Adaptive Alternating Minimization for Fitting Magnetic Resonance Spectroscopic Imaging Signals

Diana M. Sima, Anca Croitor Sava, and Sabine Van Huffel

Katholieke Universiteit Leuven, Department of Electrical Engineering, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
diana.sima@esat.kuleuven.be, anca.croitor@esat.kuleuven.be, sabine.vanhuffel@esat.kuleuven.be

Summary. In this paper we discuss the problem of modeling Magnetic Resonance Spectroscopic Imaging (MRSI) signals, in the aim of estimating metabolite concentration over a region of the brain. To this end, we formulate nonconvex optimization problems and focus on appropriate constraints and starting values for the model parameters. Furthermore, we explore the applicability of spatial smoothness for the nonlinear model parameters across the MRSI grid. In order to simultaneously fit all signals in the grid and to impose spatial constraints, an adaptive alternating nonlinear least squares algorithm is proposed. This method is shown to be much more reliable than independently fitting each signal in the grid.

1 Introduction

Magnetic Resonance (MR) is widely used in hospitals to distinguish between normal and abnormal tissue. Among the established MR techniques, Magnetic Resonance Imaging (MRI) has a high spatial resolution and is able to provide detailed pictures reflecting differences in tissue, but this technique has a low spectral resolution since it mainly represents the density of water. A second important technique is Magnetic Resonance Spectroscopy (MRS), which provides a signal from a small localized region called voxel, and has a high spectral resolution, *i.e.*, many metabolites (chemicals) are identifieble from an MR spectrum. Thirdly, Magnetic Resonance Spectroscopic Imaging (MRSI) is a multi-voxel technique that combines imaging and spectroscopy in order to provide a trade-off between spatial and spectral resolution.

An MRS signal is a complex-valued time-domain signal y induced by a population of nuclei immersed in a magnetic field after applying a radio-frequency pulse. This time-domain signal is a superposition of many exponentially decaying components. The problem of metabolite quantification amounts to fitting a certain model to the MRS signal.

M. Diehl et al. (eds.), *Recent Advances in Optimization and its Applications in Engineering*,
DOI 10.1007/978-3-642-12598-0_45,

In this paper we focus on modeling and fitting MRSI data, which is a challenging computational problem because of relatively low spectral resolution and high level of noise in the signals. To overcome low data quality, it is important to formulate appropriate constraints and to use good starting values in the nonconvex metabolite quantification optimization problems. In particular, we focus on the spatial smoothness of the nonlinear model parameters across the MRSI grid. In order to simultaneously fit all signals in the grid and to impose spatial constraints, an alternating nonlinear least squares algorithm is proposed. This method is adaptive, in the sense that each subproblem may tune some hyperparameters at run-time, instead of always keeping them fixed.

The paper is organized as follows. In Section 2, the state-of-the-art model for MRS signals, as well as details on the optimization methods used for single-voxel MRS signals, are presented. Further, we pursue in Section 3 the topic of MRSI data quantification, where we first motivate the need to impose spatial relations between the grid's signals; then, the optimization problem and solution method for simultaneous MRSI data quantification are described. Finally, numerical illustrations on simulated noisy MRSI grids are found in Section 4.

2 Metabolite quantification of MRS signals

2.1 MRS model

An MRS signal can be modeled in the time-domain as a sum of complex damped exponentials $\sum_{k=1}^{K'} a_k \exp(j\phi_k) \exp(-d_k t + 2\pi j f_k t)$, where a_k are amplitudes, ϕ_k phases, d_k damping factors and f_k frequencies, $j = \sqrt{-1}$ and t denotes a particular time instant among the discrete measuring times $t_0, \ldots, t_{m-1}$. In this parametric model, the frequencies are characteristic to the metabolites under investigation, while the amplitudes are proportional to the concentration of the respective molecule.

Due to the fact that many metabolites resonate in a well-defined pattern at more than one frequency, depending on the molecular configuration, a more sophisticated model is currently used for MRS signals,

$$\widehat{y}(t) = \sum_{k=1}^{K} a_k \exp(j\phi_k) \exp(-d_k t + 2\pi j f_k t)\, v_k(t), \tag{1}$$

where we point out that v_k, with $k = 1, \ldots, K$, denotes a pure metabolite signal, which can be measured *in vitro* or simulated using quantum mechanical knowledge. In this case the factor $\exp(j\phi_k) \exp(-d_k t + 2\pi j f_k t)$ accounts for corrections to the ideal metabolite signal v_k, such as small frequency shifts f_k, small damping corrections d_k and phase corrections ϕ_k, while a_k stands for the total amplitude of metabolite k.

2.2 Model fitting

Metabolite quantification amounts to a nonlinear least squares problem of fitting model (1) to a measured signal $y(t)$.[1] Figure 1 (left) shows a typical basis set of metabolite spectra v_k that can be used for fitting *in vivo* measured MRS signals from the human brain. Figure 1, bottom right, illustrates the fitting of a noisy signal with the metabolite basis set; to this end, the metabolite spectra are appropriately modified, as shown in Figure 1, top right, by broadening the peaks (*i.e.*, increasing d_k), by slightly shifting the spectra along the frequency axis (with f_k Hz), and by scaling each of them to an appropriate amplitude a_k.

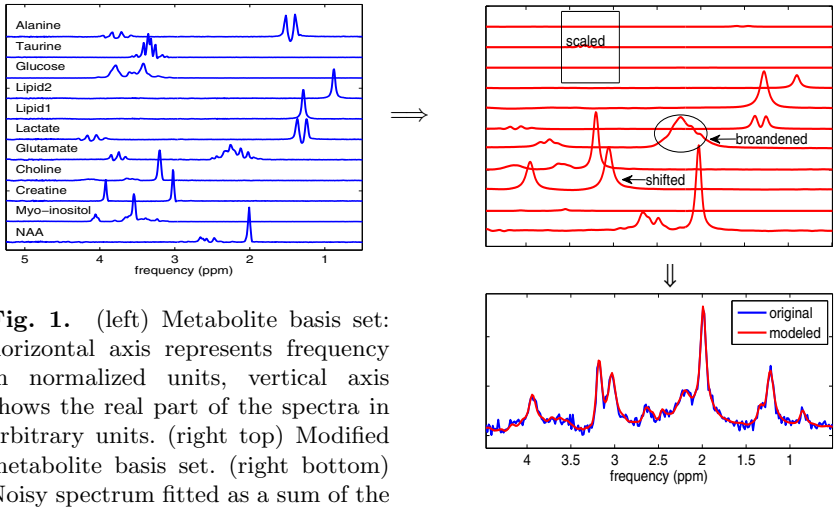

Fig. 1. (left) Metabolite basis set: horizontal axis represents frequency in normalized units, vertical axis shows the real part of the spectra in arbitrary units. (right top) Modified metabolite basis set. (right bottom) Noisy spectrum fitted as a sum of the modified metabolite basis set.

The nonlinear least squares problem mentioned above involves also bounds on the considered parameters, which come from the physical meaning of these parameters. In mathematical terms, this problem reads:

$$\min_{\substack{a_k,\phi_k,d_k,f_k \\ k=1,\ldots,K}} \|y-\widehat{y}\|^2 \quad \text{s.t.} \quad a_k \geq 0,\ \phi_k \in [0,2\pi], d_k \in (-\epsilon_d,\epsilon_d),\ f \in (-\epsilon_f,\epsilon_f) \tag{2}$$

It is important to notice the two important hyperparameters ϵ_d and ϵ_f, which specify the allowed variation of the damping corrections and of the frequency

[1] There are several acquisition conditions that lead to distortions or artifacts of the considered model (1) and for which specialized preprocessing steps exists. They will not be discussed in this paper; see, *e.g*, [12] for more details.

shifts, respectively. Since *in vivo* spectra may be quite different from each other, there are no predetermined optimal values for these hyperparameters, however such bounds are needed in order to preserve the physical meaning of each metabolite spectrum. Their chosen values might be critical for the estimated model parameters (a_k, etc.).

2.3 Variable projection approach

In the considered model (1), the *complex amplitudes* $\alpha_k = a_k \exp(j\phi_k)$ appear as coefficients of a linear combination of nonlinear functions in the parameters d_k, f_k. Thus, for any fixed values of the *nonlinear parameters* d_k, f_k, $k = 1, \ldots, K$, one can obtain corresponding optimal values for all α_k using linear least squares. The constraints $a_k \geq 0$, $\phi_k \in [0, 2\pi]$ are then readily satisfied if we take $a_k = |\alpha_k|$ and $\phi_k = \text{angle}(\alpha_k)$.

The variable projection approach [4, 11] is an optimization framework where the coefficients α_k are projected out, such as to obtain an optimization problem only in the remaining nonlinear variables. The projected functional will be denoted $\phi(\theta)$, where $\theta \in \Re^{2K}$ stands for the vector of parameters $d_1, \ldots, d_K, f_1, \ldots, f_K$. Function and Jacobian evaluations needed by optimization solvers such as Gauss-Newton, Levenberg-Marquardt, or trust region, are slightly more computationally expensive than for the original problem formulation. Still, it is well known and proven by theory [10] and practice that variable projection always converges in less iterations than the original full functional approach. This includes convergence in cases when the full functional approach diverges. Another advantage of this formulation is that no starting values are needed for the linear parameters, and that the number of parameters is halved.

The Levenberg-Marquardt algorithm [6] applied to the variable projection functional is implemented in the quantification method AQSES (Accurate Quantification of Short Echo-Time MRS Signals) [8]. The starting values for the nonlinear parameters d_k and f_k are set by default in AQSES to zero, with the motivation that d_k and f_k represent *small corrections* to the metabolite profiles in the basis set.

3 Metabolite quantification of MRSI signals

3.1 Characteristics of MRSI data

MRSI signals can be modeled with the same mathematical formulation as the MRS signals (1). A straightforward approach to quantify metabolites in a grid of MRSI voxels would be to apply a single-voxel quantification method, such as AQSES, to each signal in the grid individually. As opposed to single-voxel measurements, the MRSI signals usually have a much lower quality, due to the spatial/spectral trade-off for the available measuring time. Thus,

they are more prone to quantification errors, since metabolites present in low concentration are almost embedded in noise. Moreover, a lower spectral resolution also implies that metabolite components become more strongly overlapping in frequency.

It is obvious that supplementary information expressed as constraints on the optimization parameters would be very valuable in analysing this type of data. Since MRSI signals are obtained during a single scan using a certain acquisition protocol, many characteristics of the signals within the same grid are related [3]. Differences in the signals may appear due to two main causes: the heterogeneity of the tissue under investigation, and the magnetic field applied in the scanner, which cannot be kept perfectly constant over the whole volume under investigation.[2] In particular, the damping factors and frequency location of each individual exponential decay are directly related to the local magnetic field. Assuming there are no abrupt changes in the magnetic field, the damping and frequency parameters exhibit smooth maps over the considered MRSI grid.

3.2 Smoothness of parameter maps

Smoothness of a 2D parameter map can be locally measured at every voxel (ℓ, κ) in the grid by using the parameter value at the current location and the values in a certain neighborhood. We denote that two voxels are neighbors by $(\ell_1, \kappa_1) \sim (\ell_2, \kappa_2)$. Because MRSI grids are rather coarse, we usually focus on 3×3 regions with the current voxel (ℓ, κ) in the center. When (ℓ, κ) is on the border of the MRSI grid, only the available neighbors are used. A possible measure for the smoothness at point (ℓ, κ) is given by the first order difference norm

$$\sum_{(i,j)\sim(\ell,\kappa)} (p_{\ell\kappa} - p_{ij})^2, \tag{3}$$

where p stands for any of the parameters d_k or f_k, for any k. Second order formulas are also possible, such as the second order differences

$$(2p_{\ell\kappa} - p_{\ell-1,\kappa} - p_{\ell+1,\kappa})^2 + (2p_{\ell\kappa} - p_{\ell,\kappa-1} - p_{\ell,\kappa+1})^2, \tag{4}$$

$$(4p_{\ell\kappa} - p_{\ell-1,\kappa} - p_{\ell+1,\kappa} - p_{\ell,\kappa-1} - p_{\ell,\kappa+1})^2, \tag{5}$$

3.3 Simultaneous optimization of MRSI signals

A complete optimization problem for fitting all signals in the MRSI grid and, simultaneously, penalizing all the parameter maps for smoothness (with, *e.g.*, a penalty of type (3)) is formulated as (see also Kelm [5])

[2] Other causes of spectral differences could be differences in temperature or in pH, but we assume them constant over the grid.

$$\min_{\Theta \in I} \sum_{\ell,\kappa} \phi_{\ell\kappa}(\theta_{\ell\kappa}) + \sum_{(i,j)\sim(\ell,\kappa)} \lambda_{(i,j),(\ell,\kappa)} \|W(\theta_{\ell\kappa} - \theta_{ij})\|_2^2, \tag{6}$$

where Θ stands for the entire set of parameters $\theta_{\ell\kappa} \in \Re^{2K}$, for all voxels (ℓ, κ), and I denotes the box defined by the hyperparameters ϵ_d, ϵ_f. Moreover, the diagonal $2K \times 2K$ matrix W is used to account for different scaling of the d_k and f_k parameters in $\theta_{\ell\kappa}$, and the scalars $\lambda_{(i,j),(\ell,\kappa)}$ are regularization hyperparameters that affect the trade-off between data fitting and parameter map smoothing.

This optimization problem is highly dimensional, having $2KMN$ variables, where $M \times N$ is the grid size. (In practice we may have grids of at least 16×16 voxels and at least 10 metabolite signals in the basis set, leading to a total of more than 5000 nonlinear variables.) However, the objective function is a sum of squares, where each term contains only a few variables. Assuming all variables fixed, except for the vector $\theta_{\ell\kappa}$, we obtain tractable subproblems of the form

$$\min_{\theta_{\ell\kappa} \in I_{\ell\kappa}} \phi_{\ell\kappa}(\theta_{\ell\kappa}) + \sum_{(i,j)\sim(\ell,\kappa)} \lambda_{(i,j),(\ell,\kappa)} \|W(\theta_{\ell\kappa} - \theta_{ij})\|_2^2, \tag{7}$$

with $I_{\ell\kappa}$ denoting the box corresponding to the vector $\theta_{\ell\kappa}$. Thus, the total optimization problem (6) is a natural candidate for an alternating minimization procedure, where subproblems of the type (7) are solved for each voxel in several sweeps through the grid, until convergence.

Remark 1. In a statistical setting, this type of alternating minimization has been introduced in the field of computer vision under the name *iterated conditional modes* (ICM). An extension of ICM to MRSI data is proposed in [5] under the name block-ICM, where instead of minimizing only over $\theta_{\ell\kappa}$, each subproblem takes a set of parameters corresponding to a neighborhood of voxels as free variables.

3.4 Adaptive alternating minimization

Alternating minimization algorithms are known to converge under very mild conditions [9, 2]. Recently, convergence properties have been analyzed for the situation when the problem statement slightly changes from sweep to sweep [7]. In [7] the variables are partitioned in only two sets, while here we apply adaptive alternative minimization with a large number of subsets (one subset per voxel). Slight changes in problem formulation are expressed, in our case, as modifications of the hyperparameters of the problem. These are, essentially, the bounds on d_k and f_k, which define at each sweep $w = 1, 2, \ldots$ a box $I^w_{\ell,\kappa}$ for each voxel (ℓ, κ). Thus, the new subproblem at sweep w for the voxel (ℓ, κ) is very similar to (7), except that the box constraint may vary at each sweep, and so do the regularization factors. Updates for the box constraints, for the regularization factors and for the starting values of each subproblem are proposed in [1].

4 Numerical results

In this section we illustrate several aspects of the new method on realistically simulated signals. The simulated signals follow model (1), where 11 *in vitro* measured metabolite profiles v_k are used, and the model parameters take biologically relevant values.[3] In order to better approximate MRSI situations, we artificially smoothen the parameter maps of the nonlinear variables. We set random, but realistic values for the amplitude maps, except in the case of two metabolites: for the first, we create a smooth map and for the second a map with an abrupt change in value. This is done for the purpose of checking whether the method is able to capture such specific situations, although the amplitudes are not explicitly constrained. The signals are finally perturbed with additive white noise with various signal-to-noise ratios. The size of the simulated grids is 5×5, the considered neighborhoods are 3×3, and only maximum 4 neighbors (up, down, left, right) are considered when imposing spatial constraints. The spatial constraint in these simulations involves the second order difference (5). Results with the first order difference (3) are comparable, but a bit less suitable for this particular simulation with very smooth parameter maps for the nonlinear variables.

Figure 2 depicts the estimated amplitude values for a grid of signals, and Figure 3 the corresponding frequency shifts, when the signals contain a high level of noise (SNR = 5). We clearly see that the results of the multi-voxel approach are much closer to the true simulated values compared to the single-voxel based method AQSES. For lower noise levels the differences are not as pronounced, since in that case AQSES performs already well enough.

Further we illustrate in Figure 4 the effect of the hyperparameters ϵ_d, ϵ_f. These bounds are computed at each sweep as the median value of the corresponding parameters from the neighboring voxels plus/minus a fraction of the previous length of the interval.

A final illustration of the importance of the considered box constraints is given by the contour plot in Figure 5. All parameters are set to the optimal values computed by the new method, except for two frequency shifts, f_2 and f_4, corresponding to metabolites that partially overlap in frequency. The projected objective function, although regularized with the smoothness penalty terms, and having excellent values for 20 out of 22 model parameters, is highly nonconvex. Still, the obtained optimal solution is very close to the minimum and also close to the true simulated values.

5 Conclusions

We discussed an alternating minimization algorithm with varying values for the hyperparameters, applied to the simultaneous, spatially constrained fitting

[3] See [8] for more details on the measured metabolite profiles and on how meaningful values for the model parameters are obtained.

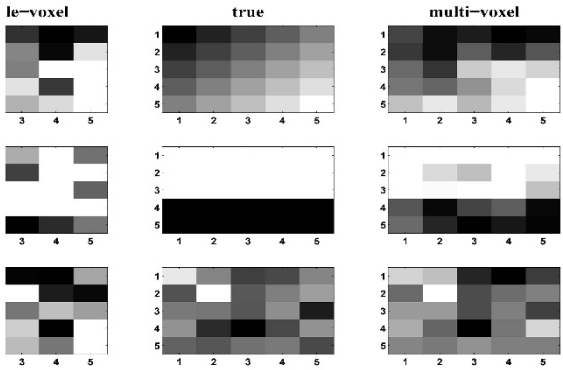

Fig. 2. Amplitude values on a 5×5 grid for a selection of 3 out of 11 metabolites (the 3 rows of grids), namely two metabolites with smooth and abrupt amplitude maps, and a third one with random entries. The middle column corresponds to the true values, while the left and right columns correspond to the estimated amplitudes provided by the single-voxel and the multi-voxel approaches, respectively.

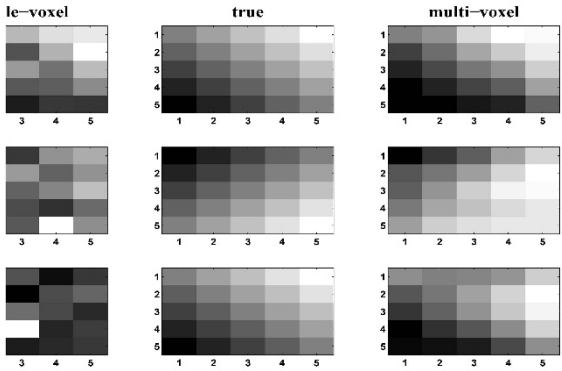

Fig. 3. Frequency values for the same example as in Figure 2. Damping maps are similar, although not shown here.

of Magnetic Resonance Spectroscopic Imaging signals. This approach is more accurate than individually fitting each signal in the grid. Still, some issues must be further studied, such as what smoothness measure is more appropriate for *in vivo* data, or how to automatically safeguard against decreasing the constraint box too much; ideas from trust region methods could be adapted for this purpose. Finally, the relevance of this approach to clinical data obtained from brain tumor patients is being evaluated in [1].

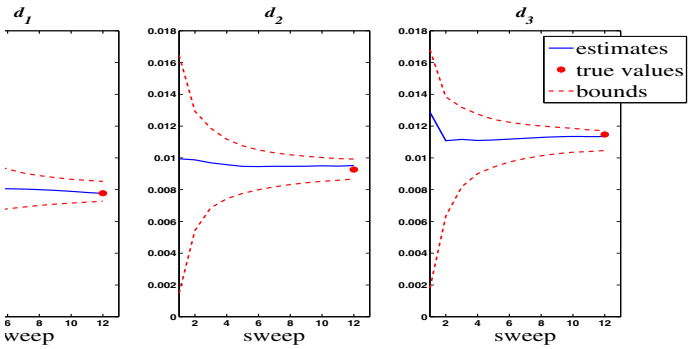

Fig. 4. Convergence of three damping estimates corresponding to the middle voxel of a 5×5 simulated MRSI grid with SNR=10. The estimated values for 12 sweeps are plotted together with the corresponding upper and lower bounds; the true values are also shown as big dots.

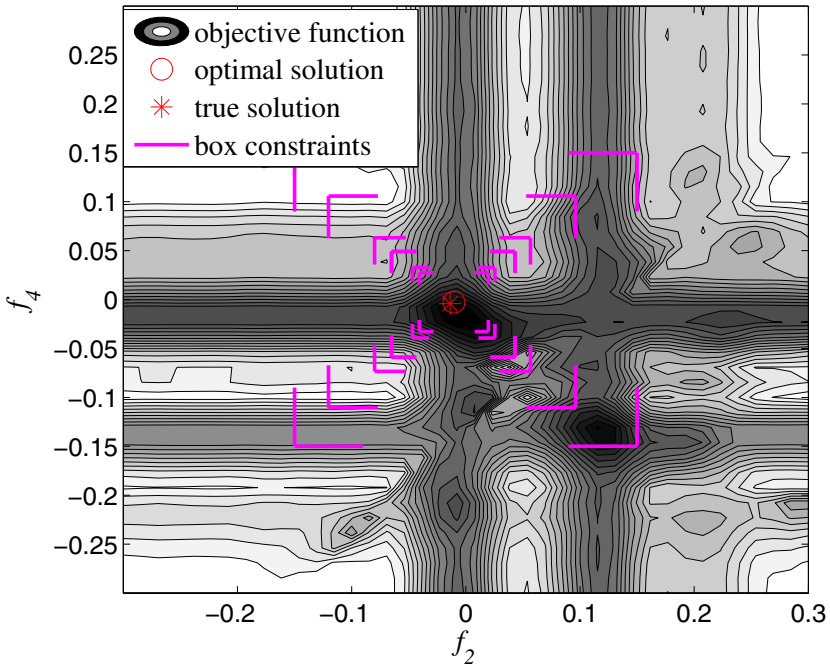

Fig. 5. Contour plot for the objective function projected onto the (f_2, f_4)-plane of a subproblem corresponding to the voxel $(1, 1)$ during the last sweep of the multi-voxel method, in a simulated MRSI grid with SNR=10. True and estimated parameters are shown as a star and a circle, respectively. The successive box constraints are also sketched.

Acknowledgments

D.M. Sima is a postdoctoral fellow of the Fund for Scientific Research-Flanders. S. Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. Research supported by

- Research Council KUL: GOA-AMBioRICS, GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), IDO 05/010 EEG-fMRI, IDO 08/013 Autism, IOF-KP06/11 FunCop, several PhD/postdoc and fellow grants;
- Flemish Government:
  - FWO: PhD/postdoc grants, projects, G.0519.06 (Noninvasive brain oxygenation), FWO-G.0321.06 (Tensors/Spectral Analysis), G.0302.07 (SVM), G.0341.07 (Data fusion), research communities (ICCoS, ANMMM);
  - IWT: TBM070713-Accelero, TBM070706-IOTA3, PhD Grants;
- Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011);
- EU: ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST200427214), FAST (FP6-MC-RTN-035801), Neuromath (COST-BM0601)
- ESA: Cardiovascular Control (Prodex-8 C90242)

References

1. A. Croitor Sava, D.M. Sima, J.B. Poullet, A. Heerschap, and S. Van Huffel, *Exploiting spatial information to estimate metabolite levels in 2D MRSI of heterogeneous brain lesions*, Tech. Report 09-182, ESAT-SISTA, K.U. Leuven, 2009.
2. I. Csiszár and G. Tusnády, *Information geometry and alternating minimization procedures*, Statistics & Decisions **Supplement Issue** (1984), no. 1, 205237.
3. R. de Graaf, *In Vivo NMR Spectroscopy. Principles and Techniques*, 2nd ed., John Wiley & Sons, 2007.
4. G.H. Golub and V. Pereyra, *Separable nonlinear least squares: the variable projection method and its applications*, Inverse Problems **19** (2003), no. 2, 1–26.
5. B. M. Kelm, *Evaluation of vector-valued clinical image data using probabilistic graphical models: Quantification and pattern recognition*, Ph.D. thesis, Ruprecht-Karls-Universität, Heidelberg, 2007.
6. J. J. Moré, *The Levenberg-Marquardt algorithm: Implementation and theory*, Numerical Analysis: Proceedings of the Biennial Conference held at Dundee, June 28-July 1, 1977 (Lecture Notes in Mathematics #630) (G. A. Watson, ed.), Springer Verlag, 1978, pp. 104–116.
7. U. Niesen, D. Shah, and G.W. Wornell, *Adaptive alternating minimization algorithms*, IEEE Transactions on Information Theory **55** (2009), no. 3, 1423–1429.
8. J.B. Poullet, D.M. Sima, A. Simonetti, B. De Neuter, L. Vanhamme, P. Lemmerling, and S. Van Huffel, *An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES*, NMR in Biomedicine **20** (2007), no. 5, 493–504.
9. M.J.D. Powell, *On Search Directions for Minimization Algorithms*, Mathematical Programming **4** (1973), 193–201.
10. A. Ruhe and P.A. Wedin, *Algorithms for separable nonlinear least squares problems*, SIAM Review **22** (1980), 318–337.
11. D.M. Sima and S. Van Huffel, *Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification*, Journal of Computational and Applied Mathematics **203** (2007), 264–278.
12. D.M. Sima, J. Poullet, and S. Van Huffel, *Exponential data fitting and its applications*, ch. Computational aspects of exponential data fitting in Magnetic Resonance Spectroscopy, Bentham eBooks, 2009.