ORIGINAL PAPER

The variable projection algorithm in time-resolved spectroscopy, microscopy and mass spectrometry applications

Katharine M. Mullen · Ivo H. M. van Stokkum

Abstract Nonlinear least squares optimization problems in which the parameters can be partitioned into two sets such that optimal estimates of parameters in one set are easy to solve for given fixed values of the parameters in the other set are common in practice. Particularly ubiquitous are data fitting problems in which the model function is a linear combination of nonlinear functions, which may be addressed with the variable projection algorithm due to Golub and Pereyra. In this paper we review variable projection, with special emphasis on its application to matrix data. The generalization of the algorithm to separable problems in which the linear coefficients of the nonlinear functions are subject to constraints is also discussed. Variable projection has been instrumental for model-based data analysis in multi-way spectroscopy, time-resolved microscopy and gas or liquid chromatography mass spectrometry, and we give an overview of applications in these domains, illustrated by brief case studies.

Keywords Variable projection · Separable nonlinear least squares ·
Spectroscopy · Microscopy · Mass spectrometry

K. M. Mullen (✉) · I. H. M. van Stokkum
Department of Physics and Astronomy, Faculty of Sciences,
Vrije Universiteit Amsterdam, de Boelelaan 1081,
1081 HV Amsterdam, The Netherlands
e-mail: kate@few.vu.nl

I. H. M. van Stokkum
e-mail: ivo@few.vu.nl

1 Introduction

An unconstrained nonlinear optimization problem

$$\min_{x \in \mathbb{R}^n} \gamma(x) \tag{1}$$

is *separable* if the variables $x \in \mathbb{R}^n$ can be partitioned into $x = (a, z)$ with $a \in \mathbb{R}^p$, $z \in \mathbb{R}^q$, $p + q = n$ such that the sub-problem

$$\min_{a \in \mathbb{R}^p} \gamma(a, z) \tag{2}$$

is *easy to solve* for fixed z. In other words, separable optimization problems are those in which it is possible to partition variables x into two sets a and z such that the variables a can be determined as a function of the variables z, written $a(z)$, in a way that is not too difficult to compute (in a sense that will be defined momentarily). A separable optimization problem can be formulated in the reduced space of z alone, as

$$\min_{z \in \mathbb{R}^q} \gamma(a(z), z) \tag{3}$$

although every evaluation of the objective function requires computation of $a(z)$. What is meant by *easy to solve* or *not too difficult to compute* is simply that solving the separated Problem (3) is in some sense preferable in comparison to solving the unseparated Problem (1), e.g., in terms of execution time, conditioning, or ease of use, since starting values for the parameters a are not required. See, e.g., [36, 38] for further introduction to separable optimization problems.

A common separable optimization problem is the task of fitting a linear combination of n_{comp} nonlinear functions to observations under least squares criteria, that is,

$$\min_{z \in \mathbb{R}^q} \left\| \psi - \sum_{i=1}^{n_{\text{comp}}} c_i(z)a[i] \right\|^2 \tag{4}$$

where ψ is an $m-$vector of observations, c_i is a nonlinear function of z that outputs an $m-$vector, $a[i] \in a$ for $i = 1, 2, \ldots, n_{\text{comp}}$, and $\| . \|$ is, as throughout, the 2-norm. This problem can be written in matrix notation as

$$\min_{z \in \mathbb{R}^q} \| \psi - C(z)a \|^2 \tag{5}$$

where column i of C corresponds to $c_i(z)$.

Golub and Pereyra [10, 11] developed an algorithm for Problem (5) termed *variable projection*, which eliminates the linear parameters a analytically from the problem, and assumes, as we do throughout, that the objective function is twice differentiable. Variable projection has been applied to the solution of instances of Problem (5), termed separable nonlinear least squares problems, in a wide variety of applications, as [12] review. Lukeman [24] traces the history of developments of algorithms for Problem (5), and describes the connection to solving systems of n_{comp} nonlinear equations written as $C(z)a + \psi = 0$.

Golub and LeVeque [9] provide an extension of the algorithm to problems in which the same nonlinear functions are used to describe each column of matrix data Ψ but the conditionally linear parameters a vary per-column, so that the associated optimization problem is

$$\min_{z \in \mathbb{R}^q} \| \Psi - C(z)E^T \|^2 \tag{6}$$

where Ψ is $m \times n$, with column j storing ψ_j, and E is $n \times n_{comp}$, with row j storing $a_j = a_j[1], a_j[2], \ldots, a_j[n_{comp}]$. Golub and LeVeque [9] and Gay and Kaufman [6] refer to Problem (6) as a separable nonlinear least squares problem with multiple right-hand sides. In the time-resolved spectroscopy and microscopy literature (e.g., [3]), Problem (6) is termed the problem of *global analysis*, referring to the assumption that the same nonlinear functions underlie each column of matrix data.

The statistical model associated with Problems (5) and (6) can be written (per-column j in the case of Problem (6)) as

$$\psi_j = C_j(z)a_j + \varepsilon \tag{7}$$

where each element of the noise vector ε represents an independent and identically distributed Gaussian random variable with expectation 0 and constant variance ($NID(0, \sigma^2)$). When matrix data is considered, the function C_j may be different for each j. Furthermore, Ψ in Eq. 6 may be a *ragged* matrix [4], meaning that the vectors ψ_i and ψ_j with $i \neq j$ may each represent a different number of observations, which possibly arise under different experimental conditions. The connection between observations ψ_j for $j = 1, 2, \ldots, n$ is via the vector of nonlinear parameters z. A given function C_j may depend on some or all of the parameters in z. In applications it is common that the observations take the form of vectors $\psi_{j_1}, \psi_{j_2}, \ldots, \psi_{j_K}$ with j representing an independent variable (like a wavelength, or a location) measured many times under different conditions, indexed $1, 2, \ldots, K$. An assumption that is often physically motivated is that the same vector of conditionally linear parameters a underlies $\psi_{j_1}, \psi_{j_2}, \ldots, \psi_{j_K}$ (as in Eq. 14). We will discuss these situations in turn.

A single residual function representing the sum of squared errors that result from using Eq. 7 to model a collection of vectors ψ_j is formed and minimized under the variants of variable projection discussed in Section 2.1. The methodology presented there has allowed the extension of variable projection to d-way datasets, with $d > 2$, and offers many possibilities for the synthesis of data arising in different experiments via a unified model having a single vector z of associated nonlinear parameters.

Here we focus on the use of variable projection for solving instances of Problems (6) and variations thereof. Section 9 describes variable projection in some detail, including implementation strategies, estimation of linear approximation standard errors, and modifications that allow for the inclusion of constraints on the linear parameters. Applications in modeling time-resolved spectroscopy (Section 3), microscopy (Section 4) and mass spectrometry

(Section 5) data are also discussed. Variable projection has been instrumental for model-based data analysis in these areas. The applications to microscopy and mass spectrometry data have been newly developed in the years since the review of [12].

2 Variable projection

Given fixed z and $C(z)$ of full rank, the problem

$$\min_{a \in \mathbb{R}^p} \| \psi - C(z)a \|^2 \tag{8}$$

is solved analytically as $a = C(z)^+ \psi$, where $C(z)^+$ is the Moore-Penrose pseudoinverse $C(z)^+ = (C(z)^T C(z))^{-1} C(z)^T$. The assumption that $C(z)$ has full rank is used throughout, though note that [11] also describe a variable projection algorithm for the case that the columns of $C(z)$ are not linearly independent. Using $a = C(z)^+ \psi$, the objective function to minimize with respect to z for Problem (5) is

$$\begin{aligned} f(z) &= \frac{1}{2} \| \psi - C(z)a \|^2 \\ &= \frac{1}{2} \| (I - C(z)C(z)^+)\psi \|^2 \\ &= \frac{1}{2} \| Q_2 Q_2^T \psi \|^2 . \end{aligned} \tag{9}$$

where the QR decomposition of $C(z) = QR = [Q_1 \ Q_2]R$ is used for numerical stability, so that Q_1 is $m \times n_{\text{comp}}$, Q_2 is $m \times (m - n_{\text{comp}})$, Q is orthogonal, and $R = \begin{bmatrix} R_{11} \\ 0 \end{bmatrix}$, with R_{11} being $n_{\text{comp}} \times n_{\text{comp}}$ and upper triangular [13]. $f(z)$ is the variable projection functional, so named because $I - C(z)C(z)^+$ is the projector on the orthogonal complement of the column space of $C(z)$. Some computational efficiency may be gained by dropping the factor Q_2, and formulating $f(z)$ as $f(z) = \frac{1}{2} \| Q_2^T \psi \|^2$ [6].

Standard algorithms for nonlinear least squares problems, namely Gauss-Newton and Levenberg-Marquardt, can be used to approach $\hat{z}$ that minimize $f(z)$. However, these algorithms require the Jacobian matrix $J = \dfrac{\partial r(z)}{\partial z}$ of the residual vector $r(z) = Q_2 Q_2^T \psi$, (or, dropping the Q_2 factor, $r(z) = Q_2^T \psi$). Two classes of approaches to obtain J may be distinguished: methods that use a finite difference approximation, and methods that rely on an analytical expression. Many widely applied implementations of nonlinear least squares allow J to be determined by either approach, as in, e.g., the modification of Levenberg-Marquardt found in MINPACK [26] or the Gauss-Newton method employed by the R function nls [37].

If a finite difference approach is used to build up an approximation of J, $r(z)$ is repeatedly evaluated for different parameter vectors $\check{z}$ obtained by

perturbing z by an amount h. Using a forward difference approximation, the ith parameter in z is incremented by h, yielding $\check{z}$, and the ith column of J is determined as $J[,i] = \frac{r(\check{z})-r(z)}{h}$, requiring $length(z)$ evaluations of $r(\check{z})$ to calculate an update of J. The associated computational expense is often more than offset by the advantage of not having to derive and compute an analytic expression for J. For example, in the applications described in Sections 3 and 4, the method of determining $C(z)$ is often at least partially numerical, and there is no closed-form expression available for even $C(z)$. In these applications obtaining a closed-form expression for the three-dimensional tensor $\frac{\partial C(z)}{\partial z}$ needed for the determination of J via an analytical expression is difficult at best, and a finite difference approximation is always used. Minimization of the variable projection functional $f(z)$ with Levenberg-Marquardt using a finite difference approximation of J was described by [23].

If an analytical expression for $\frac{\partial C(z)}{\partial z}$ is available, then the method suggested by [11] may be used to determine J per-column i as

$$J[,i] = -\left[\left(C(z)C(z)^+\frac{\partial C(z)}{\partial z[i]}C(z)^+\right) + \left(C(z)C(z)^+\frac{\partial C(z)}{\partial z[i]}C(z)^+\right)^T\right]\psi \tag{10}$$

where we follow the presentation in [12]. Kaufman [19] suggested that this expression for J could be well-approximated by only using the first term, letting

$$J[,i] = -\left(C(z)C(z)^+\frac{\partial C(z)}{\partial z[i]}C(z)^+\right)\psi, \tag{11}$$

thereby introducing only a negligible loss in accuracy and a significant savings in the number of computations required. The Kaufman approximation is discussed in detail in [1, 6, 30].

Once J or an approximation of J has been determined, the standard algorithms for nonlinear least squares calculate the gradient $\bigtriangledown$ of $f(\hat{z})$ as

$$\bigtriangledown f(z) = J^T r(z) \tag{12}$$

and also calculate an approximation for the Hessian $\bigtriangledown^2 f(z)$. $\bigtriangledown f(z)$ and $\bigtriangledown^2 f(z)$ allow determination of a direction and step size to move the current estimates $\hat{z}$ in z-space so that $f(\hat{z})$ is reduced. From the new parameter estimates, the process of determining J and making a new step in z-space is iterated until stopping criteria are met. The particulars of the standard algorithms are found in e.g., [2, 40].

Several results justify and motivate the minimization of $f(z)$ as opposed to the unreduced objective function with parameters $x = (z, a)$. Golub and Pereyra [11] give a proof that the stationary points of $f(z)$ are the same as for the unseparated problem when the rank of $C(z)$ is constant over

z-space. Therefore given $\hat{z}$ that determines either a local or global optimum in $f(z)$, we determine $\hat{a} = C(\hat{z})^{+}\psi$, and arrive at estimates $\hat{x} = (\hat{a}, \hat{z})$ that define a local or global optimum in the objective function of the unseparated problem. Asymptotic convergence analysis under the Gauss-Newton algorithm by [38] shows that variable projection has superlinear convergence whereas methods that alternate between solving the linear and nonlinear problems separately are only linearly convergent. This is because, as [44] show, the separated problem is better-conditioned than the unseparated problem. Obtaining a better-conditioned problem is indeed a primary motivation for minimizing the variable projection functional as opposed to the objective function for the unseparated problem. Krogh [20] provides simple examples in which optimization of an unseparated nonlinear least squares problem results in divergence whereas optimization of the separated problem results in convergence in a small number of iterations. Osborne [35] points to the extraordinary effectiveness of variable projection in least squares problems involving fitting the parameters of a linear combination of real or complex exponential functions, which are ubiquitous in applications and notoriously difficult [16]. Golub and Pereyra [12] suggest that this success is due to the fact that the unseparated exponential fitting problem becomes increasingly ill-conditioned as (and if) the optimal parameters are converged upon, whereas the variable projection functional does not suffer from this problem.

2.1 Implementation

Problem (6) can be reformulated as an instance of Problem (5) by letting $vec(\Psi) = (I_n \otimes C(z))vec(E^T)$, where $\otimes$ is used to denote the Kronecker product. Then variable projection as presented in Section 2 can be applied. However, as [9] realized, this does not take advantage of the special structure of Problem (6), and requires storing and operating on the matrix $(I_n \otimes C(z))$, which is large for large n. Golub and LeVeque [9] thus suggested forming the residual vector associated with Problem (6) as

$$r(z) = \begin{bmatrix} Q_2^T \Psi[, 1] \\ Q_2^T \Psi[, 2] \\ \vdots \\ Q_2^T \Psi[, n] \end{bmatrix} \tag{13}$$

from which $f(z)$ and J can be determined without ever storing or operating on $(I_n \otimes C(z))$.

Mullen and van Stokkum [28] expand on the idea of determining $r(z)$ in a partitioned manner for the description of (ragged) matrix data per-column j using Eq. 7. In the applications they consider, $C_j(z)$ varies per-column j, and the residual vector may be determined as in Eq. 13, but Q_2^T must be calculated for each matrix $C_j(z)$ in $j = 1, 2, \ldots, n$. They also observe that in practice in the physical sciences, observations often take the form of vectors $\psi_{j_1}, \psi_{j_2}, \ldots, \psi_{j_K}$

with j representing an independent variable (like a wavelength, or a location) measured many times under different conditions indexed $1, 2, \ldots, K$. Then the assumption that the same vector of conditionally linear parameters a underlies $\psi_{j_1}, \psi_{j_2}, \ldots, \psi_{j_K}$, is possible to account for by letting

$$\psi_j = \begin{bmatrix} \psi_{j_1} \\ \psi_{j_2} \\ \vdots \\ \psi_{j_K} \end{bmatrix} = \begin{bmatrix} C_{j_1}(z) \\ C_{j_2}(z) \\ \vdots \\ C_{j_K}(z) \end{bmatrix} a_j = C_j(z) a_j \tag{14}$$

where the second subscript on ψ and C is the dataset index. The residual vector associated with using Eq. 14 to model the columns $j = 1, 2, \ldots, n$ of (ragged) matrix data is also determined as in Eq. 13 with the modification that Q_2^T is re-calculated for each matrix $C_j(z)$.

2.2 Standard error estimation

The matrix of first derivatives of the model function with respect to both the nonlinear parameters z and the conditional parameters a_j is

$$J_m = \left[\frac{\partial C_j(z) a_j}{\partial x} \right] = \left[\frac{\partial C_j(z)}{\partial z} a_j \quad C_j(z) \right] \tag{15}$$

Under the assumption that the noise vector ε added to the model $\psi_j = C_j(z) a_j$ is such that every element ε_i is $NID(0, \sigma^2)$, the covariance matrix associated with both z and a is

$$cov \begin{bmatrix} z \\ a_j \end{bmatrix} = \hat{\sigma}^2 \left(J_m^T J_m \right)^{-1} \tag{16}$$

where $\hat{\sigma}^2 = RSS(\hat{x})/df$, RSS is the residual sum of squares and the degrees of freedom $df = (\sum_{j=1}^n length(\psi_j)) - length(z) - (\sum_{j=1}^n length(a_j))$[40].

After writing the residual function as outlined in Section 2.1 and using a standard non-linear least squares implementation to minimize the sum squares of the residual vector with respect to z, $cov(\hat{z})$ is often returned along with $\hat{z}$, whereas $cov(a_j)$ must be determined. Using

$$J_m^T J_m = \begin{bmatrix} \left(\frac{\partial C_j(z)}{\partial z} a_j \right)^T \frac{\partial C_j(z)}{\partial z} a_j & \left(\frac{\partial C_j(z)}{\partial z} a_j \right)^T C_j(z) \\ C_j^T(z) \frac{\partial C_j(z)}{\partial z} a_j & C_j(z)^T C_j(z) \end{bmatrix} \equiv \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{17}$$

we have, from the block matrix inversion theorem found in e.g., [40], Appendix A,

$$\left(J_m^T J_m \right)^{-1} = \begin{bmatrix} X_{11}^{-1} & X_{11}^{-1} X_{12} \\ -X_{21} X_{11}^{-1} & A_{22}^{-1} + X_{21} X_{11}^{-1} X_{12} \end{bmatrix} \tag{18}$$

where $X_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21}$, $X_{12} = A_{12}A_{22}^{-1}$, and $X_{21} = A_{22}^{-1}A_{21}$. Then $\sigma^2 X_{11}^{-1} = cov(\hat{z})$ and we are interested in determining the bottom right block. Since $A_{22}^{-1} = (C_j(z)^T C_j(z))^{-1}$, we have

$$X_{21} = A_{22}^{-1} A_{21} \tag{19}$$

$$= (C_j(z)^T C_j(z))^{-1} C_j^T(z) \frac{\partial C_j(z)}{\partial z} a_j \tag{20}$$

$$= C_j(z)^+ \frac{\partial C_j(z)}{\partial z} a_j \equiv G_j \tag{21}$$

and $X_{12} = A_{12}A_{22}^{-1} \equiv G_j^T$, where G_j consists of columns $C_j^+ \frac{dC_j}{dz_i} a_j$, for each nonlinear parameter z_i. Hence it is possible to write

$$cov(\hat{a}_j) = \sigma^2 \left(C_j^+ C_j^{+T} \right) + G_j cov(\hat{z}) G_j^T. \tag{22}$$

This expression allows determination of $cov(\hat{a}_j)$ for all $j = 1, 2, \ldots, n$ with modest memory resources even when n is large.

2.3 Modification for constraints on the conditionally linear parameters

It may be desirable to subject the linear parameters a and E in Problems (5) and (6), respectively, to constraints. Consider the case that a is constrained to non-negative values. Then Problem (8) is replaced with the non-negative least squares (NNLS) problem

$$\min_{a^* \in \mathbb{R}^p} \| \psi - C(z)a^* \|^2$$

$$\text{subject to } 0 \leq a_i^* \text{ for } i = 1, 2, \ldots, n_{\text{comp}} \tag{23}$$

Problem (23) must be solved in place of $a = C^+(z)\psi$ in the expression for $f(z)$, e.g., with the algorithm by [22], so that Eq. 2 becomes

$$f(z) = \frac{1}{2} \| \psi - C(z)a^* \|^2 \tag{24}$$

Non-negativity constraints on a arise when ψ represents count data. Then the noise statistics are often best represented using the Poisson distribution, but for high-count data may be well-approximated by additive $NID(0, \sigma^2)$ noise, so that parameter estimation may proceed by minimization of $f(z)$. When using a finite difference method to obtain J, a variable projection algorithm that adds non-negativity constraints to the conditionally linear parameters a is obtained by using the definition of $f(z)$ given in Eq. 24 in place of that given in Eq. 2 as [27] discusses. When using an analytical expression for J, an approximate expression based on the Jacobian in the absence of constraints may be used, as [43] discuss.

In the case that the constraints applied to a do not ensure non-negativity, but rather some other property, the NNLS problem used to determine a^* must be replaced with the appropriate constrained optimization problem. The only

restriction on the constraints applied is practical; the separated problem with constraints on a or E should remain easier to solve than the equivalent unseparated problem. Sima and Van Huffel [42] have described the imposition of regularization constraints on a by replacing the least squares problem $a = C(z)^{+}\psi$ in the variable projection functional with $a = (C(z)^{T}C(z) + m\lambda B)^{-1}C(z)^{T}\psi$, where the term $m\lambda B$ is used to impose a certain degree of smoothness on a.

2.3.1 Modification for equality constraints between conditionally linear parameters

Equality constraints that set $a_j[g]$ to zero for component g, or set $a_j[g] = a_j[h]$ for components g and h are often incorporated to make the estimation problem better determined or account for a priori knowledge of the system underlying the observations. Such constraints are common in time-resolved spectroscopy applications where $a_j[g]$ represents the spectrum of component g at wavelength j. In the case of equality constraints that set $a_j[g]$ to zero, we remove column $C_j[, g](z)$ and element $a_j[g]$ from the model $\psi_j = C_j(z)a_j$. This results in a model with fewer free conditionally linear parameters. We refer to such equality constraints as *zero constraints*.

For the case of equality constraints that set $a_j[g]$ to be equal to $a_j[h]$, possibly with a linear scaling factor α_i, we let column $C_j[, h](z)$ be equal to $C_j[, g](z) * \alpha_i + C_j[, h](z)$, and then remove column $C_j[, g](z)$ from the model $\psi_j = C_j(z)a_j$. This also results in a better determined model. Note that α_i may be optimized as a nonlinear parameter.

3 Spectroscopy applications

Variable projection is often used to fit the free parameters of models for measurements of light, which can often be described as a linear superposition of several non-linear functions. The algorithm was first used to estimate the parameters of a model associated with measurements of light by [39], who parametrize the decay of the luminosity of supernovae in time.

Measurements of light resolved with respect to wavelength or wavenumber and one or more independent variables, such as time, are referred to as multiway spectroscopy data. Variable projection was first applied to multiway spectroscopy data by [9], who analyze time-resolved spectroscopy measurements describing the photocycle of bacteriorhodopsin. Nagle et al. [33] also consider data on bacteriorhodopsin, and were influential in spreading the use of variable projection in the biophysics community. Part of the impact of this work can be explained by the authors' elaboration of the use of the nonlinear parameters z to represent a compartmental model for the kinetics, a theme which is further elaborated in Section 3.1. Nagle [31, 32], van Stokkum [46], Nagle et al. [34] further elaborate on the utility of variable projection for describing multiway spectroscopy data with emphasis on compartmental models for kinetics. The utility of compartmental models for describing time-resolved spectroscopy

data representative of complex systems is stressed by [3, 15]. van Stokkum and Lozier [49] describe an in-depth case study in using variable projection to fit a separable nonlinear model for 5-way spectroscopy data (resolved with respect to wavelength, time, temperature, pH and polarization) representing the photocycle of bacteriorhodopsin. van Stokkum et al. [48], van Stokkum [45] review separable nonlinear models for the description of time-resolved spectroscopy data, the free parameters of which are optimized with variable projection.

Golub and Pereyra [12] discuss at length the application of variable projection to optimizing the parameters of models for Nuclear Magnetic Resonance (NMR) spectroscopy data, which often take the form of a linear superposition of complex damped exponentials. The algorithm has been important in this application domain since the influential work of [56]. Note that while the model describes complex data, the parameter estimation problem involves a residual vector and nonlinear parameters in the real domain, as described in detail in e.g., [43] who also describe an extension to variable projection to account for non-negativity constraints on the linear coefficients, motivated by problems in modeling NMR spectra.

Multi-way spectroscopy data of all varieties can be stored as a ragged matrix Ψ, where each column represents a value of a spectroscopic property, and each row represents a value of some other independent variable. To fix ideas, consider data resolved with respect to time and wavelength, which arises in time-resolved spectroscopy experiments. The data often represents n_{comp} components, each with a distinct time profile $C[, j]$ and spectrum $E[, j]$. By the matrix analogue of the Beer-Lambert law for absorption, $C[, j]$ and $E[, j]$ often combine linearly, giving rise to the equation

$$\Psi = C E^T \tag{25}$$

where the $m \times n_{\text{comp}}$ matrix C and the $n \times n_{\text{comp}}$ matrix E represent the data in time and wavelength, respectively. It is also often the case that a parametric description of either C or E, but not both matrices, is available, so that parameter estimation requires solution of an instance of Problem (6) or generalizations thereof.

3.1 Example: Time-resolved fluorescence emission measurements of photosystem I

In order to give an idea of the possibilities that variable projection has opened up in the domain of modeling time-resolved spectroscopy data, we consider a case study in brief. Photosystem I (PS-I) is one of two photosystems in oxygenic photosynthesis, a process by which plants and green algae convert photons into chemical energy. The PS-I core is a distinct functional unit of PS-I. Gobets et al. [7] describe a system consisting of PS-I cores in a buffer excited by a short laser pulse of femtosecond duration. Measurements of the fluorescence
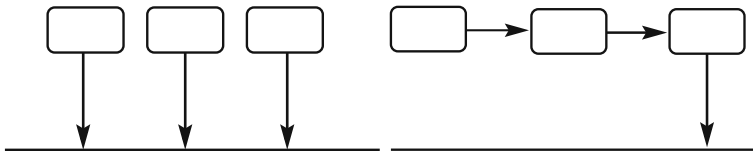
Fig. 1 Two possible compartmental models for a system represented by three components. In the *left model*, the initial excitation populates all three components, which decay in parallel. In the *right model*, the initial excitation populates the leftmost compartment, the decay of which populates the middle compartment, and so on, so that the compartments are said to be arranged sequentially

of the system at many wavelengths and times after excitation are then collected with a synchroscan streak camera in combination with a spectrograph, a technique which [47, 51] review. The observations considered here represent 48 wavelengths equidistant in the interval 626–785 nanometer (nm), and 914 timepoints in the range 0-200 picosecond (ps) after laser excitation, stored as a 914×48 matrix of data Ψ. The goal of data analysis is to describe the kinetics in terms of a model that parametrizes the formation and decay of each distinct state of the underlying system, while solving for the spectra E as linear coefficients.

In some systems, the exciting pulse results in the formation of all possible states simultaneously, as in the left panel of Fig. 1. In other systems, a single state may be formed initially, the decay of which results in the formation of a second state, and so on, for all possible states of the system, as in the right panel of Fig. 1. Alternatively, the states may be related in some other, arbitrarily complicated, way. Physical-chemical principles and ab-initio quantum mechanical calculations typically suggest a few models for the way the possible states are related and how long they persist, and selection of the model best supported by the data and the underlying theory requires estimation of the free parameters of the possible kinetic models, $C(z)$, while the spectra are treated as the linear coefficients E^T in $\Psi = C(z)E^T$. Since a fluorescence spectrum $E[, g]$ represents a count of the relative number of emission photons contributed to the data by component g at some set of wavelengths, negative values of the spectra E are not interpretable, and thus these parameters are determined using the NNLS method discussed in Section 2.3. The reason that the spectra are not often described in terms of parametric model $E(z)$ is that their fine structure is difficult to represent via a function with a small number of free parameters.

A linear compartmental model [8, 40, 48] is used to describe the relationship between states like those shown graphically in Figs. 1 and 2. The behavior of the compartmental model is given as a matrix $C(z)$ in which the concentration of a single compartment in time is represented by a column $C[, g]$, with

$$C(z) = e^{K(z)t} \oplus j(t) \tag{26}$$

where z are free parameters, K is a transfer matrix that encodes the allowable transitions between components, and uses microscopic decay rate parameters
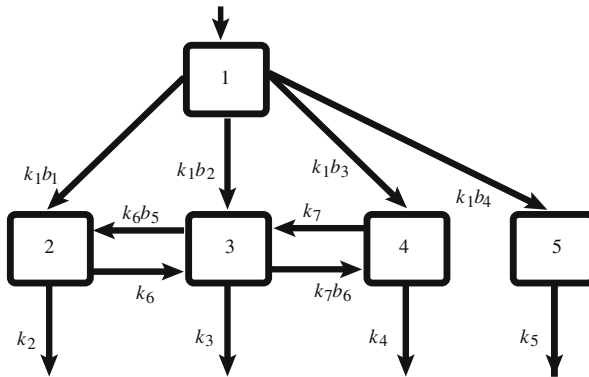
Fig. 2 A compartmental model for the kinetics of PS-I core complexes. The decay rates of the components are parametrized by k_i. Where there is more than one allowable transition out of a compartment, parameters b_i are used to scale the decay rate k_i associated with the compartment. Here values of b_1, b_2, b_3, b_4 are fixed such that $b_1 + b_2 + b_3 + b_4 = 1$ and such that the spectra estimated for compartments 2 – 5 have approximately equal area, which is a physically motivated constraint

$k \in z$ and scaling parameters $b \in z$ to describe the rate at which a state is formed and decays, and t is the vector of times that the rows of $C(z)$ represent. The vector $j(t)$ represents the proportion of the system in each compartment at the initial time point, multiplied by the instrument response function (IRF) $i(t)$. The operator $\oplus$ stands for convolution, and in the case that the IRF is described as a Gaussian distribution or other simple function, it may be performed analytically. Here the IRF is described as a Gaussian with a location parameter that is a function of the wavelength, so that the matrix $C(z)$ must be determined per-wavelength as described in Section 2.1.

The compartmental model shown in Fig. 2 was tested as a possible description for the kinetics of time-resolved spectroscopy data representing PS-I cores by fitting the free parameters of the model with variable projection. Where there is more than one allowable transition out of a compartment, parameters b_i are used to scale the decay rate k_i associated with the compartment. Here values of b_1, b_2, b_3, b_4 are fixed such that $b_1 + b_2 + b_3 + b_4 = 1$ and such that the spectra estimated for compartments 2–5 have approximately equal area, which is expected from physical first principles. The parameters b_5 and b_6 describing the equilibrium between compartments 2 and 3 and compartments 3 and 4 are only possible to estimate by adding zero constraints (as described in Section 2.3.1) to some of the values in the matrix of spectra E. Here zero constraints are applied to all wavelengths of the spectrum for compartment 1, so that it is never emissive, to wavelengths of the spectrum of compartment 2 up to 690 nm, and to wavelengths of the spectrum of compartment 3 up to 697 nm.

The concentration profiles $C(z)$ and spectra E that result from application of this model are shown in Fig. 3. Standard error estimates are shown on the spectra shown in Fig. 3 as vertical bars, and are small. In order to judge

Fig. 3 The concentration profiles $C(z)$ (*upper panel*) and spectra E (*lower panel*) that result from using the compartmental model shown in Fig. 2 to describe the kinetics of PS-I core complexes. The *legend* shown in the *upper panel* applies to the spectra as well. The concentration profile and spectrum of compartment 1 is not shown, since it is non-emissive. Standard error estimates are shown on the spectra as *vertical bars*, and are so small as to be barely visible
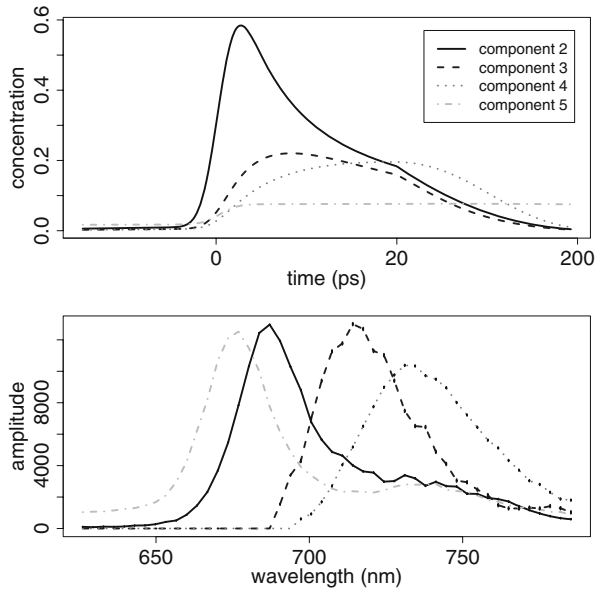
the quality of the fit, traces such as those in Fig. 4 can be inspected, and the singular value decomposition of the residual matrix can be checked for evidence of systematic structure. The fit of the model described here was deemed satisfactory. The implementation of variable projection used is from the R package **TIMP** [28], and a script to reproduce these results is included on the webpage of the package [29].
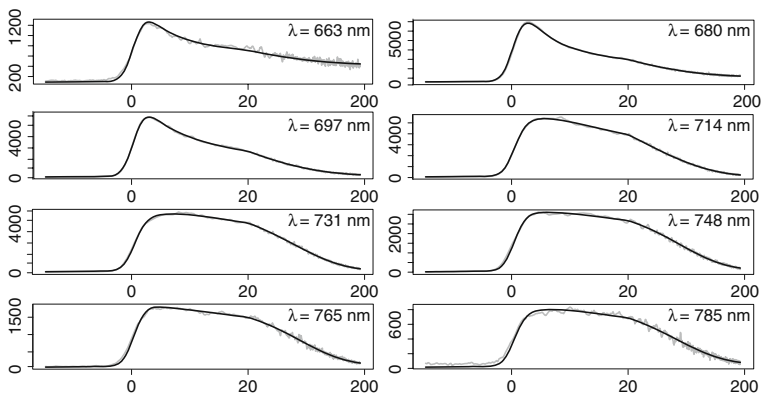
Fig. 4 Columns of a matrix of time-resolved spectroscopy data (*grey*), each representing the wavelength λ in nanometers (nm) shown in the *upper right corner* of each *panel*. The *x-axis* represents time in picoseconds. The model is shown in *black*. Free parameters were fit using variable projection

4 Microscopy applications

Applications of variable projection to modeling time-resolved microscopy data have been developed since the review of applications of the algorithm by [12], and are becoming increasingly important as the technique of Fluorescent Lifetime Imaging Microscopy (FLIM) matures. FLIM is widely applied to detect interactions between fluorescently labeled biological molecules such as proteins, lipids, DNA and RNA, and results in a count of photons detected for many time points, at many spatial locations, often with 250 nanometer spatial resolution and sub-nanosecond temporal resolution. In many FLIM experiments, proteins of interest are genetically tagged with variants of the green fluorescent protein [55].

FLIM experiments give rise to a global analysis problem when the same n_{comp} kinetic processes may be assumed to underlie the fluorescent decay at all locations. The decay of each kinetic process is exponential, but is complicated by the fact that it must be convolved with the time profile of the IRF. The IRF is often not described well by an analytical function with only a few parameters, and it is often necessary to make a measurement of the IRF time profile, and numerically convolve it with that of the exponential decay used to describe each kinetic process. Thus data analysis requires solution of an instance of Problem (6) where each kinetic process g is represented by a column of $C(z)$

$$C[, g] = e^{-k_g t} \oplus i(t) \tag{27}$$

where t is the vector of time points represented by the rows of $C(z)$, $k_g \in z$ is a parameter to be estimated, $\oplus$ indicates convolution and $i(t)$ is the measured time profile of the IRF. In many FLIM experiments, fluorescent dyes give rise to the observed fluorescence, which is measured at the maximum emission wavelength of the dye of interest, and can typically be described by two to four kinetic processes. The amplitude that each kinetic process contributes to the collected image at pixel j (that is, location j) is the row $E[j,]$ that appears in $\Psi = C(z)E^T$, where Ψ is the time-resolved image. The goal of data analysis is to estimate z and E.

Verveer et al. [57] recognized that variable projection could be applied to global analysis problems arising in FLIM data analysis, but did not implement the partitioned technique described in Section 2.1 and therefore were stymied by the large memory resources required. Laptenok et al. [21] studied the utility of variable projection for modeling FLIM data via a number of simulation studies and a control study in estimating the parameters describing the decay of Cyan Fluorescent Protein (CFP).

Note that FLIM data and fluorescence data in general represent a count of the number of photons detected at a given pixel and time. Poisson noise statistics often apply to such non-negative count data. For datasets in which the counts are large, the assumption of $NID(0, \sigma^2)$ noise made by least squares methods is acceptable. However, for datasets in which most counts are not large, optimal estimates under least squares criteria do not well-approximate the true underlying parameter values, an issue that has been studied by [25].

This issue can be addressed to some extent by weighting each data point $\Psi[i, j]$ by $\frac{1}{\sqrt{\Psi[i,j]}}$, but in order to obtain fully correct estimates it would be necessary to develop an analogue of variable projection for the Poisson noise case.

4.1 Example: detection of protein–protein interactions

In this example we consider the simultaneous analysis of multiple FLIM images. Each pixel j in each image $1, 2, \ldots, K$ is modeled using Eq. 7. The nonlinear parameters $k_g \in z$ used to describe the fluorescent decay are estimated using all data included in the simultaneous analysis. To allow a physical interpretation, the vectors $a_{j_1}, a_{j_2}, \ldots, a_{j_K}$, describing the amplitudes of the kinetic processes at pixel j in each dataset are constrained to non-negative values by the NNLS method, and the fluorescence decay rate parameters $k_g \in z$ are also constrained to non-negative values by a simple logarithmic transformation ($z_g = log(k_g)$).

The experiments giving rise to the data involve two proteins known to be homogeneously distributed in the cell nucleus. In one set of experiments, the first protein is tagged with CFP. In the second set of experiments, the first protein is tagged with CFP, and the second protein of interest is tagged with yellow fluorescent protein (YFP). When intracellular dynamics bring the proteins within 1–10 nm of each other, the CFP molecule transfers energy to the YFP molecule. This results in an increase in the decay rate of CFP, which can be observed. The general process in which excited-state energy of a donor fluorophore, like CFP, is non-radiatively transferred to a ground-state acceptor molecule, like YFP, is termed Förster Resonance Energy Transfer (FRET). FRET as measured by FLIM is extensively used to detect protein–protein interactions, as [52] review.

Since CFP acts as a donor in the CFP-YFP FRET pair, we can use the FLIM set-up to measure only the wavelength at which CFP fluoresces, and examine whether the decay rate of CFP increases in the experiment with CFP
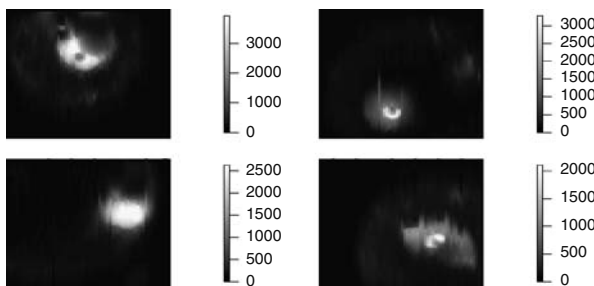
Fig. 5 Intensity images resulting from FLIM experiments that measure fluorescence from cells at the wavelength at which CFP emission is maximum. Each pixel represents the sum of the data at that pixel over the 206 time-points in the entire time-resolved FLIM dataset. The *left plots* represent CFP tagged cells; the *right plots* represent cells tagged with both CFP and YFP. Only pixels representing the cell nucleus are subject to analysis

Fig. 6 Average decay rate
per pixel for the cells shown
in Fig. 5 after parameter
estimation via global analysis.
The higher decay rates in the
datasets at right indicate
protein–protein interactions
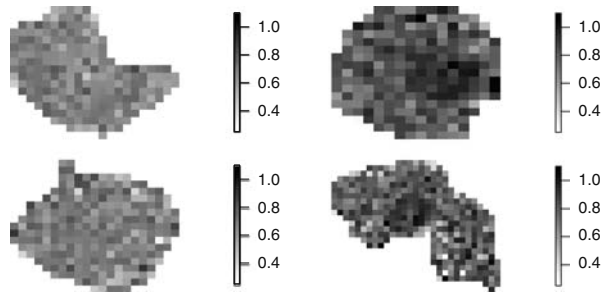
and YFP tags as compared to in the experiment with only CFP tags. Such a
decrease would be interpreted as evidence of FRET, which would imply that
the proteins are often expressed in close proximity.

For the analysis, we select those pixels that represent the nucleus. The two
CFP-only datasets whose intensity images are shown in Fig. 5 (left) are used
in a simultaneous analysis to estimate the associated decay rates. The two
CFP-YFP datasets shown in Fig. 5 (right) are analyzed together in the same
way. The decay of CFP is described by a bi-exponential decay in both pairs of
datasets.

The estimated decay rates for CFP in the first pair of datasets are $\hat{k} =
\{1.53, 0.34\}$ whereas in the second pair of datasets these decay rates are
estimated to be $\hat{k} = \{2.13, 0.44\}$. The average decay rate for pixel j is given
as

$$< k_g > = \frac{\sum_{g=1}^{n_{\text{comp}}} k_g a_j[g]}{\sum_{g=1}^{n_{\text{comp}}} a_j[g]} \tag{28}$$

Figure 6 shows the estimated average decay rate per-pixel for CFP in the CFP-
only datasets (left panels) and in the CFP-YFP datasets (right panels). Clearly,
the bi-exponential decay of CFP is significantly faster in the CFP-YFP datasets
as compared to in the CFP-only datasets. We conclude that there is evidence
of significant FRET, and that the two tagged proteins are often expressed
simultaneously in close proximity.

5 Mass spectrometry applications

The experimental techniques of gas chromatography mass spectrometry
(GC/MS) and liquid chromatography mass spectrometry (LC/MS) measure
the mass spectrum of a complex sample as it elutes from a chromatography
column (see e.g., the book by [14] for a discussion of the experimental tech-
nique). The mass spectrum of a chemical compound acts as its fingerprint: it
allows the compound to be uniquely identified. The time a chemical compound
takes to pass through a liquid or gas chromotographer, in contrast, may not be
unique. Therefore, if more than one chemical compound elutes at the same

time, the mass spectrum at those times represents a mixture of the pure mass spectra of the underlying compounds. Since the purpose of GC/MS and LC/MS experiments is typically to obtain the pure mass spectra of the underlying compounds of a sample in order to perform compound identification and possibly quantification, it is necessary to solve a component resolution problem if compounds are co-eluting.

GC/MS or LC/MS measurements of a single sample can be modeled as $\Psi = CE^T$ where C are the elution profiles of components, and E are the associated mass spectra, which are resolved with respect to mass-to-charge ratio (m/z). For the case that multiple datasets are modeled, the same mass spectra are assumed to be present in all samples $\Psi_1, \Psi_2, \ldots, \Psi_K$, though the elution profiles are usually not assumed to be constant, so that Eq. 14 applies. The inverse problem of estimating C and E, or in the multiple sample case $C_1, C_2, \ldots, C_K$ and E from Ψ or $\Psi_1, \Psi_2, \ldots, \Psi_K$, respectively, is often tackled with the multivariate curve resolution alternating least squares (MCR-ALS) algorithm, as in, e.g., [5, 17, 18, 53, 54].

MCR-ALS is a non-parametric algorithm, and the number of free parameters involved in estimating an elution profile $C[, g]$ is $length(C[, g])$. In situations where the overlap of the elution profiles is large, using a parametric description for $C[, g]$ may be desirable, as [50] have recently explored. This reduces the number of free parameters associated with estimating an elution profile dramatically, since $C[, g]$ can often be described well with an exponentially modified Gaussian with only four parameters. The problem of estimating E and the nonlinear parameters z associated with the model for the elution profiles $C(z)$ is an instance of Problem (6), which may be addressed with variable projection.

5.1 Example: component resolution in GC/MS data

We consider briefly GC/MS measurements Ψ_1 and Ψ_2 of two samples known to represent the same unknown compounds. We model the data using Eq. 14.
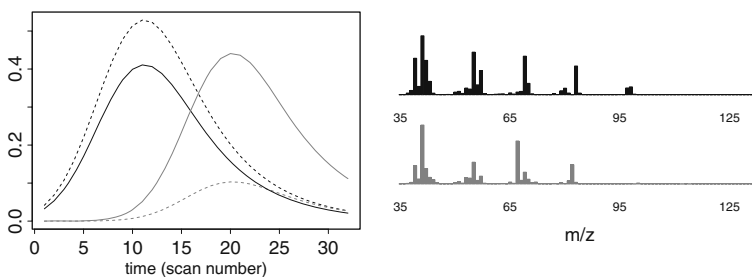
Fig. 7 (*Left*) Estimated concentration profiles for component 1 (*black*) and component 2 (*grey*). *Solid lines* are for Ψ_1, while *dashed lines* are for Ψ_2, and the profiles differ in the two datasets by amplitude only. (*Right*) Mass spectra associated with component 1 (*black*) and component 2 (*grey*), which are assumed to underlie both Ψ_1 and Ψ_2
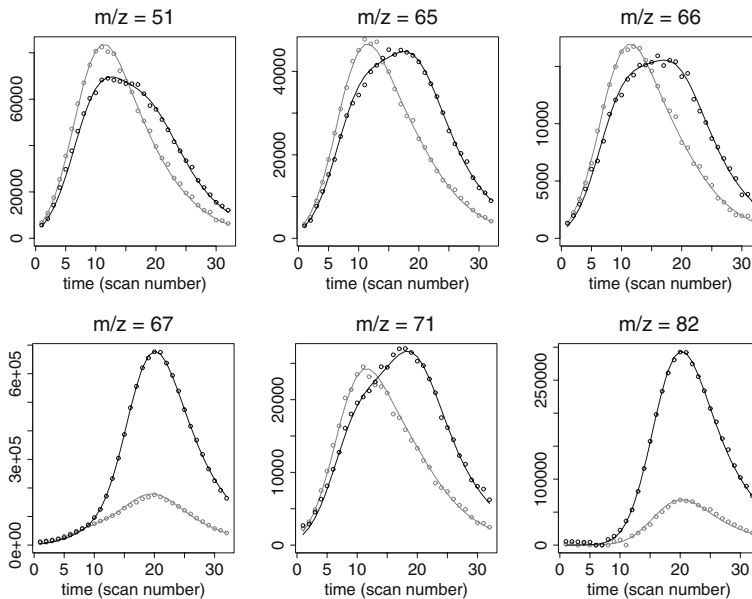
Fig. 8 Data shown as points at selected mass values from two datasets Ψ_1 (*black*) and Ψ_2 (*grey*). The fitted model associated with a simultaneous analysis of both datasets is shown as *lines*

Each column $C_j[, g](z)$ is represented by an exponentially modified Gaussian function with four parameters, for width, location, decay rate, and amplitude. The elution profiles $C_{j_1}[, g]$ and $C_{j_2}[, g]$ for component g in the two datasets at mass j are described with the same parameter values except for the amplitude, (because the elution profiles in these datasets are aligned; for datasets in which the locations of components are not aligned, the location parameter also varies per-dataset). A singular value decomposition of the data yields two singular values that clearly exceed the remaining values on a log scale; we therefore choose to model two components. It is furthermore known that the data at 38 - 44 m/z contain a large peak at every timepoint. We therefore assign these m/z values a very low weight, so that the model for $C_j(z)$ is dependent on the mass j considered. The full model specification is found on the webpage of the **TIMP** package. The estimated mass spectra and concentration profiles that result from fitting this model are shown in Fig. 7. Data and model fit at selected masses are shown in Fig. 8.

6 Future work

We see several avenues for further development of algorithms for separable nonlinear optimization problems. It would be of interest to develop methodology and software for separable optimization problems in which the noise is described by a non-Gaussian distribution. Of most practical importance would be the generalization to the Poisson noise case.

It would also be desirable to develop variable projection implementations to regularize ill-conditioned or over-parametrized models. Golub and Pereyra [12] suggest that this be accomplished via use of a truncated singular value decomposition method to obtain the Jacobian. Recent work by [41] is in this direction. However, to our knowledge no public domain general purpose optimization routines for the case that $C(z)$ is rank deficient have been made available.

7 Conclusions

We have surveyed the variable projection algorithm and its implementation with emphasis on matrix-data fitting problems. Methodology for standard error estimation was discussed, along with extensions to constrain the linear parameters.

The variable projection algorithm has been an important tool for modeling time-resolved spectroscopy data for many years, and we outlined some of the most influential work in this domain. Implementations allowing the flexible specification of models (e.g., the R package **TIMP**) are contributing to the spread of the use of the algorithm, and will continue to do so as such software further matures.

Two new application domains in which variable projection is making an impact are time-resolved fluorescence microscopy and GC/MS mass spectrometry data analysis. Datasets associated with microscopy and mass spectrometry techniques are often large, and more heuristic methods for data analysis have previously been popular largely to avoid the prohibitive execution time and memory needed for model-based analysis. Improvements in computer hardware are changing this and making model-based analysis more feasible. Here again the variable projection algorithm is posed to play a central role, since the parameter estimation task associated with many models applicable to microscopy and mass spectrometry data consists of fitting a linear combination of nonlinear functions.

References

1. Bates, D.M., Lindstrom, M.J.: Nonlinear least squares with conditionally linear parameters. In: Proceedings of the Statistical Computing Section, pp. 152–157. American Statistical Association, New York (1986)

2. Bates, D.M., Watts, D.G.: Nonlinear Regression Analysis and its Applications. Wiley, New York (1988)
3. Beechem, J., Ameloot, M., Brand, L.: Global and target analysis of complex decay phenomena. Anal. Instrum. **14**, 379–402 (1985)
4. Black, P.E.: "Ragged matrix". In: Black, P.E. (ed.) Dictionary of Algorithms and Data Structures. U.S. National Institute of Standards and Technology, Gaithersburg (2004)
5. Garrido, M., Rius, F.X., Larrechi, M.S.: Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. Anal. Bioanal. Chem. **390**, 2059–2066 (2008)
6. Gay, D., Kaufman, L.: Tradeoffs in algorithms for separable and block separable nonlinear least squares. In: Vichnevetsky, R., Miller, J.J.H. (eds.) IMACS '91, Proceedings of the 13th World Congress on Computational and Applied Mathematics, pp. 157–158. Criterion, Dublin (1991)
7. Gobets, B., van Stokkum, I.H.M., Rogner, M., Kruip, J., Schlodder, E., Karapetyan, N.V., Dekker, J.P., van Grondelle, R.: Time-resolved fluorescence emission measurements of photosystem i particles of various Cyanobacteria: a unified compartmental model. Biophys. J. **81**(1), 407–424 (2001)
8. Godfrey, K.: Compartmental Models and Their Application. Academic, London (1983)
9. Golub, G.H., LeVeque, R.J.: Extensions and uses of the variable projection algorithm for solving nonlinear least squares problems. In: Proceedings of the 1979 Army Numerical Analysis and Computers Conference, pp. 1–12 (1979)
10. Golub, G.H., Pereyra, V.: The differentiation of pseudo–inverses and nonlinear least squares problems whose variables separate. Tech. rep., Stanford University, Department of Computer Science (1972)
11. Golub, G.H., Pereyra, V.: The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. SIAM J. Numer. Anal. **10**, 413–432 (1973)
12. Golub, G.H., Pereyra, V.: Separable nonlinear least squares: the variable projection method and its applications. Inverse Probl. **19**, R1–R26 (2003)
13. Golub, G.H., van Loan, C.F.: Matrix Computations, 3rd edn. The Johns Hopkins University Press, Baltimore (1996)
14. de Hoffmann, E., Stroobant, V.: Mass Spectrometry: Principles and Applications, 3rd edn. Wiley, New York(2007)
15. Holzwarth, A.R.: Data analysis of time-resolved measurements. In: Amesz, J., Hoff, A. (eds.) Biophysical Techniques in Photosynthesis, vol. I, chap. 5, pp. 75–92. Kluwer, Dordrecht (1996)
16. Istratov, A.A., Vyvenko, O.F.: Exponential analysis in physical phenomena. Rev. Sci. Instrum. **70**(2), 1233–1257 (1999)
17. Jonsson, P., Johansson, A., Gullberg, J., Trygg, J., Jiye, A., Grung, B., Marklund, S., Sjostrom, M., Antti, H., Moritz, T.: High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. Anal. Chem. **77**(17), 5635–5642 (2005)
18. de Juan, A., Tauler, R.: Chemometrics applied to unravel multicomponent processes and mixtures revisiting latest trends in multivariate resolution. Anal. Chim. Acta **500**, 195–210 (2003)
19. Kaufman, L.: A variable projection method for solving separable nonlinear least squares problems. BIT **15**, 49–57 (1975)
20. Krogh, F.T.: Efficient implementation of a variable projection algorithm for nonlinear least squares problems. Commun. Assoc. Comput. Mach. **17**(3), 167–169, ISSN 0001-0782 (1974)
21. Laptenok, S., Mullen, K.M., Borst, J.W., van Stokkum, I.H.M., Apanasovich, V.V., Visser, A.J.W.G.: Fluorescence lifetime imaging microscopy (FLIM) data analysis with **TIMP**. J. Stat. Softw. **18**(8), 1–20 (2007). http://www.jstatsoft.org/v18/i08/
22. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems. Prentice Hall, Englewood Cliffs (1974)
23. Lawton, W.H., Sylvestre, E.A.: Elimination of linear parameters in nonlinear regression. Technometrics **13**, 461–467 (1971)

24. Lukeman, G.G.: Application of the Shen-Ypma algorithm for separable overdetermined nonlinear systems. Master's thesis, Department of Mathematics and Statistics, Dalhousie University, Canada (1999)
25. Maus, M., Cotlet, M., Hofkens, J., Gensch, T., de Schryver, F.C., Schaffer, J., Seidel, C.A.M.: An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules. Anal. Chem. **73**(9), 2078–2086 (2001)
26. Moré, J.J.: The Levenberg-Marquardt algorithm, implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis. Lecture Notes in Mathematics, vol. 630, pp. 105–116. Springer, New York (1978)
27. Mullen, K.M.: Separable nonlinear models: theory, implementation and applications in physics and chemistry. Ph.D. thesis, Department of Physics and Astronomy, Vrije Universiteit Amsterdam, The Netherlands (2008)
28. Mullen, K.M., van Stokkum, I.H.M.: **TIMP**: an R package for modeling multi-way spectroscopic measurements. J. Stat. Softw. **18**(3), 1–46 (2007). http://www.jstatsoft.org/v18/i03/
29. Mullen, K.M., van Stokkum, I.H.M.: **TIMP**: a problem solving environment for fitting separable nonlinear models in physics and chemistry applications. R package version 1.5.1. http://timp.r-forge.r-project.org/ (2008)
30. Mullen, K.M., Vengris, M., van Stokkum, I.H.M.: Algorithms for separable nonlinear least squares with application to modelling time-resolved spectra. J. Glob. Optim. **38**(2), 201–213 (2007). doi:10.1007/s10898-006-9071-7
31. Nagle, J.F.: Solving complex photocycle kinetics—theory and direct method. Biophys. J. **59**, 476–487 (1991)
32. Nagle, J.F.: Photocycle kinetics: analysis of Raman data from bacteriorhodopsin. Photochem. Photobiol. **54**, 897–903 (1991)
33. Nagle, J.F., Parodi, L.A., Lozier, R.H.: Procedure for testing kinetic models of the photocycle of bacteriorhodopsin. Biophys. J. **38**(2), 161–174 (1982)
34. Nagle, J.F., Zimanyi, L., Lanyi, J.K.: Testing BR photocycle kinetics. Biophys. J. **68**(4), 1490–1499 (1995)
35. Osborne, M.R.: Separable least squares, variable projection, and the Gauss–Newton algorithm. Electron. Trans. Numer. Anal. **28**, 1–15 (2007)
36. Parks, T.A.: Reducible nonlinear programming problems. Ph.D. thesis, Department of Computational and Applied Mathematics, Rice University, USA (1985)
37. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, ISBN 3-900051-07-0. http://www.R-project.org (2008)
38. Ruhe, A., Wedin, P.A.: Algorithms for separable nonlinear least squares problems. SIAM Rev. **22**(3), 318–337 (1980)
39. Rust, B.W., Leventhal, M., McCall, S.L.: Evidence for a radioactive decay hypothesis for supernova luminosity. Nature **262**, 118–120 (1976)
40. Seber, G.A.F., Wild, C.J.: Nonlinear Regression. Wiley, New York (2003)
41. Shen, Y., Ypma, T.J.: Solving rank-deficient separable nonlinear equations. Appl. Numer. Math. **57**(5–7), 609–615, ISSN 0168-9274 (2007)
42. Sima, D.M., Van Huffel, S.: Regularized semiparametric model identification with application to nuclear magnetic resonance signal quantification with unknown macromolecular base-line. J. R. Stat. Soc. B **68**(3), 383–409 (2006)
43. Sima, D.M., Van Huffel, S.: Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification. J. Comput. Appl. Math. **203**, 264–278 (2007)
44. Sjöberg, J., Viberg, M.: Separable non-linear least-squares minimization-possible improvements for neural net fitting. In: Proceedings of the IEEE Workshop Neural Networks for Signal Processing VII, pp. 345–354, Amelia Island, 24–26 September 1997
45. van Stokkum, I.H.M.: Global and target analysis of time-resolved spectra, Lecture notes for the Troisième Cycle de la Physique en Suisse Romande. Tech. rep., Department of Physics and Astronomy, Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands, http://www.nat.vu.nl/ivo/lecturenotes.pdf (2005)

46. van Stokkum, I.H.M., Brouwer, A.M., van Ramesdonk, H., Scherer, T.: Multiresponse parameter estimation and compartmental analysis of time resolved fluorescence spectra: Application to conformational dynamics of charge-separated species in solution. Proc. K. Ned. Akad. Wet. **96**, 43–68 (1993)

47. van Stokkum, I.H.M., Gobets, B., Gensch, T., van Mourik, F., Hellingwerf, K.J., van Grondelle, R., Kennis, J.T.M.: (Sub)-picosecond spectral evolution of fluorescence in photoactive proteins studied with a synchroscan streak camera system. Photochem. Photobiol. **82**, 380–388 (2006)

48. van Stokkum, I.H.M., Larsen, D.S., van Grondelle, R.: Global and target analysis of time-resolved spectra. Biochim. Biophys. Acta **1657**, 82–104 (2004) (Erratum **1658**, 262)

49. van Stokkum, I.H.M., Lozier, R.H.: Target analysis of the bacteriorhodopsin photocycle using a spectrotemporal model. J. Phys. Chem. B **106**(13), 3477–3485 (2002)

50. van Stokkum, I.H.M., Mullen, K.M., Mihaleva, V.V.: Global analysis of multiple gas chromatography–mass spectrometry (GC/MS) data sets: a method for resolution of co-eluting components with comparison to MCR-ALS. Chemometr. Intell. Lab. Syst. (in press)

51. van Stokkum, I.H.M., van Oort, B., van Mourik, F., Gobets, B., van Amerongen, H.: (Sub)-Picosecond Spectral Evolution of Fluorescence Studied with a Synchroscan Streak-Camera System and Target Analysis. In: Aartsma, T.J., Matysik, J. (eds.) Biophysical Techniques in Photosynthesis, Volume II, Series Advances in Photosynthesis and Respiration, vol. 26, pp. 223–240. Springer, Dordrecht, ISBN 978-1-4020-8249-8 (2008)

52. Suhling, K., French, P.M.W., Phillips, D.: Time-resolved fluorescence microscopy. Photochem. Photobiol. Sci. **4**, 13–22 (2005)

53. Tauler, R.: Multivariate curve resolution applied to second order data. Chemometr. Intell. Lab. Syst. **30**, 133–146 (1995)

54. Tauler, R., Smilde, A., Kowalski, B.: Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. J. Chemom. **9**, 31–58 (1995)

55. Tsien, R.Y.: The green fluorescent protein. Ann. Rev. Biochem. **67**, 509–544 (1998)

56. van der Veen, J.W.C., de Beer, R., Luyten, P.R., van Ormondt, D.: Accurate quantification of in vivo 31P NMR signals using the variable projection method and prior knowledge. Magn. Reson. Med. **6**, 92–98 (1988)

57. Verveer, P.J., Squire, A., Bastiaens, P.I.H.: Global analysis of fluorescence lifetime imaging microscopy data. Biophys. J. **78**(4), 2127–2137 (2000)