Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 2005-236

Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification [1]

Diana M. Sima[2] Sabine Van Huffel[2]

November 2005
Submitted to the *Journal of Computational and Applied Mathematics*.

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/sima/reports/SISTA5-236.pdf*

[2]ESAT-SCD-SISTA, K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium, Tel. 32/16/32 43 11, Fax 32/16/32 19 70, E-mails: *diana.sima@esat.kuleuven.be*, *sabine.vanhuffel@esat.kuleuven.be*. This work was supported by **Research Council KUL:** GOA-AMBioRICS, several PhD/postdoc & fellow grants; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), G.0360.05 (EEG, Epileptic), research communities (ICCoS, ANMMM); **IWT:** PhD Grants; **Belgian Federal Science Policy Office:** IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'); **EU:** BIOPATTERN, ETUMOUR.

Abstract

An application in magnetic resonance spectroscopy quantification models a signal as a linear combination of nonlinear functions. It leads to a separable nonlinear least squares fitting problem, with linear bound constraints on some variables. The Variable Projection (VARPRO) technique can be applied to this problem, but needs to be adapted in several respects. If only the nonlinear variables are subject to constraints, then the Levenberg-Marquardt minimization algorithm that is classically used by the VARPRO method should be replaced with a version that can incorporate those constraints. If some of the linear variables are also constrained, then they cannot be projected out via a closed-form expression as is the case for the classical VARPRO technique. We show how quadratic programming problems can be solved instead, and we provide details on efficient function and approximate Jacobian evaluations for the inequality constrained VARPRO method.

Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification

Diana M. Sima, Sabine Van Huffel

ESAT-SCD, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.

Abstract

An application in magnetic resonance spectroscopy quantification models a signal as a linear combination of nonlinear functions. It leads to a separable nonlinear least squares fitting problem, with linear bound constraints on some variables. The Variable Projection (VARPRO) technique can be applied to this problem, but needs to be adapted in several respects. If only the nonlinear variables are subject to constraints, then the Levenberg-Marquardt minimization algorithm that is classically used by the VARPRO method should be replaced with a version that can incorporate those constraints. If some of the linear variables are also constrained, then they cannot be projected out via a closed-form expression as is the case for the classical VARPRO technique. We show how quadratic programming problems can be solved instead, and we provide details on efficient function and approximate Jacobian evaluations for the inequality constrained VARPRO method.

Key words: magnetic resonance spectroscopy data quantification, nonlinear least squares, variable projection
1991 MSC: 90C30, 92C55, 65F99

Email addresses: diana.sima@esat.kuleuven.be (Diana M. Sima), sabine.vanhuffel@esat.kuleuven.be (Sabine Van Huffel).

1 Introduction

Separable nonlinear least squares fitting problems are minimization problems of the form

$$\min_{\mathbf{x},\mathbf{a}} \|\mathbf{y} - \Phi(\mathbf{x})\mathbf{a}\|_2^2, \tag{1}$$

where $\mathbf{y}$ is an m-dimensional given noisy data vector, $\mathbf{x}$ denotes an n-dimensional vector of nonlinear parameters, $\mathbf{a}$ is a p-dimensional vector of linear parameters, and Φ is a nonlinear function that maps a vector $\mathbf{x}$ into an $m \times p$ matrix.

In the beginning of the seventies, the VARPRO method for solving separable least squares problems appeared [7]. During the past 30 years, VARPRO received attention from theoreticians, as well as practitioners. A recent review of VARPRO and its applications is given in the paper [8]. VARPRO uses the fact that the variable $\mathbf{a}$ that appears linearly in the model function $\Phi(\mathbf{x})\mathbf{a}$ can be optimally expressed as a linear least squares solution depending on the variable $\mathbf{x}$: $\mathbf{a}^{\text{ls}}(\mathbf{x}) = \Phi(\mathbf{x})^{\dagger}\mathbf{y}$, where † denotes the Moore-Penrose pseudoinverse of a matrix. Therefore, this closed formula of $\mathbf{a}$ can be plugged in into the original minimization problem, yielding the equivalent problem only in $\mathbf{x}$:

$$\min_{\mathbf{x}} \|(I_m - \Phi(\mathbf{x})\Phi(\mathbf{x})^{\dagger})\mathbf{y}\|_2^2.$$

This problem can be solved with classical nonlinear (least squares) optimization methods, such as the Gauss-Newton or the Levenberg-Marquardt algorithms. These methods require evaluation of the Jacobian with respect to $\mathbf{x}$ of the functional inside the norm. An essential idea is found in [11] and bears the name of Kaufman's simplification: it involves computing an approximate Jacobian instead of the true Jacobian, trading-off a negligible loss of accuracy in the Jacobian for a rather important computational time saving.

An interesting extension towards constrained variable projection is the case when *separable equality constraints* appear in the problem [12]. Another important related problem deals with having two separable classes of variables, without the requirement that some of them appear linearly [21].

In this paper, we consider the classical linear/nonlinear variable separation and we analyze some extensions of the separable nonlinear least squares problem and of the VARPRO technique when (inequality) constraints appear within one or both classes of separable variables. These extensions are motivated by a biomedical application, namely the quantification of metabolite concentrations from magnetic resonance spectroscopic signals. The extensions from the classical VARPRO presented in this paper and needed in our application are shortly enumerated here, in increasing degree of difficulty:

Complex data. The VARPRO technique can be extended to work with com-

plex variables and data. This observation helps us to address the computation of the residual and of the approximate Jacobian explicitly in the complex domain. We transform to real data when solving the resulting minimization problem (only in the nonlinear variables), since all classical nonlinear minimization implementations work with real data.

Constraints on the nonlinear variables. These constraints acting only on the nonlinear variables do not affect the VARPRO idea of projecting out the linear variables. Thus, these constraints can be simply imported to the resulting minimization problem only in $\mathbf{x}$. However, the classical Gauss-Newton or Levenberg-Marquardt method that we used in the unconstrained case must be replaced with a method that can take into account constraints. In our application, we focus only on linear bounds for the variables; for this case, efficient methods are available (see the Netlib repository www.netlib.org/opt/).

Constraints on the linear variables. Imposing general constraints to the linear parameters takes away the possibility of projecting them out via a closed-form expression. Nevertheless, we would still like to keep the idea of solving an outer minimization problem only in the nonlinear variables $\mathbf{x}$, and solve the constrained linear least squares problem in $\mathbf{a}$ in an efficient manner. In our application, we have non-negativity restrictions for some of the linear variables (but linear bounds can be treated similarly). The inner problem in $\mathbf{a}$ can then be efficiently solved with a quadratic programming type of method. In any case, we expect computational advantages and possible faster convergence rate for this type of constrained VARPRO, compared to solving the initial problem (1) with additional constraints, using a general nonlinear solver that does not distinguish between linear and nonlinear variables.

In Section 2, we give a short overview of the quantification of signals in magnetic resonance spectroscopy (MRS) [5,9], with emphasis on the optimization problems that are obtained as mathematical formulations for this application. The VARPRO method was already used in MRS problems [26,27]; a historical note on the application of VARPRO to MRS data quantification can be read in Section 17 of the review paper [8]. The previous usage of VARPRO in the mentioned papers was restricted to models of the type "sum of complex damped exponentials." These models are appropriate for fitting the so-called *long echo-time* MRS signals. Nowadays, the nuclear magnetic resonance scanners record *short echo-time* signals that are richer in information because they display the responses of significantly more metabolites (chemical substances). The model given by a sum of complex exponentials no longer holds, since the response of each metabolite is spread out over the whole spectrum. However, spectra of metabolites that are relevant in the human brain can be measured separately *in vitro*. Such measurements can be grouped together in a database of metabolite signals. Then, a signal measured *in vivo* (from a region in the human brain) can be modeled as a combination of individual metabolite signals

in the database. In some cases, a *baseline* signal that accounts for the presence of some non-predominant unknown macromolecules must also be added to the model.

In Sections 3 and 4, the VARPRO extensions are presented in relation with the MRS application, and we summarize the studied cases in the following table:

Linear variables	Nonlinear variables	MRS data model	Section
complex unconstrained	real/complex (un)constrained	without/with baseline non-equal phases	3
real constrained	real/complex (un)constrained	without baseline equal phases	4.2
real partially constrained/ partially unconstrained	real/complex (un)constrained	with baseline equal phases	4.3

Finally, the numerical experiments in Section 5 aim to illustrate that an approximate Jacobian formula proposed for the constrained linear variables case yields accurate results for the MRS data quantification problem with equal phases.

2 Quantification of magnetic resonance spectroscopic signals

For the quantification of short echo-time MRS signals, we assume that we are given a "metabolite database", which is a set $\{v_k, \text{ for } k = 1, \ldots, K\}$ of complex-valued time series of length m, representing *in vitro* measured MRS responses [5]. An *in vivo* measured MRS signal y is also a complex-valued time series of length m that will satisfy the model

$$y(t) = \widehat{y}(t) + \varepsilon_t := \sum_{k=1}^{K} \alpha_k \, (\zeta_k)^t \, (\eta_k)^{t^2} \, v_k(t) + b(t) + \varepsilon_t, \quad t = t_0, \ldots, t_{m-1}, \tag{2}$$

where $\alpha_k, \zeta_k, \eta_k \in \mathbb{C}$ are unknown parameters that account for concentrations of the metabolites in the database and for the necessary corrections of the database signals v_k, due to inherent differences in the acquisition technique [18,20,14]. In fact the complex amplitudes α_k and the complex ζ_k and η_k can be written as (with $j = \sqrt{-1}$):

$$\alpha_k = a_k \exp(j\phi_k),$$

$$\zeta_k = \begin{cases} \exp(-d_k + jf_k), & \text{for Lorentzian and Voigt lineshapes;} \\ \exp(jf_k), & \text{for Gaussian lineshapes;} \end{cases}$$

$$\eta_k = \begin{cases} \exp(je_k), & \text{for Lorentzian lineshapes;} \\ \exp(-g_k + je_k), & \text{for Gaussian and Voigt lineshapes;} \end{cases}$$

where a_k are the real amplitudes, ϕ_k are the phase shifts, d_k are damping corrections, g_k are Gaussian damping corrections, f_k are frequency shifts, and e_k are eddy current correction terms [10,15].

Moreover, $b(t)$ represents the chemical part that is not modeled, which is the response of the substances that are not included in the database, and ε_t is an unknown noise perturbation with zero mean, for all t's with indices from 0 to $m-1$.

The identification of complex amplitudes α_k, and complex ζ_k's and η_k's, for $k = 1, \ldots, K$, can be accomplished by minimizing the least squares criterion: $\sum_{t=t_0,\ldots,t_{m-1}} |y(t) - \widehat{y}(t)|^2$.

The nuisance of incorporating the "baseline" term b is milded by the assumptions that can be made about it. Physiologically, the "macromolecules" (the chemicals that are left out of the database and are not relevant for the quantification) have some spectral characteristics in the MRS experiment, *e.g.*, they are broadband, low amplitude signals. In mathematical terms, we can say that b is characterized by the fact that its Fourier transformation should be a smooth function. (Actually, the Fourier transform of the unknown *continuous-time* b should be a smooth function. In practice, we work with discrete time instants and with the discrete Fourier transform.)

A semiparametric model with smoothness constraint for MRS data quantification can be designed, which takes into account the parametric part of the model, but treats the baseline nonparametrically [22,18–20,4,23]. For the nonparametric reconstruction of the baseline, we construct a basis of splines [2,3] and put the discretized splines as columns in a matrix A of size $m \times n$, with n smaller than the number of data points m. Any nonlinear function can be approximated as a linear combination of spline functions. The coefficients in this linear combination are the unknowns that must be identified. We denote these linear coefficients by $c_1, \ldots, c_n$ (or by $\mathbf{c} \in \mathbb{C}^n$, when stacked in a column vector). Thus the discretization of a nonlinear function approximated with splines can be written in matrix notation as $A\mathbf{c}$.

A regularization matrix D is defined to measure the smoothness of the base-

line in the frequency domain. We can take D as the discrete second-order differential operator, $D = \begin{bmatrix} -1 & 2 & -1 & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & -1 & 2 & -1 \end{bmatrix}$, first-order differential operator, $D = \begin{bmatrix} -1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix}$, zero-order differential operator, the identity $D = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$, or some combination.

Since the goal is to reconstruct a smooth baseline in the frequency domain, while still fitting in the time-domain, we transform the basis matrix A to the time domain, with the discrete inverse Fourier transform. Thus $\mathcal{A} := \mathcal{F}^{-1}(A)$, where the operator $\mathcal{F}^{-1}$ denotes the discrete inverse Fourier transform, applied to each column of a matrix. Finally, we consider the regularized nonlinear least squares criterion

$$\min_{\substack{\alpha_1,\ldots,\alpha_K\in\mathbb{C},\mathbf{c}\in\mathbb{C}^n \\ (\zeta_1,\ldots,\zeta_K,\eta_1,\ldots,\eta_K)\in\Omega}} \frac{1}{m}\sum_{t=t_0}^{t_{m-1}} \left| y(t) - \sum_{k=1}^{K} \alpha_k \, (\zeta_k)^t \, (\eta_k)^{t^2} \, v_k(t) - (\mathcal{A}\mathbf{c})(t) \right|^2 + \lambda^2 \mathbf{c}^H D^H D \mathbf{c}, \tag{3}$$

where Ω denotes some constrained parameter space. Ω usually involves only linear equality and inequality constraints on the real-valued parameters appearing in ζ_k, η_k. The role of possible equality constraints is to impose *prior knowledge* relationships between corresponding parameters of related metabolites. The inequality constraints are, in fact, simple bound constraints on the real-valued parameters. We expect that these corrections will be small, since all these parameters are used to *slightly correct* the metabolite signals. Otherwise, the chemical meaning of the corrected metabolite profiles will be lost.

In (3), λ is a fixed regularization (penalty) parameter, and the whole term $\lambda^2 \mathbf{c}^H D^H D \mathbf{c}$ is responsible for ensuring a certain degree of smoothness to the baseline b. The value that we give to λ controls also the degree of smoothness.

Note that in (3) the complex amplitudes α_k are free variables. This translates into the fact that, when we look at their polar representation in terms of the real amplitudes and the spectral phases, $\alpha_k = a_k \exp(j\phi_k)$, the phases $\phi_1, \ldots, \phi_K$ are free between $-\pi$ and π, and the amplitudes $a_1, \ldots, a_K$ have *positive* values. In fact, the real amplitudes are the most representative parameters for the MRS model, since they are the weights with which each metabolite appears into the quantified signal; they, thus, yield the metabolite concentrations in the given brain region, and these concentrations are indicative for the health status or tumor degree in that region [17,18].

3 Separable least squares with constraints on nonlinear variables

The motivation for the formulation in this section comes from the problem already described in Section 2, namely the MRS data quantification with non-equal phases. We show that problem (3) can be easily turned into a separable nonlinear least squares problem. The differences with respect to the classical separable problems solved by VARPRO is that the linear parameters, as well as the residual under the norm, will be complex-valued; moreover, simple equality or inequality constraints are allowed in the MRS data quantification problem formulation (3), but they may only affect the nonlinear (real-valued) variables.

We divide the analysis into two parts: quantification with or without a baseline. We start with the simple case when we ignore the baseline. This case gives us the opportunity to revisit the original ideas behind VARPRO, including computational issues such as Kaufman's simplification for Jacobian computation.

3.1 MRS data model without baseline

The nonlinear least squares problem formulation (3) simply becomes

$$\min_{\substack{\alpha_1,\ldots,\alpha_K\in\mathbb{C}\\(\zeta_1,\ldots,\zeta_K,\eta_1,\ldots,\eta_K)\in\Omega}} \quad \frac{1}{m}\sum_{t=t_0}^{t_{m-1}} \left| y(t) - \sum_{k=1}^{K} \alpha_k \ (\zeta_k)^t \ (\eta_k)^{t^2} \ v_k(t) \right|^2 . \tag{4}$$

Problem (4) is a separable problem, where linear parameters α_k can be projected out of the least squares problem, and only a smaller sized nonlinear least squares problem remains to be solved for the nonlinear variables ζ_k, η_k. For the optimization over the (possibly constrained) set of parameter values for ζ_k, η_k, we choose for an iterative minimization algorithm of the Levenberg-Marquardt type [16]. A trust-region implementation that allows imposing bound constraints on the real variables is described in [6]. Without entering into the details of such an algorithm, we continue our exposition by providing its necessary inputs: initial starting values, as well as procedures to evaluate the function value and the corresponding Jacobian, at each arbitrary set of parameter values.

We set all initial values for the real-valued nonlinear parameters to zero. This is a reasonable starting point, since it means that we start the optimization with *no spectral corrections* to the signals in the database, which corresponds to an ideal situation.

3.1.1 Function evaluation

We rewrite (4) as

$$\min_{\boldsymbol{\alpha},\mathbf{x}\in\Omega} \frac{1}{m} \left\| \mathbf{y} - \Phi(\mathbf{x})\boldsymbol{\alpha} \right\|_2^2, \tag{5}$$

where $\mathbf{y}$ is a column vector containing $y(t_0), \ldots, y(t_{m-1})$, $\boldsymbol{\alpha}$ is a complex K-dimensional column vector containing the complex amplitudes, $\mathbf{x}$ is a vector formed from all nonlinear variables (preferably, the real-valued d_k, f_k, g_k, e_k), and the $m \times K$ complex-valued matrix $\Phi(\mathbf{x})$ has elements of the form:

$$\begin{aligned} \Phi_{ik} &= (\zeta_k)^{t_i} \, (\eta_k)^{t_i^2} \, v_k(t_i) \qquad (6) \\ &= \begin{cases} \exp\left((-d_k + jf_k)t_i + je_k t_i^2\right) \; v_k(t_i), & \text{for Lorentzian lineshapes;} \\ \exp\left(jf_k t_i + (-g_k + je_k)t_i^2\right) \; v_k(t_i), & \text{for Gaussian lineshapes;} \\ \exp\left((-d_k + jf_k)t_i + (-g_k + je_k)t_i^2\right) \; v_k(t_i), & \text{for Voigt lineshapes.} \end{cases} \end{aligned}$$

As mentioned in the introduction, the optimal linear coefficients $\boldsymbol{\alpha}^{\text{ls}}(\mathbf{x})$, for some fixed values of the nonlinear coefficients, $\mathbf{x}$, can be plugged-in such that the residual that we need to compute is the following *variable projection functional*

$$\mathbf{y} - \Phi(\mathbf{x})\boldsymbol{\alpha}^{\text{ls}}(\mathbf{x}) = (I - \Phi(\mathbf{x})\Phi(\mathbf{x})^{\dagger})\mathbf{y}.$$

Clearly, we only need a basis for the column space of the matrix $\Phi(\mathbf{x})$ in order to evaluate the projection matrix $I - \Phi(\mathbf{x})\Phi(\mathbf{x})^{\dagger}$. This basis can be obtained from the QR decomposition of $\Phi(\mathbf{x})$:

$$\Phi(\mathbf{x}) = QR = \begin{bmatrix} Q_1 \; Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

where $R_1 \in \mathbb{C}^{K\times K}$, $Q_1 \in \mathbb{C}^{m\times K}$, and $Q_2 \in \mathbb{C}^{m\times(m-K)}$. Then, the residual $(I - \Phi(\mathbf{x})\Phi(\mathbf{x})^{\dagger})\mathbf{y}$ becomes $Q_2 Q_2^H \mathbf{y}$. Since only the norm of this residual is in fact needed in the optimization algorithm, we can further simplify the definition of our residual by ignoring the multiplication with Q_2 (which has orthonormal columns), and simply compute $Q_2^H \mathbf{y}$ at each function evaluation.

3.1.2 Jacobian evaluation

The gradient of the residual $Q_2^H \mathbf{y}$ is also needed by the Levenberg-Marquardt type nonlinear least squares solver. Note that in the residual $Q_2^H \mathbf{y}$, the nonlinear parameters appear implicitly through Q_2.

Consider again the original variable projection functional $f(\mathbf{x}) = (I - \Phi(\mathbf{x})\Phi(\mathbf{x})^{\dagger})\mathbf{y}$. The Jacobian of f with respect to any of the scalar variables x_k (here denoted by $\nabla_k f$) can be derived with the following manipulations:

$$\begin{aligned}\nabla_k f &= -(\nabla_k \Phi)\Phi^\dagger \mathbf{y} - \Phi(\nabla_k(\Phi^\dagger))\mathbf{y} = -(\nabla_k \Phi)\Phi^\dagger \mathbf{y} - \Phi \nabla_k \left(\left(\Phi^H \Phi\right)^{-1} \Phi^H \right) \mathbf{y} \\ &= -(\nabla_k \Phi)\Phi^\dagger \mathbf{y} + \Phi \left(\Phi^H \Phi\right)^{-1} \left[(\nabla_k \Phi)^H \Phi + \Phi^H (\nabla_k \Phi) \right] \left(\Phi^H \Phi\right)^{-1} \Phi^H \mathbf{y} \\ &= -\left(\nabla_k \Phi - (\Phi^\dagger)^H (\nabla_k \Phi)^H \Phi - \Phi \Phi^\dagger (\nabla_k \Phi) \right) \Phi^\dagger \mathbf{y}.\end{aligned}$$

Kaufman's simplification [11] proposes that only the part

$$-\left(\nabla_k \Phi - \Phi \Phi^\dagger (\nabla_k \Phi) \right) \Phi^\dagger \mathbf{y} = -(I - \Phi\Phi^\dagger)(\nabla_k \Phi)\Phi^\dagger \mathbf{y}$$

should be used to compute an approximate Jacobian, yielding a computational saving that is more important than the loss of accuracy in the Jacobian, which is negligible.

If we take into account the definition (6) of Φ for the MRS data model, the matrix $\nabla_k \Phi$ can be computed using the formulas:

$$\frac{\partial \Phi_{ik}}{\partial d_k} = -t_i \Phi_{ik}, \qquad \frac{\partial \Phi_{ik}}{\partial f_k} = j t_i \Phi_{ik}, \qquad \frac{\partial \Phi_{ik}}{\partial g_k} = -t_i^2 \Phi_{ik}, \qquad \frac{\partial \Phi_{ik}}{\partial e_k} = j t_i^2 \Phi_{ik}. \tag{7}$$

Note here that the variable d_k (or f_k, or g_k, or e_k) only appears in the column k of Φ. Therefore, all other columns of $\nabla_k \Phi$ different from the column k are identically zero:

$$\frac{\partial \Phi_{il}}{\partial d_k} = 0, \qquad \frac{\partial \Phi_{il}}{\partial f_k} = 0, \qquad \frac{\partial \Phi_{il}}{\partial g_k} = 0, \qquad \frac{\partial \Phi_{il}}{\partial e_k} = 0, \qquad \text{for } l \neq k.$$

To fix the ideas, the matrix $\nabla_k \Phi$, which represents the derivative of the matrix Φ with respect to the k^{th} variable x_k (that is either d_k, or f_k, or g_k, or e_k), is an $m \times K$ matrix that has the following structure

$$\nabla_k \Phi = \begin{bmatrix} 0 & 0 & \dfrac{\partial \Phi_{1k}}{\partial x_k} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dfrac{\partial \Phi_{mk}}{\partial x_k} & 0 & 0 \end{bmatrix}$$

Thus, the column corresponding to x_k in the approximate Jacobian equals

$$\widetilde{\nabla}_k f = -(I - \Phi\Phi^\dagger)(\nabla_k \Phi)\Phi^\dagger \mathbf{y} = -(I - \Phi\Phi^\dagger) \begin{bmatrix} \dfrac{\partial \Phi_{1k}}{\partial x_k} \\ \vdots \\ \dfrac{\partial \Phi_{mk}}{\partial x_k} \end{bmatrix} \boldsymbol{\alpha}_k^{\text{ls}}, \tag{8}$$

where we used the fact that $\boldsymbol{\alpha}^{\text{ls}} = \Phi^\dagger \mathbf{y}$. The complete approximate Jacobian $\widetilde{\nabla} f$ is obtained by putting next to each other all columns of type (8), one column for each nonlinear variable in our optimization.

For stable and efficient computation of the Jacobian, we make use of the QR decomposition of Φ, as introduced before. Thus, $\boldsymbol{\alpha}^{\text{ls}} = R_1^{-1} Q_1^H \mathbf{y}$ and $I - \Phi\Phi^\dagger = Q_2 Q_2^H$. Since we ignore the factor Q_2 in the function evaluation, we must also do the same thing in what concerns the Jacobian.

In the end, the approximate Jacobian using Kaufman's simplification is:

$$\widetilde{\nabla} f = -Q_2^H \Delta\Phi, \tag{9}$$

where $\Delta\Phi$ for the MRS data model in its most general formulation, is

$$\left[\begin{array}{c|c|c|c} \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{1k}}{\partial d_k}\ \cdot & \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{1k}}{\partial g_k}\ \cdot & \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{1k}}{\partial f_k}\ \cdot & \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{1k}}{\partial e_k}\ \cdot \\ \vdots \quad \vdots \quad \vdots & \vdots \quad \vdots \quad \vdots & \vdots \quad \vdots \quad \vdots & \vdots \quad \vdots \quad \vdots \\ \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{mk}}{\partial d_k}\ \cdot & \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{mk}}{\partial g_k}\ \cdot & \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{mk}}{\partial f_k}\ \cdot & \cdot\ \boldsymbol{\alpha}_k^{\text{ls}} \dfrac{\partial \Phi_{mk}}{\partial e_k}\ \cdot \end{array}\right],$$

with k going from 1 to K. In cases when not all variables among d_k, g_k, f_k, e_k, are estimated in the model, or when we impose *prior knowledge* in the form of linear equalities between some variables of the same sort (leading to eliminations), the formula above simplifies by deleting the not-needed columns.

3.2 MRS data model with baseline

Little is changed in the variable projection implementation, when we augment the optimization criterion (4) to the regularized version (3). In fact, using the notation from (5), the minimization (3) can be written as

$$\min_{\boldsymbol{\alpha}\in\mathbb{C}^K, \mathbf{x}\in\Omega,\, \mathbf{c}\in\mathbb{C}^n} \frac{1}{m} \left\| \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi(\mathbf{x})\boldsymbol{\alpha} + \mathcal{A}\mathbf{c} \\ \sqrt{m}\lambda D \mathbf{c} \end{bmatrix} \right\|_2^2,$$

which is also a separable nonlinear least squares problem, where the linear variables are $\boldsymbol{\alpha}$ and $\mathbf{c}$, and the nonlinear ones are $\mathbf{x}$.

For the function evaluation, we use the QR decomposition of the matrix $\begin{bmatrix} \mathcal{A} & \Phi(\mathbf{x}) \\ \sqrt{m}\lambda D & 0 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$, with $R_1 \in \mathbb{C}^{(n+K)\times(n+K)}$, and Q_1, Q_2 of appropriate sizes. The function value is $Q_2^H \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}$, and the projected linear

variables have the expression

$$\begin{bmatrix} \mathbf{c}^{\text{ls}} \\ \boldsymbol{\alpha}^{\text{ls}} \end{bmatrix} = R_1^{-1} Q_1^H \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}.$$

For approximate Jacobian evaluation, the only difference comes from the fact that we have augmented $\Phi(\mathbf{x})$ with some blocks that *do not* depend on the nonlinear parameters. This translates into the fact that the new Jacobian is also extended with zero blocks of corresponding dimensions. All completely zero columns can be ignored in the implementation.

4 Separable least squares with bound constraints on linear variables

4.1 Motivation: MRS data quantification with equal phases and non-negativity constraint for the amplitudes

In this section, we are concentrating on the problems introduced by requiring equal phase corrections for all metabolites ($\phi_1 = \ldots = \phi_K =: \phi_0$). Remember that the phases entered into the problem in the previous sections through the *complex amplitudes* $\alpha_k = a_k \exp(j\phi_k)$, which were the complex linear parameters in the VARPRO method.

Requiring equal phase corrections is a reasonable approximation, since the phase distortions between different metabolites within an *in vivo* signal are negligible. Moreover, it was noticed in experiments with the non-equal phases version presented in Section 3 that in some cases (when the metabolite database contains signals with overlapping resonant frequency regions) there is a tendency for overlapping metabolites to compensate for each other by having opposite phases. In other words, these metabolites partially cancel each other, and thus their amplitudes are unreliably computed, although the residual is small. As an illustration, see the reconstructed signal, together with the database of corrected metabolite profiles, in Figure 1. Anticipating the method in this section, Figure 2 shows the fitting results when the equal phase constraint is used. The reconstructed signals are very similar in the two figures, but, noticeably, there are no longer artifacts from interchangeable metabolites in Figure 2. All the plots show real parts of the signals in the frequency domain.

Conceptually, imposing equal phases makes the model simpler, but practically, the method for solving the problem becomes a bit more complicated. In the next subsection, we shall see that the fact that we want to have equal phases

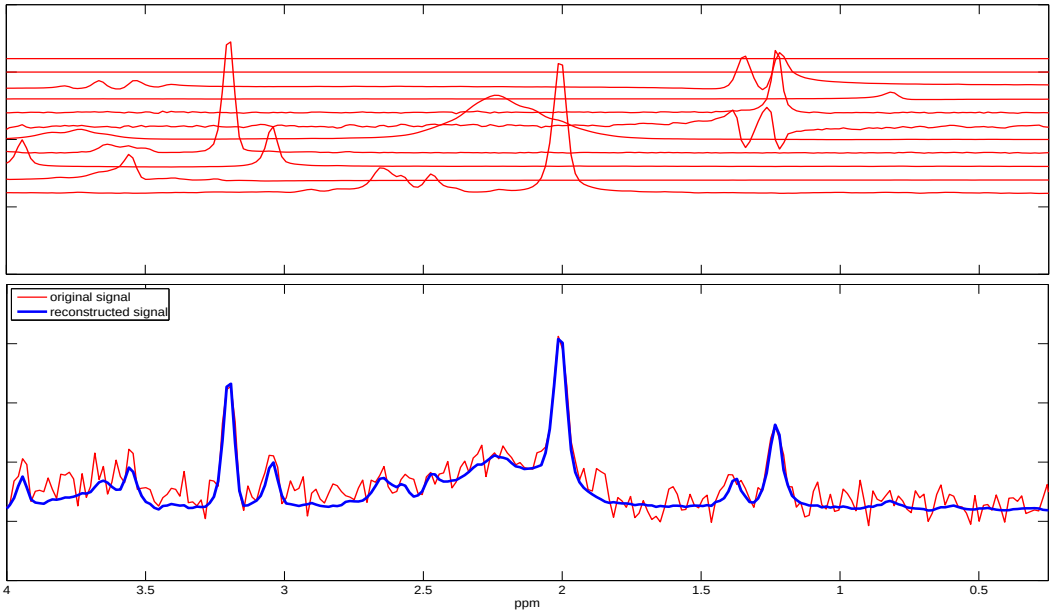


Fig. 1. Fit with non-equal phases. Upper plot shows all the corrected metabolites spectra, which should be summed up in order to yield the reconstructed signal; bottom plot shows the noisy spectrum and the reconstructed spectrum (the less noisy thick line).

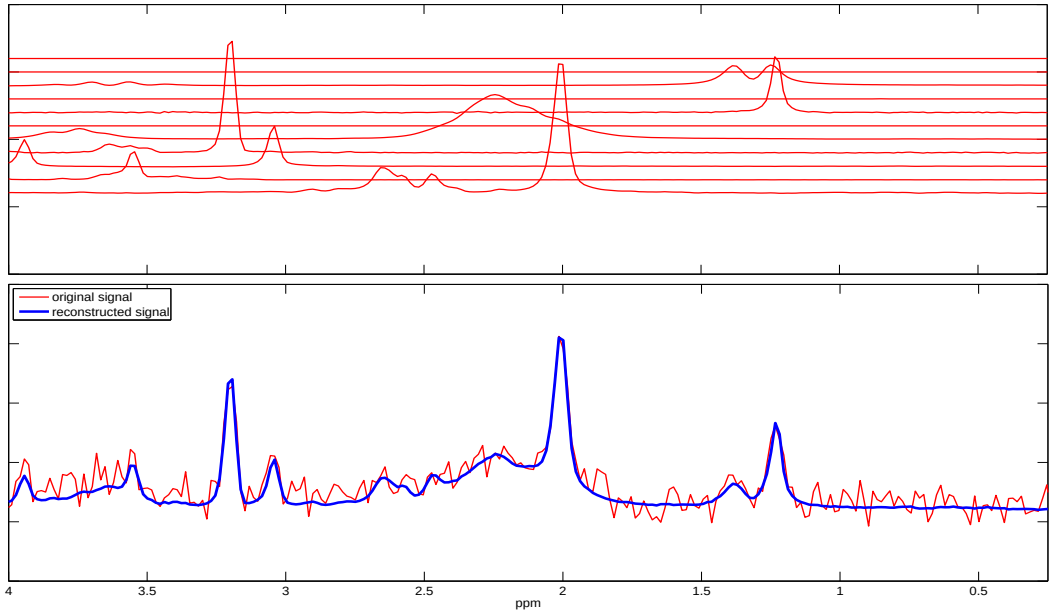


Fig. 2. Fit with equal phases. The upper and bottom plots contain the same elements as in Figure 1. Note that the canceling effect around the value 1.3 ppm is not present, as opposed to Figure 1.

implies that we can no longer use the VARPRO technique with complex linear variables. We must switch to a version where the linear variables are only the real amplitudes, and introduce the unique phase variable ϕ_0 among the

nonlinear variables. However, the real amplitudes must be non-negative in order to be meaningful.

The method developed here can be easily applied in the case when other constraints for the linear variables are imposed. One simple generalization involves linear bound constraints for each individual linear variable; the non-negativity condition is a particular case thereof.

4.2 MRS data model without baseline

4.2.1 The problem formulation is a separable nonlinear least squares with non-negative linear variables

The nonlinear least squares problem formulation (3) becomes

$$\min_{\substack{a_1,\ldots,a_K\in[0,\infty),\phi_0\in(-\pi,\pi),\\ \zeta_1,\ldots,\zeta_K,\eta_1,\ldots,\eta_K\in\Omega}} \quad \frac{1}{m}\sum_{t=t_0}^{t_{m-1}}\left|y(t)-\sum_{k=1}^{K}a_k\exp(j\phi_0)\ (\zeta_k)^t\ (\eta_k)^{t^2}\ v_k(t)\right|^2. \tag{10}$$

Problem (10) shows a mix of real and complex variables. For practical optimization, we need to transform everything to real and to optimize only with respect to real parameters:

$$\begin{aligned}\min_{\substack{a_1,\ldots,a_K\in[0,\infty),\\ \phi_0\in(-\pi,\pi),\\ (d_k,f_k,g_k,e_k)\in\Omega}} \frac{1}{m}\sum_{t=t_0}^{t_{m-1}}\Bigg\{&\left(\mathrm{real}(y(t))-\sum_{k=1}^{K}a_k\mathrm{real}\left(\exp(j\phi_0)\ (\zeta_k)^t\ (\eta_k)^{t^2}\ v_k(t)\right)\right)^2\\ &+\left(\mathrm{imag}(y(t))-\sum_{k=1}^{K}a_k\mathrm{imag}\left(\exp(j\phi_0)\ (\zeta_k)^t\ (\eta_k)^{t^2}\ v_k(t)\right)\right)^2\Bigg\},\end{aligned}$$

where ζ_k and η_k will be substituted with their formulas depending on d_k, f_k, g_k and e_k. The set Ω is used to impose simple linear (in)equality constraints on the nonlinear parameters d_k, f_k, g_k and e_k.

We rewrite the minimization problem in the following compact form, which emphasizes the fact that the parameters $a_1,\ldots,a_K$ appear linearly in the objective function:

$$\min_{\mathbf{a}\geq 0,\mathbf{x}\in\Omega}\frac{1}{m}\left\|\mathbf{y}-\Phi(\mathbf{x})\mathbf{a}\right\|_2^2, \tag{11}$$

where $\mathbf{y}$ is a column vector containing $\mathrm{real}(y(t_0))$,$\mathrm{imag}(y(t_0))$,...,$\mathrm{real}(y(t_{m-1}))$, $\mathrm{imag}(y(t_{m-1}))$, $\mathbf{a}$ is the K-dimensional column vector containing the positive amplitudes $a_1,\ldots,a_K$, $\mathbf{x}$ denotes the vector obtained from all the rest of the

real parameters d_k, f_k, g_k, e_k and ϕ_0, and the $2m \times K$ matrix $\Phi(\mathbf{x})$ has elements of the form:

$$\begin{aligned} \Phi_{(2i-1),k} &= \text{real}\left(\exp(j\phi_0)\,(\zeta_k)^{t_i}\,(\eta_k)^{t_i^2}\,v_k(t_i)\right), \\ \Phi_{(2i),k} &= \text{imag}\left(\exp(j\phi_0)\,(\zeta_k)^{t_i}\,(\eta_k)^{t_i^2}\,v_k(t_i)\right), \end{aligned} \tag{12}$$

$$\text{with} \quad \exp(j\phi_0)\,(\zeta_k)^{t_i}\,(\eta_k)^{t_i^2}\,v_k(t_i) =$$

$$\begin{cases} \exp\left(j\phi_0 + (-d_k + jf_k)t_i + je_k t_i^2\right)\ v_k(t_i), & \text{for Lorentzian lineshapes;} \\ \exp\left(j\phi_0 + jf_k t_i + (-g_k + je_k)t_i^2\right)\ v_k(t_i), & \text{for Gaussian lineshapes;} \\ \exp\left(j\phi_0 + (-d_k + jf_k)t_i + (-g_k + je_k)t_i^2\right)\ v_k(t_i), & \text{for Voigt lineshapes.} \end{cases}$$

Note that without the non-negativity constraint on $\mathbf{a}$, problem (11) is a separable least squares problem, where the optimal linear coefficients $\mathbf{a}^{\text{ls}}(\mathbf{x})$, have a closed-form expression $\mathbf{a}^{\text{ls}}(\mathbf{x}) = \Phi(\mathbf{x})^{\dagger}\mathbf{y}$.

However, the non-negative least squares problem (11), with $\mathbf{x}$ fixed, does not have closed-form solution for $\mathbf{a}$. An effective method for solving non-negative least squares (NNLS) problems is given in [13, Chapter 23]. The algorithm is very much linked to linear-quadratic programming theory and it is an iterative active set primal-dual method where convergence occurs when all elements of the dual vector become negative. Starting with a set of possible primal solutions (basis vectors), the algorithm computes an associated dual vector, and selects at each iteration the worst basis vector solution to be exchanged from the basis set, corresponding to the maximum (positive) element of the dual vector.

If, in other applications, linear bounds or other types of simple constraints on the linear parameters need to be imposed, then the non-negative least squares solver discussed above must be replaced by an appropriate method for the corresponding *constrained linear least squares* problem. For many practical cases (such as linear or quadratic inequality constraints), efficient methods and software implementations are available in the literature.

We denote the optimal non-negative solution of (11) by $\mathbf{a}^{\text{nnls}}(\mathbf{x})$. It seems preferable to optimize (11) only over the variables in $\mathbf{x}$, while estimating $\mathbf{a}$ at every iteration as $\mathbf{a}^{\text{nnls}}(\mathbf{x})$. A conceptual advantage of this approach must be emphasized: if $\mathbf{a}$ were however included together with $\mathbf{x}$ as nonlinear variables into the nonlinear least squares solver, then any value close-to-zero in $\mathbf{a}$ would cause an almost zero column in the Jacobian of the objective function, leading to possibly unreliable numerical computations. The implementation that

optimizes only on $\mathbf{x}$ and uses $\mathbf{a}^{\text{nnls}}(\mathbf{x})$ at each iteration does not suffer from this numerical problem.

The case of almost zero amplitudes is important in our application, since a practitioner might want to use a database of metabolites that has more elements than the actual number of chemicals significantly present in the signal to be quantified. Thus, identifying almost zero metabolite concentrations is an important case that should not be affected by numerical errors.

4.2.2 Function and pseudo-Jacobian evaluation

The nonlinear least squares solver needs implementations for the specific objective function and Jacobian evaluations.

As explained before, the function evaluation is performed by computing the non-negative least squares solution at the current value of $\mathbf{x}$, $\mathbf{a}^{\text{nnls}}(\mathbf{x})$; then, the residual vector is computed simply as $\mathbf{y} - \Phi(\mathbf{x})\mathbf{a}^{\text{nnls}}(\mathbf{x})$.

Unfortunately, the lack of closed-form expressions for $\mathbf{a}^{\text{nnls}}(\mathbf{x})$ implies that we do not have an expression for the true Jacobian of the residual with respect to $\mathbf{x}$. For computational efficiency, we propose to use an adequate pseudo-Jacobian, instead of numerical differentiation methods.

Consider the function $f(\mathbf{x}) = \mathbf{y} - \Phi(\mathbf{x})\mathbf{a}^{\text{nnls}}(\mathbf{x})$. The Jacobian of f with respect to any of the scalar variables x_k (here denoted $\nabla_k f$) has the formula:

$$\nabla_k f = -(\nabla_k \Phi(\mathbf{x}))\mathbf{a}^{\text{nnls}}(\mathbf{x}) - \Phi(\mathbf{x})(\nabla_k \mathbf{a}^{\text{nnls}}(\mathbf{x})). \tag{13}$$

While the matrix $\nabla_k \Phi(\mathbf{x})$ is easily computable, the gradient $\nabla_k \mathbf{a}^{\text{nnls}}(\mathbf{x})$ cannot be computed explicitly. At this point, a numerical differentiation technique can be used in order to estimate the matrix $\nabla \mathbf{a}^{\text{nnls}}(\mathbf{x})$. In Section 5, we show experimental results that are obtained with this approach for Jacobian computation, as well as with the approach described next, using another approximate Jacobian that is cheaper to compute.

Since $\mathbf{a}^{\text{nnls}}(\mathbf{x})$ is in many cases close to the least squares solution $\mathbf{a}^{\text{ls}}(\mathbf{x}) = \Phi(\mathbf{x})^{\dagger}\mathbf{y}$, we can approximate the gradient of $\mathbf{a}^{\text{nnls}}(\mathbf{x})$ with the one of $\mathbf{a}^{\text{ls}}(\mathbf{x})$, yielding:

$$\begin{aligned}
\nabla_k f &\approx -(\nabla_k \Phi)\mathbf{a}^{\text{nnls}} - \Phi \nabla_k \left(\left(\Phi^{\top}\Phi \right)^{-1} \Phi^{\top} \right) \mathbf{y} \\
&= -(\nabla_k \Phi)\mathbf{a}^{\text{nnls}} + \Phi \left(\Phi^{\top}\Phi \right)^{-1} \left[(\nabla_k \Phi)^{\top} \Phi + \Phi^{\top} (\nabla_k \Phi) \right] \left(\Phi^{\top}\Phi \right)^{-1} \Phi^{\top}\mathbf{y} \\
&\approx -\left(\nabla_k \Phi - (\Phi^{\dagger})^{\top} (\nabla_k \Phi)^{\top} \Phi - \Phi\Phi^{\dagger} (\nabla_k \Phi) \right) \mathbf{a}^{\text{nnls}}.
\end{aligned}$$

Moreover, Kaufman's simplification [11] proposes to avoid the complicated computation of $(\Phi^\dagger)^\top (\nabla_k \Phi)^\top \Phi$ and thus only the part

$$-\left(\nabla_k \Phi - \Phi\Phi^\dagger(\nabla_k \Phi)\right) \mathbf{a}^{\text{nnls}} = -(I - \Phi\Phi^\dagger)(\nabla_k \Phi)\mathbf{a}^{\text{nnls}}$$

can be used to compute an approximate Jacobian.

For more details on how to compute the elements of $\nabla_k \Phi$ we refer to Section 3 or [24]. The only addition is that we have a new column in the Jacobian, corresponding to the variable ϕ_0, which is equally treated as a nonlinear parameter. The gradient with respect to ϕ_0 is easily computable, since ϕ_0 only appears in the factor $\exp(j\phi_0)$.

In order to obtain the Jacobian matrix needed in the optimization process, all columns of the type $(\nabla_k \Phi)\mathbf{a}^{\text{nnls}}$ should first be stacked into a matrix $\Delta\Phi$. To complete the Jacobian computation, the product $(I - \Phi\Phi^\dagger) \cdot \Delta\Phi$ should be evaluated. For stable and efficient computation, we make use of the QR decomposition of Φ,

$$\Phi = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

where R is upper triangular, Q is an orthogonal matrix, $R_1 \in \mathbb{R}^{K \times K}$, $Q_1 \in \mathbb{R}^{2m \times K}$, and $Q_2 \in \mathbb{R}^{2m \times (2m-K)}$. Thus, $I - \Phi\Phi^\dagger = I - Q_1 Q_1^\top = Q_2 Q_2^\top$, and then $\left(I - \Phi\Phi^\dagger\right)\Delta\Phi = Q_2 Q_2^\top \Delta\Phi$.

4.3 MRS data model with baseline

4.3.1 The minimization problem formulation

In this case, we augment the optimization criterion (10) to the regularized version that takes into account a smooth baseline reconstructed by penalized splines:

$$\min_{\substack{a_1,\ldots,a_K \in [0,\infty),\ \phi_0 \in (-\pi,\pi), \\ \zeta_1,\ldots,\zeta_K,\eta_1,\ldots,\eta_K \in \Omega,\ \mathbf{c} \in \mathbb{C}^n}} \quad \frac{1}{m} \sum_{t=t_0}^{t_{m-1}} \left| y(t) - \sum_{k=1}^{K} a_k \ (\zeta_k)^t \ (\eta_k)^{t^2} \ v_k(t) - (\mathcal{A}\mathbf{c})(t) \right|^2 + \lambda^2 \mathbf{c}^H D^H D \mathbf{c}, \tag{14}$$

where $\mathcal{A} \in \mathbb{C}^{m \times n}$ is an inverse Fourier transformed spline matrix, the vector $\mathbf{c} \in \mathbb{C}^n$ denotes the spline coefficients, λ is a fixed regularization (penalty) parameter, and the whole penalty term $\lambda^2 \mathbf{c}^H D^H D \mathbf{c}$ is responsible for ensuring a certain degree of smoothness to the frequency-domain baseline.

Using the notation from (11) and the transformation to real of all the complex

elements (subscripted here with an r), this minimization can be written as

$$\min_{\mathbf{a}\geq 0, \mathbf{x}\in\Omega, \mathbf{c}_r\in\mathbb{R}^{2n}} \frac{1}{m} \left\| \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi(\mathbf{x})\mathbf{a} + \mathcal{A}_r\mathbf{c}_r \\ \sqrt{m}\lambda D_r \mathbf{c}_r \end{bmatrix} \right\|_2^2, \tag{15}$$

where $\mathcal{A}_r \in \mathbb{R}^{2m\times 2n}$ is obtained from $\mathcal{A}$ by unfolding all elements into real and imaginary parts and shuffling them such that each odd/even row corresponds to the real/imaginary part of an element of $y(t)$, and each odd/even column corresponds to the real/imaginary part of an element of the spline coefficient vector $\mathbf{c}$, which is also unfolded into the vector of real elements $\mathbf{c}_r \in \mathbb{R}^{2n}$. The matrix D_r is obtained from D, shuffled in the same manner.

The problem (15) is also a separable nonlinear least squares problem, where the linear variables are $\mathbf{a}$ and $\mathbf{c}_r$, and the nonlinear ones are grouped in the vector $\mathbf{x}$. However, the linear amplitudes in $\mathbf{a}$ are non-negatively constrained, while the spline parameters $\mathbf{c}_r$ are free. Moreover, the coefficient matrices that multiply $\mathbf{c}_r$ are independent of $\mathbf{x}$. This allows us to use an explicit optimal solution for $\mathbf{c}_r$, while still using a non-negative least squares solver to optimize $\mathbf{a}$ at each new $\mathbf{x}$.

Solving the optimization problem (15) is done using a nonlinear least squares minimization over $\mathbf{x}$, where for each fixed $\mathbf{x}$ (thus, at every function evaluation), the linear parameters are the optimal solutions of a minimization problem of the type:

$$\min_{\mathbf{a}\geq 0, \mathbf{c}_r\in\mathbb{R}^{2n}} \frac{1}{m} \left\| \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \mathcal{A}_r & \Phi \\ \sqrt{m}\lambda D_r & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}_r \\ \mathbf{a} \end{bmatrix} \right\|_2^2. \tag{16}$$

Using a QR factorization of the matrix $\left[\begin{smallmatrix} \mathcal{A}_r \\ \sqrt{m}\lambda D_r \end{smallmatrix}\right] = ST = \left[\begin{smallmatrix} S_1 & S_2 \end{smallmatrix}\right] \left[\begin{smallmatrix} T_1 \\ 0 \end{smallmatrix}\right]$, with $T_1 \in \mathbb{C}^{2n\times 2n}$, and S_1, S_2 of appropriate sizes, the optimal $\mathbf{c}_r$ is expressed as

$$\mathbf{c}_r^{\text{ls}} = T^\dagger S^\top \left(\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \mathbf{a} \right) = T_1^{-1} S_1^\top \left(\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \mathbf{a} \right),$$

which can be plugged in into (16) such that a non-negative least squares problem in the variable $\mathbf{a}$ only remains to be solved:

$$\min_{\mathbf{a}\geq 0} \frac{1}{m} \left\| S\left(I - TT^\dagger\right) S^\top \left(\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \mathbf{a} \right) \right\|_2^2 \iff \min_{\mathbf{a}\geq 0} \frac{1}{m} \left\| S_2^\top \left(\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \mathbf{a} \right) \right\|_2^2,$$

where we used the fact that $\left(I - TT^\dagger\right) S^T = \left[\begin{smallmatrix} 0 & 0 \\ 0 & I \end{smallmatrix}\right] S^\top = \left[\begin{smallmatrix} 0 \\ S_2^\top \end{smallmatrix}\right]$, and we ignored the multiplication with the orthogonal matrix S (since the norm is invariant

to such an operation).

4.3.2 Function and pseudo-Jacobian evaluation

To evaluate the residual needed in the nonlinear least squares algorithm, it is thus possible to ignore the orthogonal matrix S and to compute instead of $f(\mathbf{x}) = \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \mathcal{A}_r & \Phi(\mathbf{x}) \\ \sqrt{m}\lambda D_r & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}_r^{\text{ls}} \\ \mathbf{a}^{\text{nnls}} \end{bmatrix}$, directly the residual $\tilde{f}(\mathbf{x}) = S_2^\top \left(\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \mathbf{a}^{\text{nnls}} \right)$.

The Jacobian that we need to compute is

$$\nabla_{\mathbf{x}} \tilde{f} = -S_2^\top \nabla_{\mathbf{x}} \left(\begin{bmatrix} \Phi(\mathbf{x}) \\ 0 \end{bmatrix} \mathbf{a}^{\text{nnls}}(\mathbf{x}) \right) = -S_2^\top \begin{bmatrix} \nabla_{\mathbf{x}} \left(\Phi(\mathbf{x}) \mathbf{a}^{\text{nnls}}(\mathbf{x}) \right) \\ 0 \end{bmatrix}.$$

For its approximate evaluation, we need the same tricks as in the case when there is no baseline. One difference is that the new Jacobian is extended with a zero block, which comes from the fact that we augmented $\Phi(\mathbf{x})$ with some blocks that *do not* depend on the nonlinear parameters.

5 Numerical experiments

In this section, we focus on illustrating the fact that the pseudo-Jacobian introduced in Section 4.2.2 performs as good in our optimization problems as the alternative approach of using an approximate Jacobian with numerical differentiation. The latter Jacobian combines an analytical formula with the numerical differentiation of the part that does not have a closed-form expression; this is the part involving the vector $\mathbf{a}^{\text{nnls}}(\mathbf{x})$. For numerical differentiation, we choose a simple forward difference approximation, which involves computing the NNLS solution in as many vectors (neighboring the current $\mathbf{x}$) as there are elements in the vector $\mathbf{x}$.

The computation of $\nabla \mathbf{a}^{\text{nnls}}(\mathbf{x})$ using numerical differentiation and its use within the analytical formula of the full Jacobian (see (13)) is more efficient than using numerical differentiation for the full Jacobian itself. However, the pseudo-Jacobian from Section 4.2.2 is much more computationally efficient, since the NNLS solution is computed only once (at $\mathbf{x}$).

We use five data sets, each consisting of 100 simulated MRS signals. (These

sets are described in more detail in [25] (set 1, 3-6), where they were used to validate – in a clinically relevant way – the performance of the optimization method with non-equal phases from Section 3, implemented in the software package AQSES.)

- Set 1 consists of signals obtained from a metabolite database of 8 components. The model (2) with random values (but extracted from meaningful intervals) for the parameters of interest (amplitudes, damping corrections, frequency shifts, and equal phase correction ϕ_0). No baseline term and no noise are added to the simulation signals in this set.

  This set corresponds to a *zero-residual* nonlinear least squares problem.
- Set 2 from [25] is omitted as irrelevant for the comparisons herein.
- Set 3 is obtained from Set 1 by adding noise terms with a signal-to-noise ratio of 25.
- Set 4 is obtained from Set 1 by adding noise terms with a signal-to-noise ratio of 7.
- Set 5 is obtained from Set 1 by adding simulated (smooth in the frequency domain) baseline terms. This set corresponds to a zero-residual nonlinear least squares problem only in the case when the simulated baseline can be perfectly reconstructed by penalized splines.
- Set 6 is obtained from Set 5 by adding also noise terms.

In Figure 3, we show in a condensed manner the relative errors for all the simulation scenarios described above. The errors were computed with respect to the true values used for building up the simulation sets. The relative error formula is $\|\mathbf{X}^{\text{estimated}} - \mathbf{X}^{\text{true}}\|_F / \|\mathbf{X}^{\text{true}}\|_F$, where we stacked in the matrix $\mathbf{X}^{\text{true}}$ (respectively, $\mathbf{X}^{\text{estimated}}$) all the 100 vectors of true (respectively, estimated) parameters for the 100 simulation examples in each set. The norm $\|\cdot\|_F$ is the Frobenius norm.

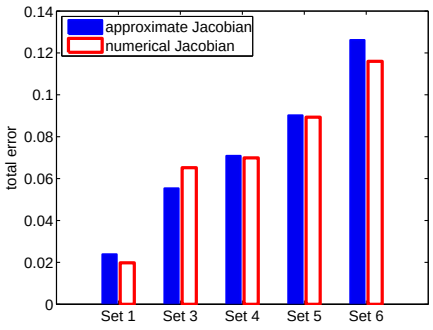


Fig. 3. Comparison of total estimation errors for the two Jacobian variants and the 5 sets of simulations.

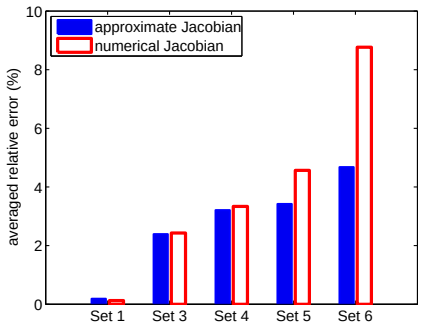


Fig. 4. Comparison of averaged relative estimation errors for the two Jacobian variants and the 5 sets of simulations.

Moreover, Figure 4 views the same results under a different error measure: the

averaged relative error, computed as the mean (over all variables in each set of 100 simulations) of individual relative square errors of the form $\left(x_k^{\text{estimated}} - x_k^{\text{true}}\right)^2 / \left(x_k^{\text{true}}\right)^2$.

Obviously, there is no loss of accuracy related to the pseudo-Jacobian approach compared with the numerical differentiation scheme.

We plot the relative errors $\left|x_k^{\text{estimated}} - x_k^{\text{true}}\right| / \left|x_k^{\text{true}}\right|$ for two individual variables (the amplitude a_1 and the phase ϕ_0) in Figures 5 and 6. The scale is logarithmic and the relative errors (in percentages) are sorted for each of the five simulation sets and the two methods under investigation. The solid lines pertain to the relative errors obtained with the pseudo-Jacobian, and the lines with different markers are the corresponding errors for the numerical differentiation-based Jacobian. The two related curves for each set are very similar. Note also that the estimation errors for the zero-residual problems in Set 1 are under 0.001% for 95% of the simulations. The errors deteriorate when the noise level is increased or when the baseline term is added. However, the errors stay under a reasonable threshold of about $1 - 5\%$.

These numerical results show that the optimization approaches proposed in Section 4 are performing well, and that we can safely make use of the easily computable pseudo-Jacobian.

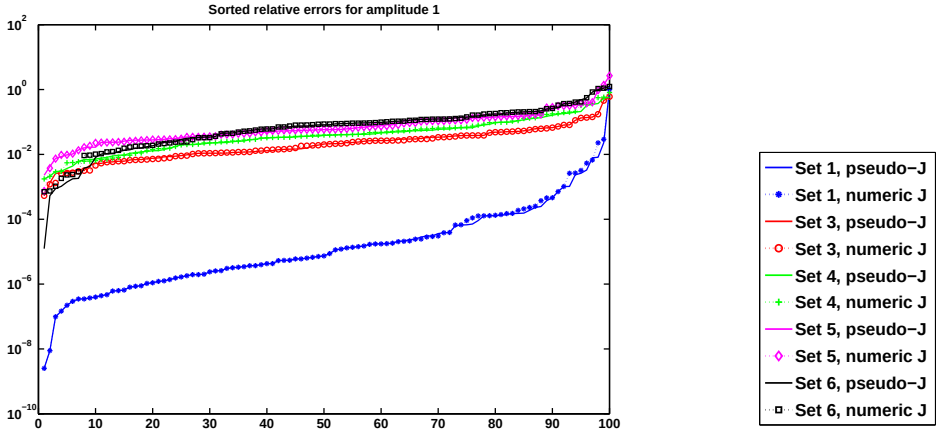


Fig. 5. Relative errors for the variable a_1 in 100 simulations for each of the five testing scenarios. The two proposed approximate Jacobians perform equally good.

6 Conclusions

We have described computational details related to the implementation of several variants of non-standard variable projection algorithms for separable

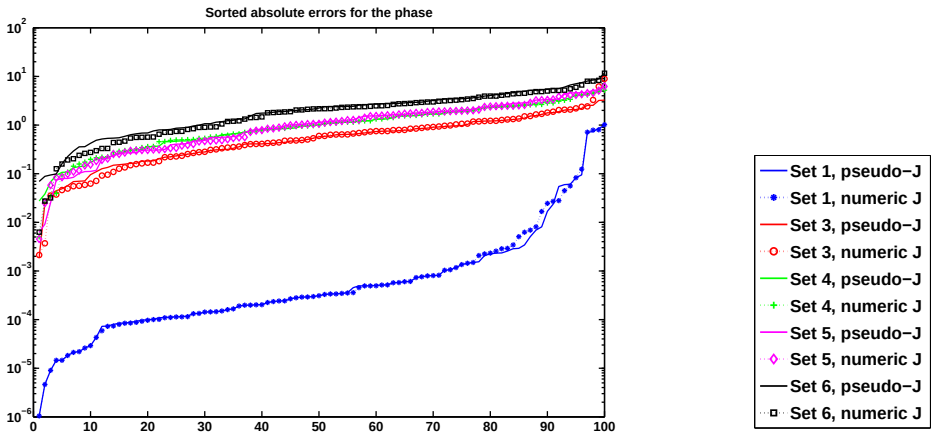


Fig. 6. Absolute errors for the variable ϕ_0 in 100 simulations for each of the five testing scenarios. The two proposed approximate Jacobians perform equally good.

nonlinear least squares. The main issues that we encountered are related to the introduction of constraints on the linear or nonlinear variables. In order to take into account the separability of the minimization criterion, the constrained (or unconstrained) linear subproblem must be solved efficiently and independently at each function evaluation of the outer nonlinear (constrained or unconstrained) minimization.

The described extensions were motivated by optimization problem formulations needed in the quantification of metabolites from short echo-time magnetic resonance spectroscopic signals. All the methods are implemented in the AQSES processing module of the AqsesGUI software package [1], which is a toolbox for analysis and visualization of short echo-time magnetic resonance spectroscopic signals. Quantification results emphasizing the clinical performance of AQSES are presented in [25].

Acknowledgments

We thank Arjan Simonetti for pointing out the need of imposing the phase constraint in order to improve the quantification method AQSES. We thank Jean-Baptiste Poullet for constructing and providing realistic simulation signals.

Diana Sima is a research assistant and Dr. Sabine Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. Research supported by

Research Council KUL: GOA-AMBioRICS, several PhD/postdoc & fellow grants;

Flemish Government: FWO: PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), G.0360.05 (EEG, Epileptic), research communities (ICCoS, ANMMM); **IWT:** PhD Grants;
Belgian Federal Science Policy Office: IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling');
EU: BIOPATTERN, ETUMOUR.

References

[1] AqsesGUI: Software package for accurate quantification of short-echo time NMR signals. www.esat.kuleuven.be/sista/members/biomed/new/, 2005. K.U. Leuven, E.E. Dept. (ESAT-SISTA), Belgium.

[2] C. de Boor. *A Practical Guide to Splines*. Springer, Berlin, 1978.

[3] P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121, 1997.

[4] C. Elster, F. Schubert, A. Link, M. Walzel, F. Seifert, and H. Rinneberg. Quantitative magnetic resonance spectroscopy: semiparametric modeling and determination of uncertainties. *Magnetic Resonance in Medicine*, 53(6):1288–1296, 2005.

[5] D. Gadian. *NMR and its applications to living systems*. Oxford Science publishers, 2nd edition, 1995.

[6] D. M. Gay. A trust-region approach to linearly constrained optimization. In D. F. Griffiths, editor, *Numerical Analysis Proceedings*, pages 72–105, Dundee, 1983. Springer-Verlag.

[7] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10:413–432, 1973.

[8] G. H. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems*, 19(2):1–26, 2003.

[9] J. C. Hoch and A. Stern. *NMR Data Processing*. John Wiley & Sons, 1996.

[10] L. Hofmann, J. Slotboom, B. Jung, P. Maloca, A. Boesch, and R. Kreis. Quantitative ^{1}H-magnetic resonance spectroscopy of human brain: Influence of composition and parameterization of the basis set in linear combination model-fitting. *Magnetic Resonance in Medicine*, 48:440–453, 2002.

[11] L. Kaufman. A variable projection method for solving separable nonlinear least squares problems. *BIT*, 15:49–57, 1975.

[12] L. Kaufman and V. Pereyra. A method for separable nonlinear least squares problems with separable equality constraints. *SIAM Journal on Numerical Analysis*, 15:12–20, 1978.

[13] C. Lawson and R. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, NJ, 1974. Republished by SIAM, Classics in Applied Mathematics Series 15, 1994.

[14] P. Lemmerling, L. Vanhamme, H. J. A. in't Zandt, S. Van Huffel, and P. Van Hecke. Time-domain quantification of short-echo-time in-vivo proton MRS. *MAGMA*, 15:178–179, 2002.

[15] I. Marshall, J. Higinbotham, S. Bruce, and A. Freise. Use of Voigt lineshape for quantification of *in vivo* ^{1}H spectra. *Magnetic Resonance in Medicine*, 37:449–459, 1997.

[16] J. J. Moré. The Levenberg-Marquardt algorithm: Implementation and theory. In G. A. Watson, editor, *Numerical Analysis: Proceedings of the Biennial Conference held at Dundee, June 28-July 1, 1977 (Lecture Notes in Mathematics #630)*, pages 104–116. Springer Verlag, 1978.

[17] K. Opstad, M. Murphy, P. Wilkins, B. Bell, J. Griffiths, and F. Howe. Differentiation of metastases from high-grade gliomas using short echo time ^{1}H spectroscopy. *Journal of Magnetic Resonance Imaging*, 20:187–192, 2004.

[18] S. W. Provencher. Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra. *Magnetic Resonance in Medicine*, 30(6), 1993.

[19] H. Ratiney, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly. Time-domain quantitation of ^{1}H short echo-time signals: Background accommodation. *MAGMA*, 16:284–296, 2004.

[20] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR in Biomedicine*, 17:1–13, 2004.

[21] A. Ruhe and P.-A. Wedin. Algorithms for separable nonlinear least squares problems. *SIAM Review*, 22:318–337, 1980.

[22] D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, 2003.

[23] D. M. Sima and S. Van Huffel. Regularized semiparametric model identification with application to NMR signal quantification with unknown macromolecular baseline. Technical Report 04-229, K.U. Leuven, E.E. Dept. (ESAT-SISTA), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, December 2004. Available from ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/dsima/abstracts/04-229.html.

[24] D. M. Sima and S. Van Huffel. AQSES$_{\mathrm{VP}}$ – description of a variable projection implementation for nonlinear least squares with linear bounds constraints, applied to accurate quantification of short-echo time magnetic resonance spectroscopic signals. Technical Report 05-120, K.U. Leuven, E.E.

Dept. (ESAT-SISTA), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, 2005. Available from ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/dsima/abstracts/05-120.html.

[25] A. Simonetti, J.-B. Poullet, D. M. Sima, B. De Neuter, L. Vanhamme, P. Lemmerling, and S. Van Huffel. An open source short echo time MR quantitation software solution: AQSES. Technical Report 05-168, K.U. Leuven, E.E. Dept. (ESAT-SISTA), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, 2005.

[26] J. van der Veen, R. de Beer, P. Luyten, and D. van Ormondt. Accurate quantification of *in vivo* PNMR signals using the variable projection method and prior knowledge. *Magnetic Resonance in Medicine*, 6:92–98, 1988.

[27] L. Vanhamme, T. Sundin, P. Van Hecke, and S. Van Huffel. MR spectroscopic quantitation: a review of time domain methods. *NMR in Biomedicine*, 14:233–246, 2001.