Separable Nonlinear Least Squares: the Variable Projection Method and its Applications

Gene Golub†Victor Pereyra‡

†Scientific Computing and Computational Mathematics, Stanford University, Stanford, CA. The work of the first author was partially supported by NSF Grant CCR9971010.
‡Weidlinger Associates, 4410 El Camino Real, Los Altos,CA. The work of the second author was partially supported by NSF SBIR Grant 0124874.

Abstract. In this paper we review 30 years of developments and applications of the Variable Projection method for solving separable nonlinear least squares problems. These are problems for which the model function is a linear combination of nonlinear functions. Taking advantage of this special structure the method of Variable Projections eliminates the linear variables obtaining a somewhat more complicated function that involves only the nonlinear parameters. This procedure not only reduces the dimension of the parameter space but also results in a better conditioned problem. The same optimization method applied to the original and reduced problems will always converge faster for the latter. We present first a historical account of the basic theoretical work, its various computer implementations and then report on a variety of applications from electrical engineering, medical and biological imaging, chemistry, robotics, vision, and environmental sciences. An extensive bibliography is included. The method is particularly well suited for solving real and complex exponential model fitting problems, which are pervasive in the applications and are notoriously hard to solve.

I. Introduction and Historical Background

1. Introduction

We consider in this paper nonlinear data fitting problems which have as their underlying model a linear combination of nonlinear functions. More generally, one can also consider that there are two sets of unknown parameters, where one set is dependent on the other and can be explicitly eliminated. Models of this type are very common and we will show a variety of applications in different fields. Inasmuch as many inverse problems can be viewed as nonlinear data fitting problems, this material will be of interest to a wide cross-section of researchers and practitioners in parameter, material or system identification, signal analysis, the analysis of spectral data, medical and biological imaging, neural networks, robotics, telecommunications and model order reduction, to name a few.

The authors published in April 1973 the paper [45] on "Differentiation of Pseudoinverses and Separable Nonlinear Least Squares". This was work initiated in 1971, motivated by and generalizing earlier work of Irwin Guttman, Victor Pereyra and Hugo Scolnik [49], which in turn elaborated and generalized work in Hugo Scolnik's Doctoral Thesis at the University of Zurich [112, 113].

Scolnik's original work dealt with the case in which the nonlinear functions depended on one variable each and were of exponential type (t_i^{α}). In Guttman, Pereyra and Scolnik [49] this was extended to general functions of one variable, while in Golub and Pereyra [45] functions of several variables were considered. In this last paper the authors also developed a differential calculus for projectors and pseudoinverses, and proved that the separation of variables approach led to the same solution set as that of the original problem. Then Ruhe and Wedin [101] extended these ideas to the more general case of two arbitrary sets of variables.

The separability aspect comes from the idea of eliminating the linear variables (i.e., either the coefficients in the linear combinations or one set of the nonlinear variables), and then minimizing the resulting *Variable Projection* (**VP**) functional that depends only on the remaining variables.

Improvements in the algorithm, most specially a simplification due to Kaufman [57], extended by Ruhe and Wedin [101] to the general nonlinear case, made the cost per iteration for the reduced functional comparable to that of the full functional, as proven by the complexity analysis of Ruhe and Wedin.

At first it was thought that the main benefit of the elimination of one set of parameters was that fewer variables needed to be estimated. Numerical experience showed, however, that in most cases the reduced problem converged in fewer iterations than if the same minimization algorithm were used for the full functional. This was later substantiated theoretically for the Gauss-Newton method by the work of Ruhe and Wedin [101], where the asymptotic rates of convergence for the Golub-Pereyra, the Kaufman variant, the NIPALS algorithm, and the unreduced functional are compared by studying the eigenvalues of the corresponding fixed point iteration functions.

Since the reduced functional was more complicated, this reduction in the number of iterations did not guarantee that the total computing time was smaller, but there were classes of problems for which the reduction was dramatic, and in fact, it was clearly beneficial to use the variable projection formulation. Exponential fitting was one of these types of problems. See for instance [45, 102].

The main contributions of the Golub-Pereyra paper were to give a clean, matrix-based formulation of the problem, a differential calculus for matrix functions, orthogonal projectors and generalized inverses, and a modern (for the early 70's) and detailed algorithm to deal with the complexities arising from the Variable Projection formulation, that included an early efficient implementation of an adaptive Levenberg-Marquardt algorithm. They also showed that the reduced problem had the same minima as the original one, provided that the linear parameters were recovered by solving an appropriate linear least squares problem.

We believe that an important part of the impact of this work has come from the fact that a usable, public computer implementation of the algorithm was made available. The 1973 paper is evenly divided between theoretical derivation and implementation details. In a Stanford report with the same name [44], we included a listing of the program **VARPRO** that actually implemented the Variable Projection algorithm and produced the numerical results in the paper. Variants of this original code are still in use in some of the applications that we mention below.

The purpose of this paper is to take the story where we left it after our second paper in 1976 [46], which already contains details on a number of related contributions, mostly clustered around the early 70's. In fact, what we want to stress here is the surprising richness of applications of this idea and its impact in a number of very different fields, with lively developments carrying over to this century.

Thus, we will classify the applications that we have collected through the years roughly by field, giving a more detailed description of the basic problem solved for some selected items and pointing out to whatever new insights of general value have been discovered. We hope that this strategy will help different practitioners find clustered the material of most interest to them, while also calling their attention to possible cross-pollination.

We do not attempt to be comprehensive, and we refer to the excellent bibliographies of many of the quoted papers for connected contributions. The selection is more by expediency than by any attempt to single out some contributions and slight others.

Some interesting trends are observed: during the first few years most of the contributions relate to enhancements, modifications, theoretical results and comparisons. Some of the early applications occur in the area of signal localization, which is still one of the most active fields today. This is not totally surprising given the recent explosion in new telecommunication and mobile applications. Another very active field is that of the modeling and interpretation of Nuclear Magnetic Resonance data, where **VARPRO** has a stellar position. A more recent interesting field of application is that of neural network training.

We will use repeatedly some acronyms that we define here for further reference: **LLS,** Linear Least Squares; **NLLS,** Nonlinear Least Squares; **SNLLS,** Separable Nonlinear Least Squares; **VP**, Variable Projection; **SVD,** Singular Value Decomposition.

2. Separable Nonlinear Least Squares and the Variable Projection Method

Given a set of observations $\{y_i\}$, a separable nonlinear least squares problems is defined in [45] as one for which the model is a linear combination of nonlinear functions that can depend on multiple parameters, and for which the ith component of the residual vector is written as

$$r_i(\mathbf{a}, \alpha) = y_i - \sum_{j=1}^{n} a_j \phi_j(\alpha; t_i).$$

Here the t_i are independent variables associated with the observations y_i, while the a_j, and the k-dimensional vector α are the parameters to be determined by minimizing the functional $||\mathbf{r}(\mathbf{a}, \alpha)||_2^2$, where $\mathbf{r}(\mathbf{a}, \alpha) = \sum_{i=1}^{m} r_i^2(\mathbf{a}, \alpha)$, and $||.||_2$ stands for the l_2 vector norm. We can write this functional using matrix notation as

$$||\mathbf{r}(\mathbf{a}, \alpha)||_2^2 = ||\mathbf{y} - \mathbf{\Phi}(\alpha)\mathbf{a}||_2^2,$$

where the columns of the matrix $\mathbf{\Phi}(\alpha)$ correspond to the nonlinear functions $\phi_j(\alpha;t_i)$ of the k parameters α evaluated at all the t_i values, and the vectors $\mathbf{a}$ and $\mathbf{y}$ represent the linear parameters and the observations respectively.

Now it is easy to see that if we knew the nonlinear parameters α, then the linear parameters $\mathbf{a}$ could be obtained by solving the linear least squares problem:

$$\mathbf{a} = \mathbf{\Phi}(\alpha)^+ \mathbf{y}, \tag{1}$$

which stands for the minimum-norm solution of the linear least squares problem for fixed α, where $\mathbf{\Phi}(\alpha)^+$ is the Moore-Penrose generalized inverse of $\mathbf{\Phi}(\alpha)$. By replacing this $\mathbf{a}$ in the original functional the minimization problem takes the form

$$\min_{\alpha} \frac{1}{2} ||(\mathbf{I} - \mathbf{\Phi}(\alpha)\mathbf{\Phi}(\alpha)^+)\mathbf{y}||_2^2, \tag{2}$$

where the linear parameters have been eliminated.

We define $\mathbf{r_2}(\alpha) = (\mathbf{I} - \mathbf{\Phi}(\alpha)\mathbf{\Phi}(\alpha)^+)\mathbf{y}$, which will be called the Variable Projection (VP) of $\mathbf{y}$. Its name stems from the fact that the matrix in parentheses is the projector on the orthogonal complement of the column space of $\mathbf{\Phi}(\alpha)$, that we will denote in what follows by $P^{\perp}_{\mathbf{\Phi}(\alpha)}$. We will also refer to $\frac{1}{2}||\mathbf{r_2}(\alpha)||_2^2$ as the Variable Projection functional.

This is a more powerful paradigm than the simple idea of alternating between minimization of the two sets of variables (such as the NIPALS algorithm of Wold and Lyttkens [139]), which can be proven theoretically and practically not to result, in general, in the same enhanced performance.

In summary, the Variable Projection algorithm consists of first minimizing (2) and then using the optimal value obtained for α to solve for $\mathbf{a}$ in (1). One obvious advantage is that the iterative nonlinear algorithm used to solve the first minimization problem works in a reduced space and in particular, fewer initial guesses are necessary. However, the main payoff of this algorithm is the fact that it always converges in fewer iterations than the minimization of the full functional, including convergence when the same minimization algorithm for the full functional diverges (see for instance [64]), i.e., the minima for the reduced functional are better defined than those for the full one.

We demonstrated also that the set of stationary points of the original and reduced functionals are the same. This theorem has been reassuring to many practitioners and has been used to derive other theoretical results in similar situations.

A different reason to use the reduced functional is to observe from the above results that the linear parameters are determined by the nonlinear ones, and therefore the full problem must be increasingly ill-conditioned as (and if) it converges to the optimal parameters. That is probably one of the reasons why the important and prevalent problem of real or complex exponential fitting is so hard to solve directly. See for instance [117] for a theoretical discussion of this issue and an interesting application to the training of nonlinear neural networks. Further comments on the basic results can also be found in the textbooks of Seber and Wild [114] and Björck [14].

3. Numerical Methods for Nonlinear Least Squares Problems

General numerical optimization methods can be used to solve **NLLS** problems, but it pays to take into consideration the special form of the goal functional (a sum of squares), just as it pays to take advantage of the special form of separable problems. We review briefly some of the basic concepts that lead to the main numerical methods used today in standard and **SNLLS** problems.

We assume in what follows that the model functions, $\phi_j(\alpha; t)$, are twice differentiable with respect to α. A fundamental quantity for any optimization method for **NLLS** that uses derivatives is the Jacobian matrix $\mathbf{J}(\alpha)$ of the vector of residuals: $\mathbf{r_2}(\alpha)$. This appears when calculating the gradient of the **VP** functional

$$\nabla \frac{1}{2} ||\mathbf{r_2}(\alpha)||_2^2 = \mathbf{J^T}(\alpha)\mathbf{r_2}(\alpha),$$

while its Hessian is given by

$$\nabla^2 \frac{1}{2} ||\mathbf{r_2}(\alpha)||_2^2 = \mathbf{J^T}(\alpha)\mathbf{J}(\alpha) + \sum_{i=1}^{n} r_i(\alpha)\nabla^2 r_i(\alpha).$$

The Gauss-Newton (GN) method for nonlinear least squares can be viewed as a quasi-Newton method in which the Hessian is approximated by $\mathbf{J^T}(\alpha)\mathbf{J}(\alpha)$, while the Levenberg-Marquardt (LM) enhancement adds a positive definite matrix and a coefficient λ in order to combat ill-conditioning. Gauss-Newton is very effective for small-residual problems, since in that case the neglected term is unimportant.

By using a trust-region strategy for step control, needed to stabilize the plain GN and LM iterations and making them more globally convergent, one obtains Moré's implementation [76]. See also [93] for an early proof of convergence of the GN algorithm, [90] for an early proof of convergence for LM, and the Golub-Pereyra paper for a detailed implementation of an adaptive LM method for **SNLLS** problems.

In [91], Pereyra also describes a detailed implementation of an **SVD**-based trust region LM algorithm for **NLLS** problems that appear in large-scale seismic inversion tasks. These are notoriously ill-conditioned problems and often outright singular. It would be worthwhile to consider such an algorithm for the regularization of ill-conditioned **SNLLS** problems.

Instead of describing the implementation in the Golub-Pereyra (GP) paper we go directly to that of Kaufman [57], as described in Gay and Kaufman [41]. In this discussion we ommit the α dependency in all the matrices in order to lighten up the notation. They first observe that GP have proven that the full Moore-Penrose pseudoinverse is not necessary to represent the orthogonal projector, and that a symmetric generalized inverse $\mathbf{\Phi}^-$ satisfying only $\mathbf{\Phi}\mathbf{\Phi}^-\mathbf{\Phi} = \mathbf{\Phi}$ and $(\mathbf{\Phi}\mathbf{\Phi}^-)^{\mathbf{T}} = \mathbf{\Phi}\mathbf{\Phi}^-$ suffices. Thus, the jth column of the Jacobian $\mathbf{J}$ can be written as

$$\mathbf{J}_{.j} = -[(\mathbf{P}^{\perp}_{\mathbf{\Phi}}\frac{\partial \mathbf{\Phi}}{\partial \alpha_j}\mathbf{\Phi}^-) + (\mathbf{P}^{\perp}_{\mathbf{\Phi}}\frac{\partial \mathbf{\Phi}}{\partial \alpha_j}\mathbf{\Phi}^-)^{\mathbf{T}}]\mathbf{y}.$$

Kaufman's simplification for the **VP** functional consists in approximating the Hessian in the Gauss-Newton method by using only the first term in the Jacobian formula: $\mathbf{L} = -(\mathbf{P}^{\perp}_{\mathbf{\Phi}}\frac{\partial \mathbf{\Phi}}{\partial \alpha_j}\mathbf{\Phi}^-)\mathbf{y}$, thus saving in the numerical linear algebra involved at the cost of marginally more function and gradient evaluations. It has been extensively demonstrated, as we indicate below, that savings of up to 25% are achieved by this simplification, making the **VPK** method as cost efficient per iteration as working with the full functional.

Kaufman's argument to justify the small impact of her simplification is that if one writes $\mathbf{J} = \mathbf{K} + \mathbf{L}$, where $\mathbf{K}_{.j} = -(\mathbf{P}^{\perp}_{\mathbf{\Phi}}\frac{\partial \mathbf{\Phi}}{\partial \alpha_j}\mathbf{\Phi}^-)^{\mathbf{T}}$, then:

$$\mathbf{J}^{\mathbf{T}}\mathbf{J} = \mathbf{K}^{\mathbf{T}}\mathbf{K} + \mathbf{L}^{\mathbf{T}}\mathbf{L} + \mathbf{K}^{\mathbf{T}}\mathbf{L} + \mathbf{K}\mathbf{L}^{\mathbf{T}} = \mathbf{K}^{\mathbf{T}}\mathbf{K} + \mathbf{L}^{\mathbf{T}}\mathbf{L},$$

since $\mathbf{K}$ lies in the null space of $\mathbf{\Phi}^{\mathbf{T}}$, while $\mathbf{L}$ lies in the range of $\mathbf{\Phi}$, so only the term $\mathbf{L}^{\mathbf{T}}\mathbf{L}\mathbf{r}(\alpha)$ is being dropped from the exact Hessian.

We point to the reference above for the linear algebra aspects involved in the Kaufman simplification, which combine to produce the quoted 25% reduction in time per iteration.

4. Variations and Related Algorithms

Bates and Linstrom [7] give an interesting statistical interpretation of both the Golub-Pereyra and the Kaufman algorithms. After some analysis and numerical comparisons they conclude that these algorithms are very attractive and provide greater stability than methods for the full functional, besides reducing the dimensionality of the optimization problem.

Later, Ruhe and Wedin [101] demonstrated the asymptotic convergence properties of the two methods, confirming the experimental results that Gauss-Newton always converges in fewer iterations for the reduced functional than for the full one. They also extended **VP** to more general separable problems, where the set of variables splits in two, and where presumably, one of the sets can be easily eliminated. They showed that **VP**, with the Kaufman simplification, had the same cost per iteration as the full functional approach.

Golub and LeVeque [43] extended **VP** to problems with multiple right-hand sides. See also [59].

Böckmann [10] has considered regularization through a trust region algorithm combined with separation of variables and has obtained excellent results in comparison with state of the art solvers applied to the unreduced problem.

In a recent M.Sc. Thesis from Dalhousie University (advisor Patrick Keast), Lukeman [72] extends the application of the Shen-Ypma algorithm [115] to overdetermined systems, establishing the connection with the Golub-Pereyra approach, and he gives a very lucid and accurate description of the early developments. See also [78].

Osborne and collaborators [85, 86] have studied through the years Prony's method, another reduction technique valid for the exponential fitting problem, and they have derived variations that make it more stable.

5. Constrained Problems

Kaufman and Pereyra [58] extended **VP** to problems with separable equality constraints of the form

$$\mathbf{H}(\alpha)\mathbf{a} = \mathbf{g}(\alpha).$$

They show how to reduce this to an unconstrained problem that can be solved by a standard **SNLLS** solver, such as **VARPRO.** They also go on to develop a more efficient algorithm that takes into account the special structure of this problem. See also [27] for further refinements.

Parks [87] considered the basic theory for separable nonlinear programming problems in the spirit of Ruhe and Wedin, that she called reducible. Then she went on to consider specific special cases, including the one corresponding to the equality constrained problem studied by Kaufman and Pereyra, that she called semilinear.

More recently, Conn et al. [25] elaborated on this idea in the context of trust-region algorithms for general nonlinear programming problems, where a subset of variables occur in some predictable functional form that can be used to optimize them more economically.

6. Various Implementations

The **VARPRO** program [44] was written by V. Pereyra. At Stanford University, John Bolstad streamlined the code and improved the documentation of **VARPRO** under the guidance of Gene Golub. He also included the Kaufman simplification and the calculation and output of the covariance matrix, a very important addition that is frequently missing in least squares codes written by numerical analysts. Another graduate student at Stanford, Randy LeVeque, wrote a modified version based on the original **VARPRO** code called **VARP2** which extends the original code to problems with multiple right-hand sides.

Both **VARPRO** and **VARP2** have been publicly available for a long time in netlib, on the port library compiled by David Gay [40]. There one can also find versions of the Gay and Kaufman implementation for unconstrained (nsf.f) and bound constrained **SNLLS** problems, which include the option of using finite differences to approximate the derivatives. A FORTRAN 90 version of **VARPRO** can be found in Alan Miller's repository [75].

Sagara and Fukushima [106] use parameter continuation to solve **SNLLS** problems and report reasonable success in increasing the domain of convergence for certain complicated exponential-polynomial fitting problems.

An advocate of **VARPRO,** Bert Rust, brought the code early on to Oak Ridge National Laboratory and the National Bureau of Standards (NBS, now known as NIST

(National Institute of Standards and Technology)), where it has been used through the years in many applications, some of which we mention below. Rall and Funderlic [97] wrote an interactive front end at Oak Ridge National Laboratory, **INVAR,** to facilitate the use of **VARPRO** and to add more statistics. They also added a finite difference evaluation of the derivatives as an option. Later this code was improved at NBS, including graphical output [140], and it is still in use at NIST.

Another group of scientists who adopted **VARPRO** were those involved in analyzing *in vivo* Magnetic Resonance Spectra (MRS); this line of research begun with the work of van der Veen, de Beer, Luyten and van Ormondt [130] at Delft University in The Netherlands.

A version of **VARPRO** can be found in MRUI (Magnetic Resonance User Interface) [81], a widely used system for MR imaging maintained by the Universidad Autonoma de Barcelona, Spain (a version in MATLAB [82] is also available). This is mostly a non-profit effort financed through grants of the European Union under the project on Advanced Signal Processing for Medical Magnetic Resonance Imaging and Spectroscopy (TMR-CT970160). More than 300 research groups in 40 countries have licensed MRUI.

7. Performance and Comparisons

In the original paper we showed the main performance characteristics of the **VP** method in several problems involving exponentials, Gaussians, and a rational model for fitting iron Mossbauer spectra, as compared to using the full functional. Four different methods were considered: PRAXIS, a general optimization code that does not use derivatives by Richard Brent [17], a Gauss-Newton method with step control, a Marquardt type code with adaptive choice of the regularization parameter that we developed, and a variable metric code by M. Osborne.

The conclusions were in line with what was known at the time and they are still valid today: Gauss-Newton was fastest when it worked, requiring a good initial estimate, while the variable metric code was not competitive. Brent's code is recommended if analytical derivatives are a problem, but otherwise it usually requires more function evaluations. Since the cost per iteration was higher for the **VP** functional, reduction in the number of iterations alone was not a guarantee of less computer time, as shown in some of the examples. Again, these results were dependent on the method used, and for the same problems with the same initial values and final error we obtained different comparative performances for different methods. In these examples, the **VP** approach consistently required less iterations, including problems in which it converged while the iteration for the full functional diverged.

For the exponential fitting problem, the **VP** approach was consistently faster, and the results for Gauss-Newton and Marquardt were comparable and best by far. Since the Kaufman simplification would give an additional 25% edge to the **VP** method, we see that by adjusting the computer times accordingly in [45], then in all cases considered the time performance of **VPK** is better than that corresponding to the minimization of the full functional. As mentioned above, this has also been confirmed theoretically in the paper by Ruhe and Wedin.

That is to say, when **VP** is combined with the Kaufman modification, the cost per iteration for the full and reduced functionals are similar, debunking the original notion that the **VP** functional was considerably more complicated and therefore more expensive to calculate than the original one.

Krogh [64], Kaufman [57], Corradi [26], Gay and Kaufman [41], and Nielsen [84], among others, had similar experiences. The most comprehensive independent studies are those of Corradi who also considered problems with noise, and Gay and Kaufman, who were concerned with proving that the Kaufman modification did not alter the robustness or number of iterations necessary. Unfortunately Corradi does not report computer times.

See also the section on Medical and Biological applications for a discussion of some detailed comparisons within Magnetic Resonance Spectroscopy applications.

II. Applications

8. Applications to Numerical Analysis, Mathematics and Optimization

One of the uses of the derivative of projectors and generalized inverses, as indicated in [45, 131], is in the study of sensitivity to errors in solving linear least squares problems. For instance, the usual perturbation analysis can be extended to the rank deficient case (for rank preserving perturbations): If $\mathbf{A}(\epsilon) = \mathbf{A} + \epsilon\mathbf{B}$ and $||\mathbf{A}|| = ||\mathbf{B}|| =1$, we can estimate the error in the least squares solution of $\mathbf{A}(\epsilon)\mathbf{x}(\epsilon) = \mathbf{b}$, for small ϵ.

By using a Taylor expansion,

$$\mathbf{A}^{+}(\epsilon) - \mathbf{A}^{+}(0) = \epsilon\mathbf{D}\mathbf{A}^{+}(0)\mathbf{B} + O(\epsilon^2),$$

and the formula from [45] for the Fréchet derivative of the generalized inverse with respect to ϵ,

$$\mathbf{D}\mathbf{A}^{+}(0) = -\mathbf{A}^{+}\mathbf{B}\mathbf{A}^{+} + \mathbf{A}^{+}\mathbf{A}^{+\mathbf{T}}\mathbf{B}^{\mathbf{T}}\mathbf{P}_{\mathbf{A}}^{\perp} + \mathbf{Q}\mathbf{B}^{\mathbf{T}}\mathbf{A}^{+\mathbf{T}}\mathbf{A}^{+},$$

where $\mathbf{Q}$ stands for the projector on the orthogonal complement of the *row* space of $\mathbf{A}$ and we have used the fact that $\mathbf{DA}(0) = \mathbf{B}$, we obtain the classical estimate,

$$||\mathbf{x}(\epsilon) - \mathbf{x}(0)|| \leq 2\epsilon||\mathbf{A}^{+}||||\mathbf{x}(\mathbf{0})|| + \epsilon||\mathbf{A}^{+}||^2||\mathbf{r}|| + O(\epsilon^2),$$

where $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$ is the residual vector. Additional perturbation analyses of this kind are considered in [118, 47, 33, 48].

Koliha [63] has extended the differentiability results of the Moore-Penrose inverse in [45] to C*-algebras.

In [123] Trosset considers the problem of computing distances between convex sets of positive definite matrices as a reducible programming problem in the sense of Park, and takes good advantage of the separability of the variables.

Some of the results of [45, 46] are used by Bramley and Winnicka [16] when solving linear inequalities and by Byers et al. [20] when looking for the nearest non-regular pencil of matrices.

Lanzkron and Rose [67] discuss in detail approximate nonlinear elimination. This is the problem of solving large-scale, sparse, nonlinear algebraic equations by a nonlinear Gauss-Seidel method. So, basically there is an external Gauss-Seidel iteration and an internal iterative solver for single nonlinear equations (or systems in the block case), and the old question of how to manage the inner iterations so that the outer iterations preserve some good properties (see [89, Section 3] for an early solution) is revisited. They also discuss the problem of slow convergence, associated perhaps with poor scaling of certain sets of variables, and consider the separation of those variables in order to improve the overall convergence. This is then connected to **SNLLS** problems and the **VP** approach.

9. Applications to Chemistry

The moment-singularity method for the calculation of densities of state in the vicinity of van Hove singularities is considered by Prais, Wheeler, and Blumstein [137, 94]. Because the asymptotic behavior of the modified moments is related to the singular behavior of the density, information about the locations and functional forms of the singularities can be determined directly from the moments themselves. In the 1974 paper a least squares method is described to calculate the locations and exponents of the singularities. This **NLLS** problem is separable and the authors report good success with the 'very fast' **VARPRO** code, when using reasonable initial estimates for the location of the singularities. Several numerical results and comparisons are reported. In the 1986 paper they refine the method and still make good use of **VARPRO,** reporting that they have found this code leading to accurate determination of the singularities when as few as 20 moments are used. Several new calculations, comparisons and validations are included.

Yoon Sup Lee and K. Koo Baeck [68] demonstrate, using a highly positive uranium ion as a test case, that the exact relationship between small and large components of a Dirac spinor in relativistic self-consistent field calculations is not fully satisfied by the kinetic balance condition, even for a two electron system. In order to obtain a basis set for a multiple electron system, the numerical atomic spinors obtained by Dirac-Hartree-Fock calculations are fitted by a given number of Slater-type functions. This **SNLLS** problem is solved using **VP.** The calculations for a uranium ion demonstrate the difference between the exponents for the small and large components, giving numerical evidence that the kinetic balancing is not an exact relationship between small and large components of spinors.

In [11], Beece et al. use a **VP** algorithm in an exponential fitting problem associated with the effect of viscosity on the kinetics of the photochemical cycle of bacteriorhodopsin. Marque and Eisenstein [73] extend this work to consider pressure effects on the photocycle of purple membrane. By considering several kinetic data sets taken at the same temperature and pressure but with different monitoring wavelengths and an exponential model, they are able to use **VARP2** to separate the variables and efficiently solve a problem with multiple right-hand sides. The first to use this method in these type of problems was Richard Lozier [71], who motivated the development of the **VARP2** extension and became a champion in this field for many years (we thank Randy LeVeque for this insight).

In [51], Holmström considers constrained **SNLLS** for chemical equilibrium analysis. He uses his own implementation of a **VP** algorithm that can be found in the LAKE program system [54]. See also the description of the MATLAB based TOMLAB system in [52]. A recent review paper with Pettersson [53] is also noteworthy. In an earlier publication with Andersson and Ruhe [2] the use of SNLS in chemical applications is described in detail.

10. Mechanical Systems

Prells and Friswell [95] consider the application of the **VP** method to the update of finite element models of mechanical systems when the forces applied to them are unknown, an inverse problem. Thus, the forces and the model parameters are estimated from observed values of the system response. The **VP** method leads to the elimination of the unknown forces and results in an extension of the output residual method, which is frequently used to solve this type of inverse problem. The method is exemplified in three applications that are discussed in detail:

- Damage detection of a multi-story building subject to wind excitation;
- Estimation of a foundation model of a rotary machine;
- Model estimation of a steel beam tested by hammer excitation.

Prells and Friswell observe that the inclusion of prior knowledge (constraints) helps stabilize some of these ill-conditioned problems, and that in general the use of **VP** is very successful. They plan to incorporate better regularization techniques and to consider industrial-scale problems in future work.

Westwick et al. [135] consider the nonlinear identification of automobile dynamics when the car is attached to a vibration test-rig. This time-dependent problem is analysed, and after some manipulations it is recognised that in the nonlinear case a **SNLLS** problem has to be solved. They use the algorithms of [117] and report excellent results in a benchmark example of a continuous-time simulation of the vibrations of a quarter-car model. This model consists of two second order linear ordinary differential equations and a nonlinear algebraic constraint that stands for the suspension spring.

Kaufman and collaborators [59, 60] consider large-scale system identification problems arising in the estimation of modal parameters from multiple-driver, multiple-receiver transfer function data in the frequency domain. These transfer functions describe the vibrations of structurally complex objects and they contain many modes. By judiciously applying the **VP** reduction they are able to make very large problems amenable to computation. This is a separable problem with multiple right-hand sides, where the nonlinear parameters are common but the linear ones are not, and therefore the separation of variables produces a tremendous reduction in the dimensionality of the problem.

Similarly, Edrissi et al. [32] have considered the parameter identification problem associated with linear, time invariant state space models by using separation of variables. They cite as the main advantages better conditioning and performance than if the full functional approach is used. Similar results are obtained by Bruls et al. [18] for the linear output error and the innovation model and Wiener system for the non-linear case. Again they claim better conditioning for the reduced approach and they apply it to two real industrial problems.

11. Neural Networks

A very interesting and important application can be found in the training of neural networks. Feedforward neural networks, such as multi-layer perceptrons or radial function networks, are nonlinear parametric models, as clearly explained by Weigl et al. [132, 133, 134] and as discussed in the excellent book by Bishop [9]. In these works, some types of neural networks are conceived as linear combinations of basis functions, for instance, sigmoids, or radial functions, with free parameters on them in the nonlinear adaptive case. Training corresponds to fitting these separable models in the least squares sense by using pre-classified data (a training set) and an optimization algorithm. Using **VARPRO** makes the training an order of magnitude faster than traditional back-propagation algorithms.

This speed in the training phase allows the researcher to design rapidly the neural network that best fit its needs. A transparent description of a simple neural network of this type is provided by Weigl et al. Given an input node, a hidden layer made of two nodes, and an output node, the various steps are summarized as follows:

- The input node accepts the input value x and sends it to the two hidden layer nodes, which are sub-indexed with i.
- The hidden nodes calculate a nonlinear function $g_i(a_i, \theta_i; x) = sigm(a_i x + \theta_i)$, where the function $sigm(z) = \frac{1}{1+e^{-z}}$ is the classic perceptron sigmoid mapping function, a_i is a scalar parameter, and θ_i is a bias.
- Each hidden layer node then sends its function value to the output node, which combines the various values linearly by multiplying them with a coefficient A_i and adding them up:

$$f(x) = \sum_i A_i g_i(a_i, \theta_i; x).$$

Given a training set $(x_j, f(x_j)),\ j = 1, ..., J$, we teach or train the neural network by finding the parameters $(\mathbf{A}, \mathbf{a}, \theta)$ that solve

$$\min_{\mathbf{A},\mathbf{a},\theta} \sum_j (f(x_j) - \sum_i A_i g_i(a_i, \theta_i; x_j))^2,$$

a **SNLLS** problem. The extension to general multi-input, multi-layer and multi-output neural networks is now straightforward.

As in other applications, one advantage of this approach is that the coefficients $\mathbf{A}$ are computed directly, without a learning step, and thus convergence is much faster. Other basis functions can be used instead of the sigmoids. See for instance [83] and [88]. An interesting extension would be to consider very large networks, such as those used in VLSI design, by using block techniques combined with separability [91].

In [34] the authors consider issues of regularization for **NLLS** problems that appear in training feedforward neural networks. It is odd that these authors, who are very familiar with the **SNLLS** technology, seem to have missed the above connection, at least in this contribution. However, some fellow Swedes have not [117] (although they do not seem to be aware of the work of Weigl et al.).

More importantly, besides using **VP** to train neural networks, Sjöberg and Viberg are the only researchers that we have seen mentioning explicitly and demonstrating the fact that, in general, the **VP** functional is better conditioned than the full functional.

12. Parameter Estimation and Approximation

This is a rich application area. Schwetlick [110] gives a comprehensive survey, including **SNLLS.** A traditional application (see [56, 111, 109]) is the fitting of data by splines with free knots. In this problem, a data set $\{x_i, y_i\}$ is given, where the x_i are abscissae and the y_i are noisy measurements of values of an unknown smooth function $f(\mathbf{x})$. One wishes to approximate $f(\mathbf{x})$ in the least squares sense by a function $s(\mathbf{x})$ belonging to a finite dimensional space. Splines provide a powerful choice for the approximation functions. They require for their definition a knot sequence $a \leq t_1 < ... < t_i < ... < t_n \leq b$, where it is advisable to take $n \ll m$.

Thus, the approximation functions are linear combinations of basis functions

$$s(x) = \sum_{j=1}^{n} \alpha_j B_j(x),$$

where the knots appear in the definition of the B-splines. See [30, 4.2] for more details.

If the knots are fixed (usually uniformly distributed), then the fitting problem,

$$\min_{\alpha} \sum_{i=1}^{m} (y_i - \sum_{j=1}^{n} \alpha_j B_j(x_i))^2,$$

is a simple **LLS** one. However, if we consider the basis function knots as unknowns, thus potentially leading to better approximation properties for functions with high variability in the interval of interest, then this is a **SNLLS** problem, which has been successfully solved using **VP.** In fact, for practical applications, constraints need to be imposed to avoid a change in the linear ordering of the knots, but the resulting problem is still separable in the sense of Kaufman and Pereyra. There is an unexploited conexion here with unequally spaced Fast Fourier, Wavelet and Radon Transforms [12].

Francos et al. [37] consider the problem of estimating discrete homogeneous random fields. Again, the separation paradigm helps significantly in solving this problem which involves models with discrete complex exponentials.

An interesting paper that connects **VP** with the manipulation of tensor products [92] is [124], where the approximation in the Frobenius norm of a rectangular matrix by a Kronecker product is discussed.

Bates and Watts [6] consider the problem of multi-response nonlinear least squares estimation $\mathbf{Z}(\theta) = \mathbf{Y} - \mathbf{H}(\theta)$, where the $N \times K$ matrix $\mathbf{Y}$ contains the N observations of K responses, and $\mathbf{H}(\theta)$ is a nonlinear model, where the P-dimensional vector of parameters θ is to be estimated. For this estimation, they chose to minimize with respect to θ the determinant of the matrix $\mathbf{Z^T Z}$

$$min_\theta |\mathbf{Z^T Z}|,$$

where $|.|$ stands for determinant. In order to apply standard optimization methods they derive formulas for the derivatives of this matrix. In fact, the elements g_{θ_p} of the gradient vector of the logarithm of the determinant

$$\nabla \frac{1}{2} \log |Z^T(\alpha) Z(\alpha)| \tag{3}$$

can be written as $g_{\theta_p} = trace(Z^+ Z_{\theta_p})$ [5], and the Hessian can then be obtained from the formulas for the differentiation of the pseudoinverse.

13. Telecommunications, Electrical and Electronic Engineering

A number of applications of **VP** are related to classical and modern telecommunication problems, many of which can be cast as **SNLLS** fitting problems for linear combinations of complex exponentials, where the linear coefficients represent amplitude, while the nonlinear ones are the phases of the signals (plane waves).

Roy and Kailath [100] describe in detail applications to practical signal processing problems. The objective there is to estimate from measurements a set of constant (time-independent) parameters upon which the received signal depends. Among these, high-resolution direction of arrival (DOA) estimation is important in many sensor systems such as radar, sonar, electronic surveillance, and seismic exploration. High-resolution frequency estimation is important in numerous applications, such as the design and control of robots and large flexible space structures. In such problems, the functional form of the underlying signals can often be assumed (e.g., narrow-band plane waves, cisoids). The quantities to be estimated are parameters in these functional descriptions, such as frequencies and directions of arrival for plane waves, or cisoid frequencies.

Several approaches have been developed through the years for solving these problems, including Capon's [23] maximum likelihood and Burg's [19] maximum entropy methods. These methods have significant limitations and Pisarenko was one of the first to consider the structure of the data model to estimate the parameters of cisoids in additive noise using a covariance approach. Schmidt [108] and Bienvenu [13] were the first to exploit correctly the measurement model in the case of a sensor array of arbitrary form. Schmidt's algorithm, MUSIC (MUltiple SIgnal Classification), which according to that author was inspired by the separation of variables technique, has been widely studied and was considered in an MIT study of that time as the most promising high-resolution algorithm. However, MUSIC's success came at a high computational cost that involved a search in parameter space and the storage of array calibration data.

Roy and Kailath developed a new algorithm, called ESPRIT, that dramatically reduced the computational cost and storage for sensor arrays that show what they call displacement invariance. These are arrays where the sensors come in matched pairs with identical displacement vectors.

We will concentrate now on the direction-of-arrival problem for plane waves (i.e., the sensors are in the far field of the source of energy), in an isotropic, homogeneous and non-dispersive medium, so that energy propagates in straight lines. Under those assumptions, the complex signal output of the kth sensor at time t can be written as

$$\mathbf{x}(t) = \sum_{i=1}^{d} s_i(t)\mathbf{a}(\theta_i),$$

where $\mathbf{a}(\theta_i) = [a_1(\theta_i e^{-j\omega_0\tau_1(\theta_i)}), ..., a_m(\theta_i e^{-j\omega_0\tau_m(\theta_i)})]$ is the array steering vector for direction θ_i. Observe that this is a separable model.

Unfortunately, many of the earlier simplified algorithms are ineffective when some of the sources are coherent. This can stem from multipath effects or it can be introduced artificially to impede detection. Kumaresan and Shaw [65] and Cadzow [21] have studied in detail the application of separation of variables to this classical problem. More recently, a number of new algorithms have been developed to consider the more challenging problem of multiple broad-band source location. A variety of least squares modeling methods provide viable means for overcoming the difficulties of coherent sources. As we have seen above, in the standard **LLS** approach, the measured sensor signals are modeled as linear combinations of the source steering vectors.

Cadzow [21] presents a method that models the signal eigenvectors. These are linear combinations of the steering vectors instead of the sensor signals, which introduces a smoothing effect and decreases the computational cost, while the use of the **VP** approach produces significant additional computational savings. As Roy and Kailath [100] indicate, **VP**-type algorithms were considered too expensive until fairly recently, thus justifying the use of the simplified **SVD** based ones. However, the increasing power of modern computers has rendered some of those arguments and simplified methods obsolete, especially in low signal-to-noise situations, where they do not work well.

Friedlander [38] has analyzed the sensitivity of the Maximum Likelihood method for the problem above. This is a separable problem and the sensitivity study involves the differentiation formulas of [45]. This analysis is valuable because the fast algorithms require a knowledge of the antenna array that is hard to come by in real situations, and thus have not been used as often as they deserve.

Talwar et al. [121, 120] have considered the problem of estimating co-channel digital signals using an antenna array when the spatial response of the array is unknown. Traditional techniques, such as MUSIC or ESPRIT, are dependent on the reliability of the array manifold. In the application the authors envision (mobile communications), the array manifold is poorly determined because of a highly variable propagation environment. They consider instead a block **SNLLS** approach, which is both fast and reliable.

Rao and Arun [98] discuss the problem of estimating closely spaced frequencies of multiple, superimposed sinusoids from noisy measurements as a **NLLS** problem. This variant of the problem discussed earlier has wide applications to radio-astronomy, interference spectroscopy, seismic data processing, and MR spectroscopy (which we discuss in detail later). Because of the cost of the computation, as compared to the simplified methods, **SNLLS** is only advisable at low signal-to-noise ratios.

Abel [1] applies separation of variables techniques to problems in underwater acoustic testing and Global Positioning System (GPS) navigation. The GPS problem involves using pseudo-range measures to determine a user's position. GPS consists of a set of satellites transmitting time-stamped signals; pseudo-ranges are formed by comparing satellite signal arrival times according to the user clock, to their transmission times according to GPS time, and then scaling by the propagation velocity. This can be stated as a **SNLLS** problem that is solved by a **VP**-like algorithm.

Zhou, Yip and Leung [144] consider the DOA problem for multiple moving targets by a passive array of sensors, a problem of great interest in communications, air traffic control, and tactical and strategical defense operations. In satellite and personal communication systems it is also advantageous to deploy sensor arrays to reject undesired signals. The classical techniques mentioned above deteriorate rapidly in the presence of moving targets, since they provide poor resolution because of the spread array spatial spectrum caused by the target motion. This deterioration increases with the number of sensors. Zhou et al. propose a maximum likelihood algorithm, where the target motion is assumed to be locally linear, which helps eliminate the spread spectrum effects and provides accurate target dynamical state estimates. Since they use the array signal model for an array of omnidirectional sensors, their approach leads to a separable problem that is solved by a **VP** method.

Liang et al. [69] consider a two-stage hybrid approach for separate co-channel interference reduction and intersymbol interference equalization in a slow Raleigh fading channel, by using a space-time filter. This design involves an optimization problem that is recognized as a (novel) separable problem and it is dealt with by **VP** techniques.

Lilleberg et al. from Nokia Mobile Phones [70] consider a near-far resistant iterative algorithm for multiuser signature sequence delay estimation. **VP** is used to separate the delay and data to be estimated, obtaining a so-called blind maximum likelihood estimator that does not require any knowledge of user amplitudes and data.

Heredia and Arce [50] have considered the splitting of a signal into a set of multilevel components as a **SNLLS** problem. They use as a comparative example a system identification problem for wave propagation through a nonlinear multi-layer channel, where they test the new concepts against Linear, Volterra, and Neural Network alternatives. They show that the realization of piecewise linear filters with unknown thresholds leads to a **SNLLS** problem. In the test problem they verify that the new approach can cope with the difficulties of the problem that trip the Volterra and Neural Network approaches.

Borden [15] considers the problem of obtaining accurate and high-resolution images of complex targets from radar scattering. Simply matching templates is not adequate when the radar images contain erroneous or surplus information. The remedy here is to try to locate some salient features of the scattering volume. In noisy and data-limited environments, the best that can be done is to try to locate the position and strength of scattering centers. Even that process may be difficult in real-life situations. Borden derives a model that includes noise and eventually reduces to a fairly complicated separable model which he solves by **VP.** He develops his own implementation of the reduction and optimization using Householder transformations and gives numerical results for some synthetic data. He concludes that this problem, previously intractable, can now be efficiently and robustly solved by taking advantage of separability. The method is robust against noise contamination and displays good super-resolution capabilities. He also indicates that by using truncated **SVD**'s one could also take care of the problem of automatically determining the correct number of scatterers, in the case that that number is overestimated in the model.

Baum et al. [8] review the singularity expansion method (SEM) for quantifying the transient electromagnetic scattering from targets illuminated by pulsed EM radiation. The SEM theory suggests that the late-time scattered field of a target, interrogated by pulsed EM radiation, can be represented as a sum of natural-resonance modes. Since the excitation-independent natural frequencies depend upon the detailed size and shape of the target, then the full complement of those frequencies is unique to a specified target and provides a potential basis for its identification. The first efforts to extract such natural frequencies from measured target pulse responses were based on Prony's method. However, in the practical low signal-to-noise environment in which this inverse problem occurs, only one or a few modes could be extracted reliably using that inherently unstable algorithm. Although several efforts have improved the reliability of Prony-based methods, realistic problems require a nonlinear approach, and since the problem is separable, **VP** has found another good application in the radar cross-section identification business.

Robertazzi and Schwartz [99] consider the problem of applying Kalman filters to nonlinear regression models. Their idea is to process the data offline in a random instead of a causal manner. They argue that the standard approach processes the data as it arrives and only provides sub-optimal solutions. They use as an example a separable regression problem and its solution through **VP** to show how bad such a sub-optimal solution can be.

In some recent publications, Escovar and Suaya [35, 36] discuss some difficult problems that arise in the design of cutting-edge microcircuits when the clock speed runs over the 1 GHz regime. Critical signals, such as clocks and high-speed data buses, have lengths and transition times comparable to the transit time of light propagation in the medium over those distances. Optimization methods applied to the layout problem based on rigorous solution to Maxwell's equations in the media provide an appealing approach to the design of high-speed data lines, which are the critical components on present day microprocessors. At certain stages of this work, separable nonlinear models appear and they are handled with **VARPRO.** Suaya [119] also considers the role of resistance, inductance and capacitance coupling, and finds again a good use for **VARPRO**.

14. Differential Equations and Dynamical Systems

Wikström [138] has considered in detail the problem of parameter estimation associated with ordinary differential equations arising in chemical kinetics, theoretical biology and population dynamics in ecology. She shows how some of these problems can be profitably attacked via constrained **SNLLS** techniques. A MATLAB toolbox has also been developed.

The parameter estimation problem considered by Wikström is the following:

$$\min_{\mathbf{y}(t,\mathbf{k}),\mathbf{k}} \frac{1}{2}\sum_{i=1}^{M}(\tilde{\mathbf{y}}_{\mathbf{i}} - \mathbf{y}(t_i,\mathbf{k}))^T \mathbf{W}_i(\tilde{\mathbf{y}}_{\mathbf{i}} - \mathbf{y}(t_i,\mathbf{k})) \tag{4}$$

$$\text{s.t.} \quad \dot{\mathbf{y}}(t,\mathbf{k}) = \mathbf{G}(t,\mathbf{y}(t))\mathbf{k}, \quad \mathbf{y}(t_0) = \mathbf{y}_0, \tag{5}$$

where $\mathbf{W}_i$ are diagonal weight matrices, $\tilde{\mathbf{y}}_{\mathbf{i}}$ are observed vector values of the solution of the differential equation $\mathbf{y}(t_i,\mathbf{k})$ at the times t_i, $\mathbf{G}(t,\mathbf{y}(t))$ is a matrix valued function, and the elements of the vector $\mathbf{k}$ are the parameters to be determined. By numerically integrating the differential equations and considering the discrete system in matrix form, Wikström obtains the following equality constrained separable NLLS problem:

$$\min_{\mathbf{y},\mathbf{k}} \frac{1}{2}||\mathbf{W}^{\mathbf{1/2}}(\tilde{\mathbf{y}} - \mathbf{y})||^2 \tag{6}$$

$$\text{s.t.} \ \mathbf{y}_i = \mathbf{y_0} + \mathbf{H_i}(\mathbf{y})\mathbf{k}, \quad \mathbf{y}(t_0) = \mathbf{y}_0, \tag{7}$$

where $\mathbf{H_i}(\mathbf{y})$ is the result of numerically integrating $\mathbf{G}(t,\mathbf{y}(t))$ from t_0 to t_i.

15. Environmental Sciences and Time Series Analysis

Bert Rust and collaborators have used **VARPRO** in a number of applications related to modeling time series in environmental application [62, 103, 104, 105]. In their most recent and most comprehensive contribution, they consider the inverse modulation of global fossil fuel production $P(t)$ by variations in Northern Hemispheric temperatures $T(t)$. They propose a number of newmodels using some additional data. As an example, a delay differential equation they propose is

$$dP/dt = (\alpha - \beta[T(t) - T(t-\tau)]/\tau)P,$$

where α, β, and the lag parameter τ are to be determined by fitting. The function $T(t)$ is obtained as an optimal smoothing cubic spline fit of the temperature data provided by Jones et al. [55], where the smoothing parameter was chosen to minimize the generalized cross-validation statistics [42]. All the fittings in the paper are done with **INVAR,** the interactive version of **VARPRO** at NIST.

See also [61] for an application that strips a time series of its trend, which can be a combination of a polynomial, exponential, autoregressive, and sinusoidal terms. As an illustration, a purely mathematical fit to the Mauna Loa carbon dioxide monthly averages from 1958 to 1976 is performed. The following model is fitted using **VARPRO**:

$$y(t) = a + de^{\alpha t} + f_1 \sin[2\pi/12(t+\gamma_i)] + f_2 \sin[2\pi/6(t+\gamma_2)] + \rho_1 \sin[2\pi/\tau_1(t+\theta_1)] + \rho_2 \sin[2\pi/\tau_2(t+\theta_2)]. \tag{8}$$

16. Robotics and Vision

In [122] Tomasi and Shi consider the problem of determining the direction of heading from image deformations, which is important in applications to robotics and vision. They show how to eliminate the unknown depth values by using **VP.** As observed earlier in other applications, they report better defined minima for the reduced functional, leading to a more reliable solution. The method also degrades gracefully with increasing uncertainty.

In related work, Chiuso et al. [24] consider the problem of estimating spatial properties of a three-dimensional scene from the motion of its projection onto a two-dimensional surface, such as the retina. They address this problem, which has been studied for more than 20 years, from the point of view of an engineer trying to use conventional algorithms that have been tested on controlled sequences of images and do not behave as advertised in real-life applications. They indicate that the problem is noise and show to handle it in a robust manner. Another problem is the trade-off between optical-flow/feature-tracking, which requires image-motion of a few pixels to work well. This makes the problem of getting structure from motion very hard, since then the average image-motion is comparable with the localization error. Again, a separation of variables argument allows them to simplify the solution and provides a more robust alternative in the presence of noise.

Radke et al. [96] consider the problem of estimating projective transformations associated with standard problems in image processing and computer vision. The estimation problem leads to the minimization of a nonlinear functional of eight parameters, which, through a separation argument, can be reduced to a problem in only two variables.

Gardner and Milanfar [39] consider several algorithms for reconstructing convex bodies from brightness functions. One of the algorithms results in a bound constrained **SNLLS** problem. Their current implementation uses the optimization toolbox from MATLAB, without taking advantage of the separability, and results in an algorithm that is too slow to be used in three dimensions. The authors conjecture that they could do much better by taking into account the separability.

Yen and Petzold have considered in a series of papers [141, 142, 143] a coordinate-splitting formulation of the equations of motion for multibody systems, which is very effective when solving certain nonlinear, highly oscillatory problems. The problem leads to an index-3 differential-algebraic system, involving generalized coordinates q and Lagrange multipliers λ :

$$\begin{aligned} \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} - \mathbf{f}(\mathbf{q}, \dot{\mathbf{q}}, t) + \mathbf{G}(\mathbf{q})^{\mathbf{T}}\lambda &= 0, \\ \mathbf{g}(\mathbf{q}) &= \mathbf{0}. \end{aligned} \tag{9}$$

It is possible to eliminate the Lagrange multipliers by choosing an appropriate annihilation matrix $\mathbf{P}(\mathbf{q})$, such that $\mathbf{P}(\mathbf{q})\mathbf{G}^{\mathbf{T}}(\mathbf{q}) = \mathbf{0}$. Writing (9) in first-order form and performing some additional manipulations yields a stabilized index-2 system. Premultiplying this system by the matrix $\mathbf{P}(\mathbf{q})$ results in the system

$$\begin{aligned} \mathbf{P}(\mathbf{q})(\dot{\mathbf{q}} - \mathbf{v}) &= \mathbf{0}, \\ \mathbf{P}(\mathbf{q})(\mathbf{M}(\mathbf{q})\dot{\mathbf{v}} - \mathbf{f}(\mathbf{v}, \mathbf{q}, \mathbf{t})) &= \mathbf{0}, \\ \mathbf{G}(\mathbf{q})\mathbf{v} &= \mathbf{0}, \\ \mathbf{g}(\mathbf{q}) &= \mathbf{0}. \end{aligned} \tag{10}$$

The results on differentiation of projectors in [45] are now combined with discretization via a k-th order BDF formula to yield a system of nonlinear difference equations.

Dutre et al. [31] discuss parallel kinematic structures encountered in the design of robotic manipulators. They consider an analytical description of the velocity closure equations and their Jacobian matrix, i.e., the linear mapping from driving joint velocities to end effector velocities or twist. Since available numerical methods are always more efficient than this analytic approach, the justification for the study is to provide more geometrical insights on the problem. The analytic expressions of the Jacobian and higher derivatives also open up new approaches in dynamics, acceleration analysis, calibration, and manipulability assessment of parallel robotic manipulators. After deriving the analytical velocity closure and the dependency matrix, the authors proceed to the calculation of the Jacobian matrix. The time derivative of the dependency matrix requires the differentiation of a pseudoinverse and the results of [45] are used to obtain closed analytical expressions.

17. Medical and Biological Imaging

Magnetic Resonance (MR) in liquids and solids was discovered over 50 years ago by Ed Purcell (Harvard) and Felix Bloch (Stanford), who shared the Nobel Prize in Physics in 1952 for their work. MR is now a fundamental analytical tool in synthetic chemistry, plays an important role in biomedical research [22], and it has revolutionized modern radiology and neurology. Its applications outside the laboratory or medical clinic are as diverse as oil-well imaging, food analysis, and the detection of explosives.

A typical MR experiment involves placing the sample under study in a strong magnetic field, which forces the magnetic moments or spins of all the nuclei in the sample to line up along the main applied field and precess around this direction. The spins precess at the same frequency but with random phases. Pulses of radiofrequent (RF) magnetic fields are then applied that disturb the spin alignments but makes the phases coherent and detectable. As this state precesses in the magnetic field, the spins emit radiofrequency radiation that can be analysed to reveal the structural, chemical and dynamical properties of the sample. The idea of applying strong RF pulses is due to H.C. Torrey and E.L. Hahn. Higher and higher magnetic fields of the order of teslas have been used in order to increase the sensitivity of the method until just recently, when researchers in the USA and Germany have shown that MR can be performed with fields in the microtesla range [74], by pre-polarizing the nuclei and using a superconducting quantum interference device (SQUID) to detect individual flux quanta. In a magnetic field of 1.8 microtesla the researchers observe proton magnetic resonance in a liquid sample at about 100 Hz with an astonishing degree of sensitivity.

In vivo MR spectroscopy has the strongest connection and owes most to the **VP** methodology. In 1988, van der Veen et al. [130], a group of researchers at Delft University and Phillips Medical Systems in The Netherlands, published a very influential paper on the accurate quantification of *in vivo* MRS data, using the **VP** method and prior knowledge (constraints). According to one of the authors, as of August 2002 this paper had had more than 190 citations. The problem here is to fit MR spectra in the time domain, using models whose parameters have physical significance. An example would be a linear combination of exponentially damped sinusoids, but other types of nonlinear functions are also considered.

Van der Veen et al. consider MRS measurements of human calf muscle and human brain tissue and compare the FFT's of the original data with a linear prediction and **SVD** decomposition, and **VARPRO** with and without prior knowledge. The best results are obtained with **VARPRO** plus prior knowledge, which was difficult or impossible to impose on previous simpler approaches. They also list as desirable features the fact that starting values of the amplitudes of the spectral components are not required and that there are no restrictions on the form of the model functions.

Some years later, a group in Leuven, Belgium [125, 126, 127], made a comparative study for this problem, including artificial noise and also using prior knowledge. They examined one of the data sets considered earlier by van der Veen et al., added different levels of white Gaussian noise: small (5%), medium (15%), and large (25%), and considered 300 runs per level in a Monte Carlo simulation.

The model, without prior knowledge, involves 256 complex data points and 11 exponentials, for a total of 44 parameters divided equally between linear and nonlinear. In order to obtain good initial values, the time signal was Fourier transformed and displayed. Then, an interactive peak-picking was performed. This provided good initial values for the frequency and the damping of each peak. When the full functional is minimized, one also needs initial values for the linear parameters, and these are obtained by solving the **LLS** problem obtained by evaluating the nonlinear part at the chosen initial values. As observed in our original paper [45], this is much better than taking arbitrary values.

They first consider three different optimization methods for the full and the reduced problem, using **VP** with the Kaufman improvement: the original **VARPRO** with the Levenberg-Marquardt implementation, a secant type code NL2SOL [29], and lmder, a modern Levenberg-Marquardt implementation from MINPACK [77]. According to their results, NL2SOL with the separated functional seems to be the most reliable, even for large levels of noise. MINPACK's routine is almost as reliable and systematically faster. We should indicate here that the reported average times for this problem are about one minute on a SUN ULTRA2 (200 MHz), so in current and future platforms the differences in performance are negligible.

The most important conclusion drawn here is that if a **VARPRO**-type code is to be used, it should include the Kaufman simplification and should take advantage of the advances in numerical optimization. The MINPACK **NLLS** solver LMDER or the Gay-Kaufman implementation are good candidates for replacement codes. This has been confirmed in [129], where the author, in the process of developing an object oriented system for the analysis of *in vivo* MR signals, made similar comparisons and came to the conclusion that **VP** did not help if used with an external **NLLS** code, but it was quite efficient if properly implemented within a modern solver such as Gay and Kaufman's NSG code.

Vanhamme et al.'s study continues to consider three different methods to obtain starting values: HSVD [66], a fully automatic parameter estimation method that combines a state-space approach with **SVD**'s; pick1, the peak-picking method described above, and pick2, a more careful (and time consuming) version of pick1. Since the influence of these different initial value choices seems to be method independent, only results for MINPACK are presented. The conclusions are that the desirable automatic procedure works very well for low and medium noise levels, but not as well for high levels of noise. The procedures pick1 and pick2 are similar in performance and reliability, with a slight edge for pick2.

Finally, they consider the effect of using prior knowledge about the problem. In prior1, the number of linear variables is reduced to 11 by noting that all the peaks have a phase of 135^0. In prior2 all the known prior knowledge is used to eliminate variables, obtaining a problem with 5 linear and 12 nonlinear parameters. Results are given for MINPACK only, and here with prior1, **VP** has a slight edge in performance for small and medium noise levels, although for medium and high level of noise there is a deterioration in the reliability. For prior2, **VP** is consistently more reliable than the full functional approach, with a slight penalty in performance (under 8 %,) which has now come down to under 10 seconds of CPU time.

As a conclusion to this study the authors suggested using the full functional instead of the reduced one and they proceeded to write their own solver **AMARES** to do that. This new solver also includes some other features special to the problem that were not present in [130]. The study of van Leeuwen indicates that with the Gay and Kaufman implementation of **VP,** the balance would clearly tilt in favor of the reduced functional, both in speed and reliability.

Both **VARPRO** and **AMARES** are currently offered in the MRUI system [81]. See also [128] for a recent review article.

Recently, de Beer et al. [28], in a multi-center study (done in the context of the European Union BIOMED 1 Concerted Action No. PL 920432) with the purpose of reducing the variations in quantitative results of the participating *in-vivo* MR groups, worked with the same data-analysis protocol. A central part of that protocol was the **VARPRO** procedure.

Sala et al. [107] study respiratory rehabilitation, including lower limb exercise training, as part of the management for patients with chronic obstructive pulmonary disease (COPD). In order to understand better the physiologic mechanisms underlying the beneficial consequences of training, they made MRS measurements in the quadriceps of 13 patients with COPD and 8 healthy patients, before and during a period of exercises. Free induction decays were analyzed in the time domain using the **VARPRO** software, by fitting single Lorentzian functions for the resonance frequencies and amplitudes. The individual half-time of phosphocreatine recovery was fitted in the time domain to an exponential function. The authors then present a detailed discusion of the physiological conclusions stemming from this study.

Barone [4] has considered the problem of fast deconvolution in the case of noisy data. This is a general problem that occurs in many applications, such as the recovery of images observed in noise through a linear system representing a physical measuring device. Speed of computation is the driving factor for this study, especially when large amounts of data are present. The problem is shown to be separable, and using a **VP** approach is considerably more efficient than a previous simulated annealing method used on the full functional. Three examples are considered, including the restoration of a Magnetic Resonance image of a human brain that includes Gibbs oscillations.

Mosher et al. [79, 80, 3] have considered the problem of modeling the spatio-temporal neuromagnetic field or magneto-encephalogram (MEG) produced by neural activity of the brain. A popular model for the neural activity produced in response to a given sensory stimulus is a set of current dipoles, where each dipole represents the primary current associated with the combined activation of a large number of neurons located in a small volume of the brain. An important problem in the interpretation of MEG data is the localization of these neural current dipoles. The key concept here is that given any arbitrary static current distribution, the magnetic field can be obtained by using the Biot-Savart law. The general model of p_r rotating dipoles and p_f fixed dipoles results in a problem with $3r$ unknown location parameters ($r = 2p_r + p_f$,) and p_f unknown constrained moment parameters. Using m SQUID (Superconducting

QUantum Interference) biomagnetometers, n samples are collected to form a spatio-temporal data matrix $\mathbf{F}$. The problem is then:

$$\min_{\mathbf{L},\mathbf{M},\mathbf{S}} ||\mathbf{F} - \mathbf{H}(\mathbf{L},\mathbf{M})\mathbf{S}||_F^2,$$

where the model $\mathbf{H}(\mathbf{L},\mathbf{M})\mathbf{S}$ involves the gain matrix $\mathbf{H}$, the time series $\mathbf{S}$, the moment orientation matrix $\mathbf{M}$, and $||.||_F$ is the Frobenius norm. This separable model is one of the most ambitious and large-scale applications of the **VP** approach that we have seen to date and it is much too complicated to be discussed here. We refer the reader to the original papers for more details.

Solving this problem directly would require finding five parameters for each of p dipoles at each point of n time snapshots, for an overall total of $5pn$ parameters. By considering the separability of the problem and using the **VP** algorithm one can reduce the dimensionality of the problem by factoring out the linear moments. Given the complexity of the model, these authors recommend the use of a derivative-free approach, such as a Nelder-Mead type algorithm. The case of rotating dipoles is most favorable, since there the nonlinear parameters are not time-dependent, and therefore the problem can be reduced to finding only $3p$ location parameters, while deferring the calculation of the $2pn$ linear parameters, to be obtained *a posteriori* with a single **LLS** solve. **SVD**'s are recommended to optimize the operation count and more importantly, to regularize the problem in the ill-conditioned or even rank-deficient case.

Silva et al. [116] also consider realistic head models, instead of spherical ones, applied to the neural source localization problem associated with epileptic foci from scalp EEG data. Usually, accurate head models are constructed from MR images of the brain, skull, and scalp. Since this is an expensive and time consuming procedure, the authors have developed a parameterized realistic head model that can be adjusted to a particular patient by using external distances between anatomic landmarks. To localize the neural sources, an inverse problem, one needs to be able to solve the direct problem of determining the scalp potentials at the electrodes, generated by a given source. The authors use the Boundary Element Method (BEM) as their forward solver and the **VP** method for the inverse problem. Extensive validation studies are reported, together with application to three clinical cases.

Westwick and Kearney [136] consider the identification of a Hammerstein model of a biological system using **VP.** Again they cite the simplification arising from the reduced dimensionality and indicate that this work can be extended to other commonly used model structures, such as the Wiener cascade.

18. Conclusions and Further Developments

We have reviewed the ideas behind the **VP** method as an efficient solution to **SNLLS** problems. We have then briefly visited a significant number of very different applications that indicate how prevalent these problems are and how successful the method has been during the last thirty years, including very recent results.

One of the most exciting new applications is to the training of some types of neural networks. It would be interesting to extend the method to large-scale problems, i.e., those with too many parameters to be solved directly by a **VP**-type code. By their nature, those problems will be sparse, although sparsity may be destroyed by the elimination of the linear variables.

What we have learned from the original results and the accumulated experience of a large number of researchers in very diverse disciplines, is that the **VP** approach is chosen not only because it is more efficient than solving the unreduced problem, but also because it is a good preconditioner and leads to problems with better defined minima.

A public domain version of **VARPRO** with the Kaufman modification and using modern optimization technology would also be welcome, together with one that uses a truncated **SVD** approach to regularize ill-conditioned or over-parameterized models.

19. Distribution of References

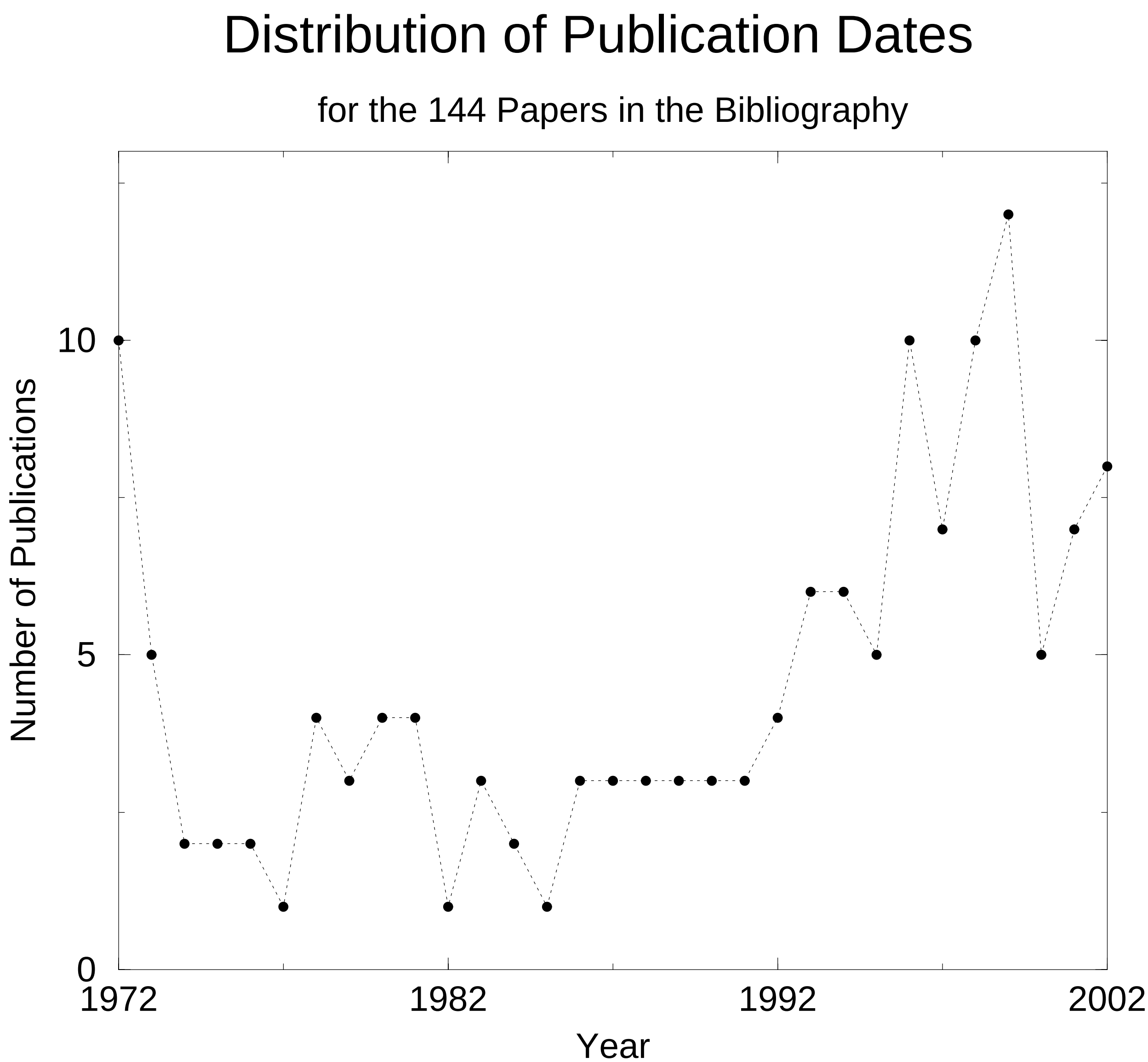
We show in the graph the distribution of cited references per year. We observe an increase of interest in the last decade, although that observation may be tainted by the fact that there are more items of recent vintages available on the Internet.

20. Acknowledgements

We thank Dr. Burt Rust for his valuable comments, for promoting the use of **VARPRO** through the institutions in which he worked, and for providing a number of interesting papers for our collection.

We also thank Drs. Linda Kaufman and Dirk van Ormondt and his colleagues for their role in developing **VARPRO** through the years.

A special thanks for Professor Michael Saunders, for his careful reading of the manuscript and perceptive comments that have helped improve an earlier version quite considerably.



t

[1] Abel, J.S., *A variable projection method for additive components with application to GPS.* IEEE Transact. on Aerospace and Electronic Systems **30:**928-930 (1994).

[2] Andersson, L., K. Holmström, and A. Ruhe, *Complex formation constants. A challenging data fitting problem from solution chemistry.* Algorithms for Approximation **10:**557-572. Clarendon Press (1987).

[3] Baillet, S., J.C. Mosher, and R.M. Leahy, *Electromagnetic brain mapping.* IEEE Signal Processing Magazine pp. 14-30, November (2001).

[4] Barone, P., *Fast deconvolution by a two-step method.* SIAM J. Sci. Comput. **21**:882-899 (1999).

[5] Bates, D.M., *The derivative of* $|\mathbf{X}^{\mathbf{T}}\mathbf{X}|$ *and its uses.* Technometrics **25:**373-376 (1983).

[6] Bates, D.M., and D.G. Watts, *A generalized Gauss-Newton procedure for multi-response parameter estimation.* SIAM J. Sci. Stat. Comp. **8:**49-55 (1987).

[7] Bates, D.M., and M.J. Lindstrom, *Nonlinear least squares with conditionally linear parameters.* Proceedings Annual Meeting, pp. 152-157. American Statistical Association, Chicago, Ill.

(1986).

[8] Baum, C.E., E.J. Rothwell, K-M. Chen, and D.P. Nyquist, *The singularity expansion method and its application to target identification.* Proc. IEEE **79:**1481-1492 (1991).

[9] Bishop, C.M., Neural Networks for Pattern Recognition. Oxford University Press (1995).

[10] Böckmann, C., *A modification of the trust-region Gauss-Newton method for separable nonlinear least squares problems.* J. Math. Systems, Estimation and Control **5**:1-16 (1995).

[11] Beece, D., S.F. Bowne, J. Czege, L. Eisenstein, H. Frauenfelder, D. Good, M.C. Marden, J. Marque, P. Ormos, L. Reinisch, and K.T. Yue, *The effect of viscosity on the photocycle of bacteriorhodopsin.* Photochemistry and Photobiology **33**:517-522 (1981).

[12] Beylkin, G., *On applications of unequally spaced Fast Fourier Transforms.* Mathematical Geophysics Summer School, Stanford University (1998).

[13] Bienvenu, G., and L. Kopp, *Adaptivity to background noise spatial coherence for high resolution passive methods.* Proc. IEEE on Acoustics, Speech, and Signal Processing, pp. 307-310 (1980).

[14] Björck, Å., Numerical Methods for Least Squares Problems. SIAM Pub., Philadelphia, PA (1996).

[15] Borden, B., *Radar scattering centre localization by subspace fitting.* Inverse Problems **17:**1483-1491 (2001).

[16] Bramley, R., and B. Winnicka, *Solving linear inequalities in a least squares sense.* SIAM J. Sci. Comp. **17**:275-286 (1996).

[17] Brent, R., Algorithms for Finding Zeros and Extrema of Functions Without Calculating Derivatives. Prentice Hall, Englewood Cliffs, NJ (1973).

[18] Bruls, J., C.T. Chou, B.R.J. Haverkamp, and M. Verhaegen, *Linear and non-linear system identification using separable least squares.* European J. of Control **5** (1999).

[19] Burg, J.P., *Maximum entropy spectral analysis.* Soc. Exploration Geophysicists (1967).

[20] Byers, R., C. He, and V. Mehrmann, *Where is the nearest non-regular pencil?* Linear Algebra and its Applications **285:**81-105 (1998).

[21] Cadzow, J.A., *Multiple source location–The signal subspace approach.* IEEE Trans. on Acoustics, Speech, and Signal Processing **38:**1110-1125 (1990).

[22] Callaghan, P., *Spectroscopy scales new peaks.* Physics World **15:**23-25 (2002).

[23] Capon, J., *High-resolution frequency-wavenumber spectrum analysis.* Proc. IEEE **57:**1408-1418 (1969).

[24] Chiuso, A., R. Brockett, and S. Soatto, *Optimal structure from motions: local ambiguities and global estimates.* Proc. IEEE CVPR'98 (1999).

[25] Conn, A.R., L.N. Vicente, and C. Visweswariah, *Two-step algorithm for nonlinear optimization with structured applications.* SIAM J. Opt. **9** (1999).

[26] Corradi, C., and L. Stefanini, *Computational experience with algorithms for separable nonlinear least squares problems.* Calcolo **XV**:317-330 (1978).

[27] Corradi, C., *A note on the solution of separable nonlinear least-squares problems with separable nonlinear equality constraints.* SIAM J. Numer. Anal. **18**:1134-1138 (1981).

[28] de Beer, R., B. Barbiroli, G. Gobbi, A. Knijn, H. Kugel, K.W. Langenberger, I. Tkac, and S. Topp, *Multicentre 1H MRS of the human brain addressed by one and the same data-analysis protocol.* Magnetic Resonance Imaging **16:**1107-1111 (1998).

[29] Dennis, J.E., D. M. Gay, and R. E. Welsch, *An adaptive nonlinear least-squares algorithm.* ACM Trans. Math. Software **7**:348-368 (1981).

[30] Dierckx, P., Curve and Surface Fitting with Splines. Clarendon Press, Oxford (1993).

[31] Dutre, S., H. Bruyninckx, and J. de Schutter, *The analytical Jacobian and its derivative for a parallel manipulator.* IEEE Conf. on Robotics and Automation, pp. 2961-2966. Albuquerque, NM (1997).

[32] Edrissi, H. E., M.H.G. Verhaegen, B.R.J. Haverkamp, and C.T. Chou, *Off- and on-line identification of discrete time LTI systems using separable least squares.* Proc. Conf. Decision and Control, Tampa, FL., IEEE cdc98 WM07 (1998).

[33] Eriksson, J., P.-Å. Wedin, and M. Gulliksson, *Regularization methods for nonlinear least squares problems. Part I: Exact rank-deficiency.* Techn. Rep. UMINF 96.04, Dept. Comp. Sc., Umea Univ., Sweden (1997).

[34] Eriksson, J., M. Gulliksson, P. Lindström, and P.-Å. Wedin, *Regularization tools for training feed-forward neural networks.* J. Optimization Methods and Software **10:**49-69 (1998).

[35] Escovar, R., *Optimal design of leading edge microprocessors.* ENCIMAG INPG, M. Sc. Thesis, Univ. Grenoble, FRANCE (2001).

[36] Escovar, R., and R. Suaya, *Optimal design of clock trees for leading edge microprocessors.* To appear in Journal of Modeling and Simulation of Microsystems (2002).

[37] Francos, J.M., A. Narasimhan, and J.W. Woods, *Maximum likelihood parameter estimation of the harmonic evanescent and purely indeterministic components of discrete homogeneous random fields.* IEEE Trans. Info. Theory **42**:916-930 (1996).

[38] Friedlander, B., *Sensitivity analysis of the maximum likelihood direction-finding algorithm.* IEEE Trans. on Aerospace and Electronic Systems **26**:953-968 (1990).

[39] Gardner, R.J., and P. Milanfar, *Reconstruction of convex bodies from brightness functions.* To appear in Discrete and Computational Geometry (2002).

[40] Gay, D.M., Port Library: http://netlib.bell-labs/netlib/master/readme.html (2002).

[41] Gay, D.M., and L. Kaufman, *Tradeoffs in algorithms for separable nonlinear least squares.* Proc. World Congress Comp. Appl. Math. pp. 157-158 (Eds. R. Vichnevetsky and J.J.H. Miller), Criterion Press, Dublin, Ireland (1991).

[42] Golub, G.H., M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter.* Technometrics **21**:215-223 (1979).

[43] Golub, G.H., and R. LeVeque, *Extensions and uses of the variable projection algorithm for solving nonlinear least squares problems.* Proc. Army Numerical Analysis and Computers Conference, pp. 1-12. US Army Res. Office (1979).

[44] Golub, G.H., and V. Pereyra, *The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate.* STAN-CS-72-261 (1972).

[45] Golub, G.H., and V. Pereyra, *The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate.* SIAM J. Numer. Anal. **10**:413-432 (1973).

[46] Golub, G.H., and V. Pereyra, *Differentiation of pseudoinverses, separable nonlinear least squares and other tales.* Generalized Inverses and Applications pp. 303-324, (Ed. M.Z. Nashed). Academic Press, New York (1976).

[47] Golub, G.H., and C.F. Van Loan, Matrix Computations, 3rd Ed. John Hopkins Univ. Press, Baltimore (1989).

[48] Gulliksson, M., and P.-Å. Wedin, *The use and properties of Tikhonov filter matrices.* SIAM J. Matrix Anal. Appl. **22**:276-281 (2000).

[49] Guttman, I., V. Pereyra, and H.D. Scolnik, *Least squares estimation for a class of nonlinear models.* Technometrics **15**:209-218 (1973).

[50] Heredia, E.A., and G.R. Arce, *Piecewise linear systems modeling based on a continuous threshold decomposition.* IEEE Trans. on Signal Processing **44** (1996).

[51] Holmström, K., *Constrained separable* **NLLS** *algorithms for chemical equilibrium analysis.* Proc. 5th Meeting, Nordic Sect., Math. Prog. Soc. (Ed. A. Lökketangen), Molde Univ. (1998).

[52] Holmström, K., *The TOMLAB optimization environment in MATLAB.* Advanced Modeling and Optimization **1**:47-69 (1999).

[53] Holmström, K., and J. Petersson, *A review of the parameter estimation problem of fitting positive exponential sums to empirical data.* App. Math. and Comp. **126**:31-61 (2002).

[54] Ingri, N., I. Andersson, L. Pettersson, L. Andersson, A. Yagasaki, and K. Holmström, *LAKE-A program system for equilibrium analytical treatment of multimethod data, especially combined potentiometric and NMR data.* Acta Chem. Scand. **50**:717-734 (1996).

[55] Jones, P.D., *Hemispheric surface air temperature variations: recent trends and update to 1987.* J. Clim. **1**:654-659 (1988).

[56] Jupp, D.L., *Approximation to data by splines with free knots.* SIAM J. Numer. Anal. **15**:328-343 (1978).

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

[58] Kaufman, L., and V. Pereyra, *A method for separable nonlinear least squares problems with separable equality constraints.* SIAM J. Numer. Anal. **15**:12-20 (1978).

[59] Kaufman, L., and G. Sylvester, *Separable nonlinear least squares with multiple right-hand sides.* SIAM J. Matrix Anal. Appl. **13**:68-89 (1992).

[60] Kaufman, L., G. Sylvester, and M.H. Wright, *Structured linear least-squares problems in system identification and separable nonlinear data fitting.* SIAM J. Optimization **4**:847-871 (1994).

[61] Kirk, B.L., and B.W. Rust, *Inductive modeling of time series. A detrending approach.* Environmetrics 81, SIAM (1981).

[62] Kirk, B.L., and B.W. Rust, *The solar cycle effect on atmospheric carbon dioxide levels.* In Weather and Climate Responses to Solar Variations. Colorado Associated University Press, Boulder, CO (1983).

[63] Koliha, J.J., *Continuity and differentiability of the Moore-Penrose inverse in C^*-algebras.* Mathematica Scandinavica **88**:154-160 (2001).

[64] Krogh, F.T., *Efficient implementation of a variable projection algorithm for nonlinear least*

squares problems. Comm. ACM **17**:167-169 (1974).

[65] Kumaresan, R., and A.K. Shaw, *Superresolution by structured matrix approximation.* IEEE Trans. on Antennas and Propagation **36**:34-44 (1988).

[66] Kung, S.Y., K.S. Arun, and D.V. Bhaskar Rao, *State-space and singular value decomposition based approximation methods for the harmonic retrieval problem.* J. Opt. Soc. Amer. **73**:1799-1811 (1983).

[67] Lanzkron, P.J., D.J. Rose, and J.T. Wilkes, *An analysis of approximate nonlinear elimination.* SIAM J. Sci. Comp. **17**:538-559 (1996).

[68] Lee, Y.S., and K.K. Baeck, *Basic set requirement for small components besides kinetic balance in relativistic self-consistent-field calculations of many electron systems.* Bull. Korean Chem. Soc. **7**:428-433 (1986).

[69] Liang, Jen-Wei, J.-T. Chen, and A.J. Paulraj, *A two-stage hybrid approach for CCI-ISI reduction with space-time processing.* Communication Letters **1** (1997).

[70] Lilleberg, J., E. Nieminen, and M. Latva-aho, *Blind iterative multiuser delay estimator for CDMA.* Proc. IEEE Int. Symp. Personal Indoor and Mobile Radio Communications (PIMRC), pp. 565-568. Taipei, Taiwan (1996).

[71] Lozier, R.H., R.A. Bogomolni, and W. Stoeckenius. Biophys. J. **15**:955-962 (1975).

[72] Lukeman, G.G., *Application of the Shen-Ypma algorithm for separable overdetermined nonlinear systems.* M. Sc. Thesis, Dalhousie Univ., Halifax, Nova Scotia (1999).

[73] Marque, J., and L. Eisenstein, *Pressure effects on the photocycle of purple membrane.* Biochemistry **23**:5556-5563 (1984).

[74] McDermott, R., A.H. Trabesinger, M. Muck, E.L. Hahn, A. Pines, and J. Clarke, *Liquid-state NMR and scalar couplings in microtesla magnetic fields.* Science **295**:2247 (2002).

[75] Miller, A.J, http://users.bigpond.net.au/amiller (2002).

[76] Moré, J.J., *The Levenberg-Marquardt algorithm: implementation and theory.* Numerical Analysis, pp. 105-116 (ed. G.A. Watson). Springer-Verlag (1978).

[77] Moré, J.J., B.S. Garbow, and K.E. Hillstrom, User Guide to MINPACK. Argonne Nat. Lab. Rep. ANL-80-74, Argonne, Ill. (1980).

[78] Morrison, D.B., *Application of the Shen-Ypma algorithm for nonlinear systems with some linear unknowns.* M. Sc. Thesis, Dalhousie Univ., Halifax, Nova Scotia (1998).

[79] Mosher, J.J., P.S. Lewis, and R.M. Leahy, *Multiple dipole modeling and localization from spatio-temporal MEG data.* IEEE Trans. Biomedical Eng. **39**:541-557 (1992).

[80] Mosher, J.J., and R.M. Leahy, *Recursive MUSIC: a framework for EEG and MEG source localization.* IEEE Trans. Biomedical Engineering **4511**:1342-1354 (1998).

[81] MRUI, Magnetic Resonance User Interface at http://carbon.usb.es/mrui (2002).

[82] Naressi, A., C. Couturier, J.M. Devos, I. Castang, R. de Beer, and D. Graveron-Demilly, *Java-based graphical user interface for the MRUI quantitation package.* Comp. in Biology and Medicine **31**:269-286 (2001).

[83] Ngia, L., *System modeling using basis functions and applications to echo cancellation.* Ph.D. Thesis, Chalmers Inst. of Techn., Sweden, NNSP2000 (2000).

[84] Nielsen, H.B., *Separable nonlinear least squares.* Techn. Report IMM-REP-2000-01, Dept. Math. Modelling, Techn. Univ. Denmark (2000).

[85] Osborne, M.R., *Some special nonlinear least squares problems.* SIAM J. Numer. Anal. **12**:571-592 (1975).

[86] Osborne, M.R., and G.K. Smyth, *A modified Prony algorithm for exponential function fitting.* SIAM J. Sci. Comp. **16**:119-138 (1995).

[87] Parks, T.A., *Reducible nonlinear programming problems.* Ph. D. Thesis, Dept. Comp. and Applied Math., Rice Univ., Houston, TX (1985).

[88] Pati, Y.C., and P.S. Krishnaprasad, *Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations.* IEEE Trans. Neural Networks **4** (1993).

[89] Pereyra, V. *Accelerating the convergence of discretization algorithms.* SIAM J. Numer. Anal. **4**:508-533 (1967).

[90] Pereyra, V. *Iterative methods for solving nonlinear least squares problems.* SIAM J. Numer. Anal. **4**:27-36 (1967).

[91] Pereyra, V., *Ray tracing methods for inverse problems.* Inverse Problems **16**:R1-R35 (2000).

[92] Pereyra, V., and G. Scherer, *Efficient computer manipulation of tensor products with applications to multidimensional approximation.* Math. Comp. **27**:595-605 (1973).

[93] Pereyra, V., and P. Zadunaisky, *On the convergence and precision of a process of successive differential corrections.* Proc. IFIP 65 (1965).

[94] Prais, M.G, and J.C. Wheeler, *Improved modified-moment-singularity method.* Physical Review A **33**:1233-1245 (1986).

[95] Prells, U., and M.I. Friswell, *Application of the variable projection method for updating models of mechanical systems.* J. of Sound and Vibration **225:**307-325 (1999).

[96] Radke, R., P. Ramadge, and T. Echigo, *Efficiently estimating projective transformations.* Proc. IEEE Int. Conf. Image Processing, 2000. Submitted for publication to IEEE Trans. Pattern Analysis and Machine Intelligence (2001).

[97] Rall, J.E., and R.E. Funderlic, *Interactive* **VARPRO (INVAR)**, *A nonlinear least squares program.* ORNL/CSD-55, Oak Ridge National Lab., TN (1980).

[98] Rao, B.D., and K.S. Arun, *Model based processing of signals: a state space approach.* Proc. IEEE **80:**283-309 (1992).

[99] Robertazzi, T.G., and S.C. Schwartz, *On applying the extended Kalman filter to nonlinear regression models.* IEEE Trans. on Aerospace and Electronic Systems **25:** 433-438 (1989).

[100] Roy, R., and T. Kailath, *ESPRIT-estimation of signal parameters via rotational invariance techniques.* IEEE Trans. on Acoustics, Speech, and Signal Processing **37:**984-995 (1989).

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

[102] Rust, B.W., M. Leventhal, and S.L. McCall, *Evidence for a radiactive decay hypothesis for supernova luminosity.* Nature **262:**118-120 (1976).

[103] Rust, B.W., and B.L. Kirk, *Inductive modelling of population time series.* Time Series and Ecological Processes, pp. 154-192 (Ed. H.H. Shugart). SIAM Pub., Philadelphia, PA (1978).

[104] Rust, B.W., and B.L. Kirk, *Modulation of fossil fuel production by global temperature variations.* Environment International **7:**419-422 (1982).

[105] Rust, B.W., and F.J. Crosby, *Further studies on the modulation of fossil fuel production by global temperature variations.* Environment International **20:**429-456 (1994).

[106] Sagara, N., and M. Fukushima, *A continuation method for solving separable nonlinear least squares problems.* J. of Computational and Applied Mathematics **10:**157-161 (1984).

[107] Sala, E., J. Roca, R.M. Marrades, J. Alonso, J.M. Gonzalez de Suso, A. Moreno, J.A. Barbera, J. Nadal, Luis de Jover, R. Rodriguez-Roisin, and P.D. Wagner, *Effects of endurance training on skeletal muscle bioenergetics in chronic obstructive pulmonary disease.* American J. Respiratory Crit. Care Med. **159:**1726-1734 (1999).

[108] Schmidt, R.O., *Multiple emitter location and signal parameter estimation.* Proc. RADC Spectrum Estimation Workshop (1979).

[109] Schutze, T., and H. Schwetlick, *Constrained approximation by splines with free knots.* BIT **37:**105-137 (1997).

[110] Schwetlick, H., *Nonlinear parameter estimation: models, criteria, and algorithms.* Numerical Analysis, pp. 164-193 (Eds. D.F. Griffiths, and G.A. Watson). Longman, Harlow, and J. Wiley, NY (1992).

[111] Schwetlick, H., and T. Schutze, *Least squares approximation by splines with free knots.* BIT **35**:1-23 (1995).

[112] Scolnik, H.D., *On the solution of non-linear least squares problems.* Ph.D. Thesis, Univ. Zurich, Switzerland (1971).

[113] Scolnik, H.D., *On the solution of non-linear least squares problems.* Proc. IFIP 71, pp. 1258-1265. North Holland, Amsterdam (1972).

[114] Seber, G.A.F., and C.J. Wild, Nonlinear Regression. Wiley Interscience, New York (1989).

[115] Shen, Y.-Q., and T.J. Ypma, *Solving N+m equations with only m nonlinear variables.* Computing **44:**259-270 (1990).

[116] Silva, C., R. Almeida, T. Oostendorp, E. Ducla-Soares, J.P. Foreid, and T. Pimentel, *Interictal spike localization using standard realistic head model: simulations and analysis of clinical data.* Clinical Neurophisiology **110:**846-855 (1999).

[117] Sjöberg, J., and M. Viberg, *Separable non-linear least squares minimization – possible improvements for neural net fitting.* IEEE Workshop in Neural Networks for Signal Processing. Amelia Island Plantation, FL (1997).

[118] Stewart, G.W., *On the perturbation of pseudo-inverses, projections and linear least squares problems.* SIAM Review **19:**634-662 (1977).

[119] Suaya, R., *Interconnect modeling for high speed digital circuits. The role of rlc coupling.* To appear in Special Issue, J. of Modeling and Simulation of Microsystems (2002).

[120] Talwar, S., *Blind space-time algorithms for wireless communication systems.* Ph. D. Thesis, SCCM, Stanford University (1996).

[121] Talwar, S., M. Viberg, and A. Paulraj, *Blind estimation of multiple co-channel digital signals arriving at an antenna array.* IEEE SP Letters **1:**29-31 (1994).

[122] Tomasi, C., and J. Shi, *Direction of heading from image deformations.* IEEE Conf. on Comp. Vision and Pattern Recognition (1993).

[123] Trosset, M.W., *Computing distances between convex sets and subsets of the positive semidefinite matrices.* Tech. Report 97-3, Dept. Comp. & Appl. Math., Rice Univ., Houston, TX (1997).

[124] Van Loan, C., and N. Pitsianis, *Approximation with Kronecker products.* Linear Algebra for Large Scale and Real Time Applications (Eds. M.S. Moonen, and G.H. Golub), pp. 293-314, Kluwer Pubs. (1993).

[125] Vanhamme, L., A. van den Boogaart, and S. Van Huffel, *Fast and accurate parameter estimation of noisy complex exponentials with use of prior knowledge.* Proc. EUPSICO-96, Trieste, Italy (1996).

[126] Vanhamme, L., A. van den Boogaart, and S. Van Huffel, *Improved method for accurate and efficient quantification of MRS data with use of prior knowledge.* J. of Magnetic Resonance **129**:35-43 (1997).

[127] Vanhamme, L. *Advanced time-domain methods for Nuclear Magnetic Resonance Spectroscopy data analysis.* PhD Thesis, Univ. of Leuven (1999).

[128] Vanhamme, L., 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).

[129] Van Leeuwen, J.P., *Object georiënteerd parameterschatten.* M. Sc. Thesis, Tech. Univ. Delft (1994).

[130] van der Veen, J.W.C., R. de Beer, P.R. 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).

[131] Wedin, P.Å., *Perturbation theory for pseudo-inverses.* BIT **13**:217-232 (1973).

[132] Weigl, K., and M. Berthod, *Neural networks as dynamical bases in function space.* Report #2124, INRIA, Prog. Robotique, Image et Vision. Sophia-Antipolis, France (1993).

[133] Weigl, K., G. Giraudon, and M. Berthod, *Application of projection learning to the detection of urban areas in SPOT satellite images.* Report #2143, INRIA, Prog. Robotique, Image et Vision. Sophia-Antipolis, France (1993).

[134] Weigl, K., and M. Berthod, *Projection learning: Alternative approach to the computation of the projection.* Proc. European Symp. on Artificial Neural Networks, pp. 19-24. Brussels, Belgium (1994).

[135] Westwick, D.T., K. George, and M. Verhaegen, *Nonlinear identification of automobile vibration dynamics.* Proc. 7th Mediterranean Conf. on Control and Automation (MED99), pp. 724-737. Haifa, Israel (1999).

[136] Westwick, D.T., and R.E. Kearney, *Identification of a Hammerstein model of the stretch reflex EMG using separable least squares.* Proc. IEEE EMBS22, Chicago, IL (2000).

[137] Wheeler, J.C., M.G. Prais, and C. Blumstein, *Analysis of spectral densities using modified moments.* Physical Review B **10**:2429-2447 (1974).

[138] Wikstrom, G., *Algorithms and software for the computation of parameters occurring in ODE-models.* Thesis, Umeå Univ., Sweden (1997).

[139] Wold, H., and E. Lyttkens, *Nonlinear iterative partial least squares (NIPALS) estimation procedures.* Bull. ISI **43**:29-51 (1969).

[140] Wolfe, C.M., B.W. Rust, J.H. Dunn, and I.E. Brown, *An interactive nonlinear least squares program.* NBS Tech. Note 1238, Gaithersburg, MD (1987).

[141] Yen, J., and L.R. Petzold, *Convergence of the iterative methods for coordinate-splitting formulation in multibody dynamics.* Tech. Report TR95-052, Dept. Comp. Sci., Univ. Minnesota (1995).

[142] Yen, J., and L.R. Petzold, *Computational challenges in the solution of nonlinear oscillatory multibody dynamical systems.* Proc. Biennial Conf. on Numer. Anal. (Eds. D.F. Griffiths, and G.A. Watson). Pittman Res. Notes in Math. #344, Addison Wesley Longman Ltd. (1996).

[143] Yen, J., and L.R. Petzold, *An efficient Newton-like iteration for the numerical solution of highly oscillatory constrained multibody dynamical systems.* SIAM J. Sci. Comput. **19**:1513-1534 (1998).

[144] Zhou, Y., P.C. Yip, and H. Leung, *Tracking the direction-of-arrival of multiple moving targets by passive arrays: algorithm.* IEEE Trans. on Signal Proc. **47**:2655-2666 (1999).