Variable Projections for Separable Nonlinear Least Squares Problems Since 2002

Victor Pereyra, Smart Fields, Stanford University, Godela Scherer, University of Reading

---

Abstract

In [11] we surveyed the Variable Projections (VARPRO) method for the numerical solution of separable nonlinear least squares problems [9] and its applications. Since that date the interest in the method has not abated and therefore in this paper we update that survey.

The source of citations is Google Scholar. We have tried to complete these citations as much as possible. There is also a significant number of citations of VARPRO that do not refer to our papers, mainly in the field of MRI in medicine and we have included some of those also in this survey. Since the list is fairly long, we do not include citations published before 2015 that have themselves only a small number of citations.

We have chosen a new style of presentation: we put the different kinds of applications in classes and sub-classes. When there is a new field of application, not considered in earlier publications [9, 11, 7, 19], we start with a brief description. Instead of collecting alphabetically all the many references at the end, as is customary, we include the references in the classes. In that way practitioners will be able to find information of their interest in a straightforward way.

Contents

1. Separable Nonlinear Least Squares and the Variable Projection Method

We consider nonlinear data fitting problems that have as their underlying model a linear combination of nonlinear functions. Models of this type are very common, as we will show in a variety of applications from 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 in parameter, material or system identification, signal analysis, analysis of spectral data, medical and biological imaging, neural networks, robotics and telecommunications, to name a few.

In April 1973 the paper "Differentiation of pseudoinverses and separable nonlinear least squares" by G. Golub and V. Pereyra [9] was published. This work was initiated in 1971, motivated by and generalizing earlier work with Irving Guttman and Hugo Scolnik [12], which in turn elaborated and generalized work in H. Scolnik's Ph D Thesis at the University of Zurich [22, 23].

Scolnik's original idea applied to so called posynomial models, i.e., polynomials basis with arbitrary exponents, $\sum_i a_i x^{\alpha_i}$. By a trivial change of variables these models can be converted to the important case of exponential data fitting, $\sum_i a_i e^{\alpha_i x}$ [8]. In [12] this was extended to general functions of one variable, while [9] considered the general multi-variable case. In this last paper the authors also considered a differential calculus for projectors and pseudoinverses and described a carefully crafted stable numerical algorithm.

Given a set of observations $\{y_i\}$, a separable nonlinear least squares problem is defined in [9] where 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 contains 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:

$$\mathbf{a} = \mathbf{\Phi}(\alpha)^{+}\mathbf{y}, \tag{1.1}$$

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

$$\min_{\alpha} \frac{1}{2}||(\mathbf{I} - \mathbf{\Phi}(\alpha)\mathbf{\Phi}(\alpha)^+)\mathbf{y}||_2^2, \tag{1.2}$$

where the linear parameters have been eliminated. Another way to see this is to observe that 1.1 is a nonlinear equality constraint. A joint method, where the two sets of variables are considered independently will not, in general, honor this constraint.

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 $\mathbf{\Phi}(\alpha)$, which 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 [27]), 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 (1.2) and then using the optimal value obtained for α to solve for $\mathbf{a}$ in (1.1). 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 [15, 20]), i.e., the minima for the reduced functional are better defined than those for the full one.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 data fitting is so hard to solve directly.

After the publication of [9] improvements to the original algorithm started to appear. One important development came from the observation of L. Kaufman [13] that a term in the derivative of the projector could be neglected for problems with small residuals without affecting the speed of convergence. This simplification made the costs of iterations for the separated and full approaches similar, showing an actual gain in computing time for the theoretically faster convergence of VARPRO.

In [3] the authors consider the problem of regularizing ill-conditioned NLLSQ and their approach leads to objective functions that change during the optimization process. One of the most important developments of

VARPRO are connected to constrained problems. Early on we contributed [10, 14] to this area by introducing so called separable equality constraints, which are of the form $H(\alpha)a = g(\alpha)$, where a, α are the linear and nonlinear parameters respectively. Observe that constraints on the nonlinear parameters do not affect the procedure: one just need to use an appropriate constrained optimization code for the reduced functional. Extensions to other types of constraints have come from practitioners in different fields whose problems were constrained and they can be found in the appropriate sections.

Some modern implementations in Matlab, Julia and C++ can be found in [3, 25, 8]. The paper [8] contains a very interesting discussion on the connection between VARPRO and the joint minimization approach, besides pointing to the excellent performance of VARPRO on some difficult affine bundle adjustment problems in computer vision that includes an scalable implementation for large problems.

Osborne [18], who has a long history of working on these problems, discusses in detail the rate of convergence of variable projections in conjunction with the Gauss-Newton algorithm and emphasizes its particular effectiveness for large data sets. Bert Rust [21], an early champion of VARPRO at NIST, discusses linear and nonlinear least squares with strong emphasis on statistics in a four part series of papers. Part 4 is dedicated to VARPRO and diverse real life applications and models are used for illustration.

Shen and Ipma [24] present a method for solving separable nonlinear least squares problem. Their technique replaces this large problem by a much smaller problem in the nonlinear variables alone. They show how Newton's method can be used to solve the latter problem, obtaining quadratic convergence even in the nonzero residual case. While the method is in principle based on the use of one particular orthonormal basis for the null-space of A^T throughout the computation, they show that one can instead use any convenient orthonormal basis for this null-space at each successive iteration point without affecting the process.

The purpose of the current paper is to tell the story of VARPRO developments and applications since the first review in 2003. We want to stress the richness and variety of applications and fields, and that the interest on the technique has not abated. We again classify the citations with regards to application field, but now we put the citations at the end of the section where they are quoted. We feel this is a more convenient organization for researchers interested in a particular application.

The common thread for the majority of the papers is the use of variable projections for separable models in a least squares context, while a small percentage uses only the results on the derivatives of pseudoinverses and projectors.

[1] Variable projection without smoothness. A Aravkin, D Drusvyatskiy,

T van Leeuwen, arXiv preprint arXiv:1601.05011v2 (2017).

[2] Modified Gram-Schmidt method-based variable projection algorithm for separable nonlinear models. Guang-Yong Chen, Min Gan, Feng Ding and C. L. Philip Chen, IEEE Transactions on Neural Networks and Learning Systems 1-9 (2018).

[3] A regularized Variable Projection algorithm for separable nonlinear least squares problems. G Y Chen, M Gan, C L P Chen, H-X Li, IEEE Transactions on Automatic Control (2018).

[4] A regularized variable projection for separable nonlinear least-squares problems. G Y Chen, M Gan, C L P Chen and H-X Li, IEEE Transactions on Automation **64**:526-537 (2019).

[5] On some separated algorithms for separable nonlinear least squares problems. M Gan, C L P Chen, G Y Chen, Long Chen, IEEE Transactions on Cybernetics 1-9 (2017).

[6] An efficient variable projection formulation for separable nonlinear least squares problems. M Gan, H X Li, IEEE Transactions on Cybernetics **44**:707-711 (2014).

[7] Least Squares Data Fitting With Applications. P C Hansen, V Pereyra, G Scherer, John Hopkins University Press (2013).

[8] Revisiting the Variable Projection method for separable nonlinear least squares problems. J H Hong, C Zach, A Fitzgibbon, IEEE Conference on Computer Vision and Pattern Recognition 5939-5947 (2017).

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

[10] Differentiation of pseudo-inverses, separable nonlinear least squares problems, and other tales. G Golub and V Pereyra, Proc. MRC Seminar on Generalized Inverses and its Applications 303-324 (1976).

[11] Separable nonlinear least squares: the variable projection method and its applications. GH Golub, V Pereyra, Inverse problems **19**: R1 (2003).

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

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

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

[15] Efficient implementation of a variable projection algorithm for nonlinear least squares problems. F T Krogh, Comm. ACM **17:**167-169 (1974).

[16] A proof of the VARiable PROjection (VARPRO) method in Hilbert space. Qiang Ning, Manuscript, U. of Illinois (2010).

[17] Variable projection for nonlinear least squares problems. D P O'Leary, B W Rust, Computational Optimization and Applications **54**:579-593 (2013).

[18] Separable least squares, variable projection, and the Gauss-Newton algorithm. M R Osborne, Electronic Transactions on Numerical Analysis (2007).

[19] Exponential Data Fitting and its Applications. V Pereyra, G Scherer, Bentham Science Publishers (2010).

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

[21] Fitting nature's basic functions Part IV: the variable projection algorithm. BW Rust, Computing in Science & Engineering **5**:74-79 (2003).

[22] On the solution of nonlinear least squares problems. H. Scolnik, Ph D Thesis, U. of Zurich, Switzerland (1971).

[23] On the solution of nonlinear least squares problems. H. Scolnik, Proc. IFIP **71**:1258-1265, North Holland, Amsterdam (1972).

[24] Solving separable nonlinear least squares problems using the QR factorization. Yunqiu Shen and Tjalling J Ypma, Journal of Computational and Applied Mathematics Volume **345**:48-58 (2019).

[25] https://www.juliaobserver.com/packages/Varpro (2016).

[26] Variable projection for affinely structured low-rank approximation in weighted 2-norms. K Usevich, I Markovsky, Journal of Computational and Applied Mathematics **272**:430-448 (2014).

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

2. Electrical Engineering

Here we find applications to motor fault diagnostics, amplifiers, communications and wind energy systems. In separate categories we put power systems, signal identification (that includes location and antennas), VLSI design and signal processing.

Gan and Li [6] consider the radial basis function network-based autoregressive model with exogenous inputs. (RBF-ARX) models have much more linear parameters than nonlinear parameters. Taking advantage of this special structure, a variable projection algorithm is proposed to estimate the model parameters more efficiently by eliminating the linear parameters through the orthogonal projection. The proposed method not only substantially reduces the dimension of parameter space of the RBF-ARX model but also results in a better-conditioned problem. In this paper, both the full Jacobian matrix of Golub and Pereyra and the Kaufman's simplification are used to test the performance of the algorithm. An example of chaotic time series modeling is presented for the numerical comparison. It clearly demonstrates that the proposed approach is computationally more efficient than the previous structured nonlinear parameter optimization method and the conventional Levenberg–Marquardt algorithm without the parameters separated. Finally, the proposed method is also applied to a simulated nonlinear single-input single-output process, a time-varying nonlinear process and a real multi-input multi-output nonlinear industrial process to illustrate its usefulness.

In [1] Alamir considers the problem of sensitivity analysis of the simultaneous estimation of state and parameters for induction motors by writing it as a separable nonlinear least squares problem.

An interesting application is found in [11], for the adaptive sparse recovery of inverse synthetic aperture radar (ISAR) of uniformly rotating targets by parametric weighted L_1 minimization. One of the steps of this complicated algorithm requires solving a separable least squares problem and the authors apply the VARPRO idea successfully. The authors of [12] consider measurement-based load modeling, especially in the presence of new loads such as power electronics-interfaced loads and electric vehicles with fast dynamics, which require fast-converging algorithms that provide the model parameters with high reliability. In the current practice, all or only a subset of the parameters of an aggregated load model are estimated using iterative optimization algorithms. Thus, the identification problem either has a high dimension, which leads to a large variance for the estimated parameters, or does not include a subset of the parameters with low sensitivity. In this paper, an efficient approach for the estimation of the composite load model parameters is proposed that addresses these issues. This method partitions the parameters into two subsets; one that appears nonlinearly in the model output, and a second set that affects the outputs linearly. Then, the optimization is performed only with respect to the nonlinear set, with the linear parameters treated as explicit functions of the nonlinear ones. This approach effectively reduces the dimension of the search space since it only includes the nonlinear parameters in the optimization, and also includes the linear parameters by computing them using linear regression at each iteration. These features lead to a much faster convergence while all of the composite load model parameters are estimated reliably. Experimental and simulation data are presented to demonstrate the performance of the proposed method.

Bouleux [3] points out that an optimal prior-knowledge method for Direction Of Arrival (DOA) estimation has been proposed. This method solely estimates a subset of DOA's taking into account known ones. The global idea is to maximize the orthogonality between an estimated signal subspace and noise subspace by constraining the orthogonal noise-made projector to only project onto the desired unknown signal subspace. To understand how this is possible requires the derivation of the variance for the DOA estimates. During the derivation, oblique projection operators and their first order derivatives are required.

[1] Sensitivity analysis in simultaneous state/parameter estimation for induction motors. M Alamir, International Journal of Control **75**:753-758 (2002).

[2] Digital predistortion of single and concurrent dual-band radio frequency GaN amplifiers with strong nonlinear memory effects. S Amin, P Händel, D Rönnow, IEEE Transactions on Microwave Theory and Techniques **65**:2453-2464 (2017).

[3] Prior knowledge optimum understanding by means of oblique projectors and their first order derivatives. G Bouleux, IEEE Signal Processing Letters **20**:205-208 (2013).

[4] Estimation of DC offset parameters using global optimization technique. F Duan, R Zivanovic, Australasian Universities Power Engineering Conference 1-4 (2008).

[5] Induction motor fault diagnostics using global optimization algorithm. F Duan, R Zivanovic, Power Engineering Conference (2009).

[6] A variable projection approach for efficient estimation of RBF-ARX model. M Gan, H X Li, H Peng, IEEE Transactions on Cybernetics **45**:462-471 (2015).

[7] Packaging RF devices and modules. Y C Lee, W Zhang, B Su, Z Feng, K C Gupta, Chong-ll Park, Encyclopedia of RF (2005).

[8] Hybrid precoding for millimeter wave MIMO systems with finite-alphabet inputs. J Jin, Y R Zheng, W Chen, C Xiao, IEEE Global Communications Conference 1-6 (2017).

[9] On the convergence of the vector-fitting algorithm. S Lefteriu, A C Antoulas, IEEE Transactions on Microwave Theory and Techniques **61**:1435-1443 (2013).

[10] Non-coherent diagonal distributed space-time block codes for amplify-and-forward half-duplex cooperative relay channels. S Ma, J K Zhang, IEEE Transactions on Vehicular Technology **60**:2400-2405 (2011).

[11] Adaptive sparse recovery by parametric weighted L_1 minimization for ISAR imaging of uniformly rotating targets. W Rao, G Li, X Wang, X G Xia, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **6**:942-952 (2013).

[12] An efficient approach for measurement-based composite load modeling. Mohammad Rasouli, Reza Sabzehgar and Hamid Reza Teymour, IEEE Energy Conversion Congress and Exposition (2018).

[13] Identification of Wind Energy Systems. GJ Van der Veen, Ph D Thesis, Technische Universiteit Delft (2013).

[14] Energy-efficient uniquely factorable constellation designs for non-coherent SIMO channels. L Xiong, J K Zhang, IEEE Transactions on Vehicular Technology **61**:2130-2144 (2012).

2.1. Power Grid

The main objective of [5] is to estimate the voltage parameters, i.e., angular frequency and phasor, from the input three-phase signals, X. In this context, the author proposes the use of a Maximum Likelihood estimator (MLE). MLE corresponds to a least square estimator when the noise is white Gaussian noise. This spectral estimation based parametric model can be decomposed into two steps: first, the estimation of angular frequency from X is the main difficult step; then, the estimated phasor can be obtained once the angular frequency is estimated. This is, of course, a separable nonlinear LSQ problem. Interesting comparisons between VARPRO, Prony, matrix pencil and ERA methods can be found in [3]. The conclusion, as in other studies, is that VARPRO is superior.

[1] Robustness of DC power networks under weight control. Q Ba, K Savla, arXiv preprint arXiv:1609.02179, (2016).

[2] Power system modal analysis tool developed for industry use. A R Borden, B C Lesieutre, J Gronquist, North American Power Symposium (2013).

[3] Variable projection method for power system modal identification. A R Borden, B C Lesieutre, IEEE Transactions on Power Systems **29**:2613-2620 (2014).

[4] Fast modal identification, monitoring, and visualization for large-scale power systems. using dynamic mode decomposition. S Mohapatra, T J Overbye, Power Systems Computation Conference 1-7 (2016).

[5] On Electric Grid Power Quality Monitoring Using Parametric Signal Processing Techniques. Z Oubrahim, Ph D Thesis, Universite de Bretagne Occidentale (2017).

[6] Maximum-likelihood frequency and phasor estimations for electric power grid monitoring. Z Oubrahim, V Choqueuse, Y Amirat, M. Benbouzid, IEEE Transactions on Industrial Informatics **14**:167-177 (2018).

[7] Disturbances classification based on a model order selection method for power quality monitoring. Z Oubrahim, V Choqueuse, Y Amirat, MEH Benbouzid, IEEE Transactions on Industrial Electronics **64**:9421-9432 (2017).

2.2. Signal Identification, Location, Antennas

Traditionally this has been a strong user of VARPRO, as we described in detail in our previous survey. The classical location algorithms MUSIC and SPRIT both owe inspiration to VARPRO. Current problems include interference cancellation, ghost signals. multi-path, noise, superimposed signal replicas, etc. This area includes many modern communication, radar and positioning systems. We comment on a sample of this large set of references.

A number of different applications to identification and other problems in signal processing will also be discussed. For instance, [39] studies the reconstruction of a multichannel sampling scheme when both gains and offsets are unknown, which appears in many practical signal processing applications.

In [28] the authors consider a very interesting and challenging problem: how to endow autonomous underwater vehicles with abilities akin to fish, in terms of sensing flows and nearby movements. To do this they add a lateral array of MEMS flow sensors. The calibration of this system leads to a SNLLS problem that is solved by VARPRO.

In [30] the authors consider the problem of direction of arrival estimation with multi-mode antennas using VARPRO, reducing significantly the complexity and allowing a practical implementation. A similar problem is considered in [44] for radio ultra-wide-band ranging.

In [5] a two-stage estimation algorithm for global system positioning is proposed. At the first stage, the authors ignore the additional path error and obtain a relatively accurate initial position. Then, based on this result they include the additional path error in the estimation problem and calculate it by the variable projection method. Numerical results show that the proposed algorithm based on the VARPRO method can effectively mitigate multi-path interference.

Aldana et al [2] consider the problem of blindly and semi-blindly obtaining the channel gains for a multiuser multi-carrier system. One of the optimization problems in channel gain estimation is a separable one and VARPRO is used advantageously.

In [33] the authors consider the problem of indoor positioning when there are multi-path arrivals by using directional antennas. In order to estimate the agent's position, they express the delays τ as a function of the agent's position p using the geometric model of the environment. With hypothesized τ, the amplitudes are estimated using least-squares, resulting in a separable problem.

In the interesting report [14] a power-law scattering model is considered for super-resolution radar. This is based on the ray theory of diffraction of J. Keller [17] and leads to a separable problem.

GPS positioning in complex environments is the theme of [5]. The objective is to diminish the influence of multi-path arrivals. Traditional algorithms have a certain deviation from the real-additional path delay. For more accurate estimation, the authors first restrict the additional path error in a relatively wide interval; then, the position estimation is converted into a constrained separable nonlinear least squares problem. Numerical results show that the proposed algorithm based on the VP method can effectively mitigate multi-path interference.

[1] Classification of non-stationary narrowband signals using segmented chirp features and hidden Gauss-Markov models. P L Ainsleigh, S G Greineder, N Kehtarnavaz, IEEE transactions on Signal Processing **53**:147-157 (2005).

[2] Channel estimation for multi-carrier multiple input single output systems using the EM algorithm. C H Aldana, E de Carvalho, J M Cioffi, IEEE Transactions on Signal Processing **51**:3280-3292 (2003).

[3] Maximum likelihood joint angle and delay estimation from multi-path and multi-carrier transmissions with application to indoor localization over IEEE 802.11ac radio. Faouzi Bellili, Souheib Ben Amor, Sofiéne Affes and Ali Ghrayeb, IEEE Transactions on Mobile Computing (2018).

[4] MIMO radar sensitivity analysis of antenna position for direction finding. H Chen, X Li, W Jiang, Z Zhuang, IEEE Transactions on Signal Processing **60**:5201-5216 (2012).

[5] A two-stage estimation algorithm based on Variable Projection method for GPS positioning. G Y Chen, M Gan, C L P Chen, L Chen, IEEE Transactions on Instrumentation and Measurement 1-8 (2018).

[6] Asymptotic properties of nonlinear weighted least squares in radar array processing. J Eriksson, M Viberg, IEEE Transactions on Signal Processing **52**:3083-3095 (2004).

[7] Gradient radial basis function based varying-coefficient autoregressive model for nonlinear and non-stationary time series. M Gan, C L P Chen, H X Li, L Chen, IEEE Signal Processing Letters **22**:809-812 (2015).

[8] Super-resolution TDOA estimation based on eigenanalysis and sequential quadratic programming. F X Ge, D Shen, Y Peng, V O K Li, IEEE Proceedings-Radar, Sonar and Navigation **151**:197-202 (2004).

[9] Super-resolution time delay estimation in multi-path environments. F X Ge, D Shen, Y Peng, V O K Li, IEEE Transactions on Circuits and Systems **54**:1977-1986 (2007).

[10] Bayesian Signal Processing Techniques for GNSS Receivers: From Multi-path Mitigation to Positioning. P Closas Gómez, Ph D Thesis, Universitat Politècnica de Catalunya (2009).

[11] A Bayesian approach to multi-path mitigation in GNSS receivers. P Closas, C Fernandez-Prades, J A Fernadez-Rubio, IEEE Journal of Selected Topics in Signal Processing **3**:695-706 (2009).

[12] Sparse antenna array design and combined range and angle estimation for FMCW radar sensors. R Feger, S Schuster, S Scheiblhofer, A Stelzer, IEEE Radar Conference 1-6 (2008).

[13] Joint channel synchronization under interference limited conditions. J Grotz, B Ottersten, J Krause, IEEE Transactions on Wireless Communications **6**:1-9 (2007).

[14] Power-law scattering models and nonlinear parametric estimation for super-resolution radar. N Hatch, D Holl Jr, D Cyganski, Lincoln Lab. MIT Tech. Report ADA427396 (2002).

[15] Robust spectral factor approximation of discrete-time frequency domain power spectra. K Hinnen, M Verhaegen, N Doelman, Automatica **41**:1791-1798 (2005).

[16] DOA Estimation of uncorrelated, partly correlated and coherent signals using alternating oblique projection. H Hou, X Mao, Radio-engineering **26**:588-600 (2017).

[17] Geometrical theory of diffraction. J Keller, Journal of the Optical Society of America (1962).

[18] On Interference Rejection in Wireless Multichannel Systems. G Klang, Ph D Thesis, Stockholm (2003).

[19] Structured semi-blind interference rejection in dispersive multichannel systems. G Klang, B Ottersten, IEEE Transactions on Signal Processing **50**:2027-2036 (2002).

[20] Using DecaWave UWB transceivers for high-accuracy multi-path-assisted indoor positioning. J Kulmer, S Hinteregger, IEEE International Conference on Communications Workshops 1239-1245 (2017).

[21] Newton-type method in spectrum estimation-based AOA estimation. J H Lee, S W Cho, H S Kim, IEICE Electronics Express **9** (2012).

[22] A consistent OFDM carrier frequency offset estimator based on distinctively spaced pilot tones. J Lei, T S Ng, IEEE Transactions on Wireless Communications **3**:588-599 (2004).

[23] Estimation Using Low Rank Signal Models. K Mahata, Ph D Thesis, Uppsala University (2003).

[24] Subspace fitting approaches for frequency estimation using real-valued data. K Mahata, IEEE Transactions on Signal Processing **53**:3099-3110 (2005).

[25] Single antenna interference cancellation for GSM by enhanced maximum SNR filter. L Mattellini, 2nd International Symposium on Wireless Communication Systems 332-336 (2005).

[26] Fast and approximative estimation of continuous-time stochastic signals from discrete-time data. M Mossberg, E K Larsson, IEEE International Conference on Acoustics, Speech, and Signal Processing (2004).

[27] Recursive identification of acoustic echo systems using orthonormal basis functions. L S H Ngia, IEEE Transactions on Speech and Audio Processing **11**:278-293 (2003).

[28] Flow vision for autonomous underwater vehicles via an artificial lateral line. N Nguyen, D L Jones, Y Yang, C Liu, EURASIP Journal on Advances in Signal Processing (2011).

[29] Joint maximum likelihood time-delay estimation for LTE positioning in multi-path channels. J A del Peral-Rosado, J A López-Salcedo, J A Seco-Granados, F Zanier, M Crisci, EURASIP Journal on Advances in Signal Processing **33** (2014).

[30] On the potential of multi-mode antennas for direction-of-arrival estimation. R Pöhlmann, S A Almasri, S Zhang, T Jost, arXiv preprint arXiv:1806.11320. Electrical Engineering and Systems Science, Signal Processing (2018).

[31] Analysis of a nonlinear least squares procedure used in global positioning systems. G W Pulford, IEEE Transactions on Signal Processing **58**:4526-4534 (2010).

[32] Joint maximum likelihood estimation of carrier frequency offset and channel in uplink OFDMA systems. M O Pun, S H Tsai, C C J Kuo, IEEE Global Telecommunications Conference **6**:3748-3752 (2004).

[33] Multi-path-assisted indoor positioning enabled by directional UWB sector antennas. M Rath, J Kulmer, M S Bakr, B Grosswindhager, K Witrisal, IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications 1-5 (2017).

[34] A simple multi-path delay estimator based on alternating projection algorithm. H Saarnisaari, V Tapio, Position, Location, and Navigation Symposium 1086-1093 (2006).

[35] Complexity reduction in the parametric estimation of superimposed signal replicas. J Selva, Signal Processing **84**:2325-2343 (2004).

[36] A new array pattern synthesis algorithm using the two-step least-squares method. Z Shi, Z Feng, IEEE Signal Processing Letters **12**:250-253 (2005).

[37] New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data. P Stoica, P Babu, J Li, IEEE Transactions on Signal Processing **2**:529-532 (2011).

[38] Efficient Multi-path Mitigation in Navigation Systems. J S Vera, Ph D Thesis, Universitat Politecnica de Catalunya (2004).

[39] Multichannel sampling with unknown gains and offsets: A fast reconstruction algorithm. M Vetterli, Y Lu, Proceedings of Allerton Conference on Communication, Control and Computing (2010).

[40] The Wald test and Cramer-Rao bound for misspecified models in electromagnetic source analysis. L J Waldorp, H M Huizenga, R P P P Grasman, IEEE Transactions on Signal Processing **53**:3427-3435 (2005).

[41] Parametric Array Calibration. S Wan, Ph D Thesis, The University of Edinburgh (2011).

[42] Least squares error detection for non-coherent cooperative relay systems. L Xiong, J K Zhang, IEEE Transactions on Vehicular Technology **61**:3677-3692 (2012).

[43] Blind estimation of finite alphabet digital signals using eigenspace-based beam-forming techniques. J L Yu, Y C Cheng, Signal Processing **84**:895-905 (2004).

[44] Impulse radio ultra-wide-band ranging based on maximum likelihood estimation. H Zhan, J Ayadi, J Farserotu, Jean-Yves Le Boudec, IEEE Transactions on Wireless **8**:5852-5861 (2009).

[45] Full diversity blind signal designs for unique identification of frequency selective channels. J K Zhang, C Yuen, IEEE Transactions on Vehicular Technology. arXiv.org > cs > arXiv:0805.3118v1 (2012).

[46] Evaluation of the LTE Positioning Capabilities in Realistic Navigation Channels. J A del Peral-Rosado, G Seco Granados, Ph D Thesis, Universitat Autònoma de Barcelona. (2014).

[47] Bayesian nonlinear filters for direct position estimation. P Closas, C Fernández-Prades, IEEE Aerospace Conference 1-12 (2010).

2.3. VLSI Design

This is an interesting new category that arose from a collaboration with R. Suaya of Mentor Graphics. In [6, 1] the authors model capacitance coupling and extraction of wires for timing and noise simulation of digital circuits. The model turns out to be separable and thus amenable to a successful application of VARPRO. This approach replaces the use of full field solvers in 3D, a daunting task.

Fast solution of the 3D field equations for EM propagation in complex media is obtained by approximating the Green function, that leads to separable nonlinear least squares problems, as is exemplified in [8]. Another similar application occurs in the consideration of gigahertz frequencies on clocked digital systems, where inductance effects become significant [2]. More elaborated use of similar ideas can be found in [3, 4, 5].

[1] Transmission line design of clock trees. R Escovar, R Suaya, Proceedings IEEE/ACM International Conference on Computer-Aided Design 334-340 (2002).

[2] Optimal design of clock trees for multi-gigahertz applications. R Escovar, R Suaya, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **23**:329-345 (2004).

[3] Analytical expressions for high-frequency VLSI interconnect impedance extraction in the presence of a multilayer conductive substrate. N Srivastava, R Suaya, K. Banerjee, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **28**:1047-1060 (2009).

[4] Efficient 3D high-frequency impedance extraction for general interconnects and inductors above a layered substrate. N Srivastava, R Suaya, K Banerjee, DATE '10 Proceedings of the Conference on Design, Automation and Test in Europe 459-464 (2010).

[5] Accurate calculations of the high-frequency impedance matrix for VLSI interconnects and inductors above a multi-layer substrate: A VARPRO success story. N Srivastava, R Suaya, V Pereyra, K Banerjee, Exponential Data Fitting and its Applications, Bentham Books, Chapter 9 (2010).

[6] Interconnect modeling for high speed digital circuits-the role of RLC coupling. R Suaya, IEEE Proc. on Modeling and Simulation (2002).

[7] Synthesis strategies based on the appropriate use of inductance effects. R Suaya, R Escovar, US Patent 7,013,442 (2006).

[8] Fast high-frequency impedance extraction of horizontal interconnects and inductors in 3-D ICs with multiple substrates. C Xu, N Srivastava, R Suaya, K. Banerjee, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **31**:1698-1710 (2012).

3. Medical and Biological

Many applications to imaging and spectroscopy in biological and medical systems can be found in this section. Again, identification problems are common. A software tool which appears in several articles is the open source package, TIMP [5]. TIMP is a problem solving environment for fitting separable nonlinear models to measurements arising in Physics and Chemistry written in R. It has been extensively tested for time-resolved spectroscopy and FLIM-FRET data (FLIM: Fluorescent Lifetime Imaging Microscopy; FRET: Förster resonance energy transfer).

In [9], a tool for the processing and automatic quality grading in the fish industry is developed based on diffuse reflectance imaging and the subsequent unmixing of the absorption spectra using a constrained least squares model to detect hemoglobin concentration. It is common for the absorption lines to have a Gaussian or Lorentzian distribution shape, so VARPRO is used in the decomposition of the spectra.

In [11], a dynamical model, described by the equations of motion of a spring-mass-damper system, was set up to estimate the impedance (force-position) for the human elbow. Then, a system identification technique based on prediction-error minimization (PEM) was developed of this non-parametric time-domain model augmented with a parametric noise model. VARPRO is used in the approximation.

The paper [4] describes a methodology to determine the Förster resonance energy transfer in live cells as measured using fluorescence imaging microscopy. The parameters of the nonlinear quantitative model can be computed using a variable projection type of algorithm. The software used is part of the open source package TIMP. In [10] VARPRO is used to fit NMR spectra to depict small and highly oblique nerves of the lumbosacral plexus. The aim is to use the methodology for diagnostics of pathologies.

The articles [2, 13, 7, 6] are all about compartmental analysis, a mathematical modeling tool that originated in pharmacokinetics and is now used widely in medical and biological applications. The compartment model is formed by separate homogeneous compounds, compartments. These may represent a certain space (blood, brain, etc) or a compound in a specific form (for example in a different chemical binding). The important point is that each compartment is assumed to be homogeneous. They interact with each other by exchanging material and for most medical and biological systems this exchange is assumed to obey a linear differential equation with constant transfer coefficients, so that the system behavior in time is modeled by a linear system of differential equations. Although the equations are linear the solution is not. It can be fitted by a linear combination of nonlinear functions (basis functions), and the parameters can then be obtained through a variable projection algorithm.

For medical imaging applications, a radioligand (a radioactive biochemical substance, in particular a radiolabelled substance) or tracer is introduced, usually intravenously. The transport and the binding rate of the tracer are assumed to depend linearly of the difference of the tracer concentration between two compartments, so defining the system of ODE in the tracer concentrations. Often the data measured are sums of the concentrations.

[1] Parameter estimation and Cramér-Rao lower bounds in non-contact temperature measurement. Dominik Exel, Bernhard Zagar, Stefan Schuster, Vera Ganglberger and Johann Reisinger, Technisches Messen **86** (2018).

[2] Generalized separable parameter space techniques for fitting 1K-5K serial compartment models. D J Kadrmas, M B Oktay, Medical Physics **40** (2013).

[3] Improved cardiac shim using field map estimate from multi-echo Dixon

method. P Kellman, D Hernando, S Shah, Proc Intl Soc Mag Reson 4522 (2009).

[4] Global analysis of Förster resonance energy transfer in live cells measured by fluorescence lifetime imaging microscopy exploiting the rise time of acceptor fluorescence. S P Laptenok, J W Borst, K M Mullen, I H M van Stokkum, A J W Visser, H van Anerongen, Physical Chemistry Chemical Physics **12**:7593-7602 (2010).

[5] TIMP: An R package for modeling multi-way spectroscopic measurements. Katharine M. Mullen, Ivo H. M. van Stokkum, Journal of Statistical Software **18**:1-46 (2007).

[6] Blind identification of the kinetic parameters in three-compartment models. D Y Riabkov, E V R Di Bella, Physics in Medicine & Biology **49** (2004).

[7] Estimation of kinetic parameters without input functions: analysis of three methods for multichannel blind identification. D Y Riabkov, E V R Di Bella, IEEE Transactions on Biomedical Engineering **49**:1318-1327 (2002).

[8] EEG Signal Processing. S Sanei, J A Chambers, Wiley (2013).

[9] Detection of blood in fish muscle by constrained spectral unmixing of hyper-spectral images. M H Skjelvareid, K Heia, S H Olsen, S Kristian, Journal of Food Engineering **212**:252-261 (2017).

[10] High isotropic resolution T2 mapping of the lumbosacral plexus with T2-prepared 3D turbo spin echo. N Sollmann, D Weidlich, B Cervantes, E Klupp, C. Ganter, H Kooijman, E J Rummeny, C Zimmer, J S Kirschke, D C Karampinos, Clinical Neuroradiology (2018).

[11] Closed-loop identification: application to the estimation of limb impedance in a compliant environment. D T Westwick, E J Perreault, IEEE Transactions on Biomedical Engineering **58**:521-530 (2011).

[12] Primed for efficient motion: ultrafast excited state dynamics and optical manipulation of a four stage rotary molecular motor. Theodore E Wiley, Arkaprabha Konar, Nicholas A Miller, Kenneth G Spears and Roseanne J Sension, J. Phys. Chem. A **122**:7548-7558 (2018).

[13] Separable parameter estimation method for nonlinear biological systems. F X Wu, L Mu, 3rd International Conference on Bioinformatics and Biomedical Engineering 1-4 (2009).

3.1. Spectroscopy

After obtaining expressions for the identification of relaxation times associated with kinetic fluorescence decay and those associated with the dynamic evolution of fluorophores (chemical compounds that can re-emit light upon light excitation), the author of [3] suggests the use of variable projection algorithms in the evaluation of photochemical bioimaging, when the fluorophores are used as the probe molecules. In these studies a multi-exponential decay surface can be ascribed to each pixel, where the fluorescence decay times and the corresponding emission or excitation wavelength dependent amplitudes can be recovered by the VARPRO algorithm.

The first part of [4] is a survey of the adaptation of the variable projection algorithm to the case of matrix data and of constraints on the linear parameters. This form of least squares approximation to fit linear combination of nonlinear functions is common in the applications considered in the paper, namely, spectroscopy, microscopy and mass spectrometry. The authors emphasize the importance of forming the residual vector in a particular manner to avoid storing and operating with tensor products, and describe a way to compute the covariance for the linear parameters using less memory resources.

The application to spectroscopy involves determining the kinetics of a compartmental model of a photo-system for conversion of photons into chemical energy, using the time-resolved fluorescence measurements. This is a separable nonlinear least squares problem with matrix data Ψ and unknown kinetic model $C(z)\,E^T$, where z are the non-linear parameters and the spectra is represented by the (non-negative) linear coefficient matrix E^T: $\Psi \simeq C(z)\,E^T$. It is solved using the software package TIMP.

In the application to microscopy the authors consider the detection of a protein-protein interaction by the simultaneous analysis of multiple FLIM images. FLIM counts the photons detected at several time intervals and over many locations. The data analysis gives rise to a separable nonlinear problem with the same nonlinear parameters but different linear ones, and multiple right hand sides (also termed global analysis problem) and with a matrix data:

$$min_{z\in\mathbb{R}^{n}} \left\|\Psi - C(z)\,E^T\right\|_2 .$$

The kinetic processes of the fluorescence decays are exponential and are represented in the columns of $C(z)$ convolved with an instrumental response function (IRF). The rows of E are the amplitudes corresponding to each kinetic process.

An additional point to consider here is the distribution of errors in the FLIM data, a count of number of photons detected at a given position and time. If the count is large, then to assume that the data errors have a Gaussian distribution with mean 0 and variance σ^2 is valid and the least squares criteria acceptable. For smaller counts, the errors have a Poisson distribution and the least squares estimates are not very good. One possible correction is to weight the data points $\Psi[i,j]$ with the factor $1/\sqrt{\Psi[i,j]}$.

The third area of application considered is mass spectrometry (MS), in conjunction with gas (GC/MS) or liquid (LC/MS) chromatography. These are analytical chemistry techniques that combine the physical separation of the different molecules in the chromatography column with the separation of the ions according to their mass-to-charge ratio in the MS step. In the GC or LC step the sample molecules passing through a column, elute or come off at different times according to their affinity with the chemical in the column. In the MS step the sample is ionized and the resulting ions separated according to mass-to-charge ratio by deflection due to a electric or magnetic field.

The GC/MS or LC/MS measurements of a sample can be modeled by $\Psi \approx CE^T$. Here C are the elution profiles and E the mass spectra resolved with respect to the mass-to-charge ratio. When several samples of the same compound are considered, the elution profiles are different but the mass spectra is considered the same. Often C can be well represented by choosing columns of exponentially modified Gaussians with 4 parameters, width, location, decay rate and amplitude. The problems can be solved with variable projections algorithms.

The performance of three gradient type algorithms: alternating least squares, VARPRO and the Kaufman simplification are compared in the case of a multi-exponential model of a photo-physical system [5]. Corroborating results of other authors, alternating least squares where the linear and nonlinear parameters are alternatively fixed and the problem minimized over the complementary set, is found to be the least efficient. The Fisher information matrix computation enables the authors to determine a lower bound for the covariance estimate of the precision of the nonlinear parameters, and conclude that in the present case the Kaufman simplification is the most cost efficient.

MEG or magneto-encephalogram is an imaging tool that can measure changes in the neural activity on a very small time scale (of the order of milliseconds). In this paper [2] the authors compare several algorithms that solve the inverse problem: given magnetic field values at a number of measurement points, reconstruct the sources, i.e., compute the location and moment parameters of the set of dipoles whose fields best approximates the data in the least squares sense.

This is a large nonlinear optimization problem with a complex objective function and many local minima. However, the model is a separable function, i.e., a linear combination, with coefficients depending on the dipole moment parameters, of nonlinear functions of the location parameters, amenable to variable projection. The authors compare several gradient-free algorithms for the reduced nonlinear problem: simulated annealing, genetic algorithms and a tabu search, and conclude that for the given problem, the best algorithm is a local genetic algorithm.

In 2D spectroscopy, contrary to conventional 1D, the third order optic signal at given population times is a function of two frequencies. In [8] the authors propose a method to analyze the 2D signals using a global analysis method based on the variable projection algorithm. To reduce the dimensionality, the 3D complex valued data arrays containing the signal as a function of the excitation and emission frequency and time, are reorganized into a matrix Y, each of its columns representing the evolution in time of a specific pair of frequencies. A multi-exponential model is then defined by $M = E\,A$, where each column of E contains a complex exponential function $E_{nk} = e^{b_n t_k}$ and the amplitudes are found in the matrix A.

The data analysis can also be extended to fit rephasing and non-rephasing data simultaneously by building appending blocks of data to form the matrix Y. As this problem is computationally challenging due to its size, the authors suggest using in the minimization step a subsample of the data in the frequency dimensions, reporting that even using only 5% of the data gives satisfactory results.

[1] Spectroscopic study of NH-tautomerism in novel cycloketo-tetraphenylporphyrins. E A Ermilov, B Büge, S Jasinski, N Jux, B. Roder, The Journal of Chemical Physics **130** (2009).

[2] A comparative study of global optimization approaches to MEG source localization. T Jiang, A Luo, X Li, F Kruggel, International Journal of Computer Mathematics **80**:305-324 (2003).

[3] Fluorescence polarization spectroscopy at combined high-aperture excitation and detection: application to one-photon-excitation fluorescence microscopy. J J Fisz, The Journal of Physical Chemistry A **111**:8606-8621 (2007).

[4] The variable projection algorithm in time-resolved spectroscopy, microscopy and mass spectrometry applications. K M Mullen, I H M Van Stokkum, Numerical Algorithms **51**:319-340 (2009).

[5] Algorithms for separable nonlinear least squares with application to modeling time-resolved spectra. K M Mullen, M Vengris, I H M van Stokkum, Journal of Global Optimization **38**:201-213 (2007).

[6] Modern stable mathematical and software-based methods for distorted spectra restoration. V S Sizikov and A V Lavrov, Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki 18 (2018).

[7] Innovative Strategies in Coherent Multidimensional Electronic Spectroscopy. A Volpato, Ph D Thesis (2017).

[8] Global analysis of coherence and population dynamics in 2D electronic spectroscopy. A Volpato, L Bolzonello, E Meneghin, E Collini, Optics Express **24**:24773-24785 (2016).

3.2. Tomography

The next articles concern tomography: the imaging by sectioning of objects, or more generally, 2 and 3D imaging of the inside of objects using different signals and the reconstruction techniques employed to recover the information about these objects from the imaging data. For some of the medical imaging methods, a radiation source is used and the data acquired are multiple 2D images or projections taken from different angles. In transmission tomography, such as computerized tomography (CT), the radiation source is outside the object: a rotating X-ray tube generates X-rays that travel through the object. The detector measures the line integral of the beam intensity and the quantity to be reconstructed is the attenuation coefficient of the medium. In emission tomography (ECT), like PET and SPECT, the radiation source is inside: a radioactive tracer is introduced into the body. In the case of SPECT, the tracers emit gamma radiation that is measured. PET tracers emit positrons that when colliding with nearby electrons produce two gamma photons. The gamma radiation is detected by a rotating gamma camera that acquires multiple 2D images from different angles. The quantity to be reconstructed in both cases is the distribution and concentration of the radioactive tracer in the different parts of the body. Finally, photo-acoustic tomography is based on the PA effect, the formation of sound waves following light absorption. The PA signals are acquired at several locations around the object using a transducer array, while the goal of the photo-acoustic imaging reconstruction is to retrieve the local pressure rise inside the tissue.

Article [6] evaluates reconstruction techniques applied to data obtained from CT scans. Incidentally, these techniques could also be used in other applications, for example seismic tomography. For medical tomography, there are two different reconstruction procedures: the direct inversion methods that reverse the Radon transform (the line integral of the beam intensity) by, for example, filtered back projection, and the algebraic reconstruction methods, more adequate when the data have not been regularly sampled.

The data produced by a CT scan of an unknown object $\mathbf{u} \in \mathbb{R}^3$ are a finite number of samples $p_i\,(s_i,\, \eta_i)_1^m$ of the intensity integral of a X-ray defined by its so called acquisition parameters, origin $\mathbf{s}$ and direction $\boldsymbol{\eta}$:

$$p\,(\mathbf{s},\, \boldsymbol{\eta}) = \int_{\mathbb{R}} \mathbf{u}\,(\mathbf{s} + t\,\boldsymbol{\eta})\, dt.$$

The reconstruction problem is then reduced to solving the linear rectangular system $W\,\mathbf{u} = \mathbf{p}$, where W is the $m \times n$ projection matrix, depending on the parameters $(s_i,\, t_i)$. This can be computed by least squares methods.

In practice though, often the geometry of the imaging system, i.e., the acquisition parameters ($\mathbf{s}$, $\boldsymbol{\eta}$) are not known accurately, for example due to faulty calibration, and this produces alignment errors. One option would be to consider the problem as a Total Least Squares problem, i.e., a linear problem where the errors are not only restricted to the observations $\mathbf{p}$ but also the matrix W is not known exactly.

Instead the authors suggest a method belonging to the class of projection matching methods. Basically they solve the nonlinear model,

$$W\left(\mathbf{a}\right)\,\mathbf{u} = \mathbf{p},$$

where the unknown $\mathbf{a} \in \mathbb{R}^l$ contains the parametrization of possible rigid motions, three shifts and three rotations, for each projection image. The authors note that this is a possible severely ill-conditioned problem and since its parameters separate, into the nonlinear $\mathbf{a}$ and the linear $\mathbf{u}$, they suggest an alternating optimization procedure. After solving analytically for the linear parameters $\mathbf{u}$, i.e., eliminating the reconstruction part of the problem, they describe several gradient-descent algorithms to solve the resulting nonlinear optimization problem, and state their convergence behavior, considering when approximate derivatives are used and in case of constraints on the alignment parameters.

The articles [1, 3, 6] consider the case of SPECT tomography. SPECT imaging methods are governed by the photon transport equation and the reconstruction involves the attenuated Radon transform. Two unknown quantities must be estimated simultaneously: the radioactive emission source (the distribution of the radionuclides) $f\left(x,E\right)$ and the photon attenuation coefficient of the tissue $\mu\left(x\right)$, from the acquired data $p\left(s,\theta\right)$ on the line defined by $\left(s,\theta\right)$. The variables x and E are position and energy. The attenuated Radon transform is:

$$R_\mu f\left(s,\theta\right) = \int_{-\infty}^{\infty} f\left(s\theta + t\theta^\perp\right) exp\left(-\int_t^{\infty} \mu\left(s\theta + \tau\theta^\perp\right) d\tau\right) dt \ = p\left(s,\theta\right).$$

Under the assumption that the emission data follow a Gaussian distribution, a nonlinear least squares problem can be defined including a regularization term to avoid irregular distributions of $\left(f,\mu\right)$,

$$\min_{f,\mu} \left\|R_\mu f - p\right\|_2^2 + \alpha I\left[f,\mu\right].$$

Bronnikov exploits the fact that the variables f and μ are respectively linear and nonlinear to design a variable projection algorithm along the lines of VARPRO that works well. On the other hand, Gourion et al. use nonlinear optimization methods directly. They also consider that a Poisson distribution is more appropriate for the data.

Finally, articles [2] and [5] study the newest field in biomedical imaging: photo-acoustic computed tomography (PACT). In this imaging technique, short laser pulses are directed at the object. The absorption of the optical energy produces local heating causing expansion of the tissue and consequent photo-acoustic wave-fields that can be measured outside the body using piezoelectric ultrasonic transducers. The signals that they receive are convolved with their acoustic-electric impulse response (EIR). In [5] the authors choose to incorporate the effect of EIR into the reconstruction. This results in an inverse model with separable linear and nonlinear parameters:

$$\mathbf{u} = \mathbf{H}\,(\mathbf{h})\boldsymbol{\theta}.$$

Here, the matrix $\mathbf{H}$ contains the approximations to the function $\mathbf{A}\,(\mathbf{r}) = \sum \theta_j \phi_j\,(\mathbf{r})$ which represents the absorbed optical density of the object under study. $\mathbf{H}$ also depends on the sampled EIR values represented by $\mathbf{h}$. The image reconstruction is now formulated as an optimization problem in $\mathbf{h}$ and $\boldsymbol{\theta}$ with regularization terms:

$$\min \left\|\mathbf{u} - \mathbf{H}\,(\mathbf{h})\,\boldsymbol{\theta}\right\|^2 + \lambda R_1\,(\boldsymbol{\theta}) + \alpha R_2\,(\mathbf{h})\,.$$

To estimate the parameters, the authors alternate the minimization between the linear and the nonlinear parameters.

[1] Numerical solution of the identification problem for the attenuated Radon transform. A V Bronnikov, Inverse Problems **15**:1315-1324 (1999).

[2] Motion estimation and correction in photo-acoustic tomographic reconstruction. J Chung, L Nguyen, SIAM Journal on Imaging Sciences **10**:216-242 (2017).

[3] The inverse problem of emission tomography. D Gourion, D Noll, Inverse Problems **18** (2002).

[4] Attenuation correction using SPECT emission data only. D Gourion, D Noll, P Gantet, A Celler, J P Esquerre, IEEE Transactions on Nuclear Science **49**:2172-2179 (2002).

[5] A constrained variable projection reconstruction method for photoacoustic computed tomography without accurate knowledge of transducer responses. Q Sheng, K Wang, T P Matthews, J Xia, L Zhu, L V Wang, M A Anastasio, IEEE Transactions on Medical Imaging **34**:2443-2458 (2015).

[6] Automatic alignment for three-dimensional tomographic reconstruction. T van Leeuwen, S Maretzke, K J Batenburg, Inverse Problems **34** (2018).

3.3. Nuclear Magnetic Resonance (NMR) Spectroscopy and Imaging

This is one of the most important applications of VARPRO as explained in [11] and we keep seen strong use as many more specific and new results come out for practical applications.

The physical phenomenon associated with NMR involves a sample that is placed in a magnetic field and is irradiated with a radio-frequency (RF) pulse of a determined resonant frequency. The nuclei in the sample emit a signal that can be recorded and interpreted. NMR can be used to identify what molecules are present because the resonant frequency (Larmor frequency) depends among other factors, on the molecules. In imaging applications, hydrogen is typically the molecule of interest.

In more detail, the $\frac{1}{2}$-protons in the hydrogen nuclei have two eigenstates ($m = \pm\frac{1}{2}$), which in the absence of a magnetic field have the same energy. After a strong external static magnetic field is applied, most of the protons fall into the lower of the two states (for most isotopes used in NMR: $m = +1/2$). When the RF pulse is applied some protons are excited back into higher energy state ($m = -1/2$) and when they decay back to the lower state an electromagnetic radiation is emitted that can be measured. In the literature, T_1 or relaxation time, defines the equilibrium recovery time needed by the sample after the RF excitations. To produce images of the interior of the sample, i.e., to localize the NMR signals, spatial variations in the magnetic field strength across the sample are generated, the gradient fields.

Already in our 2003 review article we listed a number of articles with applications in which the data obtained through NMR was evaluated using the variable projection algorithm. An important area of application is still in vivo MR spectroscopy. In [15] contrast enhanced MR imaging is used to obtain a time-series of the contrast concentration in the blood plasma and the extra-vascular, extra-cellular space (EES) of prostate tissue, both cancerous and non-cancerous. The perfusion model, i.e., the pharmacokinetic model of the passage of fluid between the capillaries and the capillary bed (EES) used is a two-compartment model (Tofts), with the contrast concentration in each voxel given (after discretization) by $\mathbf{p}_i = \mathbf{a}_i v_p + \mathbf{A}\mathbf{h}_i(k_{ep})\, K^{trans}$. Here, the linear parameters are the transfer constant K^{trans} and the vascular fraction v_p, while the nonlinear parameter is the rate constant k_{ep}. The arterial input function $\mathbf{a}$ is a known function of time. This model can be fitted voxel-wise to the MR image data in the least squares sense.

The authors have compared two different algorithms to obtain the perfusion parameters: using Levenberg-Marquardt to estimate all the parameters and applying the variable projection separation of variables idea to define a nonlinear LSQ problem in the parameter k_p, followed by the solution of the linear LSQ problem in the two linear parameters. To solve the nonlinear optimization problem in one variable they use Golden Section Search. Their accuracy and noise sensitivity comparisons included numerical simulations using parameter values in the range reported both for normal and malignant tissue to generate simulated MR signals that were later converted back into concentration functions after the determination of the parameters by the two optimization algorithms mentioned above. The results were comparable, but the VP based technique was three times as fast as LM for all the variables. More important for the medical application were the clinical trials with 20 patients. Here the LM failed to converge in approximately 15% of the tissues, including normal and cancerous, whereas the VP technique converged in 100% of cases.

In article [34], global and target analysis of the time-resolved spectra obtained in bioenergetic applications are reviewed. Spectroscopy is used here as a tool to investigate the dynamic properties of complex biological systems through the application of a short pulse of high energy that produces reactions like absorption or fluorescence. Often the data collected depend on two parameters, one variable is the wave length and the second is the time after excitation.

To analyze the measurements and estimate the physicochemical parameters, both the kinetics (the compartmental model determined by transitions between the states) and the spectra must be modeled. One assumption is that the system is separable, meaning that the spectroscopic data of a complex compound is the superposition of the spectroscopic properties of the components weighted by their concentration. The simplest unidirectional kinetic model is used in global analysis, when the response is considered to be a sum of a few (2-4) exponential decays convolved with the instrument response function (IRF), i.e., the data at different wavelengths are all approximated with a single set of exponentials. Target analysis is used when the problem requires a more complicated kinetic model involving for example forward and reversible reactions, independent decays, etc.

In both cases the parameters in the final model (both the kinetic and the spectra) must be fitted. Assuming normally distributed noise, a nonlinear least squares fit gives the maximum likelihood estimator. The nonlinear least squares approximation leads to a separable nonlinear problem that is solved using the variable projection algorithm. Results for the application to the study of ultrafast dynamics of the photoactive yellow protein are presented.

In [8] a complementary approach to global analysis is described, the so called lifetime density analysis (LDA). It is a technique used in ultrafast (femto- and pico-second) spectroscopy, when investigating energy and charge transfer in complex photosystems. Instead of using a small set of exponential decays to fit the data, one assumes that they are better represented by the integral of a continuous distribution of decays. This integral is discretized using a sum of a large ($\sim$ 100) lifetimes distributed along the time of the experiment. The resulting approximation problem in matrix notation is:

$$min_{x,\tau} \left\| D\left(\tau\right) x - A_n \right\|_2^2 .$$

Here, A_n is the (mxn) measurement matrix at the m time delays and n

wavelengths, D is the matrix of the IRF convoluted with the exponential delays and x contains the amplitudes associated with the lifetimes. Again, it is a separable nonlinear least squares problem and it is solved by the variable projection algorithm.

Contrary to the global analysis strategy, the LDA approach has a large number of parameters and overfitting must be considered. The authors suggest the use of the well-known regularization techniques, truncated SVD and Tikhonov regularization. They also discuss briefly several other methods which they have included into the software, an open source Python package with GUI, that can be downloaded from the web.

Article [20] describes a technique to obtain more accurate estimates of spectral parameters when evaluating MR spectroscopic imaging (MRSI). One of the difficulties in spectral quantification is that the problem is not well conditioned. A way to avoid the large uncertainties is to incorporate the spatial smoothness information contained in the tissue properties. So, instead of computing the spectral parameters for each voxel independently, the idea is to work with a joint formulation:

$$\left(\hat{\mathbf{a}},\, \hat{\boldsymbol{\theta}}\right) = \arg \min_{\mathbf{a},\, \boldsymbol{\theta}} \left\|\mathbf{d} - \mathbf{K}\left(\boldsymbol{\theta}\right)\, \mathbf{a}\right\|_2^2 + \mathbf{R}\left(\mathbf{a},\, \boldsymbol{\theta}\right). \tag{3.1}$$

Here $\mathbf{a}$, $\boldsymbol{\theta}$ are the linear and nonlinear parameters, $\mathbf{d}$ the data, and $\mathbf{R}\left(\mathbf{a},\, \boldsymbol{\theta}\right)$ is a two-term penalty function added for regularization:

$$\mathbf{R}\left(\mathbf{a},\, \boldsymbol{\theta}\right) = \lambda \left\|\mathbf{W}_{\boldsymbol{\theta}}\right\|_2^2 + \eta \left\|\mathscr{W}_{\mathbf{a}}\right\|_1 .$$

The terms in $\mathbf{W}_{\boldsymbol{\theta}}$ and $\mathscr{W}_{\mathbf{a}}$ are designed to impose smoothness of all the parameters across the voxels. The proposed algorithm reformulates the optimization problem (3.1) as two consecutive subproblems, the first is the nonlinear least squares problem (3.1) but only with the l_2 penalty term of $\mathbf{R}$, while the second subproblem is again equation (3.1), now only with the l_1 penalty term of $\mathbf{R}$. The first subproblem is solved using the variable projection strategy to compute an updated value for the nonlinear parameters $\boldsymbol{\theta}$, which is then fed into the second subproblem. This is then a linear least squares problem in the parameters $\mathbf{a}$ with an l_1-norm regularization term and is solved by an alternating direction method of multipliers.

The technique was tested using several data sets, both from simulated and from *in vivo* experiments, and its performance was compared with QUEST, another method. The conclusion is that the accuracy is considerably improved, which was also confirmed by a Cramer-Rao bound analysis. The drawback is the cost of the computations, $\mathcal{O}\left(P^2N^2\right)$, with P the number of voxels and N the parameters for each voxel. This compares unfavorably with a voxel by voxel approach, where the cost is $\mathcal{O}\left(PN^2\right)$.

For long *echo-time* MRS signal modeling, VARPRO is used to fit a sum of complex damped exponentials. In [21] and [28] data quantification of metabolites in the case of *short echo-time* is considered. The *in vivo* short *echo-time* MRS are richer and therefore a more efficient solution is to create first a database of *in vitro* spectra measurements v_k of the individual metabolites. The *in vivo* signals $y(t)$ are then fitted with a combination of these v_k, corrected by parameters $\alpha_{k,}$ $\varsigma_{k,}$ and η_k to be determined in order to allow for specific scans. A baseline term $b(t)$ that globally represents the signals of other non-dominant, non-specified metabolites possible present, may also be added. This term is approximated by splines. The approximation problem considered is a non-linear least squares problem with a regularizing term:

$$\min_{\boldsymbol{\alpha},\mathbf{c},\boldsymbol{\varsigma},\boldsymbol{\eta}} \sum_{t=t_0}^{t_m} \left| y(t) - \sum_{k=1}^{K} \alpha_k (\varsigma_k)^t (\eta_k)^{t^2} v_k(t) - (\mathscr{A}\mathbf{c})(t) \right|^2 + \lambda^2 \mathbf{c}^H D^H D\mathbf{c}.$$

Here, the matrix D is a discrete differential operator acting as a regularization matrix. The parameters α_k have the form $\alpha_k = a_k \exp(i\phi_k)$ with a_k the real amplitudes and ϕ_k the phase shifts. The ς_k and η_k are nonlinear in the damping correction parameters. The coefficients of the spline approximation of the baseline $\mathbf{c}$ are always linear. If the amplitude parameters α_k are linear, the problem can be computed with VARPRO for complex problems. This is also the case even when constrains on the nonlinear parameters are imposed.

However, if there are constraints on the linear parameters α_k there is a difficulty because it is not possible, as is done in the variable projection technique, to write a closed-form solution for the linear problem part. One common occurrence for the approximation in MRS data quantification is that it requires equal phase corrections ϕ_k and non-negative real amplitudes a_k.

The algorithm described in the articles approaches this case by optimizing the nonlinear problem with the nonlinear parameters augmented by a common phase correction ϕ_0. At each iteration of the nonlinear optimization an approximate value for the linear parameters is computed from the constrained linear least squares problem with the fixed nonlinear parameters of the last iteration step.

It is shown in [28], by numerical tests, that this algorithm is more efficient and more accurate than a “full” LS solution, where the linear and nonlinear parameters are optimized together. An open source software package AQSES based on the above techniques is described and tested. See [21]. In article [27] this software is further improved by taking into account the effect of inhomogeneities in the applied magnetic field and heterogeneities of the tissue. These cause distortions of the NMR signals, for example a broadening of the frequency-domain line with a consequent possible spectral overlap, thereby impeding a correct metabolite quantification.

The components of an NMR signal in the time-domain are the resonance frequencies multiplied by the natural damping function (Lorentzian, Gaussian or Voigt), in turn multiplied by an instrumental broadening function and all with an added noise. The line-shape (actually damping function because it is done in the time-domain) correction algorithm described in [27] is an iterative procedure: A first step is a nonlinear least squares approximation via VARPRO of the signals $\mathbf{y}(t)$ with a model $\bar{y}(t)$, where line-shape distortions and baseline are ignored:

$$\bar{y}(t) = \sum_{k=1}^{K} \alpha_k \exp(-d_k t - g_k t^2 + 2\pi i f_k t)\, v_k\,(t)\,.$$

Using the spectral parameters thus obtained, an undisturbed signal $\hat{\mathbf{y}}(t)$ without the damping part is constructed. Then, from the quotient ${}^{y(t)}/_{\hat{y}(t)}$ a correction of the damping function $g(t)$ is constructed. After smoothing $g(t)$, eliminating any numerical instability and noise, a nonlinear least squares approximations step via VARPRO is again performed of $[\hat{y}(t) \cdot g(t)]_{denoised}$. This process is continued until convergence of the spectral parameters or the damping function. The authors have validated the efficiency of this technique in improving the quantitation results using Monte Carlo simulations.

The articles [1, 33, 6] consider how to estimate the T_1 relaxation time parameter using the variable projection technique. The T_1 parameter is an intrinsic magnetic property of tissue and of importance in many clinical applications. For *in vivo* determination of T_1, multiple datasets of signal intensity at different timings are obtained. The signal intensity can be modeled by the rational function:

$$S(t_i,\, A,\, B,\, C,\, T_1) = \frac{A + B\, \exp(-{}^{t_i}/_{T_1})}{1 - C\, \exp(-{}^{t_i}/_{T_1})},\; i = 1, ...N.$$

If the radio-frequency pulse flip angles are almost perfect (Ernst angle), then the parameter C in the above rational model is close to zero and the model is a very simple one where the linear parameters can easily be separated from the nonlinear one. This model is used in [1] and [33]. The rational model is considered in [14] and the parameters are again determined using VARPRO.

The two articles [11] and [25] describe a software for the quantification of brain metabolites using data obtained from 2D J-resolved magnetic resonance spectroscopy, a technique developed to reduce the spectral overlap common when using clinical strength 3T NMR. The ProFit tool described in the articles implements a variable projection algorithm that also takes into account possible linear relations between parameters of different metabolites, differences between parameters of different spins in the same metabolite and the fixing of specific parameters.

The Dixon technique is a chemical shift imaging method that allows to create fat only or water only images, and can therefore be used when fat or water conceal the signal of interest. In [31], in order to reconstruct the images from a dual-echo Dixon, a voxel-wise cost function is defined in the linear parameters, water and fat magnitudes W_ν, F_ν and the nonlinear ones, initial phase shared by water and fat Φ_{0_ν} and ω_ν, the phase induced by the inhomogeneity of the static magnetic field. To determine the maximum likelihood estimates of these parameters, a variable projection technique is used to solve a simplified nonlinear least squares problem. The method is applied to *in vivo* studies of foot/ankle and CE-MRA of thighs.

The following study [6], on the effect of intra-or-extracellular water accumulation and intracellular acidification in muscles on the rate of transverse relaxation was performed using spectroscopy before and after exercise. The resulting imaging data was modeled by a sum of exponentials and the maximum likelihood fit, a nonlinear least squares approximation, was solved using VARPRO.

The subject of article [10], diffusion MRI (dMRI) uses the diffusion of water molecules in the generation of MRI tissue images. Molecular diffusion in tissues is constrained by interactions with obstacles such as macromolecules, fibers or membranes. In this paper, MIX, a method to characterize the tissue microstructure of white matter fibers is developed. It uses a multi-compartmental model (intra-axonal, extra-axonal, isotropic) and allows for multiple fiber orientations. The data are fitted in the least squares sense to a combination of exponentials with constraints on the linear coefficients. In a first step, the algorithm constructs good initial values both of the linear and the nonlinear parameters. This is accomplished by applying the variable projection technique to separate linear from nonlinear parameters disregarding for the time being the fact that there are constraints on the linear parameters. The reduced nonlinear problem is solved using a stochastic method, a genetic algorithm, and the linear problem with constrained parameters is solved by CVX (Convex programming Matlab program designed by S. Boyd). In a second step, a trust region method is used to estimate all the parameters. The method was tested on synthetic and *ex-vivo* and *in-vivo* brain data.

An interesting technique for quantitative MRI inspired by Compressed Sensing and developed recently, Magnetic Resonance Fingerprinting (MRF), is the subject of the article [36]. The aim is to acquire enough magnetic resonance signal information in a reasonable short time to be able to deduce multiple tissue-specific parameters such as T_1 and T_2 times and spin density $\rho(\mathbf{x})$. Given the time restriction, and based on the assumption that the image can be approximated by a low-dimensional model, MRF opts for an under-sampled $\mathbf{k}-$space using incoherent data acquisition schemes.

The data model has the form: $\mathbf{d} = \mathbf{F}\,\mathbf{S}\,\boldsymbol{\Phi}(\mathbf{T_1}, \mathbf{T_2})\,\boldsymbol{\rho}$, with $\boldsymbol{\Phi}$ and $\boldsymbol{\rho}$ including the parameters that are of interest, $\mathbf{S}$ contains data-acquisition information (coil sensitivities) and $\mathbf{F}$ is the Fourier encoding matrix. If the noise is Gaussian, a maximum likelihood estimation leads to a nonlinear least squares problem to determine the optimum times $\mathbf{T_1}$, $\mathbf{T_2}$ and spin density $\boldsymbol{\rho}$. One of the complications of using optimization methods to solve the problem is the fact that, since there is no analytic expression for the elements of $\mathbf{\Phi}$, then Bloch equation simulations have to be performed at each iteration. The algorithm proposed by the authors is an iterative method of a reformulated problem obtained by the introduction of auxiliary variables and the subsequent splitting into three subproblems, two of them linear least squares problems. The third is a nonlinear separable variables least squares problem solved using variable projection techniques. The time consuming Bloch equation solutions needed here are substituted by dictionary values computed in advance.

Wilson [35] considers a new method for spectral registration. In contrast to previous approaches, the registration problem is formulated as a variable-projection (VARPRO) in the frequency domain. The use of VARPRO allows the incorporation of baseline modeling, whilst also reducing the iterative optimization complexity from two parameters (phase and frequency) to one (frequency). The approach is compared with TDSR (time-domain spectral registration), and found to be more robust to large frequency shifts (>7Hz), baseline distortions and edited-MRS frequency misalignment. In his Ph D Thesis Zhou [37] discusses and compares various methods for exponential fitting associated with NMRI, including VARPRO. In [29] the authors consider an interesting application to the assessment of myofascial trigger point via MRI for patients with migraine.

[1] A robust methodology for in vivo T1 mapping. J K Barral, E Gudmundson, N Stikov, M Etezadi-Amoli, P Stoica, D G Nishimura. Magnetic Resonance in Medicine **64** (2010).

[2] Three-dimensional water/fat separation and T estimation based on whole image optimization. Application in breath-hold liver imaging at 1.5 T, J Berglund, J Kullberg, Magnetic Resonance in Medicine (2012).

[3] Contributions to Signal Processing for MRI. M Björk, Ph D Thesis, Uppsala University (2015).

[4] Clinical Magnetic Resonance Spectroscopy. E B Cady, Springer (2012).

[5] Reliability and Uncertainty in Diffusion MRI Modeling. C N Charles, Ph D Thesis, University of Sidney (2016).

[6] Intracellular acidification and volume increases explain R2 decreases in exercising muscle. B M Damon, C D Gregory, K L Hall, H J Stark, V Gulani, M J Dawson, Magnetic Resonance in Medicine **47** (2002).

[7] Simultaneous phase unwrapping and removal of chemical shift (SPURS) using graph cuts: application in quantitative susceptibility mapping. J Dong, T Liu, F Chen, D Zhou and A Dimov, IEEE Transactions on Medical Imaging **34**:531-540 (2014).

[8] PyLDM-An open source package for lifetime density analysis of time-resolved spectroscopic data. G F Dorlhiac, C Fare, J J van Thor, PLoS Computational Biology (2017).

[9] Quantification of cervical spine muscle fat: a comparison between T1-weighted and multi-echo gradient echo imaging using a variable projection algorithm (VARPRO). James M Elliott, David M Walton, Alfred Rademaker and Todd B Parrish. BMC Medical Imaging **13**:30 (2013).

[10] Microstructure imaging of crossing (MIX) white matter fibers from diffusion MRI. H Farooq, J Xu, J W Nam, D F Keefe, E Yacoub, T Georgiou, C Lenglet, Scientific Reports (2016).

[11] ProFit revisited. A Fuchs, P Boesiger, R F Schulte, A Henning, Magnetic Resonance in Medicine **71** (2014).

[12] The use of the Levenberg–Marquardt and variable projection curve-fitting algorithm in intravoxel incoherent motion method for DW-MRI data analysis. R Fusco, M Sansone and A Petrillo, Applied Magnetic Resonance, **46**:551-558 (2015).

[13] A comparison of fitting algorithms for diffusion-weighted MRI data analysis using an intravoxel incoherent motion model. R Fusco, M Sansone and A Petrillo, Magnetic Resonance Materials in Physics, Biology and Medicine **30**:113-120 (2017).

[14] Maximum likelihood estimation of T1 relaxation parameters using VARPRO. J P Haldar, J Anderson, S W Sun, Proceedings of the 15th Annual Meeting of ISMRM; Berlin, Germany. 41 (2007).

[15] Robust and efficient pharmacokinetic parameter non-linear least squares estimation for dynamic contrast enhanced MRI of the prostate. Soudabeh Kargar, Eric A Borisch, Adam T Froemming, Akira Kawashima, Lance A Mynderse, Eric G Stinson, Joshua D Trzasko and Stephen J Riederer, Magnetic Resonance Imaging Volume **48**:50-61 (2018).

[16] Over-expression of dimethylarginine dimethylaminohydrolase enhances tumor hypoxia: an insight into the relationship of hypoxia and angiogenesis in vivo. V Kostourou, H Troy, J F Murray, E R Cullis, G S J Whitley, Neoplasia, 6:401-411 (2004).

[17] Independent component analysis for automated decomposition of in vivo magnetic resonance spectra. Christophe Ladroue, Franklyn A Howe, John R Griffiths and A Rosemary Tate, Magnetic Resonance in Medicine **50** (2003).

[18] JIGSAW: Joint inhomogeneity estimation via global segment assembly for water–fat separation. W Lu, Y Lu, IEEE Transactions on Medical Imaging **30:1417-1426** (2011).

[19] Fast and accurate compensation of signal offset for T2 mapping. J Michálek, P Hanzlíková, T Trinh, D Pacík, Resonance Materials in Physics, Biology and Medicine 1-14 (2019).

[20] Spectral quantification for high-resolution MR spectroscopic imaging with spatiospectral constraints. Q Ning, C Ma, F Lam, Z P Liang, IEEE Transactions on Biomedical Engineering **64**:1178-1186 (2017).

[21] An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. J B Poullet, D M Sima, A W Simonetti, B. De Neuter, L Vanhamme, P Lemmerling, S Van Huffel, NMR in Biomedicine **20** (2007).

[22] Evaluation of T1ρ as a potential MR biomarker for liver cirrhosis: comparison of healthy control subjects and patients with liver cirrhosis, I Rauscher, M Eiber, C Ganter, P Martirosian, W Safi, A Umgetter, E J Rummeny, K Holzapfel, European Journal of Radiology **83**:900-904 (2014).

[23] Relaxation times of choline, creatine and N-acetyl aspartate in human cerebral white matter at 1.5 T. D R Rutgers, J Van der Grond, NMR in Biomedicine: **15** (2002).

[24] Accurate fatty acid composition estimation of adipose tissue in the abdomen based on bipolar multi-echo MRI. M Schneider, G Janas, F Lugauer, E Hoppe, D Nickel, B Dale, B Kiefer, A Maier and M R Bashir, Magnetic Resonance in Medicine **81**:2330-2346 (2018).

[25] ProFit: two-dimensional prior-knowledge fitting of J-resolved spectra. R F Schulte, P Boesiger, NMR in Biomedicine **19** (2006).

[26] AQSESVP – Description of a Variable Projection implementation for nonlinear least squares with linear bound constraints, applied to accurate quantification of short-echo time magnetic resonance spectroscopic signals. Diana M Sima, Sabine Van Huffel, Departement Elektrotechniek ESAT-SISTA/TR 120, Katholieke Universiteit Leuven (2005).

[27] Line-shape estimation for magnetic resonance spectroscopy (MRS) signals: self-deconvolution revisited. D M Sima, M I O Garcia, J Poullet, A Suvichakorn, J-P Antoine, S Van Huffel, D van Ormondt, Measurement Science and Technology **20** (2009).

[28] Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification. D M Sima, S Van Huffel, Journal of Computational and Applied Mathematics **203**:264-278 (2007).

[29] Quantitative magnetic resonance imaging of the upper trapezius muscles – assessment of myofascial trigger points in patients with migraine. Nico Sollmann, Nina Mathonia, Dominik Weidlich, Michaela Bonfert, Sebastian A Schroeder, Katharina A Badura, Tabea Renner, Florian Trepte-Freisleder, Carl Ganter, Sandro M Krieg, Claus Zimmer, Ernst J Rummeny, Dimitrios C Karampinos, Thomas Baum, Mirjam N Landgraf and Florian Heinen, The Journal of Headache and Pain 20:8 (2019).

[30] Time-resolved contrast-enhanced MR angiography with single-echo Dixon fat suppression. E G Stinson, J D Trzasko, N G Campeau, J F Glockner, J Houston III, P M Young, S J Riederer, Magnetic Resonance in Medicine (2018).

[31] Dual echo Dixon imaging with a constrained phase signal model and graph cuts reconstruction. E G Stinson, J D Trzasko, J G Fletcher, S J Riederer, Magnetic Resonance in Medicine **78** (2017).

[32] 31P-Magnetic resonance spectroscopy studies of nucleated and non-nucleated erythrocytes; time domain data analysis (VARPRO) incorporating prior knowledge can give information on the binding of ADP. Marion Stubbs, Aad Van den Boogaart, C Lindsay Bashford, Pedro M C Miranda, Loreta M Rodrigues, Franklyn A Howe and John R.Griffiths, Biochimica et Biophysica Acta (BBA) - General Subjects **1291**:143-148 (1996).

[33] Estimating T1 from multichannel variable flip angle SPGR sequences. J D Trzasko, P M Mostardi, S J Riederer, A Manduca, Magnetic Resonance in Medicine **69** (2013).

[34] Global and target analysis of time-resolved spectra. I H M van Stokkum, D S Larsen, R van Grondelle, Biochimica et Biophysica Acta **1657**:82-104 (2004).

[35] Robust retrospective frequency and phase correction for single-voxel MR spectroscopy. Martin Wilson, Magnetic Resonance in Medicine (2018).

[36] Maximum likelihood reconstruction for magnetic resonance fingerprinting. B Zhao, K Setsompop, H Ye, S F Cauley, L L Wald, IEEE Transactions on Medical Imaging **35**:1812-1823 (2016).

[37] Estimation and inference in problems from imaging and biophysics. Y Zhou, Ph D Thesis, U. of Delaware (2018).

3.4. Brain Imaging

This is an interesting area with great future potential as more brain research with electro-magnetic methods is performed.

In [1] the authors consider diffusion MRI to map the brain microstructure. For this purpose they introduce a novel data fitting procedure based on multi-compartment models. Their model is separable and they use VARPRO and a global search algorithm in the nonlinear phase. They show that the algorithm is faster than the commonly used linear search without separation. The accuracy and robustness is demonstrated on synthetic and real data.

Data-acquisition time length is a factor in the applicability of $in - vivo$ spectroscopic imaging. Short data-acquisition time is possible when applying echo-planar spectroscopy imaging (EPSI) but with the disadvantage of a poor signal-to-noise ratio. The approach taken in [2] to obtain a high spatiotemporal resolution is to use a hybrid technique using a first step of double-echo chemical shift imaging (CSI) followed by an EPSI step. In the CSI step, the data sets $\mathscr{D}_{1s}$ and $\mathscr{D}_{1L}$ with a limited number of spatial values but with high temporal resolution are acquired. In the second EPSI stage, a data set $\mathscr{D}_2$ with extended space coverage but limited temporal sampling is obtained. The algorithm uses the union-of-spaces idea, namely that the measured signals are a sum of nuisance, baseline (macromolecules) and metabolite signals. It also assumes partial separability of each sub-signal, i.e., that it can be modeled by a linear combination of temporal basis functions $\varphi_l\,(t)$ with spatial components $c_l\,(\mathbf{k})$. In a first preliminary step of the data processing, the nuisance signals are removed from the measured data. In a second step, the temporal basis functions of the baseline and the metabolite signals are determined from the information in the data sets obtained from CSI. Finally, the spatial components are obtained from the signals measured with EPSI. The two last steps involve the solutions of a nonlinear least squares problems and are computed using VARPRO.

In his Ph D Thesis [3], J-B Poullet is concerned with the quantification and classification of MRS data for brain tumor diagnosis. MR Spectroscopy is a complementary technique to MR Imaging for brain cancer diagnosis. It provides metabolic information that is not available with MRI. One of the steps in the long and complicated overall process involves AQSES, a time domain quantification method designed for short echo time MR spectra. The model involves a linear combination of corrected in vitro metabolic profiles and therefore VARPRO is a natural tool that is used with advantage. Although the use of VARPRO in MRS is not new, the author points out that the previous work was restricted to long echo time MRS signals and the use of complex damped exponentials. He found that the advantages of VARPRO are even larger than in the previous applications, since now there are a large number of linear parameters that get eliminated. Also, initializing the nonlinear variable to zero provides a good starting point for the optimization process.

[1] Brain microstructure mapping from diffusion MRI using least squares variable separation. H Farooq, J Xu, E Yacoub, T Georgiou, C Lenglet, CDMRI (MICCAI) (2015).

[2] High-resolution 1H-MRSI of the brain using short-TE SPICE. C Ma, F Lam, Q Ning, C L Johnson, Z P Liang, Magnetic Resonance in Medicine **77** (2017).

[3] Quantification and Classification of Magnetic Resonance Spectroscopic Data for Brain Tumor Diagnosis. J B Poullet, Ph D Thesis, Katholic University of Leuven (2008).

[4] Maximum likelihood reconstruction for magnetic resonance fingerprinting. B Zhao, K Setsompop, H Ye, S F Cauley, L L Wald, IEEE Transactions on Medical Imaging **35**:1812-1823 (2016).

4. Image Processing, Vision, Video

4.1. Blind Deconvolution

The separation of source signature and propagation effects in different types of time signals when there is no exact knowledge of either is called blind deconvolution. One important application of blind deconvolution is in reflection seismology, when separating the convolution of the source signature from the Earth's reflectivity function based on data from seismograms. In [7] the authors have described an algorithm that requires solving a nonlinear, non-convex optimization problem, to recover the sources and source durations based on the information of multiple seismograms that share another convolution element (Earth's reflectivity or Green's function).

The equations defining the unknown sources and source durations that constitute the objective function are obtained from homogeneous relations based on the redundancy of seismograms of multiple events (earthquakes), starting at a similar location with the additional condition of shortest source durations. Using the separation of variables technique, the algorithm iterates between the solution of a linear and a nonlinear optimization problem. It was tested using synthetic seismograms with good results.

The problem also appears in the processing of blurred images when trying to reconstruct the real image. Here the blurred/noisy image is considered a convolution of the exact image with a point spread function (PSF) and possible added noise, making blind deconvolution techniques necessary.

Assuming that the PSF depends non-linearly on parameters $\boldsymbol{\phi}$, the blind deconvolution model for all the pixels of an image can be written as:

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\phi})\,\mathbf{x} + \boldsymbol{\eta}.$$

Here $\mathbf{y}$ and $\mathbf{x}$ are the observed and true images pixel-wise respectively, $\mathbf{A}$ is defined by the point spread function and $\boldsymbol{\eta}$ is the noise. $\mathbf{A}$ is an ill-conditioned matrix, often sparse and/or structured; if the blur is spatially invariant and periodic boundary conditions are imposed, then A is a circulant matrix. One can obtain an approximation to the parameters $\boldsymbol{\phi}$ and the true image $\mathbf{x}$ by approximating $\mathbf{y}$ in the least squares sense. The authors in [3] use VARPRO with conjugate gradient as a inner solver for the approximation. In their particular application to astronomy imaging they are able to use a spectral decomposition of the matrix $\mathbf{A}$ to reduce the complexity of the algorithm to $\mathcal{O}(N \log N)$, where N is the number of pixels in the image.

The technique described works well if the number of nonlinear parameters $\boldsymbol{\phi}$ in the PSF is considerably smaller that the number of pixels, the dimension of $\mathbf{x}$. In principle the same methodology could be applied when the number of nonlinear parameters is very large, for example in the case of multi-frame blind deconvolution, where one starts with multiple images $\mathbf{y}_i$ of the same object. But this is very inefficient and the authors suggest instead a decoupling method. The objective function here will be the sum of the objective functions for each frame plus constraints that minimize the difference between two consecutive images $\mathbf{x}_i$

$$\|\mathbf{A}(\boldsymbol{\phi}_1)\,\mathbf{x}_1 - \mathbf{y}_1\|_2^2 + ... + \|\mathbf{A}(\boldsymbol{\phi}_m)\,\mathbf{x}_m - \mathbf{y}_m\|_2^2 + \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 + ... + \|\mathbf{x}_{m-1} - \mathbf{x}_m\|_2^2$$

For this least squares problem again the nonlinear part can be separated from the linear one so a variable projection technique es applicable. The multi-frame blind deconvolution algorithm was tested on a satellite image blurred by the PSF of three pupil phase functions and a deblurred image with a relative error of 0.122 could be obtained. One should note here that the number of nonlinear parameters in this type of blur is $\mathcal{O}\left(10^4\right)$ for an $256x256$ image. Using decoupling and taking advantage of the block structure of the Jacobian matrix for a parallel implementation, the variable projection algorithm saves considerable time.

In the following article [2] the authors developed a gradient projection algorithm to solve a nonlinear separable ill-conditioned least squares problem with inequality constraints on the linear parameters. This technique is then applied to a blind deconvolution imaging problem. The optimization method used is VARPRO with some modifications that take into consideration the non-negativity constraints on the linear parameters. First, to address the ill-conditioning of the matrix $\mathbf{A}$, a Tikhonov regularization term is introduced and the problem is reformulated as:

$$\min_{\mathbf{x},\boldsymbol{\phi}} \|\mathbf{A}(\boldsymbol{\phi})\,\mathbf{x} - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2, \text{ subject to } \mathbf{x} \geq \mathbf{0}.$$

The solution of the linear least squares subproblem can be written in closed form by incorporating a diagonal matrix D whose diagonal elements are 1 or 0 depending if the corresponding linear parameters are nonnegative or not. Hence, it is possible to write formally the nonlinear reduced cost functional in a similar form to the VARPRO algorithm. But to solve the nonlinear problem we now require an explicit knowledge of the linear parameters. Therefore the Gauss-Newton algorithm used in the solution of the nonlinear reduced problem must be modified by adding the solution of the linear least squares problem at each iteration step. To avoid local minima traps the algorithm uses an implicit filtering algorithm.

The topic discussed in [4] is *shape from defocus*, i.e., the reconstruction of the 3D geometry of a scene given several defocused images of it, taken under different optical settings. Two variables are of interest: the surface s, measured from the lens plane to the scene and the radiance r. Given PSF, the point spread function $h_s(\mathbf{x}, \mathbf{y})$, that depends on the imaging device and the scene, the defocused image $I(\mathbf{y})$ is expressed by

$$I(\mathbf{y}) = \int h_s(\mathbf{x}, \mathbf{y})\, r(\mathbf{x})\; d\mathbf{x}$$

and is measured by a charge-coupled device sensor (CCD sensor).

Given a number of images $I_i(\mathbf{y})$ of the same scene but for different PSF and recorded with errors, one can determine optimal values for the surface s and the radiance r by minimizing the $\mathcal{L}_2$-norm of the errors:

$$\sum_i \left\| I_i(\mathbf{y}) - \int h_s^i(\mathbf{x}, \mathbf{y})\, r(\mathbf{x})\; d\mathbf{x} \right\|^2 .$$

By reorganizing the data, the final problem can be written in the simpler form,

$$\hat{s}, \hat{r} = \arg\min_{s,r} \left\| \mathbf{I} - \mathbf{H_s} r \right\|^2 ,$$

where $\mathbf{H}_s$ is now a linear operator. Using an analogous technique to the one in [6] the authors extend the central result in this paper to the present linear operators case, namely the possibility to separate the nonlinear from the linear problem and to obtain therewith a reduced nonlinear problem without losing extrema of the original problem. The reduced nonlinear problem, expressed in terms of the orthogonal operator $\mathbf{H}_s^{\perp}$, is to find the local extremum of the function $\psi(s) \doteq \left\| \mathbf{H}_s^{\perp} \mathbf{I} \right\|^2$. The first step is to compute the orthogonal operator and for this only the finite dimensional range of the PSF operator $\mathbf{H}_s$ must be determined. If the PSF is known, the operator $\mathbf{H}_s$ can be expressed using its SVD decomposition and from it an expression for the orthogonal projection operator is obtained. If the PSF is not known, then the range of PSF can be obtained from a number of defocused images by solving a system of linear equations, $\mathbf{H}_s^{\perp}[\mathbf{I}_1 ... \mathbf{I}_T]$. The authors tested the algorithm both with synthetic data without noise and on real data with sensor noise.

[1] Adaptive Blind Signal and Image Processing. A Cichocki, S Amari, Wiley (2002).

[2] Constrained numerical optimization methods for blind deconvolution. A Cornelio, E L Piccolomini, J G Nagy, Numerical Algorithms **65**:23-42 (2014).

[3] An efficient computational approach for multi-frame blind deconvolution. Y W D Fan, J G Nagy, Journal of Computational and Applied Mathematics **236**:2112-2125 (2012).

[4] A geometric approach to shape from defocus. P Favaro, S Soatto, IEEE Transactions on Pattern Analysis and Machine Intelligence **27**:406-417 (2005).

[5] 3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion-Blur. P Favaro, S Soatto, Springer (2007).

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

[7] Blind deconvolution of seismograms regularized via minimum support. A A Royer, M G Bostock, E Haber, Inverse Problems **28** (2012).

4.2. Image Processing

Storing and/or transmission of the large amount of data contained in digital images, both still and moving, requires some form of image compression. One common way to approach the problem is by using block coders that subdivide the image into non-overlapping blocks of pixels and then process each block. The disadvantage of this approach is that at low bit-rates artifacts are generated at the block boundaries. In [11, 12] the authors consider ways to avoid that. The solution they study is to downsample the image, apply the coding process and at the end interpolate to obtain the original resolution. They design two algorithms: in the first they optimize the interpolation filter giving rise to a linear least squares problem and then, in a second more comprehensive algorithm, they optimize both the decimation and the interpolation filters, getting a separable nonlinear least squares problem:

$$\min_{\mathbf{f},\mathbf{g}} \|\mathbf{x} - \boldsymbol{\Phi}(\mathbf{f})\,\mathbf{g}\|_2^2 .$$

Here, $\mathbf{g}$ represents the interpolation filter, $\mathbf{f}$ the decimation filter and $\boldsymbol{\Phi}(\mathbf{f})$ the interpolation window. The nonlinear problem after separating variables is:

$$\min_{\mathbf{f}} \left\| P^{\perp}_{\boldsymbol{\Phi}(\mathbf{f})}\mathbf{x} \right\| ,$$

where $P^{\perp}_{\boldsymbol{\Phi}(\mathbf{f})}$ is the orthogonal projector. For simplicity and to reduce the size of the problem, they solve a suboptimal problem considering only low-pass filters, obtaining an univariate minimization problem solved using Brent's minimization algorithm without derivatives. Applied to several classic compression benchmark problems the results are encouraging.

In [9] the authors consider the determination of the shape of a polygon P with vertices $z_1, z_2, ..., z_N$ from its moments $\tau_0, \tau_1, ..., \tau_M$, defined by a special case of Davis's theorem:

$$\tau_k = k(k-1) \int\int_P z^{k-2} dxdy = \sum_{n=1}^{N} a_n z_n^k,$$

where the coefficients a_n are:

$$a_n = \frac{i}{2}\left(\frac{\bar{z}_{n-1} - \bar{z}_n}{z_{n-1} - z_n} - \frac{\bar{z}_n - \bar{z}_{n+1}}{z_n - z_{n+1}}\right). \tag{4.1}$$

The question is: given M complex moments $\{\tau_k\}_{k=0}^{M}$ with $(\tau_0 = \tau_1 = 0)$, is it possible to recover the polygon vertices when the number of vertices N is known but not its order?

The paper suggests the use of an hybrid technique to obtain the values of the vector of vertices $\mathbf{z}$ and to use either a Prony or a matrix pencil type algorithm to compute a good initial estimate for the nodes of the polygon. In a next stage they use VARPRO to determine the best values for the nodes z_i from the nonlinear least squares problem:

$$\min_{\mathbf{z}} \sum_{k=0}^{M} \left| \tau_k - \sum_{n=1}^{N} a_n(\mathbf{z})\, z_n^k \right|^2 .$$

Since the order of the vertices z_i is important (and unknown *a priori*), they impose the constraints (4.1) and use the technique described by Kaufman-Pereyra [6]. The tests show a noticeable improvement with respect to the values obtained using only the Prony or matrix pencil algorithms of the first stage.

The following two articles belong to geometric tomography. This is the branch of Mathematics that studies the recovery of information about geometric objects given for example its orthogonal projection onto subspaces. It has important applications in computerized tomography, electron microscopy and computer vision.

Article [3] considers algorithms for the reconstruction of origin-symmetric convex bodies from the brightness function they generate, i.e., from the function defining the volume of its projections onto hyperplanes. Theoretically, the reconstruction is based on two important results, Aleksandrov's projection theorem: "Any two origin-symmetric convex bodies in $\mathbb{R}^n$ that have the same brightness function must be equal." And Minkowski's existence theorem: "Given m real numbers a_j and unit direction vectors v_j that span $\mathbb{R}^n$

and satisfy the linear combination $\sum_{j=1}^{m} a_j v_j = o$, there is a convex polytope P in $\mathbb{R}^n$ with dimension n whose facets have volumes a_j and normals v_j." Here, o denotes the origin.

The goal of the algorithms that are developed in this paper is to construct an origin-symmetric convex polytope P_k from the given brightness functions of the convex body in k directions. This is done in two stages: computation of the surface areas of the polytope and reconstruction of the polytope P_k given its surface areas.

For the first stage, the input are k mutually non-parallel directions u_i and corresponding brightness function values $b\,(u_i)$. Assuming the number of facets known, its volumes a_j and normals v_j can be computed by solving a separable nonlinear least squares problem with constraints on both the linear and the nonlinear parameters. The authors suggest a possible application of the VARPRO technique for its solution.

A similar, though 2D and noisy problem, is considered in [9]. Given noisy support function measurements of an unknown planar convex origin-symmetric body, the article computes an origin-symmetric convex polygon that approximates the shape. The extended Gaussian image (EGI) of the polygon defined by the length of its edges and the outer normal angles can be obtained by solving a separable nonlinear least squares problems with nonnegative constraints on the linear edge length parameters and origin-symmetry conditions on the nonlinear angle parameters.

In article [8] the authors propose a solution to an identification problem where the data have been generated by scanning in time, for example, biomedical or astrophysical images. The readouts of the scanning sensors form a time-series, with each readout a linear combination of pixels. It is possible, that due for example to an offset in the timing system or a delay of the signal processing, there is a slight shift in the time intervals. A possible model for the image estimation problem where data and readouts are assumed stacked in vectors is $d = RPm + n$. Here, d is the data vector, m, the image vector, P the pointing matrix depending on the sensors and the observation protocol, R is the shift matrix and n the error. A least squares approximation gives raise to a nonlinear separable optimization problem with m the linear parameters and R containing the nonlinear parameters.

Applying the variable separation and taking into account the fact that the matrix R is unitary, the reduced nonlinear problem has the simple form:

$$\hat{R} = argmin_R \left\| \Pi^{\perp} R^H d \right\|^2 ,$$

where the matrix Π represents the orthogonal projector onto the signal subspace constituted by $span\,(P)$. In some applications the model can be simplified by considering that all the time series have the same shift τ. Then the problem is reduced to minimize a function of τ and is solved in the paper by a Newton method. The authors compared the cost of this approach to the one of applying the alternating LS method to the original nonlinear least squares problem and tested their algorithms on PACS, the Photodetector Array Camera and Spectrometer onboard the ESA Herschel satellite.

A shape and motion recovery problem in computer vision is considered in [4]. Given the time-series of images of a moving rigid body taken by a fixed camera, the problem is to determine the depth in 3D of the object and its linear and angular velocity. The author chooses to resolve this problem by minimizing the difference between the observed snd the predicted features. The nonlinear objective function is:

$$J(\omega, v, a) = \sum_i \|u_i - A_i(-[q_i]\,\omega + a_i v)\|^2, \text{ subject to } \|v\| = 1,$$

Here, ω, v are the angular and linear velocities, a_i the depth, q_i, u_i the optical flow measurements. This is an optimization problem in the manifold $\mathbb{R}^3 \times S^2 \times R$, where S^2 is the unit sphere in $\mathbb{R}^3$. Using relations between a_i,ω, v obtained from the geometric first-order necessary conditions for an extremum, the author proves that there are two possible reduced cost functions $J_2(\omega, v)$ over $\mathbb{R}^2 \times S$ and $J_3(v)$ over S^2 that have the same local extrema and minima as $J(\omega, v, a)$ and are at least as well conditioned as J. He compares the use of Newton and Steepest Descent methods for the three optimization problems, J, J_2, and J_3 and his experiments show that the convergence when using J_3 is the fastest.

The applications studied in [5] of the estimation of homogeneous vectors in Metric Vision and the one in [7] of color filter design with application to digital photography draw on the theorem in the Golub-Pereyra paper that the minima obtained by minimizing the full separable nonlinear problem can also be computed via the separation of the variables, thereby reducing the nonlinear least squares problem to a better conditioned one with a smaller number of parameters.

[1] Reliable camera pose and calibration from a small set of point and line correspondences: A probabilistic approach. T Chaperon, J Droulez, G Thibault, Computer Vision and Image Understanding **115**:576-585 (2011).

[2] Shape from moments-an estimation theory perspective. M Elad, P Milanfar, G H Golub, IEEE Transactions on Signal Processing **52**:1814-1829 (2004).

[3] Reconstruction of convex bodies from brightness functions. R J Gardner, P Milanfar, Discrete and Computational Geometry **29**:279-303 (2003).

[4] Geometric algorithms for least squares estimation of 3-D information from monocular image. Y Han, IEEE Transactions on Circuits and Systems for Video Technology **15**:269-282 (2005).

[5] Direct estimation of homogeneous vectors: An ill-solved problem in computer vision. M Harker, P O'Leary, Computer Vision, Graphics and Image Processing, Lectures Notes in Computer Sciences **4338**, Springer (2006).

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

[7] Designing color filter arrays for the joint capture of visible and near-infrared images. Y M Lu, C Fredembach, M Vetterli, S Susstrunk, 16th IEEE International Conference on Image Processing 3797-3800 (2009).

[8] Least squares time-series synchronization in image acquisition systems. L Piazzo, M C Raguso, L Calzoletti, IEEE Transactions on Image Processing **25**:4458-4468 (2016).

[9] Shape estimation from support and diameter functions. A Poonawala, P Milanfar, R J Gardner, Journal of Mathematical Imaging and Vision **24**:229-244 (2006).

[10] Improved high-definition video by encoding at an intermediate resolution. A Segall, M Elad, P Milanfar, R Webb, Proceedings Visual Communication and Image Processing SPIE **5308** (2004).

[11] Optimal framework for low bit-rate block coders. Y Tsaig, M Elad, G H Golub, Proceedings International Conference on Image Processing **3** (2003).

[12] Variable projection for near-optimal filtering in low bit-rate block coders. Y Tsaig, M Elad, P Milanfar, G Golub, IEEE Transactions on Circuits and Systems for Video Technology **15**:154-160 (2005).

4.3. Vision

In Computer Vision, structure-from-motion (SFM) aims to extract the 3D geometry of a scene and the camera motion from a given set of images recorded with a moving camera. The idea is to factorize the measurement matrix $\mathbf{M}$ into a product of two smaller matrices $\mathbf{U}\mathbf{V}^{\mathbf{T}}$, where $\mathbf{U}$ contains the information of the scene and $\mathbf{V}$ that of the camera. In principle the factorization of the measurement matrix can be done via the SVD decomposition. However, it is often the case that the measurement matrix has missing elements, for example if a scene feature is not visible on some image frame, and thus another solution algorithm must be found.

Some algorithms recover the information by minimizing in the least squares sense a nonlinear cost function where the parameters are the distances of the scene features from the camera centre and the linear and angular velocity components. In [7] the authors using the techniques for separable nonlinear least squares problems prove that it is possible to optimize first a reduced cost function where the linear velocity v is fixed and that, except for singularities, the local minima of the reduced and the full cost functional are the same. One of the difficulties when the images stem from forward motion, the case considered in this article, is that the translation gives rise to singularities in the cost functional. The resulting large number of "artificial" local minima, due to the discontinuities, which therefore appear in the least squares retroprojection error, makes optimization algorithms too slow for some applications like navigation. The solution suggested in the article is to impose a bound on the depth of the scene that makes the least squares error continuous. For this modified cost functional, the aforementioned equivalence results for minima of the full and reduced cost functionals are still valid. This allows the use of the optimization algorithms, with the addition of a small modification restricting the depth, for the solution of the SFM problem in real time.

In the next three articles [3, 4, 5] the authors consider the last stage of the SFM reconstruction approach, the bundle adjustment, i.e., the refinement of the position of the predicted set of 3D scene points and the camera parameters. In [3] they consider the case when an affine camera model is assumed, so that the objective function is linear in both blocks of variables, $\mathbf{U}$ and $\mathbf{V}$. They apply a meta-algorithm that allows to run different minimization algorithms starting from a set of random initial points until the same "best-so-far" optimum is computed twice. They show that the implementation that uses the VARPRO algorithm performs best: "New implementations of standard variable projection (VARPRO) algorithms which offer an order of magnitude performance improvement over the best previously available algorithms."

In both, [5, 4], the more complicated case of projective bundle adjustment is analyzed, where the camera model is now projective (pin hole model). Thus the minimization problem is nonlinear in both blocks of variables (this can be interpreted as a nonlinear factorization problem). The algorithm suggested by the authors is the so called Nonlinear VARPRO extension for this non-separable problem. Note that the difficulty here is that both sets of variables $\mathbf{U}$ and $\mathbf{V}$ are nonlinear, so it is not possible to use a closed form for the minimization in the linear set as is done in VARPRO. For this Nonlinear VARPRO extension one iteration of the algorithm consists of an inner minimization over one set of nonlinear variables say $\mathbf{V}$ given a fixed set of the variables $\mathbf{U}$. Then, a minimization over $\mathbf{U}$ is computed using the Jacobian of a reduced problem obtained through a local linearization in $\mathbf{V}$ (calculated via a Gauss-Newton step).

The authors observe that VARPRO applied to geometric vision problems has a much higher probability of reaching a global optimum (termed success rate in this paper) than using a second order method on the full problem (joint optimization, i.e., without eliminating one set of unknowns). VARPRO, which also turns out to be closely connected to joint optimization, has generally no trouble improving the current solution.

[1] Enforcing consistency constraints in uncalibrated multiple homography estimation using latent variables. W Chojnacki, Z L Szpak, M J Brooks, A van den Hengel, Machine Vision and Applications **26**:401-422 (2015).

[2] Learning shape from defocus. P Favaro, S Soatto, European Conference on Computer Vision 735-745 (2002).

[3] Secrets of matrix factorization: approximations, numerics, manifold optimization and random restarts. J Hyeong Hong, A Fitzgibbon, Proceedings of the IEEE International Conference on Computer Vision (ICCV) 4130-4138 (2015).

[4] pOSE: Pseudo object space error for initialization-free bundle adjustment. J H Hong and C Zach, The IEEE Conference on Computer Vision 1876-1885 (2018).

[5] Projective bundle adjustment from arbitrary initialization using the variable projection method. J H Hong, C Zach, A Fitzgibbon, R Cipolla, European Conference on Computer Vision 477-493 (2016).

[6] Guaranteed ellipse fitting with a confidence region and an uncertainty measure for centre, axes, and orientation. Z L Szpak, W Chojnacki, A van den Hengel, Journal of Mathematical Imaging and Vision **52**:173-199 (2015).

[7] Moving forward in structure from motion. A Vedaldi, G Guidi, S Soatto, IEEE Conference on Computer Vision and Pattern Recognition 1-7 (2007).

4.4. Robotics

There is a surprising number of different applications to robotics, with identification a predominant one, as in other fields. This is highlighted in [6] where a method based on the inverse dynamic identification model and least squares minimization (IDIM-LS) is listed alongside the alternative instrumental variable method (IV) developed in the paper.

In [13] the authors develop an identification procedure for the dynamics, friction and flexibility parameters of multi-body dynamic robot models. The particular manipulator model studied is a three-mass flexible robot arm, where the input is the torque generated by an electrical motor and the output is the motor velocity. The parameters are determined from the minimization in l_2 of the prediction error of the output. To improve the convergence to the global minimum, the procedure is split into three stages of which the first one is the identification of the rigid-body dynamics and the friction parameters. The resulting least squares problem has separable variables: the nonlinear parameters appear in the Striebeck friction effect and the linear parameters are the moments of inertia and a friction coefficient. The problem is solved using the VARPRO technique for experimental data from an industrial robot, with good correspondence to Frequency Response Function measurements.

The next article [4] focus the attention on the static friction in Robotics, an important issue in the performance of a robot manipulator. The friction model, based on empirical observations, is a function with separable parameters: the linear ones are Coulomb, static and viscous friction, while the nonlinear ones are the Striebeck parameters. Variable projection is used to solve the identification cost function in the least squares sense.

The aim of the following two articles [1, 2] is the replication of human-like regulation mechanisms in robots to enable them to operate in an unstructured environment and under unpredicted interaction situations. They consider briefly the performance during the point to point free space movement. Once the robot is in contact with the environment they propose an impedance controller which emulates the human common mode stiffness CMS, meaning that the correlation in stiffness changes in different joints. The authors design a controller that regulates the CMS across the joints and in addition further minimizes the error between desired and realized endpoint stiffness. This part of the algorithm requires the derivative of the pseudo-inverse of the manipulator Jacobian and in order to compute it they use the formula developed in Golub-Pereyra. This formula is also used in a case study in [7] and in [10].

In [5] a particular space-state system, a LPV, is used as a nonlinear model of an arm-driven inverted pendulum. The dynamics of the system vary as a function of time-dependent scheduling parameters; the mathematical model is a system of difference equations with matrix coefficients whose elements are expressed as spline functions of the scheduling parameters. Given measurements of input-output data, these parameters, and hence the behavior of the system, can be determined by the solution of a nonlinear least squares problem. Experimental results show that the spline-based model is a more accurate representation compared with a LPV using polynomial functions of the scheduling parameters, or moreover, using a linear time-invariant model.

The papers [8, 9] consider the simultaneous mapping of an environment and the tracking of an agent (possibly a robot) in it (SLAM or simultaneous localization and mapping). The measurement model used is linear in the robot and landmark positions and nonlinear in the robot orientation. Given a noisy function of measurements these parameters can be determined by the solution of a sparse nonlinear separable least squares problem. The authors develop an algorithm based on the variable projection idea, adapted to take advantage of the sparsity that makes it scalable. After stacking the measurement and state vectors, $\mathbf{z}$ and $\mathbf{x}(\mathbf{p}, \boldsymbol{\theta})$, where $\mathbf{p}$ are the linear and $\boldsymbol{\theta}$ the nonlinear variables, the problem is

$$\min_{\mathbf{x}} \|\mathbf{z} - H_1(\boldsymbol{\theta})\,\mathbf{p} + H_2\boldsymbol{\theta}\|^2 .$$

To retain sparsity they iterate until convergence:

1. One step of Gauss-Newton for the complete nonlinear problem that has a sparse matrix structure, to compute only a correction for the nonlinear parameters $\boldsymbol{\theta}$.
2. One step of the linear sparse system expressing the analytic solution of the linear parameters $\mathbf{p}$ in terms of the updated value of $\boldsymbol{\theta}$.

They state that their algorithm is equivalent, in infinite arithmetic, to the Kaufman modification of VARPRO, and refer to the proof of an equivalent algorithm (by Barham and Drane, 1972) given by T. A. Parks (1985). The code is available online and used successfully on both real and synthetic datasets.

The problem considered in [11] models the contact dynamics of robot interactions with the environment, using the contact dynamics equations based on the surface compliance approach, i.e., objects are rigid and the contact forces are functions of local interference distance between the objects. In addition to systematic measurement errors of the forces involved in robot interaction, there are also geometric uncertainties caused for example by geometric simplifications. The measured force/moment vector acting onto an object depends on the normal/friction forces, contact positions, normal and tangential contact vectors and the interference distance.

A nonlinear separable least squares problem is set up using measurements of the force/moment vector to determine the unbiased estimates for the normal and friction constants (linear parameters) and the uncertainty variables (nonlinear parameters). This problem was solved using VARPRO and applied to identify the contact parameters from experimental data measured with the Special Purpose Dexterous Manipulator Task Verification Facility at the Canadian Space Agency.

[1] Tele-impedance: exploring the role of common-mode and configuration-dependent stiffness. A Ajoudani, M Gabiccini, N Tsagarakis, A Albu-Scaffer, A Bicchi, 12th IEEE-RAS International Conference on Humanoid Robots 363-369 (2012).

[2] Human-like impedance and minimum effort control for natural and efficient manipulation. A Ajoudani, M Gabiccini, N. Tsagarakis, A. Bicchi, IEEE International Conference on Robotics and Automation 4499-4505 (2013).

[3] Adaptive cancelation of self-generated sensory signals in a whisking robot. S R Anderson, M J Pearson, A Pipe, T Prescott, P Dean, J Porrill, IEEE Transactions on Robotics 2**6**:1065-1076 (2010).

[4] Static friction in a robot joint—modeling and identification of load and temperature effects. A C Bittencourt, S Gunnarsson, Journal of Dynamic Systems, Measurements and Control **134**:51013-51023 (2012).

[5] Closed-loop identification of LPV models using cubic splines with application to an arm-driven inverted pendulum. S Boonto, H Werner, Proceedings of the 2010 American Control Conference 3100-3105 (2010).

[6] A generic instrumental variable approach for industrial robot identification. A Janot, P O Vandanjon, M. Gautier, IEEE Transactions on Control System Technology **22**:132-145 (2014).

[7] Dynamics and control of an anthropomorphic compliant arm equipped with friction clutches. N Kashiri, M Laffranchi, D G Caldwell, N G Tsagarakis, IEEE/ASME Transactions on Mechatronics **21**:694-707 (2016).

[8] Exploiting the separable structure of SLAM. K Khosoussi, S Huang, G Dissanyake, Proceedings, Robotics: Science and Systems Conference (2015).

[9] A sparse separable SLAM back-end. K Khosoussi, S Huang, G Dissanayake, IEEE Transactions on Robotics **32**:1536-1549 (2016).

[10] Predictive modeling by infinite-horizon constrained inverse optimal control with application to a human manipulation. Marcel Menner, Peter Worsnop, Melanie N. Zeilinger, arXiv.org > cs > arXiv:1812.11600 (2018).

[11] Identification of contact parameters from stiff multi-point contact robotic operations. D Verscheure, I Sharf, H Bruyninckx, J Swevers, J De Schutter, The International Journal of Robotics Research (2009).

[12] Multivariable Frequency-Domain Identification of Industrial Robots. E Wernholt, Ph D Thesis, Linkopin University (2007).

[13] Nonlinear identification of a physically parameterized robot model. E Wernholt, S Gunnarsson, SYSID, Newcastle, Australia (2006).

5. Geophysics, Petroleum Engineering

This is an area of particular interest to one of the authors, who has worked many years in the area of exploration geophysics, mainly on geological modeling, seismic ray tracing and inverse problems [28, 26, 27]. In the past few years VARPRO has been discovered as one of the keys for solving the fundamental earth imaging problem using full waveform inversion [32]. This is a notoriously expensive procedure (requires many solutions of the wave equation in 3D) and it leads also to multimodal optimization problems, where local optimization algorithms have difficulties in avoiding sub-optimal minima. It turns out that several different applications within this area can be stated as separable problems, making them amenable to the use of VARPRO, which as we have seen, tends to deliver much better behaved optimization problems in the reduced space of the nonlinear variables. It turns out that Bill Symes had foreseen this early on with his related method of differential semblance [34].

For instance, in [1] the authors tackle the well-known global convergence issue associated to any full waveform inversion (FWI) approach by solving an extended-image space least-squares migration problem to remove any local minima present in the FWI objective function. They discuss the connection between the reflectivity and migration velocity inversion and show the importance of combining the two problems using one objective function. Moreover, they show the full separability of the two inverse problems by using the variable projection method. Furthermore, in [3], the same authors indicate that the main issue inherent to full waveform inversion (FWI) is its inability to correctly recover the Earth's subsurface seismic parameters from inaccurate starting models. This behavior is due to the presence of local minima in the FWI objective function. To overcome this problem, they propose a new objective function in which they modify the nonlinear modeling operator of the FWI problem by adding a correcting term that ensures phase matching between predicted and observed data. This additional term is computed by demigrating an extended model variable, and its contribution is gradually removed during the optimization process while ensuring convergence to the true solution. Since the proposed objective function is quadratic with respect to the extended model variable, they make use of the variable projection method. They refer to this technique as full waveform inversion by model extension (FWIME) and illustrate its potential on two synthetic examples for which FWI fails to retrieve the correct solution.

In [8] the authors consider planar waves events recorded in a seismic array that can be represented as lines in the Fourier domain. However, in the real world, seismic events usually have curvature or amplitude variability, which means that their Fourier transforms are no longer strictly linear but rather occupy conic regions of the Fourier domain that are narrow at low frequencies but broaden at high frequencies, where the effect of curvature becomes more pronounced. One can consider these regions as localized "signal cones". In this work, the authors consider a space–time variable signal cone to model the seismic data. The variability of the signal cone is obtained through scaling, slanting, and translation of the kernel for cone-limited (C-limited) functions (functions whose Fourier transform lives within a cone) or C-Gaussian function (a multivariate function whose Fourier transform decays exponentially with respect to slowness and frequency), which constitutes their dictionary. The authors find a discrete number of scaling, slanting, and translation parameters from a continuum by optimally matching the data. This is a non-linear optimization problem, which is solved by a fixed-point method that utilizes a variable projection method with l_1 constraints on the linear parameters and bound constraints on the non-linear parameters.

[1] Tomographic full waveform inversion via iterative least-squares migration by variable projection. G B Barnier, E B Biondi and B B Biondi, First EAGE/SBGf Workshop on Least-Squares Migration (2018).

[2] A modified approach for tomographic full waveform inversion using variable projection. G Barnier, E Biondi, B Biondi, 80th EAGE Conference and Exhibition Extended Abstract (2018).

[3] Full waveform inversion by model extension. G Barnier, E Biondi and B Biondi, SEG Technical Program Expanded Abstracts (2018).

[4] Joint Model and Minimal-Source Full Waveform Inversion for Seismic Imaging Under General Anisotropic Sources (Thesis Proposal). M J Bencomo, The Rice Inversion Project, Trip15 (2016).

[5] Representation and Estimation of Seismic Sources via Multipoles. M J Bencomo, Ph D Thesis, Rice University (2017).

[6] Fractal heterogeneities in sonic logs and low-frequency scattering attenuation. T J Browaeys, S Fomel, Geophysics **74** (2009).

[7] Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation. K Chen, M D Sacchi, Geophysical Prospecting **65** (2017).

[8] A model-based data-driven dictionary learning for seismic data representation. Can Evren Yarman, Rajiv Kumar and James Rickett, Geophysical Prospecting 66:98-123 (2018).

[9] Multicomponent seismic data registration for subsurface characterization in the shallow Gulf of Mexico. F Sergey Fomel, Milo M Backus, Michael V DeAngelo, Paul E Murray, Bob A Hardage, Offshore Technology Conference, Houston, Texas (2003).

[10] Reducing the cost of extended waveform inversion by multi-scale adaptive methods. L Fu, W W Symes, SEG Technical Program Expanded Abstracts 1127-1131 (2015).

[11] A discrepancy based penalty method for extended waveform inversion. L Fu and W W Symes, The Rice Inversion Project, TRIP16 (2017).

[12] An adaptive multi-scale algorithm for efficient extended waveform inversion. L Fu, W W Symes, Geophysics **82**:183-197 (2017).

[13] A robust approach for extended waveform inversion without low-frequency data. L Fu, William W Symes, SEG 2017 Workshop: Full-Waveform Inversion and Beyond, Beijing, China 38-42 (2017).

[14] Fast extended waveform inversion using Morozov's discrepancy principle. L Fu, W Symes, The Rice Inversion Project, TRIP16 (2017).

[15] A discrepancy-based penalty method for extended waveform inversion. L Fu, W W Symes, Geophysics **82** (2017).

[16] Frequency selected envelope inversion: a new approach for doing multiscale full waveform inversion in the time domain. Z Gao, Z Pan, J Gao, SEG 2017 Workshop: Full-Waveform Inversion and Beyond 13-17 (2017).

[17] Wavelet estimation and seismic deconvolution using separable nonlinear least squares. M Gharibi, N Shadmanaman, K Kazemi, P P Zagros, M Mirzakhanian, SEG Technical Program Expanded Abstracts 3487-3491 (2015).

[18] Dimensionality reduction and uncertainty quantification in seismic waveform inversion. Adriaan Graas, Master Thesis, Utrech University (2018).

[19] Born waveform inversion via variable projection and shot record model extension. Y Huang, W W Symes, SEG Technical Program Expanded Abstracts (2015).

[20] Matched source waveform inversion: space-time extension. G Huang, W Symes, R Nammour, SEG Technical Program Expanded Abstracts 1426-1431 (2016).

[21] Volume source-based extended waveform inversion. Guanghui Huang, Rami Nammour, William W Symes, Geophysics **83**:369-387 (2018).

[22] A comparison of seismic velocity inversion methods for layered acoustics. T Van Leeuwen, W A Mulder, Inverse Problems **26** (2009).

[23] Application of the variable projection scheme for frequency-domain full-waveform inversion. M Li, J Rickett, A Abubakar, Geophysics **78**:249-257 (2013).

[24] Resolution enhancement of seismic data using spectral modeling based on dominant Ricker components and separable nonlinear least squares. M G Mamasani, N S Manaman, K Kazemi, M Miezakhanian, The Leading Edge **36**:480-486 (2017).

[25] Source-independent Hessian-free Gauss-Newton full-waveform inversion. Wenyong Pan, Kris Innanen, CREWES Research Report **27** (2015).

[26] 3D block travel time inversion. V Pereyra, SEG 66 Annual Meeting, Houston. Extended Abstracts (1996)

[27] Fast wave propagation by model order reduction. V Pereyra and B Kaelin. ETNA **30**:406-419 (2008).

[28] Computational methods for inverse problems in geophysics: inversion of travel time observations, V Pereyra, H B Keller and W H K Lee, Physics of the Earth and Planetary Interiors **21**:120-125 (1980).

[29] Geophysical Data Registration Using Modified Plane-Wave Destruction Filters. M D Phillips, Ph D Thesis, The University of Texas at Austin (2017).

[30] Seismic time-lapse image registration using amplitude-adjusted plane-wave destruction. M Phillips, S Fomel, SEG Technical Program Expanded Abstracts 5476-5478 (2016).

[31] Estimation of time-shifts between time-lapse seismic images using spectral decomposition. M Phillips, S Fomel, SEG Technical Program Expanded Abstracts (2017).

[32] The variable projection method for waveform inversion with an unknown source function. J Rickett, Geophysical Prospecting **61**:874-881 (2013).

[33] Full-waveform inversion of passive seismic data for sources and velocities. J Sun, Z Xue, T Zhu, S Fomel, N Nakata, SEG Technical Program Expanded Abstracts 1405-1410 (2016).

[34] Inversion of reflection seismograms by differential semblance analysis: algorithm structure and synthetic examples. W W Symes, Michel Kern, Geophysical Prospecting **42**:565-614 (1994).

[35] Source estimation with surface-related multiples—Fast ambiguity-resolved seismic imaging. N Tu, A Aravkin, T van Leeuwen, T Lin, F J Hermann, Geophysical Journal International **205**:1492-1511 (2016).

[36] Simultaneous full-waveform inversion for source wavelet and earth model. K Wang, J R Krebs, D Hinkley, A Baumstein SEG Technical Program Abstracts 2537-2541 (2009).

[37] Quantification of long-range persistence in geophysical time series: conventional and benchmark-based improvement techniques. A Witt, B D Malamud, Surveys in Geophysics **34**:541-651 (2013).

[38] Source estimation for wave equations with uncertain parameters. Sergiy Zhuk, Stephen Moore, Alberto Costa Nogueira Junior, Andrew Rawlinson, Tigran Tchrakian, Lior Horesh, Aleksandr Aravkin and Albert Akhrie. IBM Research Report (2016).

6. Numerical Analysis

These are application of VARPRO to general numerical analysis topics. Most papers are connected to polynomials, as in surrogates, interpolation, cubatures, conformal mappings and common divisors. For instance in [2] the authors observe that the problem of obtaining surrogates for expensive functions by using polynomials can be stated as a ridge approximation that turns out to be a separable problem and therefore is amenable to variable projections. In [3, 5, 6] the formula for the derivative of the pseudo-inverse is used.

Hale [1, 6] considers the use of conformal mapping for polynomial approximation. In some cases this leads to a separable problem and VARPRO is used. In [7] the author applies VARPRO to the problem of computing cubature rules. Finally, in [8] the authors consider the problem of finding, for a given N-tuple of polynomials (real or complex), the closest N-tuple that has a common divisor of degree at least d. They develop optimization methods based on the variable projection principle both for image and kernel representation. In the sub-sections below we include more specific sub-fields.

[1] On the Use of Conformal Maps to Speed up Numerical Computations. N Hale, Ph D Thesis, Univ. Oxford, England (2009).

[2] Data-driven polynomial ridge approximation using variable projection. J M Hokanson, P G Constantine, SIAM Journal on Scientific Computing **40**:A1566-A1589 (2018).

[3] Differentiating the pseudo determinant. A Holbrook, Linear Algebra and its Applications **548**:293-304 (2018).

[4] Polynomial homotopy method for the sparse interpolation problem, Part I: Equally spaced sampling. L Jiao, B Dong, J Zhang, B Yu, SIAM Journal on Numerical Analysis **54**:462-480 (2016).

[5] A formula for the Fréchet derivative of a generalized matrix function. V Noferini, SIAM Journal on Matrix Analysis and Applications **38**:434-457 (2017).

[6] Discrete polynomial moments for real-time geometric surface inspection. P L O'Leary, M J Harker, Journal of Electronic Imaging **18**:013015 (2009).

[7] Efficient computation of cubature rules with application to new asymmetric rules on the triangle. S A Papanicolopulos, Journal of Computational and Applied Mathematics **304**:73-83 (2016).

[8] Variable projection methods for approximate (greatest) common divisor computations. K Usevich, I Markovsky, Theoretical Computer Science **681**:176-198 (2017).

6.1. Exponential Fitting

This is a very classical problem that is notoriously difficult to solve. Even today, one of the most used methods is that of Prony, dating to the eighteen century, which in his original form is unstable and does not work in the presence of noise. The exponential fitting problem appears in diverse applications such as magnetic resonance spectroscopy, mechanical resonance, chemical reactions, system identification and radioactive decay. In [8] we have collected the work of several authors that discuss these and other applications and they all agree that VARPRO and some modified and more stable versions of Prony's method are the algorithms of choice.

In his Ph D Thesis, Hokanson [2] discusses and analyses stable algorithms, such as variants of the classical method of Prony and variable projections. In [1] Halliday considers the problem of determining the underlying parameters of general signal models through the application of maximum likelihood estimation theory for models whose variables separate. He considers exponentials and Bessel basis functions.

In [7], the authors study the problems of time-resolved spectroscopy, microscopy and mass spectrometry, which are modeled by a sum of complex exponentials. Finally, in [10] the authors compare several methods for exponential data fitting, including modification of Prony's method, VARPRO and their own methods. Unfortunately this extensive comparison only considers models with up to 3 exponentials, but in any case provides valuable information about the relative performance of the methods, both in time, accuracy and stability.

[1] Maximum likelihood estimation of structural wave components from noisy data. P J Halliday, K Grosh, The Journal of the Acoustical Society of America **111**:1709 (2002).

[2] Numerically Stable and Statistically Efficient Algorithms for Large Scale Exponential Fitting. J M Hokanson, Ph D Thesis, Rice University (2013).

[3] Fast minimum uncertainty estimates for the exponential fitting problem. J M Hokanson, arXiv preprint arXiv:1508.05890 (2015).

[4] Projected nonlinear least squares for exponential fitting. J M Hokanson, SIAM Journal on Scientific Computing **39**:A3107-A3128 (2017).

[5] A necessary and sufficient criteria for the existence of the least squares estimate for a 3-parametric exponential function. D Jukić, Applied Mathematics and Computation **147**:1-17 (2004).

[6] A genetic algorithms based technique for computing the nonlinear least squares estimates of the parameters of sum of exponentials model. S Mitra, A Mitra, Expert Systems with Applications **39**:6370-6379 (2012).

[7] Sum-of-exponentials models for time-resolved spectroscopy data. K M Mullen, I H M van Stokkum, Exponential Data Fitting and its Applications, Bentham Science Publishers 110-144 (2010).

[8] Exponential Data Fitting and its Applications. V Pereyra, G Scherer (Editors), Bentham Science Publishers (2010).

[9] Estimation of parameters of impulse responses of mechanical systems by modified Prony method. V Slivinskas, V Šimonytė - Solid State Phenomena **113**:190-194 (2006).

[10] Optimal approximation with exponential sums by a maximum likelihood modification of Prony's method. Ran Zhang and Gerlind Plonka, Manuscript na.math.uni-goettingen.de (2018).

6.2. Optimization

We present here an eclectic mixture of applications in optimization. The common thread is VARPRO. An interesting contribution is found in [3], where a modern implementation of VARPRO in MATLAB is offered. They include comments on constrained problems and also go against previous consensus by advising not to neglect the Kaufman term in the Jacobian since that may make the algorithm less robust, specially away from the solution. More extensive statistical calculations are also included. It would be interesting to have an implementation of VARPRO that senses the need to include the Kaufman term so that a more robust algorithm arises without penalizing those applications for which the extra term is not necessary.

The authors of [1, 4] use the result about the smoothness of the pseudo-inverse in regions of constant rank.

[1] Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs. N Boumal, V Voroninski, A S Bandeira, arXiv preprint arXiv:1804.02008 (2018).

[2] A stable primal–dual approach for linear programming under nondegeneracy assumptions. M Gonzalez-Lima, H Wei, H Wolkowicz, Computational Optimization and Applications **44**:213 (2009).

[3] Variable projection for nonlinear least squares problems. D P O'Leary, B W Rust, Computational Optimization and Applications **54**:579-593 (2013).

[4] Smoothed analysis of the low-rank approach for smooth semidefinite programs. Thomas Pumir, Samy Jelassi and Nicolas Boumal, arXiv:1806.03763 (2018).

[5] Extended Gauss-Newton and Gauss-Newton-ADMM algorithms for low-rank matrix optimization. Q Tran-Dinh, Z Zhang - arXiv preprint arXiv:1606.03358 (2016).

[6] Solving semidefinite programs using preconditioned conjugate gradients. H Wolkowicz, Optimization Methods and Software **19**:653-672 (2004).

6.3. Linear Algebra

This subsection collects mostly matrix applications, including pseudo-inverses, projectors, Hankel and lower rank approximations. The books [1, 3, 4, 5] describe different aspects of separable problems and variable projections.

[1] Generalized Inverses: Theory and Applications. A Ben-Israel, TNE Greville (2003).

[2] On one-sided (B, C)-inverses of arbitrary matrices. J Benitez, E Boasso, H Jin, arXiv preprint arXiv:1701.09054 (2017).

[3] Structured Matrix Nearness Problems: Theory and Algorithms. R Borsdorf, Ph D Thesis, Univ. Manchester, England (2012).

[4] Generalized Inverses of Linear Transformations. S L Campbell, C D Meyer, SIAM Publications (2009).

[5] Projectors and Projection Methods. A Galántai, Springer (2013).

[6] Computing the nearest rank-deficient matrix polynomial. M Giesbrecht, J Haraldson, G Labahn, Proceedings of the 2017 ACM on International Symposium on Symbolic and Algebraic Computation 181-188 (2017).

[7] Computing lower rank approximations of matrix polynomials. M Giesbrecht, J Haraldson, G Labahn, arXiv preprint arXiv:1712.04007 (2017).

[8] Gene Howard Golub, 1932–2007. D P O'Leary, Linear Algebra and Its Applications **428**:2405-2409 (2008).

6.4. Nonlinear Equations

Here we collect a few applications to nonlinear problems, including Newton's method, robust estimators and bifurcation. Deuflhard in his book [1] discusses in detail separable problems, VARPRO and its variants. In [2] the authors use the formulas of differentiation of pseudo-inverses and projectors in the context of providing robust and accurate estimates for linear regression problems when both the measurement vector and the coefficient matrix are structured and subject to errors or uncertainty. In [4] the authors consider parametrized separable problems and study their bifurcation diagrams. In particular they handle the rank-one deficiency case at a bifurcation point, without loosing the separability properties of the general problem. The survey paper [5] discusses VARPRO applied to separable problems.

[1] Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms. P Deuflhard, Springer (2011).

[2] Structured least squares problems and robust estimators. M Pilanci, O Arikan, M C Pinar, IEEE Transactions on Signal Processing **58**:2453-2465 (2010).

[3] Solving separable nonlinear equations with Jacobians of rank deficiency one. Y Q Shen, T J Ypma, International Conference on Computational and Information Science CIS 99-104 (2004).

[4] Rank deficiencies and bifurcation into affine subspaces for separable parameterized equations. Y Q Shen, T Ypma, Mathematics of Computation **85**:271-293 (2016).

[5] Recent advances in numerical methods for nonlinear equations and nonlinear least squares. Y X Yuan, Numerical Algebra, Control & Optimization **1**:15-34 (2011).

6.5. Least Squares

Penalized, linear, total and alternating least squares are the problems considered in this section. Bates and DebRoy [1] deal with the problem of linear mixed models via a penalized least squares method. They make good use of the formulas for the derivatives of projectors and pseudo-inverses to obtain analytical expressions for the profiled log-likelihood. In [4] the authors use the derivative of projectors that appear in a block Gauss-Seidel method that they study.

[1] Linear mixed models and penalized least squares. D M Bates, S DebRoy, Journal of Multivariate Analysis **91**:1-17 (2004).

[2] Linear Least Squares Problems. Å Björck, Numerical Methods in Matrix Computations, Springer (2015).

[3] Structured total least squares. P Lemmerling, S Van Huffel, Total Least Squares and Errors-in-Variables 79-91, Springer (2002).

[4] Alternating least squares as moving subspace correction. I Oseledets, M Rakhuba, A Uschmajew, arXiv preprint arXiv:1709.07286. SIAM Journal on NA **56**:3459-3479 (2018).

6.6. Splines

One of the original applications of Variable Projections was to the fitting of splines with variable knots. Here we have some additional extensions. Dertimanis et al [2] and Molinari et al [3] introduce a data-driven uncertainty quantification scheme that relies on B-spline functions. Instead of predefining basis functions according to the statistics of the uncertain input, as in conventional polynomial chaos expansion (PCE)-based implementations, the method introduced herein takes advantage of the increased flexibility of B-splines, which are adaptable to a given input–output data set. Parameter estimation is effectively dealt through a Separable Nonlinear Least Squares (SNLS) procedure that allows for simultaneous estimation of both the B-splines' free knots and the corresponding coefficients of projection.

In [4] the authors consider the bi-variate problem that leads to a separable problem with tensor product structure. Spiriti et al [5] consider penalized splines and the usual VARPRO reduction, but since the problem may have multiple minima they consider search methods and parallelization to solve the projected nonlinear system. The authors of [6] remark that a lack of accurate and fast reconstruction models hinders the development of intelligent sampling techniques for surface reconstruction. In their paper, a smart surrogate model based on free-knot B-splines and variable projections is used for intelligent surface sampling design with the aid of uncertainty modeling.

[1] Total least squares fitting of Bézier and B-spline curves to ordered data. C F Borges, T Pastva, Computer Aided Geometric Design **19**:275-289 (2002).

[2] Data-driven uncertainty quantification of structural systems via B-spline expansion. V K Dertimanis, M D Spiridonakos and E N Chatzi, Computers & Structures **207**:245-257 (2018).

[3] Bounded optimal knots for regression splines. N Molinari, J F Durand, R Sabatier, Computational Statistics & Data Analysis **45**:159-178 (2004).

[4] Bivariate free knot splines. T Schütze, H Schwetlick, BIT Numerical Mathematics **43**:153-178 (2003).

[5] Knot selection for least-squares and penalized splines. S Spiriti, R Eubank, P W Smith, D Young, Journal of Statistical Computation and Simulation **83**:1020-1036 (2012).

[6] Uncertainty-guided intelligent sampling strategy for high-efficiency surface measurement via free-knot B-spline regression modeling. Jian Wang, Luca Pagani, Liping Zhou, Xiaojun Liu, Wenlong Lu, Richard Leach and Xiangqian (Jane) Jiang, Precision Engineering (2018).

7. Parameter Estimation, Approximation, Statistics

This is a large category that we have subdivided in appropriate subsections. It is interesting that here, as in other classes, there are included a number of Ph D Thesis, showing active new research and applications. Acosta and Vallejos [1] extended the known methodology for effective sample size computations for general spatial regression models. The approach, equipped with powerful computational machinery is appropriate for large spatial datasets and it provides formulas for a number of spatial processes. The methodology can easily be extended to more general models, such as a separable nonlinear model of the form

$$Y = X(\alpha)\beta + \epsilon.$$

In [5], Hokanson and Magruder develop a nonlinear least squares approach for constructing rational approximations with respect to the l_2 norm. They explore this approach using two parameterizations of rational functions: a ratio of two polynomials and a partial fraction expansion. In both cases, they show how one can use Variable Projection (VARPRO) to reduce the dimension of the optimization problem. Some of these references use the formula for the differentiation of the pseudo-inverse in various contexts [10, 7]. In [8] the authors present an interesting discussion of the large sample properties of separable problems in the complex valued case.

[1] Effective sample size for spatial regression models. Jonathan Acosta and Ronny Vallejos, Electronic Journal of Statistics **12**:3147–3180 (2018).

[2] Spectral Analysis of Non-uniformly Sampled Data and Applications. P Babu, Ph D Thesis, Uppsala Univeristet, Sweden (2012).

[3] Improved multivariate normal mean estimation with unknown covariance when p is greater than n. D Chételat, M T Wells, The Annals of Statistics **6**:3137-3160 (2012).

[4] Modified F tests for assessing the multiple correlation between one spatial process and several others, P Dutilleu B Pelletier, G Alpargu, Journal of Statistical Planning and Inference **138**:1402-1415 (2008).

[5] Least squares rational approximation. Jeffrey M Hokanson and Caleb C Magruder, arXiv:1811.12590 [math.NA] (2018).

[6] Geometric Bayes. A Holbrook, UC Irvine Electronic Theses and Dissertations (2018).

[7] Communication-efficient algorithms for distributed stochastic principal component analysis. D Garber, O Shamir, N Srebro, arXiv preprint arXiv:1702.08169 (2017).

[8] Large sample properties of separable nonlinear least squares estimators. K Mahata and T Soderstrom, IEEE Transactions on Signal Processing **52**:1650-1658 (2004).

[9] Maximum Likelihood Estimation and Inference: with Examples in R, SAS and ADMB. R B Millar, Wiley (2011).

[10] Chaos, Observability and Symplectic Structure in Optimal Estimation. D Rey, Ph D Thesis, UC San Diego, CA (2017).

[11] A novel joint diagonalization approach for linear stochastic systems and reliability analysis. F Wang, C Li, J Feng, S Cen, D R J Owen, Engineering Computations **29** (2012).

[12] Structured Higher-Order Algorithmic Differentiation in the Forward and Reverse Mode with Application in Optimum Experimental Design. S Walter, Ph D Thesis, Humboldt University, Berlin (2012).

7.1. Modeling, Identification

A large number of applications involving modeling and system identification by linear combinations of parametrized basis functions is included here. The practical application of FS–TARMA identification requires expertise on part of the user, in particular because model structure (including functional subspace) estimation is a rather complicated problem [16]. The aim of [17] is to develop a novel complete approach that largely circumvents this drawback. The approach uses – for the first time – regression type methods for the simultaneous estimation of the necessary functional subspaces and model coefficients of projection. This is accomplished through proper parametrization and a Variable Projection scheme. The effectiveness of the proposed approach is assessed via its application to the identification of the time–varying dynamics of a laboratory pick–and–place mechanism from a vibration response data record.

The Generalized Instrumental Variable Estimator (GIVE) has been introduced in Soderstrom [13] as a class of estimators based on the bias–eliminating principle and it contains many previously known methods as special cases.

In its most general form, one uses a θ-dependent weighting matrix $W(\theta)$, and the problem is then indeed a separable nonlinear least squares problem. See also [15]. In [11] the authors consider radial function basis for data driven uncertainty quantification.

In [19] the authors consider a nonlinear optimization-based identification procedure for fully parameterized multivariable state-space models. The method can be used to identify linear time-invariant, linear parameter-varying, composite local linear, bilinear, Hammerstein and Wiener systems. On the other hand, Xu et al [22] consider a time-partitioned piecewise affine output error (PWA-OE) model for batch processes, in which the time index is used to simplify the partition of regression domain by utilizing the repetitive nature of batch processes. The identification problem involves both continuous and discrete variables, and the derivative information on discrete variables is unavailable. Thus, an identification algorithm based on separable nonlinear least-squares is developed to reduce the complexity of nonlinear minimization.

[1] Fitting linear mixed-effects models using lme4. D Bates, M Mächler, B Bolker, S Walker, arXiv preprint arXiv:1406.5823 (2014).

[2] Approximation of a multidimensional dependency based on linear expansion in a dictionary of parametric functions. M G Belyaev, E V Burnaev, Informatics and its Applications **7**:114-125 (2013).

[3] Adaptive post-processing internal models design for MIMO minimum-phase nonlinear systems. M Bin, L Marconi, arXiv preprint arXiv:1805.05629 (2018).

[4] An iterative Kalman smoother/least-squares algorithm for the identification of delta-ARX models. M A Chadwick, S R Anderson, V Kadirkamanathan, International Journal of Systems Science **41**:839-851 (2010).

[5] Identificação de Sistemas Utilizando a Parametrização MOLI. Patrícia Gomes Saraiva, Master Thesis, U. Porto, Portugal (2018).

[6] Identification of spatially distributed discrete-time state-space models. A Haber, M Verhaegen, IFAC Proceedings **45**:410-415 (2012).

[7] Statistical Methods for Constructing Mathematical Models. DV Ivanov, EP Melisheva, Science Book Publishing House (2014).

[8] Optimal Design of Experiments with Mixtures. R H H Khashab, Ph D Thesis, U. of Southhampton (2018).

[9] Small Dispersion Asymptotics in Stratified Models. X Mei, Ph D Thesis, Northwestern University, Evanston, Ill (2017).

[10] Hierarchical-likelihood approach for nonlinear mixed-effects models. M Noh, Y Lee, Computational Statistics & Data Analysis **52**:3517-3527 (2008).

[11] Structured parameter optimization method for the radial basis function-based state-dependent autoregressive model. H Peng, T Ozaki, V Haggan-Ozaki, Y Toyoda, International Journal of Systems Science **33**:1087-1098 (2002).

[12] On new parametrization methods for the estimation of linear state–space models. T Ribarits, M Deistler, B Hanzon, International Journal of Adaptive Control and Signal Processing **18** (2004).

[13] A generalized instrumental variable method for errors-in-variables. T Söderstrom, Automatica, **47**:1656–1666, (2011).

[14] A unified framework for EIV identification methods when the measurement noises are mutually correlated. T Söderström, R Diversi, U Soverini, Automatica **50**:3216-3223 (2014).

[15] A unified framework for EIV identification methods in the presence of mutually correlated noises. T Söderström, R Diversi, U Soverini, 19th World Congress IFAC Proceedings (2014).

[16] FS–TARMA models for non–stationary vibration analysis: An overview and comparison. M D Spiridonako, M.D. and S D Fassois, Proc. 15th IFAC Symposium on System Identification (SYSID), Saint–Malo, France (2009).

[17] Output–only identification of time–varying structures via a complete FS–TARMA model approach. M D Spiridonakos, S D Fassois, Proceedings of the 4th International Operational Modal Analysis Conference (IOMAC 2011) Istanbul, Turkey (2011).

[18] Modeling and Identification of Linear Parameter-Varying Systems. R Tóth, Lecture Notes in Control and Information Sciences **403** (2010).

[19] Identification of fully parameterized linear and nonlinear state-space systems by projected gradient search. V Verdult, N Bergboer, M Verhaegen, IFAC Proceedings (2003).

[20] Filtering and System Identification: a Least Squares Approach. M Verhaegen, V Verdult, Cambridge University Press (2007).

[21] Maximum likelihood estimation of linear SISO models subject to missing output data and missing input data. R Wallin, A Hansson, International Journal of Control **87**:2354-2364 (2014).

[22] Time-partitioned piecewise affine output error model for batch processes. Zuhua Xu, Yaobo Huang, Jun Zhao, Chunyue Song and Zhijiang Shao, Ind. Eng. Chem. Res. **57**:1560–1568 (2018).

[23] Identification of systems with slowly sampled outputs using LPV model. W Yan, Y Zhu, L Zhu, X Liu, Computers & Chemical Engineering **112**:316-330 (2018).

[24] Estimation of nonlinear ARX models. Y Zhu, Decision and Control, 2002, Proceedings of the 41st IEEE Conference on Decision and Control **4** (2002).

7.2. Hammerstein Models

This subsection refers to the use of nonlinear models that originate from Volterra series methods. Since those are problematic to identify, block structured systems are introduced as simplified alternatives, where the structures can be exploited to improve the identification. Among these we find the Hammerstein and Wiener models and their combinations. It turns out that since these models combine linear and nonlinear parts, they are amenable to a Variable Projection treatment.

In [4] the authors discuss a Kautz-basis-expansion Hammerstein system identification method where separable least squares are adopted to estimate linear and nonlinear parameters. Elden and Ahmadi-Asl [5] consider bilinear tensor least squares problems that occur in applications such as Hammerstein system identification and social network analysis. A linearly constrained problem of medium size is considered and nonlinear least squares methods of Gauss–Newton-type are applied to numerically solve it. The problem is separable and the variable projection method can be used. Perturbation theory is presented and used to motivate the choice of constraint. Numerical experiments with Hammerstein models and random tensors are performed, comparing the different methods and showing that a variable projection method performs best.

In [12] the authors focus on the digital predistortion for the linearization of power amplifiers (PAs) with intrinsic hard nonlinearities, such as Doherty PAs, envelope tracking (ET) PAs, which are widely used in third generation (3G) communication systems. Conventional Volterra-series-based predistortion with high degree nonlinearities suffers from the difficulty of real-time implementation and numerical instability. Accordingly, a block-oriented Hammerstein predistortion with a cubic-spline static nonlinearity is proposed, and the separable nonlinear least squares (SNLS) method is used to identify the Hammerstein coefficients, which significantly reduces the dimension of the search space. A similar application can be found in [12], where a Hammerstein predistorter model for power amplifier (PA) linearization is proposed. The key feature of the model is that cubic splines, instead of conventional high-order polynomials, are utilized as the static nonlinearities, due to the fact that the splines are able to represent hard nonlinearities accurately and circumvent the numerical instability problem simultaneously. The predistorter is implemented on the indirect learning architecture, and the separable nonlinear least squares (SNLS) Levenberg-Marquardt algorithm is adopted given that the separation method reduces the dimension of the nonlinear search space and thus greatly simplifies the identification procedure.

In [1, 3], Ase and Katayama deal with identification of Hammerstein-Wiener systems, or NLN systems, in which a linear subsystem is sandwiched by two nonlinearities. Then, initializing by an estimated linear model, they apply separable least-squares to optimize the mean square error of the OE model, where a version of the DDLC-based gradient search is employed.

[1] A subspace-based identification of Wiener–Hammerstein benchmark model. H Ase, T Katayama, Control Engineering Practice **44**:126-137 (2015).

[2] Identification of Hammerstein-Wiener systems in closed-loop. Hajime Ase and Tohru Katayama, Proceedings of the 49th ISCIE International Symposium on Stochastic Systems Theory and Its Applications Hiroshima (2017).

[3] Identification of Hammerstein-Wiener systems using subspace method and separable least-squares. H Ase, T Katayama, Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications 40-47 (2017).

[4] Kautz basis expansion-based Hammerstein system identification through separable least squares method. C M Cheng, X J Dong, Z K Peng, W M Zhang and G Meng, Mechanical Systems and Signal Processing **121**:929-941 (2019).

[5] Solving bilinear tensor least squares problems and application to Hammerstein identification. Lars Eldén and Salman Ahmadi-Asl. Numerical Linear Algebra with Applications **26** (2019).

[6] Identification of IIR Wiener systems with spline nonlinearities that have variable knots. M C Hughes, D T Westwick, IEEE Transactions on Automatic Control **50**:1617-1622 (2005).

[7] Identification of time-varying Hammerstein systems using a basis expansion approach. B I Ikharia, D T Westwick, Canadian Conference on Electrical and Computer Engineering (2006).

[8] On the identification of Hammerstein systems with time-varying parameters. B I Ikharia, D T Westwick, Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE 6475-6478 (2007).

[9] Linear approximation and identification of MIMO Wiener–Hammerstein systems. T Katayama, H Ase, Automatica **71**:118-124 (2016).

[10] Quasi-convexity analysis of the Hammerstein model. M Rasouli, D Westwick, W Rosehart, Automatica **50**:277-281 (2014).

[11] Global identification of wind turbines using a Hammerstein identification method. G van der Veen, J W van Wingerden, M. Verhaegen, IEEE Transactions on Control Systems Technology **21**:1471-1478 (2013).

[12] Adaptive predistortion using cubic spline nonlinearity based Hammerstein modeling. X Wu, J Shi, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences **E95-A**:542-549 (2012).

[13] Estimation of an N–L–N Hammerstein–Wiener model. Y Zhu, Automatica **38**:1607-1614 (2002).

7.3. Error in Variables

The least-squares method generally produces biased parameter estimates when the observed input-output data are corrupted with noise. If the noise acting on both the input and output is white, and if the noise variances are known, or if estimates of the noise variances are available, then the principle of biased-compensated least squares (CLS) can be used to obtain consistent estimates. An extended version of CLS is shown to be a separable nonlinear least squares problem. The errors-in-variables (EIV) framework concerns static or dynamic systems whose input and output variables are affected by noise that is mostly assumed to be additive. These models play an important role in several engineering applications, such as, time series modeling, direction-of-arrival estimation, blind channel equalization and many other signal and image processing problems.

In [8] the authors consider the problem of dynamic errors-in-variables identification in order to avoid possible divergence of the iteration-type bias-eliminating algorithms in the case of high noise. The bias-eliminating problem is reformulated as a minimization problem associated with a concentrated loss function. A variable projection algorithm is proposed to efficiently solve the resulting minimization problem. See also [2].

[1] A separable nonlinear least-squares approach for identification of linear systems with errors in variables. M Ekman, M Hong, T Söderström, IFAC Proceedings **39**:178-183 (2006).

[2] Errors-in-variables system identification using structural equation modeling. D Kreiberg, T Söderström, F Yang-Wallentin, Automatica **66**:218-230 (2016).

[3] Novel algorithms based on conjunction of the Frisch scheme and extended compensated least squares. T Larkowski, J G Linden, B Vinsonneau, K J Burnham, Conference on Control (2008).

[4] Algorithms for recursive/semi-recursive bias-compensating least squares system identification within the errors-in-variables framework. J G Linden, T Larkowski, K J Burnham, International Journal of Control **85**:1625-1643 (2012).

[5] System identification for the errors-in-variables problem. T Söderstrom, UKACC International Conference on Control 19-32 (2010).

[6] Errors-in-variables identification using a generalized instrumental variable estimation method. T Söderström, 49th IEEE Conference on Decision and Control (CDC) (2010).

[7] Convergence properties of bias-eliminating algorithms for errors-in-variables identification. T Söderström, M Hong, W X Zheng, International Journal of Adaptive Control and Signal Processing **19** (2005).

[8] A covariance matching approach for identifying errors-in-variables systems. T Söderström, M Mossberg, M Hong, Automatica **45**:2018-2031 (2009).

8. Mechanical Systems

A large number of applications involving mechanical systems are presented here. In the Master Thesis [1], the authors consider a new method to estimate the unknown pitch actuator gain for a wind turbine system. The wind speed is included in an augmented system state. Earlier was shown that a Kalman filter can be rewritten over a moving time window both as a linear and a nonlinear LMS algorithm. A nonlinear LMS problem with a particular structure can be defined over a moving time window w, which can be solved by the separable least squares method.

The authors of [5] are motivated by the idea of turbo-machinery active subspace performance maps. In this paper they study dimension reduction in turbo-machinery 3D CFD simulations. First, they show that these subspaces exist across different blades—under the same parametrization—largely independent of their Mach or Reynolds number. This is demonstrated via a numerical study on three different blades. Then, in an attempt to reduce the computational cost of identifying a suitable dimension reducing subspace, they examine statistical sufficient dimension reduction methods, including sliced inverse regression, sliced average variance estimation, principal Hessian directions and contour regression. Unsatisfied by these results, they evaluate a new idea based on polynomial variable projection—a non-linear least squares problem. Their results using polynomial variable projection clearly demonstrate that one can accurately identify dimension reducing subspaces for turbo-machinery functionals at a fraction of the cost associated with prior methods. They apply these subspaces to the problem of comparing design configurations across different flight points on a working line of a fan blade. They demonstrate how designs that offer a healthy compromise between performance at cruise and sea-level conditions can be easily found by visually inspecting their subspaces.

[1] Fault Detection and Identification for Wind Turbine Systems: a Closed-Loop Analysis. S Donders, V Verdult, M Verhaegen, Master's Thesis, University of Twente (2002).

[2] Dynamic torque calibration by means of model parameter identification. L Klaus, B Arendacká, M Kobusch, T Bruns, Acta Imeko **4**:39-44 (2015).

[3] Unbalanced estimation using linear and nonlinear regression. P Nauclér, T Söderström, Automatica **46**:1752-1761 (2010).

[4] Equations of motion for general constrained systems in Lagrangian mechanics. F E Udwadia, A D Schutte, Acta Mechanica **213**:111-129 (2010).

[5] Supporting Multi-point Turbo-machinery Dimension Reduction via Polynomial-based Variable Projection S Yuchi, P Seshadri, G Parks and S Shahpar, Manuscript, Rolls-Royce (2018).

8.1. Vibrations

In [3] the authors consider the modeling and identification of non-stationary random vibration signals in various applications. A recent development in this area is the postulation of FS–TARMA models with complex exponential or spline basis functions, accompanied by the development of a Separable Nonlinear Least Squares (SNLS) method which achieves simultaneous estimation of the model coefficients of projection and the basis functions themselves [4]. This method drastically simplifies the identification procedure; results from application case studies are very promising.

In [2], the authors investigate the nonparametric estimation of the frequency dependent complex modulus of a viscoelastic material. The strains due to flexural wave propagation in a bar specimen are registered at different cross sections. The time domain data is transformed into frequency domain using discrete Fourier transform and a separable nonlinear least squares algorithm is then employed to estimate the complex modulus at each frequency. Inherent numerical problems due to associated ill-conditioned matrices are treated with special care, while an analysis of the quality of the nonlinear least squares estimate is also carried out.

Functional Series Time-dependent Autoregressive Moving Average (FS-TARMA) models are characterized by time varying parameters which are projected onto selected functional subspaces. They offer parsimonious and effective representations for a wide range of non-stationary random signals where the evolution in the dynamics is of deterministic nature. Yet, their identification remains challenging, with a main difficulty pertaining to the determination of the functional subspaces. In [5] the authors overcome this challenge via the introduction of the novel class of Adaptable FS-TARMA (AFS-TARMA) models, that is, models with basis functions properly parametrized and directly estimated based on the modeled signal. Model identification is effectively dealt with through a SNLS based estimation procedure that decomposes the problem into two simpler subproblems: a quadratic one and a reduced-dimensionality non-quadratic constrained optimization one.

In [6] the authors observe that a discrete-time Linear Parameter-Varying (LPV) model can be seen as the combination of local LTI (Linear Time Invariant) models, together with a scheduling signal dependent function set that selects one of the models to describe the continuation of the signal trajectories at every time instant. An identification strategy of LPV models is proposed that consists of the separate approximation of the local model set and the scheduling functions. The local model set is represented as a linear combination (series expansion) of Orthonormal Basis Functions (OBFs). First the OBFs that guarantee the least asymptotic worst-case modeling error for the local models are selected through the Fuzzy Kolmogorov c-Max approach. With the resulting OBFs, the weighting functions are identified through a separable least-squares algorithm.

[1] Maximum likelihood estimation of flexural wave-numbers in lightly damped plates. C R Halkyard, Journal of Sound and Vibration **300**:217-240 (2007).

[2] On the use of flexural wave propagation experiments for identification of complex modulus. K Mahata, S Mousavi, T Söderstrom, M. Mossberg, U. Valdek, L. Hillstron, IEEE Transactions on Control Systems Technology **11**:863-874 (2003).

[3] Non-stationary random vibration modeling and analysis via functional series time-dependent ARMA (FS-TARMA) models: a critical survey. M D Spiridonakos, S D Fassois, Mechanical Systems and Signal Processing **47**:175-224 (2014).

[4] FS–TARMA models for non–stationary vibration analysis: an overview and recent advances. M D Spiridonakos, S D Fassois. Proceedings of the 17th International Congress on Sound and Vibration (ICSV 17) (2010),

[5] Adaptable functional series TARMA models for non-stationary signal representation and their application to mechanical random vibration modeling. M D Spiridonakos, S D Fassois, Signal Processing **96**:63-79 (2014).

[6] Flexible model structures for LPV identification with static scheduling dependency. R Tóth, P S C Heuberger, P M J Van den Hof, Proceedings of the 47th IEEE Conference on Decision and Control ThB03.6 (2008).

8.2. Control

Several contributions to system control and identification can be found in this Section. Some only use the formulas for derivatives of projectors [9, 1, 6], but others concern themselves with separable models.

In [7] a novel identification algorithm for a class of non-linear, possibly parameter varying models is proposed. The algorithm is based on separable least squares ideas. These models are given in the form of a linear fractional transformation (LIFT), where the 'forward' part is represented by a conventional linear regression and the 'feedback' part is given by a nonlinear map, which can take into account scheduling variables available for measurement. The non-linear part of the model can be parameterized according to various paradigms, such as neural networks (NN) or general nonlinear autoregressive exogenous (NARX) models. The estimation algorithm exploits the separability of the criterion used to estimate the parameters. When using a NN, the results about the derivative of the pseudoinverse facilitates the calculation of the Frechet derivative needed to implement a separable least squares algorithm.

The Ph D Thesis [5] considers the problem of identifying nonlinear systems with errors-in-variables by using separable NLLS. A bias-eliminating approach, based on a compensated least- squares (CLS) solution of an overdetermined system of equations and separable nonlinear LS is used. On the other hand, the Ph D Thesis [8] considers the identification of parameters in a Hammerstein-Wiener separable model for control valves.

[1] Elements of Robustness and Optimal Control for Infrastructure Networks. Q Ba, Ph D Thesis, University of Southern California (2018).

[2] On decentralized robust weight control for DC power networks. Q Ba, K Savla, American Control Conference (ACC) (2016).

[3] Switching strategy based on homotopy continuation for non-regular affine systems with application in induction motor control. A Borisevich, G Schullerus, arXiv preprint arXiv:1203.5919 (2012).

[4] Regularized data-driven construction of fuzzy controllers. M Burger, J Haslinger, U Bodenhofer, H W Engl, Journal of Inverse and Ill-Posed Problems **10** (2002).

[5] Modeling and Control of Bilinear Systems: Application to the Activated Sludge Process. M Ekman, Ph D Thesis, University of Uppsala (2005).

[6] Algebraic invariance conditions in the study of approximate (null) controllability of Markov switch processes. D Goreac, M Martinez, Mathematics of Control, Signals, and Systems **27**:551-558 (2015).

[7] Identification of non-linear parametrically varying models using separable least squares. F Previdi, M Lovera, International Journal of Control **77**:1382-1392 (2004).

[8] Identificação de Processos Não-Lineares e Quantificação de Atrito em Válvulas de Controle. R A Romano, Ph D Thesis, Escola Politécnica, San Pablo, Brazil (2010).

[9] Optimal Nonlinear Control and Estimation Using Global Domain Linearization. L A Wendt, Ph D Thesis, University of Illinois, Urbana-Champaign (2017).

9. Machine Learning

A very current field where Variable Projections has diverse applications to classification, Markov decision processes, computer vision, discriminant analysis, clustering and image super-resolution.

Molinari [2, 3] uses splines with variable knots to represent functional data curves. This is one of the classical applications of VARPRO and the author takes good advantage of it. Chatterjee and Milanfar [1] consider separable models for patch denoising, by learning a best basis for each cluster of similar patches and then they apply VARPRO for optimization.

In [5] the authors propose a novel computationally efficient single image super-resolution method that learns multiple linear mappings (MLM), to directly transform low-resolution feature subspaces into high-resolution ones. The problem is separable with multiple right-hand sides and VARPRO is used successfully.

[1] Clustering-based denoising with locally learned dictionaries. P Chatterjee, P Milanfar, IEEE Transactions on Image Processing **18**:1438-1451 (2009).

[2] Free knot splines for supervised classification. N Molinari, Journal of Classification **24**:221-234 (2007).

[3] Choice of B-splines with free parameters in the flexible discriminant analysis context. C Reynès, R Sabatier, N Molinari, Computational Statistics & Data Analysis **51**:1765-1778 (2006).

[4] Invariant Representations and Learning for Computer Vision. A Vedaldi, Ph D Thesis UCLA (2008).

[5] Learning multiple linear mappings for efficient single image super-resolution. K Zhang, D Tao, X Gao, X Li, Z. Xiong, IEEE Transactions on Image Processing **24**:846-861 (2015).

9.1. Neural Networks

This is an area of renovated interest in Machine Learning. In [17] the author rediscovers the fact that single layer perceptrons are separable non-linear models and thus amenable to training via VARPRO [11]. Given the current rage about multilayer networks (DNN), it might be of interest to extend these results to that case. In [10] there is a first attempt to do just that.

Mizutani and Demmel [9] describe in detail an economical trust-region implementation of VARPRO in the framework of a so-called block-arrow least squares (BA) algorithm for a general multiple-response nonlinear model. They then present numerical results using an exponential-mixture benchmark, seven-bit parity, and color reproduction problems; in some situations, VARPRO enjoys quick convergence and attains high classification rates, while in some others VARPRO works poorly. This observation motivates them to investigate original VARPRO's strengths and weaknesses compared with other (full-functional) approaches. To overcome the limitations of VARPRO, they suggests how VARPRO can be modified as a Hessian matrix-based approach that exploits negative curvature when it arises. For this purpose, an economical BA algorithm is very useful in implementing such a modified VARPRO, especially when the given model is expressed as a multi-layer (neural) network.

One of the most interesting contributions is [16], which proves that the reduced VARPRO functional is always better conditioned than the original one for the full problem and that the Levenberg-Marquardt version for separable problems usually converges orders of magnitude faster for notoriously ill-conditioned single hidden layer NN. Early references to this application are [12, 18, 19]. We have also pointed out the usefulness of VARPRO in training single hidden layer networks in [11], which was re-enforced later on by [2, 3]. It is worth insisting on this fact: it is not only the reduction in the number of variables that makes VARPRO powerful, but rather its regularization effect. Highly non-convex multi-modal problems become much better defined in the reduced form for the nonlinear parameters.

Several authors discuss B-splines NN and their training by VARPRO [4, 14]. In [6] the authors consider representation discovery in reinforcement learning (RL) posing basis adaptation as a nonlinear separable least-squares value function approximation solvable by VARPRO.

[1] Reduction of adjusting weights space dimension in feedforward artificial neural networks training. S L Blyumin, P V Saraev, Proceedings 2002 IEEE International Conference on Artificial Intelligence Systems (2002).

[2] Variable projection method and Levenberg-Marquardt algorithm for neural network training. C T Kim, J J Lee, H Kim, 32nd Annual Conference on IEEE Industrial Electronics, IECON (2006).

[3] Training two-layered feedforward networks with variable projection method. C T Kim, J J Lee, IEEE Transactions on Neural Networks **71**:3640-3643 (2008).

[4] Extending the functional training approach for B-splines. C L Cabrita, A E Ruano, P M Ferreira, L T Koczy. The 2012 International Joint Conference on Neural Networks (IJCNN) (2012).

[5] On the optimal structure design of multilayer feedforward neural networks for pattern recognition. D Chen, P Burrell, International Journal of Pattern Recognition and Artificial Intelligence **16**:375-398 (2002).

[6] Basis adaptation for sparse nonlinear reinforcement learning. S Mahadevan, S Giguere, N Jacek, Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (2013).

[7] Fuzzy mixtures of complementary local experts: towards neuro-fuzzy modular networks. E Mizutani, K Nishio, IEEE International Conference on Fuzzy Systems FUZZ-IEEE '02 (2002).

[8] Multi-illuminant color reproduction for electronic cameras via CANFIS neuro-fuzzy modular network device characterization. E Mizutani, K Nishio, IEEE Transactions on Neural Networks **13**:1009-1022 (2002).

[9] On improving trust-region variable projection algorithms for separable nonlinear least squares learning. E Mizutani, J Demmel. The 2011 International Joint Conference on Neural Networks (2011).

[10] A generalized learning paradigm exploiting the structure of feedforward neural networks. R Parisi, E D Di Claudio, G Orlandi, B D Rao, IEEE Transactions on Neural Networks **7**:1450-1460 (1996).

[11] Variable projection neural network training. V Pereyra, G Scherer and F Wong, Applied Mathematics and Computers in Simulation **73**:231-243 (2006).

[12] A new formulation of the learning problem of a neural network controller. A E Ruano, D I Jones, P. J Fleming, Proceedings of the 30th IEEE Conference on Decision and Control (1991).

[13] Towards a more analytical training of neural networks and neuro-fuzzy systems. A E Ruano, C L Cabrita, P M Ferreira, Signal Processing (WISP) (2011).

[14] Supervised training algorithms for B-spline neural networks and neuro-fuzzy systems. A E Ruano, C Cabrita, J V Oliveira, International Journal of Systems Science **33**:689-711 (2002).

[15] Training neural networks and neuro-fuzzy systems: A unified view. A E Ruano, P M Ferreira, C Cabrita, S Matos, IFAC Proceedings (2002).

[16] Separable non-linear least-squares minimization - possible improvements for neural net fitting. J Sjoberg, M Viberg, Neural Networks for Signal Processing VII (1997).

[17] Learning algorithms for neural networks and neuro-fuzzy systems with separable structures. B A Skorohod, Cybernetics and Systems Analysis **51**:173-186 (2015).

[18] Application of projection learning to the detection of urban areas in SPOT. K Weigl, G Giraudon, M Berthod, Report 2143, INRIA, France (1993).

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

10. Mathematics

In [1] the authors describe and analyze an algorithm for computing the homology (Betti numbers and torsion coefficients) of basic semi-algebraic sets that works in weak exponential time. For this purpose they use the formulas of differentiation of the pseudo-inverse.

The goal of tensor completion is to fill in missing entries of a partially known tensor (possibly including some noise) under a low-rank constraint. This may be formulated as a least-squares problem. The set of tensors of a given multilinear rank is known to admit a Riemannian manifold structure, thus methods of Riemannian optimization are applicable [2]. For this the differentiation of the pseudoinverse is required.

[1] Computing the homology of basic semi-algebraic sets in weak exponential time. P Bürgisser, F Cucker and P Lairez, Journal of the ACM 66:1-30 (2018).

[2] A Riemannian trust-region method for low-rank tensor completion. Gennadij Heidel and Volker Schulz, Numerical Linear Algebra with Applications 25 (2018).

10.1. Differential Equations and Dynamical Systems

Applications to direct and inverse problems involving differential equations are presented here, including several on model order reduction. The dynamic mode decomposition (DMD) has become a leading tool for data-driven modeling of dynamical systems, providing a regression framework for fitting linear dynamical models to time-series measurement data. The method is akin to POD for model order reduction. In [1] the authors present an algorithm for computing an optimized version of the DMD for data that may be collected at unevenly spaced sample times. The primary computational tool at the heart of these algorithms is the variable projection method. To apply variable projection, the DMD is rephrased as a problem in exponential data fitting (specifically, inverse differential equations), an area of research which has been extensively developed and has many applications. A careful discussion and some impressive numerical examples are included.

In [5] the authors consider a set of response observations for a parametrized dynamical system. This is an interesting application that combines separable least squares with model order reduction. The authors seek a parametrized dynamical model that will yield uniformly small response error over a range of parameter values yet has low order. Frequently, access to internal system dynamics or equivalently, to realizations of the original system, is either not possible or not practical since only response observations over a range of parameter settings might be known. Respecting these typical operational constraints, the authors propose a two phase approach that first encodes the response data into a high fidelity intermediate model of modest order, followed then by a compression stage that serves to eliminate redundancy in the intermediate model. For the first phase, they extend non-parametric least-squares fitting approaches so as to accommodate parameterized systems. This results in a (discrete) separable least-squares problem formulated with respect to both frequency and a parameter that identifies "local" system response features. The second phase uses an H2-optimal model reduction strategy accommodating the specialized parametric structure of the intermediate model obtained in the first phase. The final compressed model inherits the parametric dependence of the intermediate model and maintains its high fidelity, while generally having dramatically smaller system order. A variety of numerical examples demonstrating this approach are provided. Also [11] considers model order reduction, this time through projected nonlinear least squares that leads to a separable complex NLLSQ problem.

Harker and Rath [4] describe a new method for identifying the system parameters of a dynamic system in state-space form by minimizing the least-squares error of the measured system output. The variable projection method is used to eliminate the necessity of estimating the system states, reducing the system identification cost function to a function of only the system parameters. In [11] the authors consider the mesh-less solution of PDE's in irregular domains by using radial functions and Galerkin collocation. They investigate the problem of the adaptive calculation of the basis function centers, which will not be coinciding with the collocation points. This leads to a SNLLSQ problem that is solved with VARPRO.

[1] Variable projection methods for an optimized dynamic mode decomposition. T Askham, J N Kutz, SIAM Journal on Applied Dynamical Systems **17**:380-416 (2018).

[2] Scientific Computing with Ordinary Differential Equations. P Deuflhard, F Bornemann, Springer (2012).

[3] Gewöhnliche Differentialgleichungen. P Deuflhard, F Bornemann. de Gruyter Studium (2013).

[4] An ODE-based method for computing the approximate greatest common divisor of polynomials. Antonio Fazzi, Nicola Guglielmi and Ivan Markovsky, Numerical Algorithms (2018).

[5] Empirical least-squares fitting of parametrized dynamical systems. Alexander Grimm, Christopher Beattie, Zlatko Drmaˇc and Serkan Gugercin, arXiv.org > math > arXiv:1808.05716 (2018).

[6] Conformal maps to multiply slit domains and applications. N Hale, T W Tee, SIAM Journal on Scientific Computing **31**:3195-3215 (2009).

[7] Global least squares for time-domain system identification of state-space models. M Harker and G Rath, 7th Mediterranean Conference on Embedded Computing (2018).

[8] Optimal model reduction using projected nonlinear least squares. J M Hokanson and C C Magruder, arXiv t arXiv:1811.11962 (2018).

[9] Identification of fractional linear dynamical systems with autocorrelated errors in variables by generalized instrumental variables. D V Ivanov, I L Sandler, E V Kozlov, IFAC-PapersOnLine, **51**:580-584 (2018).

[10] Improved estimation performance using known linear constraints. K Mahata, T Söderström, Automatica **40**:1307-1318 (2004).

[11] Radial function collocation solution of partial differential equations in irregular domains. V Pereyra, G Scherer, P. González Casanova, Int Journal of Computing Science and Mathematics **1:**28-41 (2007).

[12] Data-driven operator inference for nonintrusive projection-based model reduction. B Peherstorfer, K Willcox, Computer Methods in Applied Mechanics and Engineering **306**:196-215 (2016).

[13] An analysis of separable least squares data driven local coordinates for maximum likelihood estimation of linear systems. T Ribarits, M Deistler, B Hanzon, Automatica **41**:531-544 (2005).

[14] Numerical Computation 1: Methods, Software and Analysis. C W Ueberhuber, Springer (2012).

10.2. Inverse Problems

The interesting paper [2] discusses the problem of separable nonlinear inverse problems and compressed sensing in the case of deterministic settings. The authors apply their results to heat-source localization and estimation of brain activity from electroencephalography data.

In some inverse problems defined by models that include partial differential equations, a part of the boundary conditions are unknown and are to be estimated from experimental measurements [9, 12, 13]. It has been shown in a previous contribution that the solution of the inverse Richards' problem can allow estimating percolation rates at the bottom of landfills through the use of measurements at the surface only. This can be a useful complement of the information furnished by the vadose measurement system, pointing to the possible presence of biases of in-situ equipment and making it possible to use inexpensive mobile equipment to carry out surface measurements. In this article the authors consider a generalization that allows the presence of unknown non-linear parameters, such as the effective hydraulic conductivity and the root uptake coefficients. This is accomplished using the method of separation of variables in the resulting estimation problem. Thanks to the linearity of the model, all these conditions can be expressed as linear functions of the unknown lower boundary condition. Otherwise, the relevant non-linear parameters are to be estimated from the data as well. Obviously, the correlation between the linear parameters contained in the unknown lower boundary conditions and the non-linear parameters can reduce the reliability of the monitoring procedure and hence the necessity of limiting the number of the latter.

The comprehensive paper [5] considers large-scale inverse problems in image processing that require regularization in order to compute meaningful solutions. This book chapter treats three common methods based on a linear model, a separable nonlinear model and a general nonlinear model. Inverse problems are ubiquitous in imaging applications, including deconvolution (or, more generally, deblurring), super-resolution (or image fusion), image registration, image reconstruction, seismic imaging, inverse scattering and radar imaging. These problems are referred to as large-scale because they typically require processing a large amount of data (the number of pixels or voxels in the discretized image) and systems with a large (e.g., 10^9 for a 3D image reconstruction problem) number of equations. After a detailed discussion of the application of regularized linear methods, fully nonlinear and separable models are considered. For the separable case the linear problem can be solved using the techniques previously discussed. In the application section the separable approach is used for multi-frame blind deconvolution. Also, in [3, 10] these authors with J T Slagel consider sampling methods for massive scale linear and separable inverse problems.

The interesting papers [4, 6] propose an alternative to VARPRO for separable inverse problems: first linearize the whole problem and then project. This approach gives more flexibility and it turns out to be more efficient as shown in several nontrivial examples.

Many inverse problems include nuisance parameters which, while not of direct interest, are required to recover primary parameters. Structure present in these problems allows efficient optimization strategies. A well known example is variable projection, where nonlinear least squares problems which are linear in some parameters can be very efficiently optimized [1].

[1] Estimating nuisance parameters in inverse problems. A Y Aravkin, T Van Leeuwen, Inverse Problems **28** (2012).

[2] Sparse recovery beyond compressed sensing: separable nonlinear inverse problems. Brett Bernstein, Sheng Liu, Chrysa Papadaniil, Carlos Fernandez-Granda, arXiv:1905.04627v1 (2019).

[3] Iterative sampled methods for massive and separable nonlinear inverse problems. Julianne Chung, Matthias Chung, J Tanner Slagel, International Conference on Scale Space and Variational Methods in Computer Vision 119-130 (2019).

[4] An efficient iterative approach for large-scale separable nonlinear inverse problems. J Chung, J G Nagy, SIAM Journal on Scientific Computing **31**:4654-4674 (2010).

[5] Large-scale inverse problems in imaging. J Chung, S Knepper, J G Nagy, Handbook of Mathematical Methods in Imaging 43-86 (2014).

[6] LAP: a linearize and project method for solving inverse problems with coupled variables. James L Herring, James G Nagy and Lars Ruthotto, arXiv preprint arXiv:1705.09992 (2017).

[7] A Unified Framework for Identifiability Analysis in Bilinear Inverse Problems. Y Li, Ph D Thesis, University of Illinois (2015).

[8] The inverse problem of bioelectricity: an evaluation. A van Oosterom, Medical & Biological Engineering & Computing **50**:891-902 (2012).

[9] Use of inverse modeling techniques for the estimation of heat transfer coefficients to fluids in cylindrical conduits. A P Reverberi, B Fabiano,

V G Dovì, International Communications in Heat and Mass Transfer **42**:25-31 (2013).

[10] Row-Action Methods for Massive Inverse Problems. J Tanner Slagel, Ph D Thesis, Virginia Polytechnic Institute and State University (2019).

[11] Discrete spectrum reconstruction using integral approximation algorithm. V Sizikov, D Sidorov, Applied Spectroscopy **71**:1640-1651 (2017).

[12] A general method for the solution of inverse problems in transport phenomena. M Vocciante, A P Reverberi, V G Dovì, Chemical Engineering Transactions **43**:1615-1620 (2015).

[13] Generalization of the solution of the inverse Richards' problem. M Vocciante, A P Reverberi, V G Dovì, Chemical Engineering Transactions **52**:1285-1290 (2016).

11. Physics

The paper [2] investigates the approach of pure SU(2) lattice gauge theory with the Wilson action, to its continuum limit using the deconfining phase transition, the gradient flow and the cooling flow to set the scale. The elimination of the linear variables stabilizes the Levenberg-Marquardt procedure considerably.

In [3] the authors consider two-color QCD with two flavors of quarks as a possible theory of composite dark matter and use lattice field theory methods to investigate nuclear spectroscopy in the spin J = 0 and J = 1 multi-baryon sectors. In order to extract the energies and thereby the energy shifts, a separable model is used in which the linear variables are eliminated.

The work in [5] shows an interesting connection between NMR spectroscopy (a bestseller of VARPRO) and lattice quantum chromodynamics. From this connection and somewhat fortuitously, surged a collaboration [7, 8].

[1] Least squares fitting with one explicit parameter less. B A Berg, Computer Physics Communications **200**:254-258 (2016).

[2] Deconfinement, gradient, and cooling scales for pure SU (2) lattice gauge theory. B A Berg, D A Clarke, Physical Review D **95**:094508 (2017).

[3] Dark nuclei. II. Nuclear spectroscopy in two-color QCD. F W Detmold, M McCullough, A Pochinsky, Physical Review D **90**:114506 (2014).

[4] An artificial photosynthetic model based on a molecular triad of boron dipyrromethene and phthalocyanine. E A Ermilov, J Y Liu, R Menting, Y S Huang, B Roder, D K P Ng, Physical Chemistry Chemical Physics **18**:10964-10975 (2016).

[5] What can lattice QCD theorists learn from NMR spectroscopists? G T Fleming, QCD and Numerical Analysis **III**:143-152 (2005).

[6] Evolutionary fitting methods for the extraction of mass spectra in lattice field theory. G M von Hippel, R Lewis, R G Petry, Computer Physics Communications **178**:713-723 (2008).

[7] Excited state effective masses. George T Fleming, Saul D Cohen, Huey-Wen Lin,Victor Pereyra, Proceedings XXV International Symposium on Lattice Field Theory (2007)

[8] Excited-state effective masses in Lattice QCD. G T Fleming, S D Cohen, H-W Lin, V Pereyra, Physical Reviews D **80**:074506 (2009).

[9] Convolution-based one and two component FRAP analysis: theory and application. A Tannert, S Tannert, S Burgold, M Schaefer, European Biophysics Journal **38**:649-661 (2009).

[10] Geometry of nonlinear least squares with applications to sloppy models and optimization. M K Transtrum, B B Machta, J P Sethna, Physical Review E **83** (2011).

12. Optics

The rapid worldwide evolution of the LED (light emitting diode) industry has resulted in the implementation of LED elements in all kind of luminaries. This technology reduces energy consumption while at the same time offers endless possibilities for light engine design. It is well-known that the luminous intensity pattern of LEDs can be represented as a sum of cosine-power functions

$$I(\theta, a, b, c) = I_{max} \Sigma_{j=1}^{n} a_j \cos^{c_j}(\theta - b_j),$$

where ϑ is the polar angle and a_j, b_j and c_j are the function coefficients [2]. This is a separable model and using the variable projection procedure has been possible to find desired solutions in a fraction of the time needed for conventional discrete optimization methods.

The authors of [1] exploit the spatiotemporal correlation in adaptive optics using data-driven H2-optimal control. Adaptive optics (AO) is a well-established technique for real-time compensation of the optical wavefront distortions introduced by a turbulent medium. It has found widespread application in ground-based astronomical imaging, where it is used to counteract the devastating effect of atmospheric turbulence on the angular resolution. A data driven separable model is considered and the fit uses VARPRO.

[1] Exploiting the spatiotemporal correlation in adaptive optics using data-driven H2-optimal control. K Hinnen, M Verhaegen, N Doelman, JOSA A **24**:1714-1725 (2007).

[2] Improved approximation of spatial light distribution. D Kaljun, T Novak, J Žerovnik, PLOS 12 (2017).

[3] Optimum selection of input polarization states in determining the sample Mueller matrix: a dual photo-elastic polarimeter approach. D Layden, M F G Wood, I A Vitkin, Optics Express **20**:20466-20481 (2012).

[4] Towards unsupervised fluorescence lifetime imaging using low dimensional variable projection. Y Zhang, A Cuyt, W Lee, G L Bianco, G Wu, Y Chen, D D Li, Optics Express **24**:26777-26791 (2016).

13. Chemistry

These are applications to various topics in chemistry, including gas chromatography, spectroscopy, photo-induced processes and sequential energy and charge transfer. Several of these applications fit parameters in separable modeling problems. In [1] the authors use the formula for differentiation of the pseudo-inverse. The talk [7] uses separation of variables for models of permeation of gases in polymers.

A notable contribution is [6], where the author considers in detail the application of Variable Projections to a number of important problems in Physics and Chemistry. She also is the implementor of a version of VARPRO in the R language, including the problem of multiple right hand sides. An application of separation of variables combined with a genetic algorithm for multi-exponential fluorescence decay surface calculation can be found in [4].

[1] Path optimization by a variational reaction coordinate method. II. Improved computational efficiency through internal coordinates and surface interpolation. A B Birkholz, H B Schlegel, The Journal of Chemical Physics **144**:184101 (2016).

[2] Spectroscopic study of electron and energy transfer in novel silicon phthalocyanine—boron dipyrromethene triads. E A Ermilov, J Y Liu, D K P Ng, B Röder, Physical Chemistry Chemical Physics **11**:6230-6440 (2009).

[3] Switching the photo-induced processes in host–guest complexes of β-cyclodextrin-substituted silicon (iv) phthalocyanines and a tetrasulfonated porphyrin. E A Ermilov, R Menting, J T F Lau, X Leng, B. Roder, D Ng, Physical Chemistry Chemical Physics (2011).

[4] Combined genetic algorithm and multiple linear regression (GA-MLR) optimizer: application to multi-exponential fluorescence decay surface. J J Fisz, The Journal of Physical Chemistry A **110**:12977-12985 (2006).

[5] Sequential energy and charge transfer processes in mixed host–guest complexes of subphthalocyanine, porphyrin and phthalocyanine chromophores. R Menting, D K P Ng, B Röder, E A Ermilov, Physical Chemistry Chemical Physics (2012).

[6] Separable Nonlinear Models: Theory, Implementation and Applications in Physics and Chemistry. K M Mullen, Ph D Thesis, Vrije University (2008).

[7] Permeation of gases in polymers: parameter identification and nonlinear regression analysis. R Scheichl, M H Klopffer, Z Benjelloun-Dabaghi, B. Flaconneche, Journal of Membrane Science **254**:275-293 (2005).

[8] The influence of solvent polarity and metalation on energy and electron transfer in porphyrin–phthalocyanine heterotrimers. S Tannert, E A Ermilov, J O Vogel, M T M Choi, D K P Ng, The Journal of Physical Chemistry **111**:8053-8062 (2007).

[9] Signal amplification between Gβγ release and PI3Kγ-mediated PI(3,4,5)P3 formation monitored by a fluorescent Gβγ biosensor protein and repetitive two component total internal reflection/fluorescence redistribution after photobleaching analysis. A Tannert, P Voigt, S Burgold, S Tannert, M. Schaefer, Biochemistry **47**:11239-11250 (2008).

13.1. Gas Chromatography

In this Section we find inverse problems using reversed flow gas chromatography that are amenable to Variable Projections. Most of these chemistry papers are behind a paying wall and thus they are inaccessible to us and therefore we cannot comment in more detail.

[1] Impact of the BEA zeolite morphology on isobutane adsorption followed by reversed-flow inverse gas chromatography. N Batalha, A Soualah, L Pinard, Y Pouilloux, F Lemos, T Belin, Journal of Chromatography A **1260**:206-214 (2012).

[2] Interactions of H2 on the isobutane adsorption over bifunctional catalyst PtHBEA revealed by reversed-flow inverse gas chromatography. T Belin, N Batalha, L Pinard, F Lemos, Y Pouilloux, The Journal of Physical Chemistry **119**:1791-1799 (2015).

[3] New approach to the acidity characterization of pristine zeolite crystals by ethylene using reversed-flow inverse gas chromatography (RF-IGC). M A Benghalem, A Astafan, L Pinard, T Jean Daou, Thomas Belin, The Journal of Physical Chemistry **121**:2738-2747 (2017).

14. Other Sciences

In this section we collect an array of applications in environmental sciences, astronomy, geodesy and geostatistics, computer sciences, economics, planning, aeronautics and neurosciences.

An important problem in the *environmental sciences* is the quantification of soil contamination using surface measurements. In article [25], one of a set by the same research group, the authors develop a method that can be used to determine the time dependent values of the flux potential, pressure head, and water content, as well as some of the soil properties, such as the soil porosity or the hydraulic conductivity, etc., in the region of interest and importantly, under unknown conditions on the lower boundary of the contaminated sector. The physical model used for the water percolation in the vadose or unsaturated zone can be described by the time dependent Richards equation in one space variable z, with appropriate initial and boundary conditions on the surface and the lower boundary of the integration region. This last unknown condition is assumed to be a piecewise constant function in each time interval $t_{i-1} \leq t \leq t_i$. After a Kirchhoff transformation, the differential equation problem can be formally resolved and the general solution, involving a linear parameter c_i and nonlinear parameters corresponding to the soil properties can be expressed for each time interval $[t_{i-1}, t_i]$. Given the vector of experimental data from the surface measurements, a nonlinear least squares problem can be formulated in the linear parameters c_i that define the lower boundary conditions and the nonlinear soil parameters. This problem is solved using VARPRO.

The aim of articles [6] and [20] is the forecast of highly nonlinear time series that in [6] have seasonal, non-stationary characteristics. In order to take these into account, the functional coefficients of the autoregressive model are approximated using a weighted gradient radial basis function three-layer network. Or, in other words, the time series is approximated by a linear combination of nonlinear functions. Hence, the variable projection algorithm VARPRO can be used. The technique was tested with good results on retail sales data. In [20], the functional coefficients of the autoregressive model have the form of a constant plus an exponential function, $c_i + \pi_i \exp\left(-\gamma_i\, y_{i-1}^2\right)$, where the linear parameters c_i, π_i and the nonlinear γ_i have to be determined. Again, VARPRO is used to solve the nonlinear least squares problem. The next two papers [4, 12] consider a direct method for time series forecasting using Gaussian Process (GP) models. The authors use the idea of separation of linear/nonlinear variables to train the GP models. The estimation of the nonlinear parameters is done by a genetic algorithm.

Stellar winds are flows ejected from the upper atmosphere of a star. With their moments and energies they affect the physics of stellar atmospheres and influence the evolution of stars and galaxies. They allow spectroscopic studies of the most luminous stellar objects, even of distant galaxies, thereby enabling the observer to obtain quantitative information of their host galaxies. The determination of the wind properties such as velocities, moments, energy and mass-loss rates in first stars driven by Carbon-Nitrogen-Oxygen cycles is an important problem in astronomy [15]. These winds can only be simulated numerically since there are no available observations. A first test is to check if the winds are possible on first stars, by comparing a calculated radiative force against the gravitational force. For those stars for which it is shown that the winds exist, the hydrodynamic equations are solved to predict the wind-mass loss rate. The mass-loss rate M can be fitted in the least squares sense by the following non-linear function in the variables: luminosity L, mass fraction of heavier elements (heavier than Helium and Hydrogen) Z, and effective temperature Teff:

$$M = \alpha_0 L^{\alpha_1} 10^{\frac{\alpha_2(logZ+\alpha_3 logL)}{logZ(\alpha_4+\alpha_5 logL)+1}+\alpha_6 logT_{eff}}.$$

The values of the parameters α_i can be obtained using VARPRO.

Another application refers to gravitational lenses. A gravitational lens is a distribution of matter (a cluster of galaxies, for example), between a distant light source and an observer that bends the electromagnetic radiation passing through its gravitational field. This deformation leads to a folding of the wave-front, and thus the observer is hit several times by the wave-front, seeing multiple images at different times, i.e., there is a time delay between the pulses. A measurement of the time delay leads to a scaling of the model of the gravitational lens system. Article [24] compares three estimator techniques to measure time delays based on the data from optical light curves of lensed quasars. All three rely on iterative non-linear optimization algorithms. The estimators functions take n light curves, (n=2 or 4) as input and return n corresponding time shifts τ, one for each curve. These time shifts directly translate into time delay estimations between each pair of curves. The first estimator technique chooses a free-knot splines function to fit a single continuous model to all data points of the light curves, simultaneously adjusting time and magnitude shifts between these curves so as to minimize a $\chi 2$ fitting statistic between the data and the model. As usual, VARPRO allows to separate the linear from the non-linear part of the free-knot spline approximation problem.

Classical areas where the original least squares problem arose are geodesy and geostatistics. For instance, in [21] the authors use the derivative of the pseudoinverse formula for the Jacobian of the geodesic equation.

In [10] SNLLS is used for rational approximation of transfer matrices for the identification of rotor wake inflow, while in [1] the authors consider the problem of optimal design of a supersonic jet. Here, VARPRO is used to train neural networks that provide surrogate functions for some of the expensive steps in the modeling processes that are within the optimization loop.

An important problem in motor *neurosciences* is the accurate description of movements [7]. This would allow, for example, to characterize developmental coordination or to diagnose disorders like stages in Parkinson's disease, or autism spectrum disorder, etc. One technique used to describe an action is the SB-ST method. Here, a set of key postures of the movement (for example joint rotations) are represented by vectors thereby forming the spatial basis, SB. Then their ST profiles, representing the trajectories in time of these postures are defined. A nonlinear Gaussian model is then fitted to the spatiotemporal ST profiles, with the parameters in it representing how much (control) and when (coordination) the postures are being recruited. The parameters in the nonlinear model are computed via VARPRO.

[1] Aircraft design optimization. J J Alonso, P LeGresley, V Pereyra, Mathematics and Computers in Simulation **79**:1948-1958 (2009).

[2] Algebraic Geodesy and Geoinformatics. J L Awange, E W Grafarend, B Paláncz, P Zaletnyik, Springer (2010).

[3] Geospatial Algebraic Computations. Theory and Applications. J L Awange, B Paláncz, 3rd edn. Springer (2016).

[4] Discriminative non-negative matrix factorization for multiple pitch estimation.. N Boulanger-Lewandowski, Y Bengio, P Vincent, 13th International Society for Music Information Retrieval Conference 205-210 (2012).

[5] Estimating parameters of a sine wave by separable nonlinear least squares fitting. K F Chen, IEEE Transactions on Instrumentation and Measurement **59**:3214-3217 (2010).

[6] Generalized exponential autoregressive models for nonlinear time series: stationarity, estimation and applications. G Chen, M Gan, G Chen, Information Sciences **438**:46-57 (2018).

[7] Action Compositionality with Focus on Neurodevelopmental Disorders. L Claudino, Ph D Thesis, University of Maryland (2016).

[8] Estimating frequency statistics through distinct count measurements. E Cohen, arXiv preprint arXiv:1607.06517 (2016).

[9] Spatio-Temporal Heterogeneity: Concepts and Analyses. P Dutilleul, Cambridge University Press (2011).

[10] Managing short-term efficiency and long-term development through industrialized construction. P E Eriksson, S Olander, H Szentes, K Widen, Construction Management and Economics **32**:97-108 (2014).Identification of rotor wake inflow finite-state models for flight dynamics simulations. M Gennaretti, R Gori, J Serafini, F Cardito, G Bernardini, CEAS Aeronautical Journal **8**:209-230 (2017).

[11] State-space rotor aeroelastic modeling for real-time helicopter flight simulation. R Gori, F Pausilli, M D Pavel, M. Gennaretti, Advanced Materials Research **1016**:451-459 (2014).

[12] Time series forecasting using multiple Gaussian process prior model. T Hachino, V Kadirkamanathan, Symposium on Computational Intelligence and Data Mining 604-609 (2007).

[13] Multiple Gaussian process models for direct time series forecasting. T Hachino, V Kadirkamanathan, IEEJ Transactions on Electrical and Electronic Engineering **6**:245-252 (2011).

[14] Application of a time-accurate aeroelastic coupling framework to flutter-constrained design optimization. Jan F Kiviaho, Kevin Jacobson, Marilyn J Smith and Graeme Kennedy, Multidisciplinary Analysis and Optimization Conference (2018).

[15] CNO-driven winds of hot first stars. J Krtička, J Kubat, Astronomy & Astrophysics (2009).

[16] Plug-and-Play Monitoring and Performance Optimization for Industrial Automation Processes. H Luo, Springer (2016).

[17] Fitting the linear model of coregionalization by generalized least squares. B Pelletier, P Dutilleul, G Larocque, J W Fyles, Mathematical Geology **36**:323-343 (2004).

[18] Modeling of Lithuanian speech diphthongs. G Pyž, V Šimonytė, V Slivinskas, Informatica **22**:411-434 (2011).

[19] Second-order analysis of point patterns on a network using any distance metric. S Rakshit, G Nair, A Baddeley, Spatial Statistics **22**:129-154 (2017).

[20] Forecasting seasonal time series using weighted gradient RBF network based autoregressive model. W Ruan, Q Z Sheng, P Xu, N K Tran, N J G Falkner, X Li, W E Zhang, Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (2016).

[21] Optimization over geodesics for exact principal geodesic analysis. S Sommer, F Lauze, M Nielsen, Advances in Computational Mathematics **40**:283-313 (2014).

[22] Smooth sigmoid surrogate (SSS): An alternative to greedy search in recursive partitioning. X Su, J Kang, L Liu, Q Yang, J Fan, RA Levine, Statistics and Data Mining (2016).

[23] Topology design for optimal network coherence. T Summers, I Shames, J Lygeros, F Dorfler, European Control Conference (ECC) 1-6 (2015).

[24] COSMOGRAIL: the COSmological Monitoring of GRAvItational Lenses-XI. Techniques for time delay measurement in presence of microlensing. M Tewes, F Courbin, G Meylan, Astronomy & Astrophysics (2013).

[25] Improved remediation processes through cost-effective estimation of soil properties from surface measurements. M Vocciante, A P Reverberi, L Pietrelli, V G Dovi, Journal of Cleaner Production **167**:680-686 (2017).

[26] Adjoint-based airfoil optimization with adaptive isogeometric discontinuous Galerkin method. K Wang, S Yu, Z Wang, R Feng and T Liu, Computer Methods in Applied Mechanics and Engineering 344:602-625 (2019).

[27] Computing the Helmert transformations. G A Watson, Journal of Computational and Applied Mathematics **197**:387-394 (2006).

[28] Energy-efficient full diversity collaborative unitary space-time block code designs via unique factorization of signals. D Xia, J K Zhang, S Dumitrescu, IEEE Transactions on Information Theory **59**:1678-1703 (2013).