between -1.56 kHz and -1.29 kHz and the lower bound between -2.5 kHz and -2 kHz. The sensitivity of the RRMSE of the β-ATP peaks w.r.t. the choice of these parameters in AMARES$_{FREQ}$ is shown in the top of Fig. 6.17. The results are displayed for the lowest ATP value and the polynomial order is $P = 1$. A polynomial order of $P = 1$ is needed to get

reasonable results - low bias and standard deviation - for AMARES$_{FREQ}$. The choice of the region included in Eq. (5.7) is in this case important since the model is only valid in a rather small frequency region. Higher polynomial orders however lead to a large increase in standard deviation, thereby worsening the results even more.

The influence of the specified upper and lower bounds in the filter design scheme was tested. The lower bound of the frequency region to be suppressed was varied between -1.56 kHz and -1.29 kHz and the upper bound between 1.6 kHz and 2.5 kHz. The sensitivity of the RRMSE of the β-ATP peaks w.r.t. the choice of these parameters in AMARES$_f$ is shown in the bottom of Fig. 6.17. The results are displayed for the lowest ATP value. The choice of lower and upper bound in the AMARES$_f$ filter design scheme is not critical. Only if the lower bound is chosen close to the peaks of interest, there is a degradation of the amplitude estimates.

The influence of the choice of the model order in AMARES$_H$ and the automatic order selection method were investigated and the results of this analysis are shown in Fig. 6.18 for the β-ATP peaks. For AMARES$_H$ the choice of the model order has a strong influence on the accuracy of the amplitude estimates of the β-ATP peaks (Fig. 6.18). Slight undermodeling ($K = 8$) gives the same results as using the correct model order. Overmodeling, on the other hand, leads to much worse results. A similar phenomenon occurs when HSVD is used to model the water peak in proton spectra. A wrong choice of model order in that application also leads to a deterioration of parameter accuracy (see also chapter 8). The MDL criterion for automatic order estimation works very well for ATP values between 40 and 100 % of the original value, but breaks down for values below 20%, i.e. in low SNR cases.

In Fig. 6.19 the 4 FS parameter estimation methods are compared. The results of AMARES$_{FREQ}$ corresponding to $P = 1$ are shown. A lower bound of -1.29 kHz and an upper bound of 1.6 kHz were used in AMARES$_f$. The model order in AMARES$_H$ is equal to 11, the theoretical model order. The results for AMARES$_H$ and AMARES$_f$ are very similar if the right model order is chosen in AMARES$_H$. AMARES$_w$ and AMARES$_{FREQ}$ perform worse for all ATP amplitudes. Again the bias is effectively removed for all methods but the standard deviation of AMARES$_w$ and AMARES$_{FREQ}$ is higher.

The gain in calculation time in this particular example is less pronounced than in the ^{13}C example. In the ^{13}C example 2 out of 15 peaks are quantified. Since no prior knowledge is used in the ^{13}C example, AMARES$_f$ minimizes a cost function with 8 variables while AMARES used on the entire signal minimizes a cost function with 60 variables. In this particular example, since we impose a lot of prior knowledge, the number of variables to be fitted by AMARES used on the entire signal is 20, compared to 8 variables for AMARES$_f$. Quantifying the entire spectrum using AMARES takes about 1.5 s, while analyzing the

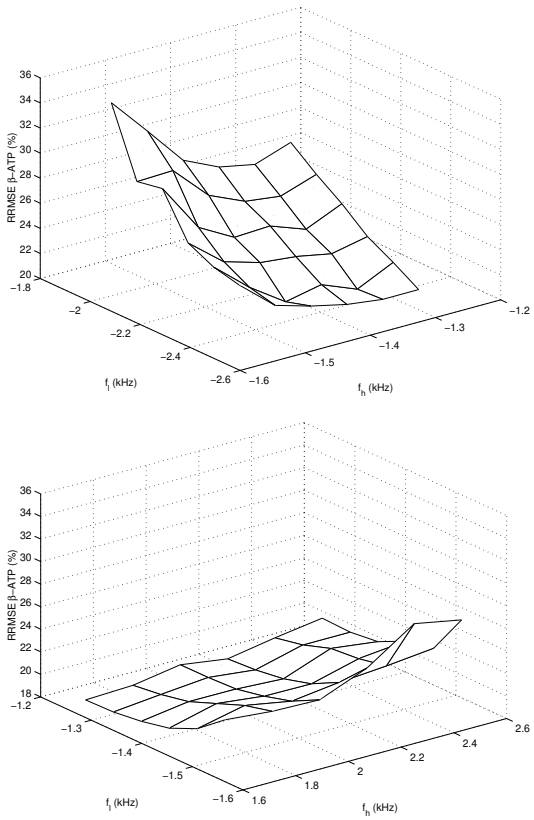


Figure 6.17: RMSE of the amplitude estimate of β-ATP in case their concentration has dropped to 20 % of their original value. Top: as a function of the frequency region included in Eq. (5.7) (AMARES$_{FREQ}$). The results are displayed for $P = 1$. Bottom: as a function of the choice of the lower and upper bounds in the filter design scheme.

β-ATP region only with AMARES$_f$ takes 0.3 s. The filter design in this case took also less than 1 second. Overall, the FIR filter approach still leads to a slight reduction in calculation time.

6.3 Conclusions

In this chapter FS quantification of biomedical MRS data was studied. The influence of nuisance peaks is in most cases not negligible and has to be taken care of prior to parameter estimation. A new method based on maximum-

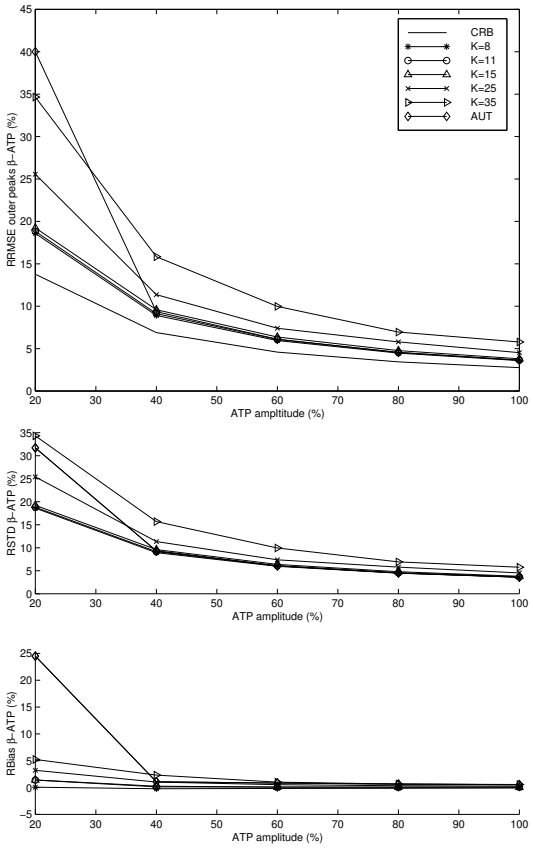


Figure 6.18: RRMSE, RSTD and RBias of the estimates of the amplitudes of the β-ATP peaks for different choices of model order (K) in AMARES$_H$. AUT denotes the results obtained by the automatic model order selection criterion. The results are obtained for ATP amplitudes ranging between 20 to 100 % of the maximum value. The corresponding SNR of the two outer peaks in the triplet varies linearly between -4 dB and 10 dB. CRB denotes the theoretical lower bound.

phase FIR filtering is presented and compared with other commonly used FS methods: time-domain weighting AMARES$_w$, HSVD filtering AMARES$_H$ and fitting in the frequency domain using a polynomial baseline AMARES$_{FREQ}$.

The time-domain weighting procedure AMARES$_w$ can be used to reduce the estimator bias if the peaks of interest are well separated from the nuisance peaks. The price is however a loss of SNR thereby increasing the estimator variance. Moreover, the method breaks down when the peak of interest is close to the nuisance peaks.

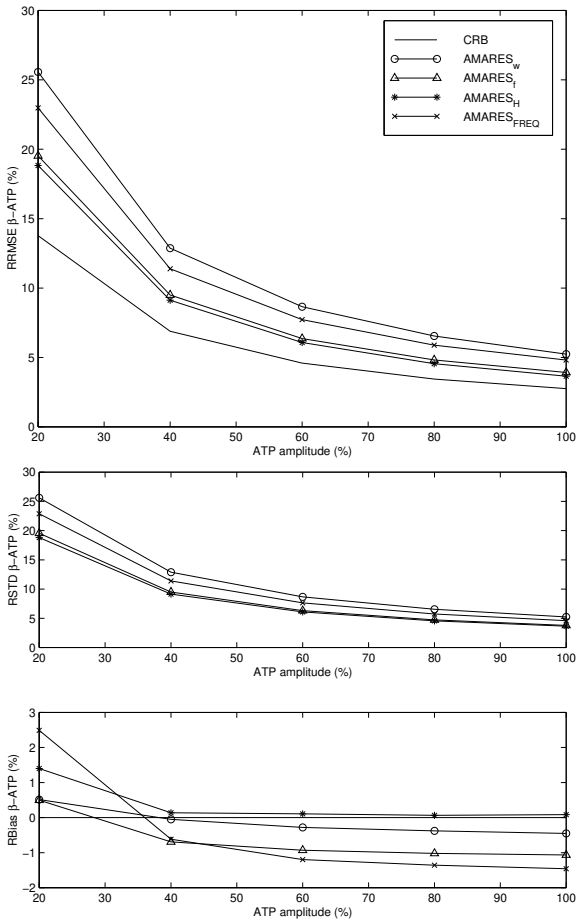


Figure 6.19: RRMSE, RSTD and RBias of the β-ATP amplitude estimates for 4 different FS estimation methods. The results are obtained for ATP amplitudes ranging between 20 to 100 % of the maximum value. The corresponding SNR of the two outer peaks in the triplet varies linearly between -4 dB and 10 dB. The polynomial order used in AMARES$_{FREQ}$ is 1 for all β-ATP amplitudes. Top: RRMSE. Middle: RSTD. Bottom: RBias. CRB denotes the theoretical lower bound.

If the appropriate model order is chosen, the HSVD filtering method AMARES$_H$ is seen to be a very accurate technique for FS estimation. A drawback of the method is the sensitivity to the chosen model order in more difficult cases. The use of the MDL criterion to automatically determine the model order was examined. MDL works well for a reasonable SNR but breaks down when the SNR decreases. A drawback of the use of MDL is that an entire singular value

decomposition of a matrix needs to be computed, which is time consuming.

Analyzing only a part of the spectrum is straightforward to implement for frequency-domain fitting methods (e.g. AMARES$_{FREQ}$) suggesting that these methods are suitable for FS estimation. However, the situation is complicated by the influence of the nuisance peaks which have to be modeled in some way. Using a polynomial to model the "tails" of the nuisance peaks is one way of reducing this influence. If the polynomial badly models the baseline in the region of interest, bias is introduced in the estimates and the choice of polynomial order in combination with the size of the frequency region is seen to be critical. These extra baseline model parameters also increase the variance of the estimates.

If the upper and lower bounds of the frequency region(s) to be suppressed are chosen with some care, AMARES$_f$ is able to keep the bias and the standard deviation of the parameters of interest low. The automatic filter design scheme makes the method easy to use and the computational complexity of AMARES$_f$ is low.

Two important MRS applications were examined. The first one is the extraction of two close peaks, remotely located from the other peaks in a ^{13}C spectrum. Since this is a rather easy example, all FS methods perform very similarly. The use of the FIR filter approach leads in this case to a reduction in computational time when compared to fitting the entire spectrum with AMARES. In the ^{31}P example, the parameters of the β-ATP peaks were extracted. The FIR filter method leads to the best results. HSVD filtering performs only well for an appropriate choice of model parameters and the accuracy of the results is seen to be very sensitive to the model order. The choice of the polynomial order and frequency region included in the fit were rather critical for this example.

In summary, the low computational complexity, ease of use and high parameter accuracy make AMARES$_f$ an attractive tool for FS parameter estimation.

Chapter 7

Fast removal of residual water in proton spectra

As explained in section 1.6 a residual water peak always remains present in the proton spectra. This residual water can not be described by an analytical function, mainly because of magnetic field inhomogeneity and lineshape distortions caused by suppression techniques. Since no model function is available for the water signal, a method like e.g. AMARES can not be used without first removing the disturbing peak. In chapter 5 it is explained how a black-box method like e.g. HSVD can be used to eliminate the water signal from proton spectra. The main drawback with HSVD, however, is the large computational load associated with the SVD of the data matrix. Efficiency is of primary importance in MR spectroscopic imaging. A single metabolite image typically requires $32 \times 32 = 1024$ *times the removal of the water resonance. The main objectives of this chapter are:*

- *To show how the SVD can be replaced by a more efficient low-rank algorithm which yields significant time savings in this application.*

- *To investigate the accuracy of this new method w.r.t. the use of HSVD in this context.*

The work described in this chapter is based on [51, 144].

7.1 Derivation of the new algorithm

7.1.1 HSVD - Toeplitz version

In this chapter a slightly different variant of the HSVD algorithm 2.1.1 is used. The data are here arranged in a Toeplitz instead of a Hankel matrix. The procedure to determine the parameters can be derived in the same way as was done in section 2.1.2.2.2. For completeness, the modified version is detailed below.

Algorithm 7.1.1 (HSVD - Toeplitz version)

Input: data samples $y_n, n = 0, \ldots, N-1$, model order K.
Output: $\hat{a}_k, \hat{d}_k, \hat{f}_k, \hat{\phi}_k, k = 1, \ldots, K$.

Step 1. Arrange the data points $y_n, n = 0, \ldots, N-1$ in a $(M \times L)$ Toeplitz matrix $\mathbf{T}$ as follows

$$\mathbf{T} = \begin{bmatrix} y_{M-1} & y_M & \cdots & y_{N-1} \\ y_{M-2} & y_{M-1} & \cdots & y_{N-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_0 & y_1 & \cdots & y_{L-1} \end{bmatrix}, \; M \geq K, \; L \geq K, \; N = M + L - 1. \tag{7.1}$$

Step 2. Compute the SVD of the Toeplitz matrix $\mathbf{T}$, i.e.,

$$\mathbf{T} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \text{ where } \mathbf{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_p), \sigma_1 \geq \ldots \geq \sigma_p \; ,$$

$p = \min(M, L)$. Choose $\mathbf{T}$ as square as possible [138], i.e., $M = L(+1) = N/2$, to have the highest accuracy.

Step 3. Truncate $\mathbf{T}$ to a matrix $\mathbf{T}_K$ of rank K

$$\mathbf{T}_K = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^H,$$

$\mathbf{U}_K$ and $\mathbf{V}_K$ are respectively the first K columns of $\mathbf{U}$ and $\mathbf{V}$, $\mathbf{\Sigma}_K$ is the $K \times K$ upper-left submatrix of $\mathbf{\Sigma}$. The model order K is chosen equal to the number of sinusoids that comprise the measured signal. In case a water peak is present the non-lorentzian lineshape of this peak has to be taken into account. In practical situations only a few lorentzians are taken to describe the water region [135]. See chapters 6 and 8 for a discussion on the influence of the choice of the model order.

Step 4. Compute the least squares solution $\hat{\mathbf{E}}$ of the following (incompatible) system

$$\mathbf{V}_K^{\uparrow} \approx \mathbf{V}_{K\downarrow}\mathbf{E}^H,$$

where $\mathbf{V}_K^{\uparrow}$ and $\mathbf{V}_{K\downarrow}$ are derived from $\mathbf{V}_K$ by omitting its first and last row respectively. Once $\mathbf{E}$ is estimated, its K eigenvalues give the signal pole estimates

$$\hat{z}_k = e^{(-\hat{d}_k + j2\pi\hat{f}_k)\Delta t}, \quad k = 1, \ldots, K.$$

From these signal poles it easy to obtain estimates of the dampings d_k and frequencies f_k.

Step 5. Finally, fill in the estimates $\hat{z}_k, k = 1, \ldots, K$ in the N model equations and compute the least squares solution $\hat{c}_k = \hat{a}_k e^{j\hat{\phi}_k}$, $k = 1, \ldots, K$, of

$$y_n \approx \sum_{k=1}^{K} c_k \hat{z}_k^n \, , n = 0, \ldots, N-1.$$

In this way estimates for the amplitudes a_k and phases ϕ_k are obtained.

The computationally most intensive part of the algorithm is the computation of the SVD of the $M \times L$ matrix $\mathbf{T}$, which requires $\mathcal{O}(ML^2 + L^3)$ flops ($\mathcal{O}(\cdot)$ denotes the order of magnitude). The least squares solution $\hat{\mathbf{E}}$ can be computed efficiently by making use of the Sherman-Morrison matrix-inversion formula [60]. As can be seen from the above algorithm, the full SVD is not required. Instead, only the first K columns of $\mathbf{V}$ are required, which estimate the signal subspace. Since K is usually much smaller than L, a lot of computational effort in computing a full SVD is wasted. Therefore a new matrix decomposition recently introduced in numerical linear algebra [50] is presented here. These so-called low-rank revealing decompositions only compute approximations to the desired signal subspace, resulting in considerable computational savings.

7.1.2 HLR

A rank revealing ULV (RR ULV) decomposition of $\mathbf{T}$ is of the following form

$$\mathbf{T} = \breve{\mathbf{U}}\breve{\mathbf{L}}\breve{\mathbf{V}}^H, \quad \breve{\mathbf{L}} = \begin{pmatrix} \breve{\mathbf{L}}_K & \mathbf{0} \\ \breve{\mathbf{H}} & \breve{\mathbf{E}} \end{pmatrix}, \tag{7.2}$$

where $\breve{\mathbf{L}}_K$ is a lower triangular $K \times K$ matrix whose singular values approximate the first K singular values of $\mathbf{T}$. The 2-norm of $(\breve{\mathbf{H}}, \breve{\mathbf{E}})$ is of the same order

of magnitude as the $(K+1)$th singular value of $\mathbf{T}$. $\breve{\mathbf{U}}$ and $\breve{\mathbf{V}}$ are unitary matrices. $\breve{\mathbf{V}}_K$, consisting of the first K columns of $\breve{\mathbf{V}}$, approximates the signal space. Special *low-rank* revealing (LRR) decompositions have been developed [50] to handle the case the dimension of the signal subspace is small, i.e., $K << L$. A LRR algorithm can be used to compute $\breve{\mathbf{V}}_K$ and $\breve{\mathbf{V}}_K$ can be used as an estimate for the signal subspace instead of computing a full SVD and truncating $\mathbf{V}$ to rank K. Several versions of LRR algorithms have been developed [50, 51]. In case $K << L$ these algorithms are able to compute $\breve{\mathbf{V}}_K$ much more efficiently than their SVD counterparts. For some of the LRR algorithms, the Toeplitz structure of the data matrix can only be exploited in the initial stage of the algorithm. The QR-based LRR algorithm [51] starts with a QR factorization. The rank-revealing steps are then applied to the triangular factor. The Toeplitz structure of $\mathbf{T}$ can only be exploited in the computation of the QR factorization. The Householder-based LRR algorithm avoids the initial QR factorization and works directly on the matrix. The algorithm requires first singular vector estimation, followed by Householder transformations. This vector can be estimated using the bidiagonal Lanczos method. For the calculation of the first singular vector fast Toeplitz matrix-vector products using FFT [141] can be used. After applying the Householder transformations to the original matrix, the Toeplitz structure of the matrix is destroyed and no fast structure exploiting methods can be used. The algorithm presented here, originally developed by R. Fierro, has the same properties as the algorithms published in [50], but makes optimal use of the Toeplitz structure of the original data matrix and therefore results in the largest computational savings [51]. It is called the Product Form LULV algorithm (PFLULV). The Toeplitz structure of the original matrix is preserved throughout all steps of the algorithm and fast Toeplitz matrix-vector products can be performed via FFT any time the Lanczos method is used. Below a brief description of the algorithm is given.

First, an estimate $\mathbf{u}_{est}^{(1)}$ of the left singular vector belonging to the largest singular value of $\mathbf{T}$ is estimated. Therefore the Lanczos method [60] is used. The stopping criterion for the Lanczos iterations is based on the current singular value estimate δ_j computed during the j-th iteration and the method proceeds until $|\delta_j - \delta_{j-1}| < \eta\delta_j$ where $\eta \geq 0$ is a threshold, or a maximum number of iterations is reached. For MRS applications the threshold can be set to $1e-03$ and the maximum number of iterations to 15, as done in this chapter. Based on $\mathbf{u}_{est}^{(1)}$ an estimate of the largest singular value $\sigma_{est}^{(1)}$ of $\mathbf{T}$ and corresponding right singular vector $\mathbf{v}_{est}^{(1)}$ are computed, i.e., $\sigma_{est}^{(1)} = \|\mathbf{T}^H\mathbf{u}_{est}^{(1)}\|$ and $\mathbf{v}_{est}^{(1)} = (\mathbf{T}^H\mathbf{u}_{est}^{(1)})/\sigma_{est}^{(1)}$. The vectors $\mathbf{u}_{est}^{(1)}$ and $\mathbf{v}_{est}^{(1)}$ are then reduced to a complex unit vector by *complex* Householder transformations [60] represented by Householder matrices $\mathbf{P}^{(1)}$ and $\mathbf{Q}^{(1)}$

$$\mathbf{P}^{(1)}\mathbf{u}_{est}^{(1)} = [e^{j\psi}, 0, \ldots, 0]^T \text{ and } \mathbf{Q}^{(1)}\mathbf{v}_{est}^{(1)} = [e^{j\varphi}, 0, \ldots, 0]^T. \qquad (7.3)$$

In general the complex Householder matrix $\mathbf{P}$ is of the form

$$\mathbf{P} = \mathbf{I} - 2\mathbf{z}\mathbf{z}^H/\mathbf{z}^H\mathbf{z},$$

where $\mathbf{z} = \mathbf{x} + sign(Re(x_1))\|\mathbf{x}\|[1, 0, \ldots, 0]^T$, $\mathbf{x}$ is the vector to be reduced to a complex unit vector.

Since $\mathbf{P}^{(1)}$ and $\mathbf{Q}^{(1)}$ are hermitian and unitary it can easily be deduced from Eq. (7.3) that the matrices have the following structure

$$\mathbf{P}^{(1)} = [\mathbf{u}_{est}^{(1)} e^{-j\psi}, \mathbf{P}_2^{(1)}] \text{ and } \mathbf{Q}^{(1)} = [\mathbf{v}_{est}^{(1)} e^{-j\varphi}, \mathbf{Q}_2^{(1)}].$$

It is then easy to see that elements 2 through n of the first row of $\mathbf{L} = \mathbf{P}^{(1)}\mathbf{T}\mathbf{Q}^{(1)}$ are equal to zero

$$\mathbf{u}_{est}^{(1)^H}\mathbf{T}\mathbf{Q}_2^{(1)} = \sigma_{est}^{(1)}(\mathbf{v}_{est}^{(1)^H}\mathbf{Q}_2^{(1)}) = \sigma_{est}^{(1)}(\mathbf{0}) = \mathbf{0}.$$

Throughout the algorithm the relation $\mathbf{T} = \breve{\mathbf{U}}^{(i)}\mathbf{L}\breve{\mathbf{V}}^{(i)^H}$ has to hold. Therefore $\breve{\mathbf{U}}^{(1)} = \mathbf{P}^{(1)^H}$ and $\breve{\mathbf{V}}^{(1)} = \mathbf{Q}^{(1)}$. The matrices $\breve{\mathbf{U}}^{(1)}$ and $\breve{\mathbf{V}}^{(1)}$ need however not be computed explicitly. Indeed, as shown above, the associated Householder vectors are therefore used and stored instead as shown in the outline of PFLULV below. The same procedure is then repeated on the submatrix $\mathbf{L}(2 : M, 2 : L) = \mathbf{P}_2^{(1)}\mathbf{T}\mathbf{Q}_2^{(1)}$, which, however, is not explicitly formed. The notation $\mathbf{L}(i : M, j : L)$ is used here to denote a submatrix of $\mathbf{L}$ consisting of the rows i to M and columns j to L. In this way the original Toeplitz structure of $\mathbf{T}$ is preserved, allowing fast matrix-vector multiplications using the FFT to be performed each time $\mathbf{u}_{est}^{(i)}$ is estimated using the Lanczos procedure. Another big advantage of HLR is the fact that the $(M \times L)$ Toeplitz matrix need not be stored in computer memory since matrix-vector multiplications with $\mathbf{T}$ or $\mathbf{T}^H$ which need to be performed during the course of the algorithm are entirely defined by the original data vector $[y_0, \ldots, y_{N-1}]$ [141].

Algorithm 7.1.2 (PFLULV)

Input: Toeplitz matrix $\mathbf{T}$, order K.
Output: $\breve{\mathbf{V}}_K$.

Step 1. Initialize $\breve{\mathbf{U}}^{(0)} \leftarrow [\,]$ and $\breve{\mathbf{V}}^{(0)} \leftarrow [\,]$, where $[\,]$ denotes the empty matrix.

Step 2. For $i = 1 : K$

Step 2.1. Compute estimate $\mathbf{u}_{est}^{(i)}$ of $\mathbf{L}(i : M, i : L) = \left(\prod_{j=1}^{i-1} \mathbf{P}_2^{(j)}\right)^H \times \mathbf{T} \times \prod_{j=1}^{i-1} \mathbf{Q}_2^{(j)}$.

Compute $\sigma_{est}^{(i)} = \|\mathbf{L}(i:M, i:L)^H \mathbf{u}_{est}^{(i)}\|$ and $\mathbf{v}_{est}^{(i)} = (\mathbf{L}(i:M, i:L)^H \mathbf{u}_{est}^{(i)})/\sigma_{est}^{(i)}$

Step 2.2. Determine the Householder vectors $\mathbf{z}_u$ en $\mathbf{z}_v$ which determine Householder matrices $\mathbf{P}^{(i)}$ and $\mathbf{Q}^{(i)}$ such that $\mathbf{P}^{(i)}\mathbf{u}_{est}^{(i)} = (e^{j\psi}, 0, \ldots, 0)^T$ and $\mathbf{Q}^{(i)}\mathbf{v}_{est}^{(i)} = (e^{j\varphi}, 0, \ldots, 0)^T$. Omit the first column of $\mathbf{P}^{(i)}$ and $\mathbf{Q}^{(i)}$ to define the related submatrices $\mathbf{P}_2^{(i)}$ and $\mathbf{Q}_2^{(i)}$.

Step 2.3. Append the Householder vectors $\mathbf{z}_u$ and $\mathbf{z}_v$

$$\breve{\mathbf{U}}^{(i)} = \begin{bmatrix} \breve{\mathbf{U}}^{(i-1)}, & \begin{bmatrix} \mathbf{0} \\ \mathbf{z}_u \end{bmatrix} \end{bmatrix}, \quad \breve{\mathbf{V}}^{(i)} = \begin{bmatrix} \breve{\mathbf{V}}^{(i-1)}, & \begin{bmatrix} \mathbf{0} \\ \mathbf{z}_v \end{bmatrix} \end{bmatrix},$$

$\mathbf{0}$ represents the zero vector/matrix of appropriate dimensions.

End (For)

Step 3. Use [60, Algorithm 5.1.2] to accumulate the K columns of $\breve{\mathbf{V}}^{(K)}$ to obtain the estimate $\breve{\mathbf{V}}_K$ for the signal subspace as follows

$$\begin{bmatrix} \mathbf{I}_{K-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{(K)} \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q}^{(2)} \end{bmatrix} \mathbf{Q}^{(1)} \begin{bmatrix} \mathbf{I}_K \\ \mathbf{0} \end{bmatrix}$$

$\mathbf{I}_K$ represents the identity matrix of size $K \times K$.

The algorithm HLR is obtained by replacing the SVD and subsequent truncation of $\mathbf{V}$ to $\mathbf{V}_K$ in Algorithm 7.1.1 by algorithm 7.1.2. The outline of HLR is given below.

Algorithm 7.1.3 (HLR)

Input: data samples $y_n, n = 0, \ldots, N-1$, model order K.
Output: $\hat{a}_k, \hat{d}_k, \hat{f}_k, \hat{\phi}_k, k = 1, \ldots, K$.

Step 1. See *Step 1* of Algorithm 7.1.1.

Step 2. Compute $\breve{\mathbf{V}}_K$ using Algorithm 7.1.2.

Step 3-4. See *Step 4-5* of Algorithm 7.1.1 with $\mathbf{V}_K$ replaced by $\breve{\mathbf{V}}_K$.

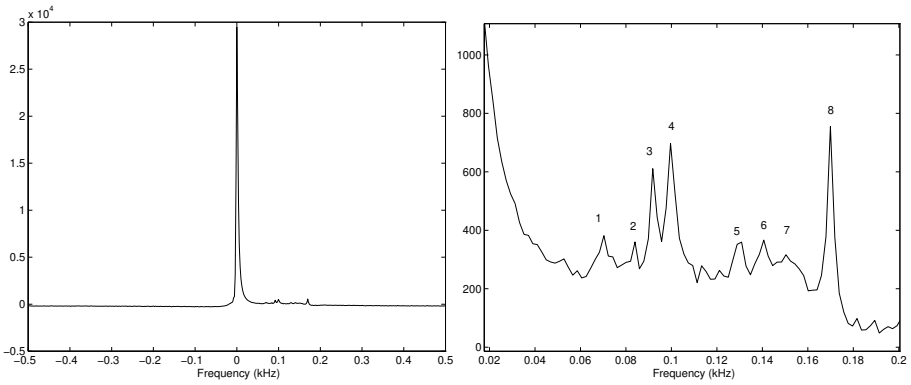


Figure 7.1: Real, phased noisy simulated ^{1}H spectrum ($\sigma = 0.8$). In the left figure the entire signal is represented, the figure to the right zooms in on the region of interest.

7.2 Simulations

In this section the accuracy and the efficiency of HLR compared to HSVD when used to preprocess ^{1}H spectra are addressed. To this end a Monte Carlo study is performed. The simulation signal used is derived from an *in-vivo* ^{1}H NMR echo signal and was previously used in [103]. From the noiseless signal 400 noisy realizations were generated with noise standard deviation σ (both on the real and imaginary parts). One of the used noisy simulation signals is displayed in Fig. 7.1.

The following signal processing protocol is applied to all simulation signals:

1. The entire signal is fitted using HLR or HSVD, and the model order $K = 10$ is used. To this end the 512 data points are arranged in a 257×256 Toeplitz matrix.

2. The peaks with frequencies belonging to the water region (defined as the region below 20 Hz) are used to reconstruct the water peak.

3. The reconstructed signal is subtracted from the original signal.

4. The residual signal is quantified using AMARES and estimates of the metabolites of interest are obtained.

HLR and HSVD are compared using the RMSE of the final parameter estimates. In Fig. 7.2 the RMSE of the amplitudes of peak 1 and 2 obtained

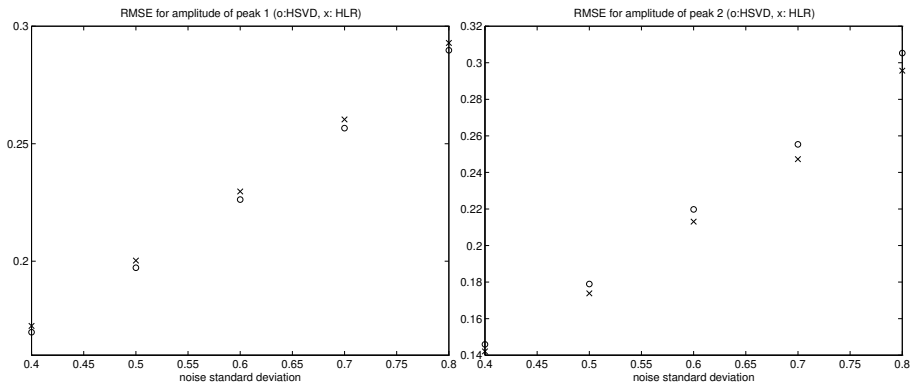


Figure 7.2: RMSE for amplitudes of peak 1 and 2 of the simulation signal for different values of the noise standard deviation after removal of the water peak and quantification of the residual metabolites using AMARES. The crosses and the circles denote that the water peak was removed using HLR and HSVD respectively. In the water removal preprocessing step the model order used was 10 and the data were arranged in a 257×256 data matrix.

by removing the water via HSVD and HLR and subsequent quantification of the residual signal with AMARES are compared. For peak 1 the RMSE obtained using HSVD preprocessing is slightly lower than that obtained using HLR preprocessing. For peak 2 the situation is reversed. For the other peaks the situation is similar–the obtained RMSE of all parameters using HSVD-preprocessing and HLR preprocessing are almost the same and neither of the two methods leads to an overall lowest RMSE.

It can be concluded that both methods, when used to subtract the water peak, lead to a comparable parameter accuracy of the metabolites of interest obtained after the final parameter estimation with a NLLS algorithm.

As a measure of efficiency the number of flops (obtained by MATLAB) required by PFLULV are compared to the number of flops required by the SVD for a 257×256 Toeplitz data matrix. The overall number of flops required by HSVD and HLR is also reported. The results for different noise levels are displayed in Table 7.1.

To get an idea of the difference in actual execution time between HSVD and HLR both algorithms were implemented in Fortran 77 making use of the BLAS and LINPACK libraries (available from netlib [53]). The timing experiments were performed on a SUN ULTRA 2 (200 Mhz). The results are displayed in Table 7.2. Although the timing results are compiler and computer dependent, Table 7.2 indicates that the gain in efficiency in terms of actual CPU times is

	$\sigma = 0.4$	$\sigma = 0.5$	$\sigma = 0.6$	$\sigma = 0.7$	$\sigma = 0.8$
$\frac{fl_{\mathrm{SVD}}}{fl_{\mathrm{LRR}}}$	57.4	57.2	57.0	56.8	56.2
$\frac{fl_{\mathrm{HSVD}}}{fl_{\mathrm{HLR}}}$	47.0	46.9	46.9	46.8	46.6

Table 7.1: Comparison between SVD and LRR, HSVD and HLR in terms of flops. In the first row the ratio of the number of flops (in MATLAB) needed by SVD to that required by the LRR algorithm PFLULV are displayed as a function of the noise level. In the second row the ratio of the number of flops required by HSVD to that required by HLR are displayed as a function of the noise level. The imposed model order K is 10 and the size of the data matrix used is 257×256.

	$\sigma = 0.4$	$\sigma = 0.5$	$\sigma = 0.6$	$\sigma = 0.7$	$\sigma = 0.8$
$\frac{t_{\mathrm{SVD}}}{t_{\mathrm{LRR}}}$	42.3	45.5	41.8	39.9	39.8
$\frac{t_{\mathrm{HSVD}}}{t_{\mathrm{HLR}}}$	37.4	40.2	36.9	35.2	35.2

Table 7.2: Comparison between SVD and LRR, HSVD and HLR in terms of CPU times. In the first row the ratio of the CPU time needed by SVD to that required by the LRR algorithm PFLULV are displayed as a function of the noise level. In the second row the ratio of the CPU time required by HSVD to that required by HLR are displayed as a function of the noise level. The imposed model order K is 10 and the size of the data matrix used is 257×256.

a factor 35 to 40 in this particular case ($K = 10$, a 257×256 data matrix). The CPU time needed by HLR to quantify one of these simulation signals is of the order of 0.3 s as opposed to 11 s for HSVD. The subsequent analysis using AMARES takes about 2 s per signal. This shows that preprocessing using HSVD takes more time than the actual quantification using AMARES, which is of course unacceptable, especially when a large amount of data has to be analyzed.

7.3 Application to *in-vivo* MRS image

This fast algorithm is particularly useful when a lot of data needs to be processed as is the case in MRSI. To illustrate the gain in efficiency in the analysis of *in-vivo* MRS signals the performance of HLR and HSVD to analyze part of a MRSI data set are compared. The data set under investigation was measured at the University of Alabama at Birmingham and provided by Dr. J.A. den Hollander, Center for NMR Research and Development, University of Alabama at Birmingham. The measurements were performed on a 1.5 T ACS/S15 Philips

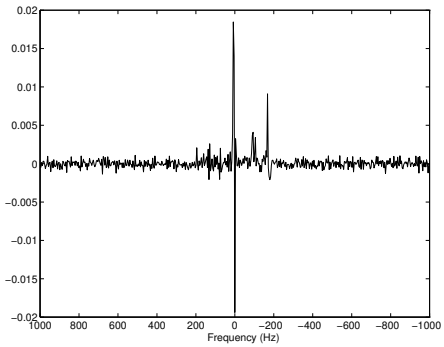


Figure 7.3: Real, phased *in-vivo* ^{1}H spectrum of the MRSI data set. The water region is defined as all peaks with a frequency higher than -78 Hz.

Gyroscan, using a protocol developed by Philips Medical Systems [32].

To make sure that only signals containing metabolites of interest were processed, 20 signals from the middle region of the image were selected and preprocessed. In these signals contributions from water had to be removed. If the entire image is to be processed the protocol explained in [32] can be used.

An exponentially damped sinusoid is assumed to contribute to the water peak if it has a frequency higher than -78 Hz. See Fig. 7.3 for a representative signal.

The number of data points used in the preprocessing stage was 512. Five exponentials were enough to remove the contributions from water. Fig. 7.4 displays the signal of Fig. 7.3 after preprocessing with HLR (the imposed model order was 5).

To show however the influence of the chosen model order on the efficiency of the algorithm, model orders of ten and twenty were also used. The total CPU time needed to preprocess twenty signals using HLR and a model order of 5 is only 3.15 s compared to 259.6 s using HSVD. The use of HLR in this case represents a reduction of a factor of 82.4 in CPU time. As illustrated in Table 7.3, the gain in efficiency decreases as more exponentials are estimated.

To show the influence on the number of data points used, the same analysis for different model orders is done using only 256 data points. The results are displayed in Table 7.4. As expected, the gain in efficiency decreases when the number of data points decreases.

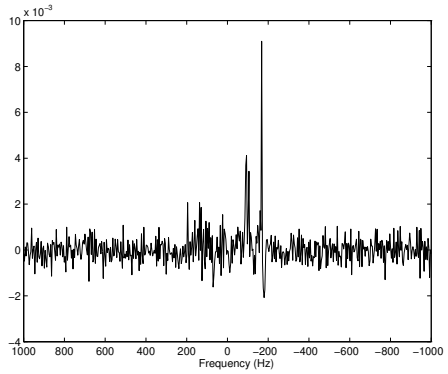


Figure 7.4: Real, phased *in-vivo* ^{1}H signal of Fig. 7.3 after removal of the water contributions using HLR and a model order of 5.

	$K = 5$	$K = 10$	$K = 20$
t_{HSVD}	259.6 s	254.3 s	263.7 s
t_{HLR}	3.15 s	7.0 s	19.8 s
$\frac{t_{\text{HSVD}}}{t_{\text{HLR}}}$	82.4	36.3	13.3

Table 7.3: Total CPU time needed by HSVD and HLR to preprocess 20 signals of a MRS image consisting of 512 data points each for imposed model orders of 5, 10 and 20. The ratios between the needed CPU times are also depicted.

	$K = 5$	$K = 10$	$K = 20$
t_{HSVD}	20.9 s	20.1 s	20.4 s
t_{HLR}	2.16 s	4.07 s	10.77 s
$\frac{t_{\text{HSVD}}}{t_{\text{HLR}}}$	9.7	5.0	1.9

Table 7.4: Total CPU time needed by HSVD and HLR to preprocess 20 signals of a MRS image consisting of 256 data points each for imposed model orders of 5, 10 and 20. The ratios between the needed CPU times are also depicted.

7.4 Conclusions

In this chapter HLR is presented as an alternative to HSVD for the removal of the water peak in ^{1}H spectra. HLR uses a low-rank revealing decomposition to extract the signal subspace instead of a full SVD as done in HSVD. This results in a considerable improvement in efficiency without affecting the accuracy of the parameters of interest. The latter can be estimated after the preprocessing

stage by means of an interactive algorithm like AMARES.

The gain in efficiency depends on both the number of data points of the signal and the imposed model order and is more pronounced when the number of data points becomes larger and the model order smaller. For a typical signal consisting of 512 data points and a model order of 10, the CPU time required by HLR is less than that required for HSVD by a factor of about 35 to 40.

A newer version of the algorithm presented here has been developed by R. Fierro very recently[1]. Instead of using $\mathbf{T}$, the method works with the matrix $\mathbf{T}^H\mathbf{T}$ to which the symmetric Lanczos method can be applied, resulting in an even bigger reduction in computational complexity.

[1] Information obtained by private communication with R. Fierro.

Chapter 8

Application of FS estimation in proton spectra

Since the removal of residual water is such an important area, this chapter is devoted to the comparison of the HSVD-based filter method and the FIR filter method which were seen to be the most accurate choices in the general context of FS parameter estimation. The goals of this chapter are:

- *To show how the automatic FIR filter design scheme can be optimized for the removal of residual water to make the procedure more computationally efficient.*

- *To compare the FIR filter approach with the HSVD-based approach in terms of accuracy and efficiency in this particular context.*

In section 8.1 it is shown how the general filter design scheme of section 6.1.4 can be adapted for optimal use in this application. A theoretical complexity analysis is also provided. In section 8.2 an interpretation of other FIR filters appeared in literature is given. In section 8.3 the FIR filter method and the HSVD version are compared with each other in terms of the influence of the (user) parameters involved with the respective approaches on the parameter accuracy and computational efficiency.

This chapter is based on work presented in [121].

8.1 Optimized filter design scheme

The filter design scheme makes use of the assumption that the high-energy water peak consists mainly of one exponentially damped sinusoid,

$$y_w(n) = a_w e^{j\phi_w} e^{(-d_w + j2\pi f_w)n\Delta t}$$

and that estimates of a_w, d_w, f_w and the noise standard deviation σ can easily be found. It is important to note that this assumption is merely used to estimate the energy content and the frequency localization of the water peak and does not impose restrictions on the actual shape of the water signal.

To obtain the estimates, the frequency domain magnitude of the water peak, s_0, is calculated as

$$s_0 = \max_l |\frac{\mathcal{Y}_l}{\sqrt{N}}|,\ l = 0, \ldots, N-1.$$

The estimate of the water peak frequency is given by the corresponding frequency

$$\hat{f}_w = \frac{1}{2\pi} \arg\ \max_l |\mathcal{Y}_l|.$$

The width f_{wp}, of the water peak (in Hz) at half height $s_0/2$, gives the following estimate of the water peak damping

$$\hat{d}_w = \pi f_{wp}.$$

The estimate of the water peak amplitude is given by

$$\hat{a}_w = s_0 \hat{d}_w.$$

Finally an estimate of the noise standard deviation can be calculated as the standard deviation of the last samples of the original signal. Based on these estimates an automatic procedure to design the FIR filter is proposed. An outline is given below.

Algorithm 8.1.1 (Filter design scheme - removal of residual water)

Input: data samples $y_n, n = 0, \ldots, N-1$, approximate frequency, f_1, of the metabolite of interest that lies closest to the water peak.
Output: the filter coefficients **h**.

Step 1. The estimates of the noise standard deviation σ, the damping d_w, frequency f_w and amplitude a_w of the water peak are calculated as described above.

Step 2. In order to suppress the water signal below the noise level, the suppression sup_0 is chosen equal to

$$sup_0 = \frac{\hat{\sigma}}{2\hat{a}_w}$$

and starting values for the filter order (e.g. $M = 50$) and passband ripple (e.g. $r = 5\%$) are determined.

Step 3. The suppression sup_0 is corrected by a function of the damping of the water signal

$$sup = sup_0 \left[\frac{M}{\sum_{m=0}^{M-1} e^{-\hat{d}_w m \Delta t}} \right] = sup_0 \left[\frac{M(1 - e^{-\hat{d}_w M \Delta t})}{1 - e^{-\hat{d}_w \Delta t}} \right].$$

The reason for this correction is that the filter suppression is given for the magnitude response of the filter. The damped sinusoids in the water signal will be less suppressed due to the weighting of the filter coefficients by the damping term. The $\left[\frac{M(1-e^{-\hat{d}_w M \Delta t})}{1-e^{-\hat{d}_w \Delta t}} \right]$ term approximately compensates for this loss.

Step 4. A linear-phase filter (using the constrained LS algorithm of [112]) is designed with the lowest possible cut-off frequency f_c, that fulfills the following constraints

$$|\mathbf{h}^T \mathbf{g}(0)| \leq sup$$

$$|\mathbf{h}^T \mathbf{g}(\hat{d}_w)| \leq sup,$$

$\mathbf{g}$ is the Fourier vector defined in Eq. (6.5).

Step 5. If this is not possible for

$$f_c < f_1$$

where f_1 is the frequency of the metabolite of interest that lies closest to the water peak, the filter order M is increased by 10 and the process is restarted from *Step 4*.

Step 6. The linear-phase FIR filter is transformed to a minimum-phase filter by spectral factorization and the filter coefficients are reordered to obtain the final maximum-phase filter.

The only parameter the user has to specify using this scheme is the approximate frequency, f_1, of the peak of interest that lies closest to the water peak. Based on this specification a filter with an as narrow as possible stop band is designed. It was seen in section 6.1.4 that a narrow stop band is advisable to have a high group delay in the pass band.

The computational complexity of the above quantification scheme is an important issue. The parameter estimation in the first step above introduces only a modest computational load. The filter design using the automatic scheme is performed by a constrained LS fit which has a computational complexity of $\mathcal{O}(M^3)$ flops per iteration. The number of iterations is dependent on the filter specifications and can not be known beforehand. However, the algorithm is known to be efficient in the sense that the required number of iterations is low. The spectral factorization can be solved by finding the roots of the filter polynomial of length M. The standard solution to solve this problem involves finding the eigenvalues of a $M \times M$ matrix leading to a computational complexity of $\mathcal{O}(M^3)$ flops. The filtering is included in the AMARES$_f$ algorithm and thereby the computational burden is increased by $\mathcal{O}(NM)$ flops compared to the standard AMARES algorithm. In summary the most computationally intensive parts in the quantification scheme are the filter design step and the spectral factorization. However, note that for applications such as spectroscopic imaging where a large number of spectra has to be processed and where the water contributions between spectra are similar it is sufficient to design the filter only once. This will decrease the computational burden significantly.

8.2 Interpretation of other FIR filters appeared in literature

The removal of residual water in ^{1}H MRS has received much attention in the literature. An interpretation of some of the often used filters which have been developed over the years in the framework of FIR filters is given below.

The water peak is normally located at zero frequency (or can be frequency shifted to zero frequency since complex data are used) and the early techniques were developed with this in mind. In [78], first and second order differentiation were used to suppress the water peak. The first order derivative used for filtering in [78] corresponds to a first order *high-pass* FIR filter with coefficients

$$h_0 = -1 \quad h_1 = 1, \tag{8.1}$$

while the second order derivative is equivalent to a second order high-pass filter with coefficients

$$h_0 = 1 \quad h_1 = -2 \quad h_2 = 1. \tag{8.2}$$

A visual inspection of the filtered data spectrum shows that in most cases the water peak is successfully removed. However, from a filtering point of view, these are very basic filters which will strongly influence the metabolite peaks making an accurate quantitation using the filtered signal impossible. It is therefore desirable to design filters which minimize the influence on the peaks of interest. A technique that is more elaborate in that sense was proposed in [83] where the main idea consists of using a *low-pass* FIR filter of relatively high order to suppress all metabolite resonances at higher frequencies. The filtered signal ideally consists of the pure water signal which can be subtracted from the original signal. Is is worth noting that low-pass filtering the data to get a pure water signal and thereafter subtracting this signal from the data is equivalent to using a corresponding *high-pass* FIR filter to suppress the water signal. Using a filter $\{h_m^{Mar}\}$ and subtracting the filtered data from the original data sequence is *exactly* equal to using the corresponding filter

$$h_m^{Tot} = \begin{cases} 1 - h_m^{Mar} & \text{if } m = (M-1)/2 + 1 \\ -h_m^{Mar} & \text{if } m \neq (M-1)/2 + 1 \end{cases} \tag{8.3}$$

operating directly on the original data sequence. The filters proposed in [83] were sine-bell-shaped

$$h_m = cos\left(\frac{(m - (M-1)/2)\pi}{M}\right) \qquad m = 0, 1 \ldots, M \tag{8.4}$$

or Gaussian-shaped

$$h_m = e^{-(4m-(M-1))^2/M^2} \qquad m = 0, 1 \ldots, M \tag{8.5}$$

(were M in both cases is assumed to be odd). The filter order M was proposed to be chosen in the interval 17-65. The relatively high filter order, compared to the simpler first and second order derivative filters described above, results in an efficient water signal suppression and a small influence on the peaks of interest. The filters proposed in [83] are *linear-phase* filters that have the advantage of not altering the phase of the filtered signal components. The symmetric linear-phase filters have a time-delay equal to half the filter length. This gives a reason to alter the filtering convolution sum

$$y_{f_n} = \sum_{m=0}^{M-1} h_m y_{n-m+(M-1)/2} \qquad n = 0, \ldots, N-1 \tag{8.6}$$

(assuming M is odd). An obvious problem with this definition is that the first and last $(M-1)/2$ values can not be calculated correctly since no data are available for $n < 0$ and $n > N-1$. In [83] this problem was solved by replacing the first and last $(M-1)/2$ filtered data samples with data obtained from a linear extrapolation of the filtered signal. The extrapolation will introduce errors and could be improved using more refined techniques. In [116] it was

proposed to use a linear prediction method which models the water signal as a sum of damped exponentials. Note that the interpretation of the procedure as a high-pass filtering operation is not exactly valid due to the extrapolation of the filtered signal but it makes the comparison with the derivative filters above and other high-pass filtering methods more transparent.

Although the method in [83] efficiently manages to suppress the water signal a few problems still exists. First there is the issue of filter design. The estimation results are strongly influenced by the users choice of filter type and filter order and very few guidelines have been provided for making that choice. Furthermore, the filter influence on the peaks of interest is not taken into account limiting the accuracy of the parameter estimates. Finally the extrapolation of the data (using linear extrapolation or linear prediction methods) introduces distortions in the first and last $(M-1)/2$ samples. All these problems are taken into account by the maximum-phase FIR filters presented in chapter 6. These are high-pass filters designed to suppress the water signal as opposed to the low-pass filters in [83] designed to suppress the high frequency metabolite peaks. The extrapolation of data is avoided by the use of so called *maximum-phase* filters. This type of filters have a high time-delay which is close to the filter order M. This makes it possible to discard the initial $M-1$ filtered samples with minimal loss of accuracy.

8.3 Numerical validation

In this section the proposed FIR filter based suppression technique is evaluated in order to determine the sensitivity of the final parameter estimates w.r.t. the choice of the filter design parameters. The method is compared to the HSVD-based method. The water suppression abilities of both methods are investigated as a function of the noise level and varying frequency distances between the metabolite peaks and the water peak. A comparison of the computational complexity between the methods is also included.

First the water suppression abilities of the FIR filter based method are illustrated by applying the method to an *in-vitro* proton MRS signal. The *in-vitro* proton MRS signal is acquired from a water solution of 100mM Cr (CH2 singlet; CH3 singlet), 100mM acetate (CH3 singlet), 50mM t-butyl alcohol (3×CH3 singlet) and 10mM TSP (3×CH3 singlet). A single-voxel signal from a spherical phantom was acquired at 1.5T (Vision, Siemens) using the STEAM sequence (TR/TE/TM=20000/20/30 ms). The result is displayed in Fig. 8.1. The top left figure shows the magnitude spectrum of the eddy current corrected signal and the top right figure displays the real, phased spectrum. In the bottom figure the real, phased FIR filtered signal is displayed.

In the figure it can be seen that the water signal, including the tilted baseline

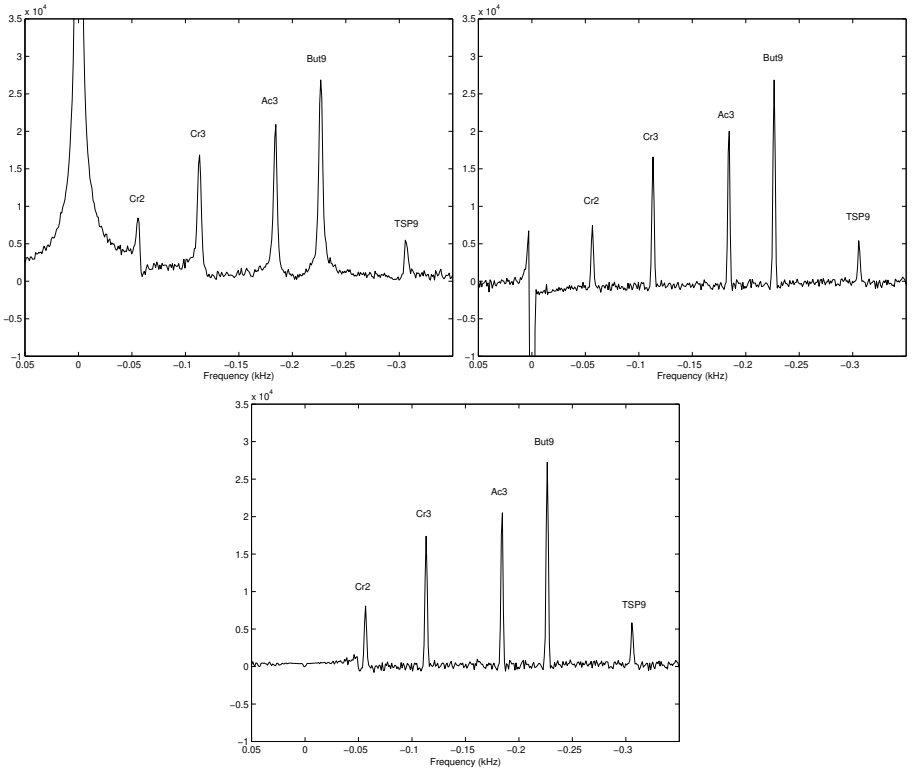


Figure 8.1: Water suppressed proton spectrum from a water solution of 100mM Cr, 100mM acetate, 50mM t-butyl alcohol and 10 mM TSP. A single-voxel signal from a spherical phantom was acquired at 1.5T using the STEAM sequence (TR/TE/TM=20000/20/30 ms). Top left: magnitude spectrum of eddy current corrected signal. Top right: phased, real spectrum of eddy current corrected signal. Bottom: real, phased FIR filtered spectrum.

distorting the nearby peaks, is perfectly removed by the filter obtained by the above filter design scheme.

The simulation signals are derived from the phantom signal described above in the following way. The acquired phantom signal was quantified with HSVD using a high model order ($M = 100$). The water signal was subsequently reconstructed with all the exponentially damped sinusoids with frequencies between $[-30, 30]$ Hz and with amplitudes above the estimated noise level $\hat{\sigma} \simeq 7.5$. The parameters of the seven peaks used to reconstruct the water resonance are found in Table 8.1. Five metabolite peaks were added as exponentially damped sinusoids with frequency, phase and damping close to what was measured in the phantom experiment. The amplitudes of the peaks were chosen to be ap-

f_{w_k} (Hz)	d_{w_k} (Hz)	θ_{w_k} (°)	a_{w_k} (a.u.)
8.48	5.10	-90.88	15.01
5.25	8.28	-45.19	64.74
2.16	10.51	-2.95	321.25
0.18	12.45	179.97	1142.30
0.17	4.24	-170.39	251.92
-3.09	6.79	36.77	201.11
-6.31	4.00	81.11	12.30

Table 8.1: Estimated water signal parameters used in the reconstruction of the water peak, $\theta_{w_k} = \phi_{w_k} * 180/\pi$ expresses the phase in degrees.

Peak k	f_k (Hz)	d_k (Hz)	θ_k (°)	a_k (a.u.)
1	-61	7	0	20
2	-118	7	0	30
3	-189	7	0	20
4	-231	7	0	20
5	-311	7	0	20

Table 8.2: Metabolite parameters used in the simulated signals, $\theta_k = \phi_k * 180/\pi$ expresses the phase in degrees.

proximately equal to the estimated TSP9 amplitude and set equal for all peaks except for the two Cr peaks whose 2:3 ratio was kept. In Table 8.2 the exact parameters used in the simulation examples are given. The added complex noise is white and circular Gaussian distributed. The noise standard deviation was varied to simulate a number of SNRs.

8.3.1 Influence of FIR filter parameters on estimation precision

The simulation signal described above is used here to examine the influence of the filter design procedure on the final parameter estimates as a function of the SNR. In Fig. 8.2 the RRMSE results obtained from 400 simulation runs are compared to the CRB for the amplitude estimates of peak1 and peak4. The estimation results for peak2 to peak5 are practically equal and therefore only the results of peak4 are shown. In addition to the automatic filter design procedure proposed in section 8.1 three different filters are chosen to investigate

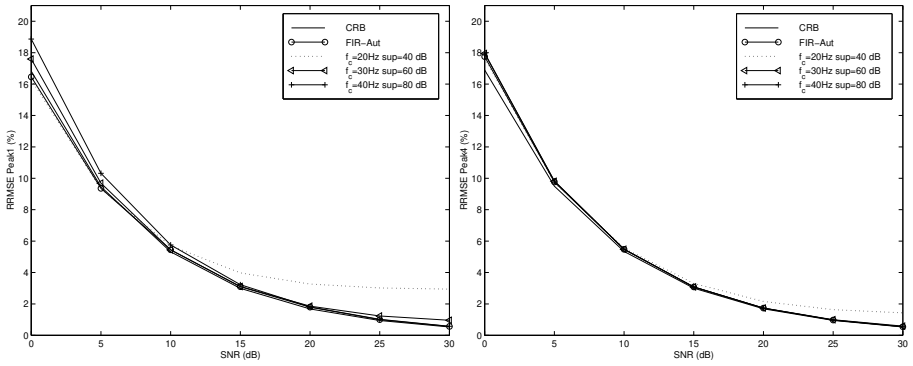


Figure 8.2: CRB and RRMSE of amplitude estimates as a function of SNR obtained from 400 simulation runs using different FIR filter parameters. Left: peak1. Right: peak4.

the filter influence on the parameter estimates. The filters have the same order, $M = 50$ and pass band ripple, $r = 0.01$, but different values of suppression: 40, 60 and 80 dB with corresponding cut-off frequencies: 20, 30 and 40 Hz respectively. Note that the influence of the filter order M is very modest. The order has to be chosen high enough to fulfill the requirements for the stop band suppression and cut-off frequency. Furthermore the pass-band ripple r is not critical either since this effect is taken into account in the estimation phase. As expected, the results for peak1 are more sensitive to the filter design procedure than the peaks further away from the water peak. The results for the different filters show that a low suppression and cut-off frequency is desirable for low SNR while a higher suppression is needed to sufficiently suppress the water peak for high SNR. These results are consistent with the theoretical reasoning that the filter should be designed to decorrelate the noise term (including the water signal) with the signals of interest as much as possible. A suitable level of the suppression must therefore be used not to deteriorate the final estimates. The automatic filter design procedure matches the filter characteristics with the water signal using the simple estimates described above. The resulting filter has a suitable suppression level and cut-off frequency leading to good results for all SNRs in the example.

8.3.2 Influence of HSVD parameters on estimation precision

For the HSVD filter method described in section 5.3.1.2 two parameters (M and f_c) have to be defined. The simulation signal is intended to examine

the sensitivity of the final parameter estimates to the choice of these user parameters. In Fig. 8.3 the RRMSE results obtained from 400 simulation runs for the HSVD method are displayed. The HSVD method was applied using different model orders (M = 10, 12, 20, 30 and 40) and cut-off frequencies (f_c = 15, 35 and 55 Hz). Only the results for f_c = 35 Hz for each model order are shown since simulations showed that the choice of cut-off frequency has a minor influence on the final parameter estimates. In this simulation example the water peak is exactly modeled by seven exponentially damped sinusoids. This implies that the correct order of the signal subspace is equal to 12 (7 sinusoids used to model the water signal and 5 to model the metabolite peaks). The results show that the correct choice of the model order gives the best results for all SNRs. It is interesting to note the relatively large variations in the final estimates for peak1 resulting from under- or overestimation of the order of the signal subspace. A low model order (i.e. $K = 10$) gives good results for low SNR while the results are very poor for high SNR. Overmodeling (i.e. $K > 12$) on the other hand gives poor estimates for low SNR. These results can be explained in a similar way as was done for the filter method above. Due to undermodeling the water signal is not completely removed which deteriorates the accuracy of the estimates at high SNR. This corresponds to the results obtained above using filters with too low suppression. On the other hand, overmodeling leads to modeling parts of the noise by damped sinusoids. Subtraction of these sinusoids from the original signal adds new features to the signal. These artificially introduced sinusoids can have a relatively high amplitude when the background noise is strong (low SNR) and influence the parameter estimates significantly. This can be compared to using an unnecessary high suppression FIR filter as described above. The simulation example is probably unrealistically simple for high SNR in which case the HSVD method has no problem in finding good estimates of the seven damped sinusoids used in the water signal reconstruction. However, the example shows that the model order selection for the HSVD method does influence the final parameter estimates and should be made with some care, thereby confirming the results obtained in chapter 6. The use of the MDL criterion as explained in section 5.3.1.2 was also examined in this specific application. Also here the method breaks down when the SNR becomes low. In low SNR cases the MDL criterion returns a much too low model order estimate (K =2-3). The HSVD method is applied with the correct model order ($K = 12$) in the following.

8.3.3 Comparison of FIR filter method and HSVD method w.r.t. estimation precision

In this simulation example the accuracy of the FIR filter method using the automatic filter design scheme is compared with the HSVD method using the correct model order ($K = 12$). The influence on the estimation accuracy as a

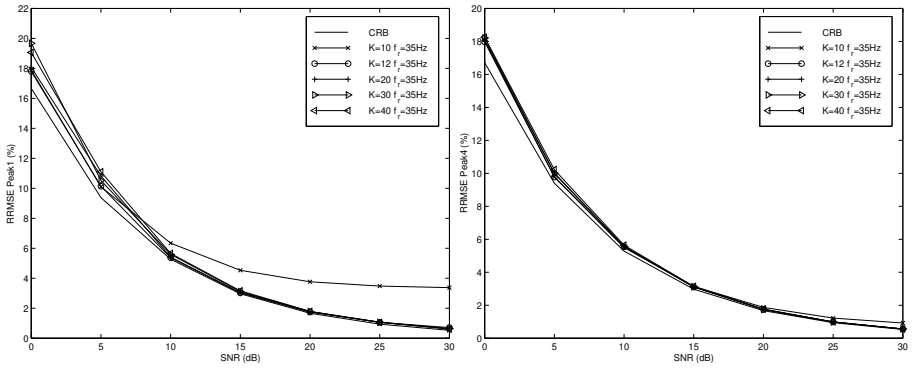


Figure 8.3: CRB and RRMSE of amplitude estimates as a function of SNR obtained from 400 simulation runs using HSVD with different model orders. Left: peak1. Right: peak4.

function of the position of the metabolite peaks is examined by modifying the basic experiment described above. The frequency of peak1 is set to different values (61, 51, 41 31 and 21 Hz) while the frequencies of the other peaks are left unchanged. 400 simulation runs with four different noise levels are used to quantify the estimation errors. In Fig. 8.4 the RRMSE for peak1 and peak4 are displayed as a function of the frequency for peak1. The results for peak1 show that the estimates are degrading when the metabolite peak is closer to the water peak. The reason is that the distortions introduced in the frequency region of the removed water signal by both methods correlate the noise term with nearby peaks. There is a difference between the distortions introduced by the two methods. The FIR filter uses a linear combination of the data samples to suppress the water peak. The distortions are in this case introduced by the inevitable suppression of the previously white noise in the water signal frequency region. The distortion introduced by the HSVD method is due to the estimation errors of the parameters of each sinusoid used to reconstruct the water signal. The subtraction of the water signal reconstruction will not lead to a perfect cancellation of the water signal and thereby introduce new features distorting the spectrum. These distortions are relatively small for high SNR where the HSVD procedure finds an excellent fit to the water signal but for realistic SNR levels the distortions have an effect on the final estimates as can be seen in this example. In this example the HSVD method is outperformed by the FIR filter method for lower SNRs which indicates that the distortions introduced by filtering are less important than the distortions introduced by the HSVD method.

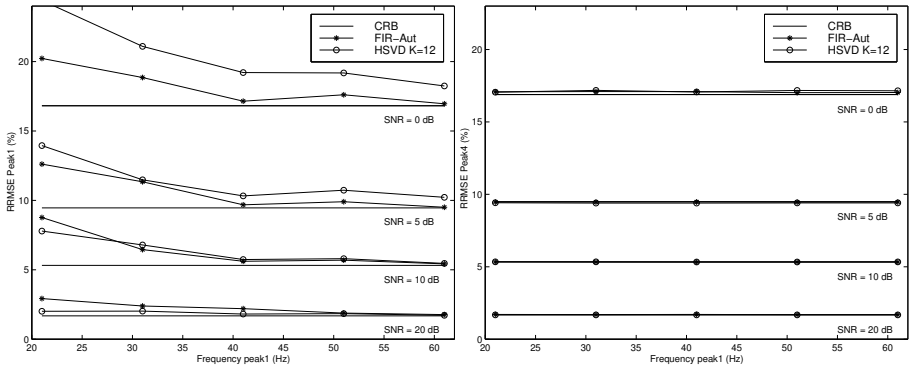


Figure 8.4: CRB and RRMSE of amplitude estimates as a function of the frequency of peak1. The results are obtained from 400 simulations runs for four different SNRs using the FIR filter method and the HSVD method. Left: peak1. Right: peak4.

8.3.4 Comparison of FIR filter, HSVD and HLR methods for water removal w.r.t. computational complexity

In this simulation example the computational complexity of the FIR filter method is compared to the HSVD and HLR methods. The comparison is performed by calculating the number of flops used in matlab when applying the water suppression techniques to the simulation signal described above for $N = 512$, 1024 and 2048 data points respectively.

The results for the FIR filter method are given in Table 8.3. The computational load is divided into three parts: the linear-phase filter design (constrained LS), the spectral factorization and the filtering operation. Spectral factorization is seen to be the most computationally intensive step. The filter design procedure can be made more efficient by using alternative minimum-phase filter design techniques that circumvent the spectral factorization (see e.g. [35] and references therein). Note also the low computational load associated with the actual filtering procedure. This implies that in applications, such as spectroscopic imaging in which the filter design only has to be performed once, the computational complexity of the FIR filter method can be decreased by an order of magnitude. The results for the HLR and HSVD methods are given in Table 8.4. Comparing the results of these methods to the total computational load of the FIR filter method shows that HLR requires 12 times more computations ($N = 1024$) than the automatic FIR filter scheme (including the filter design steps) even for a correct choice of model order ($K = 12$). When the correct model order is unknown a sufficiently high model order should be used since undermodeling could lead to loss of accuracy. The computational load

N	M	Const.LS	Spe.Fac.	Filtering	Total
512	30	0.13	0.41	0.13	0.66
	50	0.40	2.05	0.21	2.66
	70	0.76	5.00	0.29	6.06
	Aut	1.01	2.05	0.21	3.27
1024	30	0.13	0.41	0.25	0.79
	50	0.40	2.05	0.42	2.87
	70	0.76	5.00	0.58	6.35
	Aut	1.01	2.02	0.42	3.44
2048	30	0.13	0.41	0.51	1.05
	50	0.40	2.05	0.84	3.29
	70	0.76	5.01	1.16	6.93
	Aut	1.01	2.02	0.84	3.87

Table 8.3: Flops ($\times 10^6$) required by the FIR filter method for water removal. The FIR filter computations are divided into the design of the linear-phase filter by constrained LS (Const.LS), spectral factorization (Spe.Fac.) and the filtering operation (Filtering). M is the filter length and N is the number of data points.

of HLR is 80 times ($N = 1024$) higher than that of the automatic FIR filter scheme for a model order of K=30. The computational load of HSVD is 1656 ($N = 1024$) times higher than for the FIR filter method even for a low model order ($K = 12$).

In summary this simulation example shows that the FIR filter method, including the filter design step, is computationally always at least one order of magnitude more efficient than the fastest available algorithms based on the HSVD method for water removal. The difference is even higher if a high model order is required to reconstruct the water signal or when a large number of data points are used.

8.4 Conclusions

The FIR filter method is compared numerically to the HSVD method for water peak removal in a number of simulation examples. The performance of the methods is examined as a function of the metabolite frequencies and the results show that the FIR filter method using the automatic design scheme slightly

N	M	HLR	HSVD
512	10	14.45	733.0
	12	21.21	734.2
	20	58.77	741.2
	30	137.2	755.6
1024	10	33.24	5697
	12	40.95	5700
	20	120.4	5113
	30	275.6	5740
2048	10	79.16	45086
	12	92.95	45091
	20	244.1	45116
	30	544.2	45167

Table 8.4: Flops ($\times 10^6$) required by the HLR and HSVD methods for water removal. K is the model order and N is the number of data points.

outperforms the HSVD method in most cases. The use of the automatic filter design scheme leads to small estimation errors and ensures the reproducibility of the results. The HSVD method is seen to be sensitive to the choice of the model order. The FIR filter method is also seen to be computationally always at least one order of magnitude more efficient than the HLR method for water removal presented in the previous chapter. The good performance and ease of use of the FIR filter method combined with the low computational complexity motivates the use of the proposed method as an alternative to the often used HSVD method.

Chapter 9

Conclusions and new challenges

As the last pages of this text approach, it is time for a look over the shoulder but even more important for a look into the future. Section 9.1 summarizes the main conclusions and original contributions of this thesis. Section 9.2 provides some possible roads to the ultimate accurate, efficient and automatic MRS data analysis.

9.1 Conclusions

This thesis treats different aspects involved in the quantitation of MRS signals. In the following the main conclusions and original contributions are presented.

- The different methods that exist to extract the parameters from a MRS signal have been explained and connections between methods have been pointed out. Typically, time-domain methods are divided into black-box methods and interactive methods. Interactive methods fit a model function to the data and minimize a cost function. Linear prediction and state-space methods are two variants of the black-box methods which do not involve the minimization of a cost function. Algorithms which fall into this latter class of methods only approximately solve the linear prediction or state-space problem setting. It was shown here how the linear prediction and state-space problems can be solved without introducing approximations leading to ML parameter estimates. However, then again the problems reduce to the solution of a minimization problem. It was also

shown that model fitting in the frequency and time domain are equivalent from a theoretical point of view. The relation with frequency-domain techniques used in the control and identification literature was pointed out.

- AMARES (*a*dvanced *m*ethod for *a*ccurate, *r*obust and *e*fficient *s*pectral fitting) was developed and shown to outperform the popular VARPRO method in two ways. First of all an appropriate choice of the functional and the nonlinear least squares algorithm was made based on an extensive simulation study. Secondly, AMARES allows to include more prior knowledge about the signal parameters, the model function (Lorentz, Gauss and Voigt) and the type of signal (FID or echo). This results in increased accuracy and user flexibility. Overall an algorithm is obtained which outperforms VARPRO in terms of accuracy, robustness and flexibility.

- In biochemical studies MRS signals are in many cases acquired consecutively to monitor metabolic changes over time. Often information concerning the time evolution of some of the parameters in these time series is present. AMARES$_{ts}$, an extension of AMARES, that allows taking into account in a statistically optimal way the common information present in spectra of a time series was presented. The method performs very well in practical situations. AMARES$_{ts}$ leads to improved and more robust estimates than the ones obtained by processing the signals individually with AMARES since in the latter case the information present between the spectra can not be taken into account. AMARES$_{ts}$ was compared to methods that appeared previously in the literature: AMARES$_{sts}$, HTLSsum and HTLSstack. AMARES$_{sts}$ makes suboptimal use of the common information present in spectra and is restricted in the sense that the types of prior knowledge that can be imposed in a straightforward way are rather limited. In those situations where the method can be used, it performs better than AMARES applied on every signal and slightly worse than AMARES$_{ts}$. Although HTLSsum and HTLSstack do not perform as well in terms of precision as AMARES$_{ts}$ and need some more fine tuning, they perform better than HTLS applied on every signal individually and are promising black-box techniques to further automate the MRS data processing.

- Various techniques that have been used in the past to obtain estimates of parameters of selected peaks in the presence of unknown or uninteresting spectral features separated in frequency from the metabolites of interest have been investigated and compared with each other. This problem is denoted by frequency-selective (FS) parameter estimation. The properties of time-domain weighting AMARES$_w$, HSVD filtering AMARES$_H$ and fitting in the frequency domain using a polynomial baseline AMARES$_{FREQ}$ were investigated since up till now no systematic

study analyzing the final parameter accuracy of the parameters of interest had been performed. The different problems associated with each of the mentioned methods were pointed out. AMARES$_H$ is seen to be the best method of the currently used FS methods but the computational complexity of the method is rather high. This is a serious drawback in applications such as MRS imaging where large amounts of signals have to be processed.

- In the context of FS parameter estimation, AMARES$_f$, a new technique based on the use of maximum-phase FIR filters in combination with AMARES was presented and compared with the methods above. In terms of final parameter accuracy and ease of use AMARES$_f$ outperforms the others.
- In view of the high computational complexity of the HSVD method, HLR, a fast alternative based on a low-rank revealing decomposition, was developed for use in removal of residual water in proton spectra. The gain in efficiency depends on both the number of data points of the signal and the imposed model order and is more pronounced when the number of data points becomes larger and the model order smaller. For a typical signal consisting of 512 data points and a model order of 10, the method using the fast decomposition is about 35 to 40 times faster than HSVD. This (fast) HSVD-based filtering was compared to AMARES$_f$ in the context of water removal in proton spectra. It was seen that the FIR filter method outperforms the black-box based method in terms of efficiency, ease of use and accuracy.

9.2 New challenges

Although a lot of progress has been made, there are still many aspects that need further fine-tuning and/or investigation.

- The error bounds on the estimated parameters are currently computed by evaluating the Cramér-Rao bounds in the finally computed parameters. It is not clear how good these error estimates reflect the actual errors. A controlled experimental study is needed to investigate this.
- In chapter 4 it was shown that the use of AMARES$_{ts}$ can provide a substantial gain in parameter accuracy if the signals of the time series are analyzed simultaneously taking into account the information present between the signals. However, the complexity of the optimization problem and the storage requirements for classical optimization routines increases dramatically. Alternative optimization routines should be tested out in this context.

- So far AMARES$_{ts}$ has only been applied to MRS time series. The possible advantages of using the method in MRSI should also be investigated.
- In chapter 5 an overview of methods that have been developed over the years to deal with model imperfections was given. However, most evaluations of the methods have been based only on visual inspection of the spectrum after correction. A systematic study of how the different methods affect the final parameter accuracy was not carried out. A detailed study using both simulations and an extensive experimental study is needed.
- This thesis focused on FS parameter estimation. A problem which is more difficult to solve is the case in which the nuisance peaks are not separated in frequency nor in damping from the peaks of interest. A typical example is the presence of a macromolecular baseline in short-echo proton signals. Since more and more spectroscopists start measuring short-echo proton signals a good approach to get reliable parameter estimates of the metabolites of interest is urgently needed.
- Up till now the methods like VARPRO and AMARES which are able to provide accurate parameter estimates under more demanding situations are still very much interactive. With ever increasing amounts of data to be processed there is a need to automate these methods. Possible ways to achieve this are the (supervised) use of starting values provided with black-box methods, the use of a prior knowledge and starting value data base for specific applications.
- This thesis concerns the analysis of 1D experiments, i.e. the individual signals are described as a sum of Lorentzian, Gaussians, ... However, in 2D experiments, the model function comprises products of sinusoids. Many of the methods described in this thesis can be extended to the processing of such signals, opening up a new area of research.

Appendix A

Equivalence between time- and frequency domain model fitting

The DFT used in Eq. (1.8) is defined as follows

$$\begin{aligned} \overline{\mathcal{Y}}_l &= \sum_{n=0}^{N-1} \bar{y}(n) e^{\frac{-j2\pi nl}{N}}, \\ l &= 0, \ldots, N-1. \end{aligned}$$

In matrix-vector notation this becomes

$$\overline{\boldsymbol{\mathcal{Y}}} = \mathbf{E}^H \bar{\mathbf{y}}, \tag{A.1}$$

with

$$\begin{aligned} \overline{\boldsymbol{\mathcal{Y}}} &= [\overline{\mathcal{Y}}_0, \ldots, \overline{\mathcal{Y}}_{N-1}]^T \\ \bar{\mathbf{y}} &= [\bar{y}_0, \ldots, \bar{y}_{N-1}]^T \\ \mathbf{E}_{rs} &= e^{\frac{j2\pi rs}{N}}, \quad r = 0, \ldots, N-1,\ s = 0, \ldots, N-1. \end{aligned}$$

Similarly,

$$\boldsymbol{\mathcal{Y}} = \mathbf{E}^H \mathbf{y}, \tag{A.2}$$

with

$$\boldsymbol{\mathcal{Y}} = [\mathcal{Y}_0, \ldots, \mathcal{Y}_{N-1}]^T,$$

$$\mathbf{y} \quad = \quad [y_0, \ldots, y_{N-1}]^T.$$

It is easy to verify that the matrix $\mathbf{E}$ satisfies the following property $\mathbf{E}^H\mathbf{E} = \mathbf{E}\mathbf{E}^H = N\mathbf{I}_N$, where $\mathbf{I}_N$ denotes the $(N \times N)$ identity matrix. Using this property, we can rewrite Eq. (A.1) and Eq. (A.2) as

$$\begin{aligned} \bar{\mathbf{y}} &= \frac{1}{N}\mathbf{E}\overline{\boldsymbol{\mathcal{Y}}} \\ \mathbf{y} &= \frac{1}{N}\mathbf{E}\boldsymbol{\mathcal{Y}}. \end{aligned}$$

Eq. (2.3) can then be rewritten as follows

$$\begin{aligned} \sum_{n=0}^{N-1} |y(n) - \bar{y}(n)|^2 &= (\mathbf{y} - \bar{\mathbf{y}})^H(\mathbf{y} - \bar{\mathbf{y}}) \\ &= \frac{1}{N^2}(\mathbf{E}\boldsymbol{\mathcal{Y}} - \mathbf{E}\overline{\boldsymbol{\mathcal{Y}}})^H(\mathbf{E}\boldsymbol{\mathcal{Y}} - \mathbf{E}\overline{\boldsymbol{\mathcal{Y}}}) \\ &= \frac{1}{N^2}(\boldsymbol{\mathcal{Y}} - \overline{\boldsymbol{\mathcal{Y}}})^H\mathbf{E}^H\mathbf{E}(\boldsymbol{\mathcal{Y}} - \overline{\boldsymbol{\mathcal{Y}}}) \\ &= \frac{1}{N}(\boldsymbol{\mathcal{Y}} - \overline{\boldsymbol{\mathcal{Y}}})^H(\boldsymbol{\mathcal{Y}} - \overline{\boldsymbol{\mathcal{Y}}}) \\ &= \sum_{l=0}^{N-1} \frac{1}{N}|\mathcal{Y}(\nu_l) - \overline{\mathcal{Y}}(\nu_l)|^2. \end{aligned}$$

The last equation is equal to Eq. (2.40) up to a scaling constant N, thereby proving the equivalence of the two criteria.

Appendix B

Publication list

B.1 Articles in international journals

1. L. Vanhamme, A. van den Boogaart, and S. Van Huffel. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J. Magn. Reson.*, 129:35–43, 1997.

2. Š. Mierisová, A. van den Boogaart, I. Tkáč, P. Van Hecke, L. Vanhamme, and T. Liptaj. New approach for quantitation of short echo time *in vivo* ^{1}H MR spectra of brain using AMARES. *NMR Biomed*, 11:32–39, 1998.

3. L. Vanhamme, R.D. Fierro, S. Van Huffel, and R. de Beer. Fast removal of residual water in proton spectra. *J. Magn. Reson.*, 132:197–203, 1998.

4. T. Sundin, L. Vanhamme, P. Van Hecke, I. Dologlou, and S. Van Huffel. Accurate quantification of ^{1}H spectra: from finite impulse response filter design for solvent suppression to parameter estimation. *J. Magn. Reson.*, 139:189–204, 1999.

5. L. Vanhamme, S. Van Huffel, P. Van Hecke, and D. van Ormondt. Time-domain quantification of series of biomedical magnetic resonance spectroscopy signals. *J. Magn. Reson.*, 140:120–130, 1999.

6. L. Vanhamme, T. Sundin. Overview of time-domain quantitation methods. Invited paper, to appear in *NMR Biomed.*.

7. L. Vanhamme, T. Sundin, P. Van Hecke, S. Van Huffel, and R. Pintelon. Frequency-selective quantification of biomedical magnetic resonance spectroscopy data. *Accepted for publication in J. Magn. Reson. (also ESAT-SISTA Report TR 98-121, ESAT Laboratory, K.U. Leuven, Belgium)*, October 1999.

B.2 Articles in conference proceedings

1. S. Van Huffel, P. Lemmerling, and L. Vanhamme. Fast algorithms for signal subspace fitting with Toeplitz matrices and applications to exponential data modeling. In *SVD and Signal Processing III: Algorithms, Architectures and Applications* (M. Moonen, B. De Moor, Eds.), pages 191–198, Elsevier Science Publisher BV (North-Holland), 1995.

2. L. Vanhamme, A. van den Boogaart, and S. Van Huffel. Fast and accurate parameter estimation of noisy complex exponentials with use of prior knowledge. In *Proc. EUSIPCO-96*, pages 407–410, Trieste, Italy, September 10-September 13 1996.

3. A. van den Boogaart, S. Cavassila, L. Vanhamme, J. Totz, P. Van Hecke. A complete software package for MR signal processing. In *Proceedings 18th Annual Intern. Conf. IEEE Eng. in Medicine and Biology Society*, paper number 405, Amsterdam, The Netherlands, October 31-November 3 1996.

4. R.D. Fierro, L. Vanhamme, and S. Van Huffel. Total least squares algorithms based on rank-revealing complete orthogonal decompositions. In *Recent advances in Total Least Squares techniques and Errors-in-Variables Modeling* (S. Van Huffel, ed.), pages 99–116, Siam, 1997.

5. L. Vanhamme and S. Van Huffel. Multichannel quantification of biomedical magnetic resonance spectroscopic signals. In *Proceedings SPIE conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations VIII*, pages 237–248, San Diego, USA, July 22- July 24 1998.

B.3 Book chapters

1. L. Vanhamme. Prior knowledge for AMARES: Inputnv and AMARES parameter estimation. In A. van den Boogaart, *Spectral Analysis with the Magnetic Resonance User Interface MRUI*. Delft University Press, 1997.

B.4 Abstracts international conferences

1. S. Van Huffel, L. Vanhamme, and A. van den Boogaart. Fast and accurate exponential data fitting with use of prior knowledge. In *Abstracts Householder Symposium on Numerical Linear Algebra XIII*, Pontresina, Switzerland, June 17 - June 21 1996.

2. L. Vanhamme, S. Van Huffel, and A. van den Boogaart. Improved fitting method for parameter estimation of MRS signals with use of prior knowledge. In *Proceedings Fifth Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM 97)*, page 1414, Vancouver, Canada, April 12 - April 18 1997.

3. L. Vanhamme, S. Van Huffel, and R.D. Fierro. Fast low rank-revealing ULV algorithms for Toeplitz matrix approximations with applications in Magnetic Resonance Spectroscopy. In *Abstracts Siam's 45th Anniversary Meeting*, Stanford University, California, USA, July 14 - July 18 1997.

4. A.J. van den Bergh, L. Vanhamme, S. Van Huffel, and A. Heerschap. Improving the SNR of glycogen detection in ^{13}C MRS by multiexponential signal fitting using prior knowledge. In *Abstracts 14th Annual Meeting European Society for Magnetic Resonance in Medicine and Biology (ESMRMB 97)*, page 190, Brussels, Belgium, September 18 - September 21 1997.

5. Š. Mierisová, A. van den Boogaart, I. Tkáč, P. Van Hecke, L. Vanhamme, V. Mlynárik, and T. Liptaj. Construction of prior knowledge for quantitation of *in vivo* short echo time ^{1}H MR spectra of brain. In *Abstracts 14th Annual Meeting European Society for Magnetic Resonance in Medicine and Biology (ESMRMB 97)*, page 190, Brussels, Belgium, September 18 - September 21 1997.

6. P. Van Hecke, L. Vanhamme, P. Malyszewska and S. Van Huffel. Time domain removal of the water signal in biomedical proton NMR spectra: Accuracy aspects. In *Abstracts joint 29th AMPERE-13th ISMAR International conference*, Berlin, Magnetic Resonance and Related Phenomena, Vol. II, eds. D. Ziessow, W. Lubitz, and F. Lendzian, pages 780-781, August 2 - August 7 1998.

7. S. Van Huffel, L. Lagae, L. Vanhamme, P. Van Hecke and D. van Ormondt. Improving blackbox MRS data quantification by means of prior knowledge. In *Abstracts 15th Annual Meeting European Society for Magnetic Resonance in Medicine and Biology (ESMRMB 98)*, page 13, Geneva, Switzerland, September 17 -September 20 1998.

8. L. Vanhamme, P. Van Hecke, P. Malyszewska, S. Van Huffel, and D. van Ormondt. Time domain removal of the water signal in biomedical proton NMR spectra. In *Abstracts 15th Annual Meeting European Society for Magnetic Resonance in Medicine and Biology (ESMRMB 98)*, page 210, Geneva, Switzerland, September 17 -September 20 1998.

9. L. Vanhamme, S. Van Huffel, and P. Van Hecke. Maximum likelihood fitting of time series using an extension of AMARES. In *Abstracts 15th Annual Meeting European Society for Magnetic Resonance in Medicine*

and Biology (ESMRMB 98), page 210, Geneva, Switzerland, September 17 -September 20 1998.

10. L. Vanhamme S. Van Huffel, and P. Van Hecke. Extensions of AMARES to quantitate series of biomedial MRS signals. In *Proceedings Seventh Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM 99)*, page 1558, Philadelphia, USA, May 22 - May 28 1999.

11. T. Sundin, L. Vanhamme, P. Van Hecke, and S. Van Huffel. ^{1}H MRS quantitation using FIR filters for solvent suppression. In *Proceedings Seventh Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM 99)*, page 588, Philadelphia, USA, May 22 - May 28 1999.

12. T. Sundin, L. Vanhamme, P. Van Hecke, S. Van Huffel. FIR-filter based frequency selective quantitation of biomedical MRS signals. In *Proceedings Seventh Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM 99)*, page 1557, Philadelphia, USA, May 22 - May 28 1999.

Bibliography

[1] T.J. Abatzoglou and J.M. Mendel. Constrained total least squares. In *Proc. of IEEE International Conf. on Acoustics, Speech & Signal Processing*, pages 1485–1488, Dallas, 1987.

[2] T.J. Abatzoglou, J.M. Mendel, and G.A. Harada. The constrained total least squares technique and its applications to harmonic superresolution. *IEEE Trans. Signal Processing*, 39:1070–1086, 1991.

[3] F. Abildgaard, H. Gesmar, and J.J. Led. Quantitative analysis of complicated nonideal fourier transform NMR spectra. *J. Magn. Reson.*, 79:78–89, 1988.

[4] M. Andrec and J.H. Prestegard. A Metropolis Monte Carlo implementation of Bayesian time-domain parameter estimation: application to coupling constant estimation from antiphase multiplets. *J. Magn. Reson.*, 130:217–232, 1998.

[5] D. Barache, J-P. Antoine, and J-M. Dereppe. The continuous wavelet transform, an analysis tool for NMR spectroscopy. *J. Magn. Reson.*, 128:1–11, 1997.

[6] H. Barkhuysen, R. de Beer, W.M.M.J. Bovée, and D. van Ormondt. Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure. *J. Magn. Reson.*, 61:465–481, 1985.

[7] H. Barkhuysen, R. de Beer, and D. van Ormondt. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *J. Magn. Reson.*, 73:553–557, 1987.

[8] P. Barone, L. Guidoni, R. Ragona, V. Viti, E. Furman, and H. Degani. Modified Prony method to resolve and quantify in vivo ^{31}P NMR spectra of tumors. *J. Magn. Reson.*, Series B, 105:137–146, 1994.

[9] R. Bartha, D.J. Drost, and P.C. Williamson. Factors affecting the quantification of short echo in-vivo ^{1}H MR spectra: prior knowledge, peak elimination, and filtering. *NMR Biomed.*, 12:205–216, 1999.

[10] K.L. Behar, D.L. Rothman, D.D. Spencer, and O.A.C. Petroff. Analysis of macromolecule resonances in ^{1}H NMR spectra of human brain. *Magn. Reson. Med.*, 32:294–302, 1994.

[11] Z. Bi, A.P. Bruner, J. Li, K.N. Scott, Z.-S. Liu, C.B. Stopka, H.-W. Kim, and D.C. Wilson. Spectral fitting of NMR spectra using an alternating optimization method with a priori knowledge. *J. Magn. Reson.*, 140:108–119, 1999.

[12] W.M.M.J. Bovée. Longitudinal and transverse NMR relaxation in a proton AB system. *J. Magn. Reson.*, 24:327, 1976.

[13] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34:1081–1089, 1986.

[14] G.L. Bretthorst. Bayesian analysis I. Parameter estimation using quadrature NMR models. *J. Magn. Reson.*, 88:533–551, 1990.

[15] G.L. Bretthorst. Bayesian analysis II. Signal detection and model selection. *J. Magn. Reson.*, 88:552–570, 1990.

[16] G.L. Bretthorst. Bayesian analysis III. Applications to NMR signal detection, model selection, and parameter estimation. *J. Magn. Reson.*, 88:571–595, 1990.

[17] G.L. Bretthorst. Bayesian analysis IV. Noise and computing time considerations. *J. Magn. Reson.*, 93:369–394, 1991.

[18] G.L. Bretthorst. Bayesian analysis V. Amplitude estimation for multiple well-separated sinusoids. *J. Magn. Reson.*, 98:501–523, 1992.

[19] K. Bruynseels, N. Gillis, P. Van Hecke, W. Stalmans, and F. Vanstapel. Metabolic effects of the fructose analogue 2,5-anhydro-d-mannitol in perfused rat liver. In *Proc. of the Society of Magnetic Resonance, Third Scientific Meeting*, page 1658, Nice, France, August 1995.

[20] S. Cavassila, S. Deval, C. Huegen, D. van Ormondt, and D. Graveron-Demilly. The beneficial influence of prior knowledge on the quantitation of in vivo magnetic resonance spectroscopy signals. *Investigative Radiology*, 34(3):242–246, 1999.

[21] S. Cavassila, B. Fenet, A. van den Boogaart, C. Remy, A. Briguet, and D. Graveron-Demilly. ER-Filter: a preprocessing technique for frequency-selective time-domain analysis. *J. Magn. Reson. Anal.*, 3:87–92, 1997.

[22] H. Chen. *Subspace-based parameter estimation of exponentially damped sinusoids with application to nuclear magnetic resonance spectroscopy data.* PhD thesis, Dept. of Electr. Eng., Katholieke Universiteit Leuven, May 1996.

[23] H. Chen, S. Van Huffel, C. Decanniere, and P. Van Hecke. A signal enhancement algorithm for time-domain data MR quantification. *J. Magn. Reson.*, Series A, 109:46–55, 1994.

[24] H. Chen, S. Van Huffel, A.J.W. Van den Boom, and P.P.J. Van den Bosch. Extended HTLS methods for parameter estimation of multiple data sets modeled as sums of exponentials. In *Proc. 13th Int. Conf. on Digital Signal Processing (DSP97)*, pages 1035–1038, Santorini, Greece, July 1997.

[25] H. Chen, S. Van Huffel, A.J.W. Van den Boom, and P.P.J. Van den Bosch. Subspace-based parameter estimation of exponentially damped sinosoids using prior knowledge of frequency and phase. *Sign. Proc.*, 59:129–136, 1997.

[26] H. Chen, S. Van Huffel, D. van Ormondt, and R. de Beer. Parameter estimation with prior knowledge of known signal poles for the quantification of NMR spectroscopy data in the time domain. *J. Magn. Reson.*, Series A, 119:225–234, 1996.

[27] H. Chen, S. Van Huffel, and J. Vandewalle. Improved methods for exponential parameter estimation in the presence of known poles and noise. *IEEE Trans. Signal Process.*, 45(5):1390–1393, 1997.

[28] S.C. Chen, T.J. Schaewe, R.S. Teichman, M.I. Miller, S.N. Nadel, and A.S. Greene. Parallel algorithms for maximum-likelihood nuclear magnetic resonance spectroscopy. *J. Magn. Reson.*, Series A, 102:16–23, 1993.

[29] A.R. Conn, N. Gould, and Ph.L. Toint. *LANCELOT, A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*. Springer Verlag, Series on Computational Mathematics 17, Berlin, 1992.

[30] R.J.T. Corbett. How to perform automated curve fitting to *in vivo* ^{31}P magnetic resonance spectroscopic data. *MAGMA*, 1:65–76, 1993.

[31] C.J. Craven and J.P. Waltho. The action of time-domain convolution filters for solvent suppression. *J. Magn. Reson.*, Series B, 106:40–46, 1995.

[32] R. de Beer, F. Michiels, D. van Ormondt, B.P.O. van Tongeren, P.R. Luyten, and H. van Vroonhoven. Reduced lipid contamination in in vivo ^{1}H MRSI using time-domain fitting and neural network classification. *Magnetic Resonance Imaging*, 11:1019–1026, 1993.

[33] R. de Beer and D. van Ormondt. *NMR Basic Principles and Progress*, volume 26, chapter Analysis of NMR data using time-domain fitting procedures, pages 201–248. Springer-Verlag, Berlin/Heidelberg, 1992.

[34] A.A. de Graaf, J.E. van Dijk, and W.M.M.J. Bovée. QUALITY: Quantification improvement by converting lineshapes to the Lorentzian type. *Magn. Reson. Med.*, 13:343–357, 1990.

[35] R.A. de Graaf. *In vivo NMR spectroscopy.* Wiley, England, 1998.

[36] B. De Moor. Structured total least squares and L_2 approximation problems. *Linear Algebra and its Applications, Special Issue on Numerical Linear Algebra Methods in Control, Signals and Systems, (Van Dooren, Ammar, Nichols, and Mehrmann, eds.)*, 188-189:163–207, July 1993.

[37] B. De Moor. Total least squares for affinely structured matrices and the noisy realization problem. *IEEE Trans. Signal Processing*, 42:3004–3113, November 1994.

[38] C. Decanniere, P. Van Hecke, F. Vanstapel, H. Chen, S. Van Huffel, C. van der Voort, B.P.O. van Tongeren, and D. van Ormondt. Evaluation of signal processing methods for the quantification of strongly overlapping peaks in ^{31}P NMR spectra. *J. Magn. Reson.*, Series B, 105:31–37, 1994.

[39] G. Della Lunga and R. Basosi. A simple method for baseline correction in EPR spectroscopy. 2. The use of cubic spline functions. *J. Magn. Reson.*, Series A, 112:102–105, 1995.

[40] G. Della Lunga, R. Pogni, and R. Basosi. A simple method for baseline correction in EPR spectroscopy. *J. Magn. Reson.*, Series A, 108:65–70, 1994.

[41] J.E. Dennis, D.M. Gay, and R.E. Welsch. An adaptive nonlinear least-squares algorithm. *ACM TOMS*, 7(3):348–368, 1981.

[42] J.E. Dennis, D.M. Gay, and R.E. Welsch. Algorithm 573 NL2SOL - an adaptive nonlinear least-squares algorithm. *ACM TOMS*, 7(3):369–383, 1981.

[43] J.E. Dennis and R.B. Schnabel. *Numerical methods for unconstrained optimisation and nonlinear equations.* Prentice Hall, Englewood Cliffs, 1983.

[44] M. Deriche and X. Hu. Elimination of water signal by postprocessing. *J. Magn. Reson.*, Series A, 101:229–232, 1993.

[45] F.S. DiGennaro and D. Cowburn. Parametric estimation of time-domain NMR signals using simulated annealing. *J. Magn. Reson.*, 96:582–588, 1992.

[46] A. Diop, A. Briguet, and D. Graveron-Demilly. Automatic in vivo NMR data processing based on an enhancement procedure and linear prediction method. *Magn. Reson. Med.*, 27:318–328, 1992.

[47] A. Diop, W. Kölbel, D. Michel, A. Briguet, and D. Graveron-Demilly. Full automation of quantitation of in vivo NMR by LPSVD(CR) and EPLPSVD. *J. Magn. Reson.*, Series B, 103:217–221, 1994.

[48] I. Dologlou, J.-C. Pesquet, and J. Skowronski. Projection based rank reduction algorithms for multichannel modelling and image compression. *Signal Processing*, 48:97–109, 1996.

[49] I. Dologlou, S. Van Huffel, and D. van Ormondt. Frequency-selective MRS data quantification with frequency prior knowledge. *J. Magn. Reson.*, 130:238–243, 1998.

[50] R.D. Fierro and P.C. Hansen. Low-rank revealing UTV decompositions. *Numerical Algorithms*, 15:37–55, 1997.

[51] R.D. Fierro, L. Vanhamme, and S. Van Huffel. Total least squares algorithms based on rank-revealing complete orthogonal decompositions. In S. Van Huffel, editor, *Recent advances in Total Least Squares techniques and Errors-in-Variables Modeling*. Siam, 1997.

[52] R. Fletcher. *Practical methods of optimization*. Wiley, 1987.

[53] ftp: ftp.netlib.org.

[54] D.G. Gadian. *NMR and its applications to living systems*. Oxford Science publishers, second edition, 1995.

[55] D.M. Gay and L. Kaufman. Tradeoffs in algorithms for separable nonlinear least squares. In R. Vichnevetsky and J.J.H. Miller, editors, *Proc. of the 13th World Congress on Computational and Applied Mathematics, IMACS '91 (also AT&T Numerical Analysis Manuscript 90-11)*, pages 157–158. Criterion Press, Dublin, 1991.

[56] H. Gesmar, J.J. Led, and F. Abildgaard. Improved methods for quantitative spectral analysis of NMR data. *Prog. in NMR Spectrosc.*, 22:255–288, 1990.

[57] A. Gibbs and G.A. Morris. Reference deconvolution. Elimination of distortions arising from reference line truncation. *J. Magn. Reson.*, 91:77–83, 1991.

[58] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, San Diego, 1981.

[59] G.H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10(2):413–432, April 1973.

[60] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Univerity Press, 1993.

[61] J.D. Gorman and A.O. Hero. Lower bounds for parametric estimation with constraints. *IEEE Trans. Inform. Theory*, 26:1285–1301, November 1990.

[62] J.-P. Grivet. Accurate numerical approximation to the Gauss-Lorentz lineshape. *J. Magn. Reson.*, 125:102–106, 1997.

[63] O. Herrmann and H. Schüßler. Design of nonrecursive digital filters with minimum phase. *Electronics Letters*, 6:329–330, 1970.

[64] Y. Hiltunen, M. Ala-Korpela, J. Jokisaari, S. Eskelinen, K. Kiviniitty, M. Savolainen, and Y. Kesäniemi. A lineshape fitting model for ^{1}H NMR spectra of human blood plasma. *Magn. Reson. Med.*, 21:222–232, 1991.

[65] Y. Hiltunen, E. Heiniemi, and M. Ala-Korpela. Lipoprotein-lipid quantification by neural-network analysis of ^{1}H NMR data from human blood plasma. *J. Magn. Reson.*, Series B, 106:191–194, 1995.

[66] L. Hofmann, J. Slotboom, C. Boesch, and R. Kreis. Model fitting of ^{1}H-MR spectra of the human brain: incorporation of short-T_1 components and evaluation of parametrized vs. non-parametrized models. In *Proc. 7th Scientific Meeting ISMRM*, page 586, Philadelphia, USA, May 1999.

[67] Y. Hua and T.K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-38:814–824, 1990.

[68] J. Kaartinen, Š. Mierisová, J.M.E. Oja, J.-P. Usenius, R.A. Kauppinen, and Y. Hiltunen. Automated quantification of human brain metabolites by artificial neural network analysis from in vivo single-voxel ^{1}H NMR spectra. *J. Magn. Reson.*, 134:176–179, 1998.

[69] L. Kaufman. A variable projection method for solving seperable nonlinear least squares problems. *BIT*, 15:49–57, 1975.

[70] U. Klose. In vivo proton spectroscopy in presence of eddy currents. *Magn. Reson. Med.*, 14:26–30, 1990.

[71] A. Knijn, R. de Beer, and D. van Ormondt. Frequency-selective quantification in the time domain. *J. Magn. Reson.*, 97:444–450, 1992.

[72] P. Koehl. Linear prediction spectral analysis of NMR data. *Prog. in NMR Spectrosc.*, 34:257–299, 1999.

[73] P. Koehl, C. Ling, and J.F. Lefèvre. Automatic phase correction of NMR spectra: statistics and limits. *J Chim Phys*, 92:1929–1938, 1995.

[74] W. Kölbel and H. Schäfer. Improvement and automation of the LPSVD algorithm by continuous regularization of the singular values. *J. Magn. Reson.*, 100:598–603, 1992.

[75] R. Kumaresan and D.W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Accoustics, Speech and Signal Processing*, ASSP 30(6):833–840, December 1982.

[76] R. Kumeresan, C.S. Ramalingam, and D. van Ormondt. Estimating the parameters of NMR signals by transforming to the frequency domain. *J. Magn. Reson.*, 89:562–567, 1990.

[77] 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. Am.*, 73(12):1799–1811, 1983.

[78] Y. Kuroda, A. Wada, T. Yamazaki, and K. Nagayama. Postacquistion data processing method for suppression of the solvent signal. *J. Magn. Reson.*, 84:604–610, 1989.

[79] J.H.J. Leclerc. Distortation-free suppression of the residual water peak in proton spectra by postprocessing. *J. Magn. Reson.*, Series B, 103:64–67, 1994.

[80] P. Lemmerling. *Structured Total Least Squares: analysis, algorithms and applications.* PhD thesis, Dept. of Electr. Eng., Katholieke Universiteit Leuven, May 1999.

[81] Y.-Y. Lin, P. Hodgkinson, M. Ernst, and A. Pines. A novel detection-estimation scheme for noisy NMR signals: Applications to delayed acquisition data. *J. Magn. Reson.*, 128:30–41, 1997.

[82] Z.-S. Liu, J. Li, and P. Stoica. RELAX-based estimation of damped sinusoidal signal parameters. *Signal Processing*, 62(3):311–321, 1997.

[83] D. Marion, M. Ikura, and A. Bax. Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J. Magn. Reson.*, 84:425–430, 1989.

[84] I. Marshall, J. Higinbotham, S. Bruce, and A. Freise. Use of Voigt lineshape for quantification of in vivo ^{1}H spectra. *Magn. Reson. Med.*, 37:651–657, 1997.

[85] Y.-L. Martin. A global approach to accurate and automatic quantitative analysis of NMR spectra by complex least-squares curve fitting. *J. Magn. Reson.*, Series A, 111:1–10, 1994.

[86] A.A. Maudsley. Spectral lineshape determination by self-deconvolution. *J. Magn. Reson.*, Series B, 106:47–57, 1995.

[87] A.A. Maudsley, Z. Wu, D.J. Meyerhoff, and M.W. Weiner. Automated processing for proton spectroscopic imaging using water reference deconvolution. *Magn. Reson. Med.*, 31:589–595, 1994.

[88] G.J. Metzger, M. Patel, and X. Hu. Application of genetic algorithms to spectral quantification. *J. Magn. Reson.*, Series B, 110:316–320, 1996.

[89] R.A. Meyer, M.J. Fisher, S.J. Nelson, and T.R. Brown. Evaluation of manual methods for integration of in vivo phosphorus NMR spectra. *NMR Biomed.*, 1(3):131–135, 1988.

[90] Š. Mierisová, A. van den Boogaart, I. Tkáč, P. Van Hecke, L. Vanhamme, and T. Liptaj. New approach for quantitation of short echo time *in vivo* ^{1}H MR spectra of brain using AMARES. *NMR Biomed.*, 11:32–39, 1998.

[91] M.I. Miller and A.S. Greene. Maximum-likelihood estimation for nuclear magnetic resonance spectroscopy. *J. Magn. Reson.*, 83:525–548, 1989.

[92] J.J. Moré. The Levenberg-Marquardt algorithm: implementation and theory. In G.A. Watson, editor, *Numerical Analysis, Lecture notes in mathematics 630*, pages 105–116. Springer-Verlag, Berlin, 1978.

[93] J.J. Moré, B.S. Garbow, and K.E. Hillstrom. User guide for minpack-1. Technical Report Argonne National Laboratory Report ANL-80-74, Argonne, IL, 1980.

[94] G.A. Morris. Compensation of instrumental imperfections by deconvolution using an internal reference signal. *J. Magn. Reson.*, 80:547–552, 1988.

[95] R.V. Mulkern and J.L. Bowers. Calculating spectral modulations of AB systems during PRESS acquisitions. *Magn. Reson. Med.*, 30:518–519, 1993.

[96] S.J. Nelson and T.R. Brown. The accuracy of quantification from 1D NMR spectra using the PIQABLE algorithm. *J. Magn. Reson.*, 84:95–109, 1989.

[97] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer-Verlag, New York, 1999.

[98] A.V. Oppenheim and R.W. Schafer. *Discrete-time signal processing.* Prentice Hall, Eaglewood Cliffs, New Jersey, 1989.

[99] R. Ordidge and I. Cresshull. The correction of transient B_0 field shifts following the application of pulsed gradients by phase correction in the time domain. *J. Magn. Reson.*, 69:151–155, 1986.

[100] M.R. Osborne. Some aspects of non-linear least squares calculations. In F.A. Lootsma, editor, *Numerical methods for non-linear optimization.* Academic Press, London, 1972.

[101] W.W.F. Pijnappel, A. van den Boogaart, R. de Beer, and D. van Ormondt. SVD-based quantification of magnetic resonance signals. *J. Magn. Reson.*, 97:122–134, 1992.

[102] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van hamme. Parametric identification of transfer functions in the frequency domain - a survey. *IEEE Trans. Automat. Contr.*, 39:2245–2260, 1994.

[103] F. Podo, W.M.M.J. Bovée, J.D. de Certaines, O. Henriksen, M.O. Leach, and D. Leibfritz, editors. *Eurospin annual 1994*. Istituto superiore di sanità, 1994.

[104] J.G. Proakis and D.G. Manolakis. *Digital signal processing: principles, algorithms and applications*. Prentice Hall, Uppers Saddle River, New Jersey, 1989.

[105] S. Provencher. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn. Res. Med.*, 30:672–679, 1993.

[106] B.D. Rao. Relationship between matrix pencil and state space based harmonic retrieval methods. *IEEE Trans. Acoust., Speech, Signal Processing*, 38(1):177–179, 1990.

[107] C. Raphael. Analysis of phosphorus magnetic resonance spectra using hidden Markov models. *J. Magn. Reson.*, Series B, 104:256–265, 1994.

[108] J. Raz, T. Chenevert, and E.J. Fernandez. A flexible spline model of the spin echo with applications to estimation of the spin-spin relaxation time. *J. Magn. Reson.*, Series A, 111:137–149, 1994.

[109] J.B. Rosen, H. Park, and J. Glick. Total least norm formulation and solution for structured problems. *SIAM Journal on Matrix Anal. Appl.*, 17(1):110–128, 1996.

[110] A. Ruhe and P.A. Wedin. Algorithms for separable nonlinear least squares problems. *Siam Review*, 22(3):318–336, 1980.

[111] K. Sekihara and N. Ohyama. Parameter estimation for in vivo magnetic resonance spectroscopy (MRS) using simulated annealing. *Magn. Reson. Med.*, 13:332–339, 1990.

[112] I.W. Selesnick, M. Lang, and C.S. Burrus. Constrained least square design of FIR filters without specified transition bands. *IEEE Trans. Signal Processing*, 44(8):1879–1892, 1996.

[113] H. Serrai, L. Senhadji, J.D. de Certaines, and J.L. Coatrieux. Time-domain quantification of amplitude, chemical shift, apparant relaxation time T_2^*, and phase by wavelet-transform analysis. Application to biomedical magnetic resonance spectroscopy. *J. Magn. Reson.*, 124:20–34, 1997.

[114] J. Slotboom, C. Boesch, and R. Kreis. Versatile frequency domain fitting using time domain models and prior knowledge. *Magn. Reson. Med.*, 39:899–911, 1998.

[115] J. Slotboom, L. Hofmann, C. Boesch, and R. Kreis. Two-dimensional fitting with prior knowledge constraints: the solution for glutamate/glutamine quantitation at 1.5T? In *Proc., ISMRM*, page 1561, 7th Scientific Meeting, Philadelphia, USA, 22-28 May 1999.

[116] G. Sobering, M. Kienlin, C. Moonen, P. van Zijl, and A. Bizzi. Post-acquisition reduction of water signals in proton spectroscopic imaging of the brain. In *Proc. SMRM, 10th Annual Meeting*, volume 1, page 771, San Francisco, 1991.

[117] D. Spielman, P. Webb, and A. Macovski. Water referencing for spectroscopic imaging . *Magn. Reson. Med.*, 12:38–49, 1989.

[118] D.S. Stephenson. Linear prediction and maximum entropy methods in NMR spectroscopy. *Prog. in NMR Spectrosc.*, 20:515–626, 1988.

[119] P. Stoica and R. Moses. *Introduction to spectral analysis.* Prentice Hall, Uppers Saddle River, New Jersey, 1997.

[120] M. Stubbs, A. van den Boogaart, C.L. Bashford, P.M.C. Miranda, L.M. Rodrigues, F.A. Howe, and J.R. Griffiths. ^{31}P magnetic resonance spectroscopy studies of nucleated and non-nucleated erythrocytes; time domain analysis (VARPRO) incorporating prior knowledge can give information on the binding of ADP. *Biochim. Biophys. Acta*, 1291:143–148, 1996.

[121] T. Sundin, L. Vanhamme, P. Van Hecke, I. Dologlou, and S. Van Huffel. Accurate quantification of ^{1}H spectra: from finite impulse response filter design for solvent suppression to parameter estimation. *J. Magn. Reson.*, 139:189–204, 1999.

[122] J. Tang and J.R. Norris. LP-ZOOM, a linear prediction method for local spectral analysis of NMR signals. *J. Magn. Reson.*, 79:190–196, 1988.

[123] R.B. Thompson and P.S. Allen. Quantification of the coupled ^{1} H metabolites using PRESS - numerical modeling and basis function calculation. In *Proc., ISMRM*, page 1569, 7th Scientific Meeting, Philadelphia, USA, May 1999.

[124] R.B. Thompson and P.S. Allen. The response of coupled spins to the STEAM sequence - improving quantification. In *Proc. 7th Scientific Meeting ISMRM*, page 587, Philadelphia, USA, May 1999.

[125] R.B. Thompson and P.S. Allen. Sources of variability in the response of coupled spins to the PRESS sequence and their potential impact on metabolite quantification. *Magn. Reson. Med.*, 41:1162–1169, 1999.

[126] C.F. Tirendi and J.F. Martin. Quantitative analysis of NMR spectra by linear prediction and total least squares. *J. Magn. Reson.*, 85:162–169, 1989.

[127] Ph.L. Toint. On large-scale nonlinear least squares calculations. *SIAM J. Sci. Statist. Comput.*, 8:416–435, 1987.

[128] Ph.L. Toint. VE10AD: A routine for large-scale nonlinear least squares. *Harwell Subroutine Library*, 1987.

[129] S. Umesh and D.W. Tufts. Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to NMR spectroscopy data. *IEEE Trans. Signal Processing*, 44(9):2245–2259, 1996.

[130] A.J. van den Bergh. *In vivo ^{13}C MR spectroscopy for human investigations.* PhD thesis, Dept. of Radiology, Katholieke Universiteit Nijmegen, April 1999.

[131] A. van den Boogaart. *The use of signal processing techniques to obtain biochemically relevant parameters from magnetic resonance data sets.* PhD thesis, CRC Biomedical Magnetic Resonance Research Group, Division of Biochemistry, Dept. Cellular and Molecular Sciences, University of London at St. George's Hospital Medical School, June 1995.

[132] A. van den Boogaart. *MRUI manual v96.3: A user's guide to the Magnetic Resonance User Interface software package.* Delft University Press, 1997.

[133] A. van den Boogaart, M. Ala-Korpela, J. Jokisaari, and J.R. Griffiths. Time and frequency domain analysis of NMR data compared: an application to 1D ^{1}H spectra of lipoproteins. *Magn. Reson. Med.*, 31:347–358, 1994.

[134] A. van den Boogaart, F.A. Howe, L.M. Rodrigues, M. Stubbs, and J.R. Griffiths. In vivo ^{31}P MRS: absolute concentrations, signal-to-noise and prior knowledge. *NMR Biomed.*, 8:87–93, 1995.

[135] A. van den Boogaart, D. van Ormondt, W.W.F. Pijnappel, R. de Beer, and M. Ala-Korpela. Removal of the water resonance from ^{1}H magnetic resonance spectra. In J.G. McWhirter, editor, *Mathematics in Signal Processing III*, pages 175–195. Clarendon Press, Oxford, 1994.

[136] M. van der Graaf, G.J. Jager, and A. Heerschap. Removal of the outer lines of the citrate multiplet in proton MR spectra of the prostatic gland by accurate timing of a PRESS pulse sequence. *MAGMA*, 5:65–69, 1997.

[137] J.W.C. van der Veen, R. de Beer, P.R. Luyten, and D. van Ormondt. Accurate quantification of in vivo ^{31}P NMR signals using the variable projection method and prior knowledge. *Magn. Res. Med.*, 6:92–98, 1988.

[138] S. Van Huffel, H. Chen, C. Decanniere, and P. Van Hecke. Algorithm for time-domain NMR data fitting based on total least squares. *J. Magn. Reson.*, Series A, 110:228–237, 1994.

[139] S. Van Huffel, H. Park, and J. B. Rosen. Formulation and solution of structured total least norm problems for parameter estimation. *IEEE Trans. Signal Processing*, 44(10):2464–2474, 1996.

[140] S. Van Huffel and D. van Ormondt. Subspace-based exponential data modeling using prior knowledge. In *Proc. of the IEEE Benelux Chapter Signal Processing Symposium (IEEEBSPS)*, pages 211–214, Leuven, Belgium, March 1998.

[141] C.F. Van Loan. *Computational Frameworks for the Fast Fourier Transform.* SIAM, 1992.

[142] D. van Ormondt, R. de Beer, A.J.H. Mariën, J.A. den Hollander, P.R. Luyten, and J.W.A.H. Vermeulen. 2D approach to quantitation of inversion-recovery data. *J. Magn. Reson.*, 88:652–659, 1990.

[143] L. Vanhamme. Prior knowledge for AMARES: Inputnv and AMARES parameter estimation. In A. van den Boogaart, editor, *Spectral Analysis with the Magnetic Resonance User Interface MRUI.* Delft University Press, 1997.

[144] L. Vanhamme, R.D. Fierro, S. Van Huffel, and R. de Beer. Fast removal of residual water in proton spectra. *J. Magn. Reson.*, 132:197–203, 1998.

[145] L. Vanhamme, T. Sundin, P. Van Hecke, S. Van Huffel, and R. Pintelon. Frequency-selective quantification of biomedical magnetic resonance spectroscopy data. *Accepted for publication in J. Magn. Reson.*, 1999.

[146] L. Vanhamme, A. van den Boogaart, and S. Van Huffel. Fast and accurate parameter estimation of noisy complex exponentials with use of prior knowledge. In *Proc. EUSIPCO-96*, Trieste, Italy, September 10-September 13 1996.

[147] L. Vanhamme, A. van den Boogaart, and S. Van Huffel. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J. Magn. Reson.*, 129:35–43, 1997.

[148] L. Vanhamme and S. Van Huffel. Multichannel quantification of biomedical magnetic resonance spectroscopic signals. In *Proc. SPIE conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations VIII*, pages 237–248, San Diego, USA, 22-24 July 1998.

[149] L. Vanhamme, S. Van Huffel, and A. van den Boogaart. Improved fitting method for parameter estimation of MRS signals with use of prior knowledge. In *Proc. ISMRM 97, 5th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 1414, Vancouver, Canada, September 10-September 13 1997.

[150] L. Vanhamme, S. Van Huffel, P. Van Hecke, and D. van Ormondt. Time-domain quantification of series of biomedical magnetic resonance spectroscopy signals. *J. Magn. Reson.*, 140:120–130, 1999.

[151] P. Webb, D. Spielman, and A. Macovski. Inhomogeneity correction for *in vivo* spectroscopy by high-resolution water referencing. *Magn. Reson. Med*, 23:1–11, 1992.

[152] P.G. Webb, N. Sailasuta, S.J. Kohler, T. Raidy, R.A. Moats, and R.E. Hurd. Automated single-voxel proton MRS: Technical development and multisite verification. *Magn. Reson. Med*, 31:365–373, 1994.

[153] K. Young, V. Govindaraju, B.J. Soher, and A.A. Maudsley. Automated spectral analysis I: formation of a priori information by spectral simulation. *Magn. Res. Med.*, 40:812–815, 1998.

[154] K. Young, B.J. Soher, and A.A. Maudsley. Automated spectral analysis II: application of wavelet shrinkage for characterization of non-parameterized signals. *Magn. Res. Med.*, 40:816–821, 1998.

[155] G. Zhu, W.Y. Choy, and B.C. Sanctuary. Spectral parameter estimation by an iterative quadratic maximum likelihood method. *J. Magn. Reson.*, 135:37–43, 1998.

[156] G. Zhu, D. Smith, and Y. Hua. Post-acquisition solvent suppression by singular-value decomposition. *J. Magn. Reson.*, 124:286–289, 1997.

Leentje Vanhamme was born on September 25, 1972 in Tienen, Belgium. She received the degree of electrotechnical-mechanical engineering from the Katholieke Universiteit Leuven, Belgium in July 1994. After an apprenticeship at IBM Almaden Research Center, California, USA, she started working as a researcher at the Electrical Engineering Department ESAT of the K.U. Leuven in January 1995, first on a K.U. Leuven grant and from October 1995 to October 1999 on an I.W.T. (Flemish Institute for Scientific and Technological Research in Industry) grant.