50+ Years of Numerical Analysis: A Personal History

Stanford University

December 20, 2018

Abstract

A rapid view of some of my most important work as decided by my peers. From the ten most cited papers I extract a few themes and give a short description and comments on applications.

1 Introduction

For the IX Pan-American Workshop on Computational and Applied Mathematics held in Varadero, Cuba I was asked to give a historical paper on my experiences in NA through the last half century. A number of years back I had given a similar talk (Gene Golub's Linear Algebra seminar) that concentrated on themes that touched Stanford and more specifically SCCM (Scientific Computing and Computational Mathematics), the predecessor of the current ICME (Institute on Computational Mathematics and Engineering). Its purpose was to inform the students on what had happened before they came to Stanford.

In this work the emphasis will be different. I chose to let my peers decide what is important and in what order. To do that I took my ten most cited papers, according to Google Scholar (see 1.1) and extracted a few themes that I discuss in some detail below.

The main themes that emerge from this list are:

- Variable Projections for separable nonlinear least squares (VARPRO)
- Fast Vandermonde solvers and multi-dimensional extensions (DVAND)
- Deferred corrections for differential equations (PASVA series)
- Seismic ray tracing, geological modeling and inverse problems (RAY3D)
- Least squares data fitting with applications (a book)

Ten most cited papers (as of December 2018)

1. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. G Golub, V. Pereyra, SIAM J. Numerical Analysis **10**:413-432 (1973). [1437]

2. Separable nonlinear least squares: the variable projection method and its applications. G Golub, V Pereyra, Inverse Problems **19**:R1 (2003). [546]

3. Solution of Vandermonde systems of equations. A. Bjorck, V Pereyra, Math. Comp. **24**:892-903 (1970). [425]

4. An adaptive finite difference solver for nonlinear two-point boundary value problems with mild boundary layers. M Lentini, V. Pereyra, SIAM J. Numerical Analysis **14**:91-111 (1977). [286]

5. Pasva3: An adaptive finite difference Fortran program for first order nonlinear ordinary boundary problems. V Pereyra, Codes for Boundary Value Problems in Ordinary Differential Equations 67-88 (1979). [216]

6. Mesh selection for discrete solution of boundary problems in ordinary differential equations. V Pereyra, G Sewell, Numerische Mathematik **23**:261-268 (1974). [207]

7. Solving two-point seismic ray-tracing problems in a heterogeneous medium: Part1. A general adaptive finite difference method. V Pereyra, WHK Lee, HB Keller, Bulletin of the Seismological Society of America **70**:79-99 (1980). [160]

8. Least Squares Data Fitting with Applications. PC Hansen, V Pereyra, G Scherer. John Hopkins University Press (2013). [115]

9. On improving an approximate solution of a functional equation by deferred corrections. V Pereyra, Numerische Mathematik **8**:376-391 (1966). [112]

10. A variable order finite difference method for multipoint boundary value problems. M Lentini, V Pereyra, Mathematics of Computation **28**:981-1003 (1974). [111]

Figure 1.1: Ten most cited papers. Number of citations in [.]

An interesting fact is that most of these works are connected to public domain developed software that have been widely used in a large variety of applications. I include in parenthesis the name or family of relevant codes.

2 Variable Projections for Separable Nonlinear Least Squares

In least squares data fitting one has a parametrized model and a set of observations. The purpose is to find the best model (parameters) that makes the difference between the data and the model least in the l_2 norm. In the present case the model is nonlinear and more specifically it is a linear combination of nonlinear parametrized functions of the form:

$$F(a, w) = \Phi(w)a = \Sigma_i a_i f_i(w).$$

So, the optimization problem is:

$$min_{a,w} \|y - F(a, w)\|_2^2, \tag{2.1}$$

where y is the vector that contains the data. The critical observation is that for fixed w (2.1) is a linear least squares problem whose solution can be written explicitly as:

$$a = \Phi^+ y, \tag{2.2}$$

where Φ^+ stands for the pseudo-inverse of Φ. If we replace this in (2.1) we will have a problem only on the nonlinear parameters:

$$min_w \|(I - \Phi\Phi^+)y\|_2^2.$$

The matrix in this expression depends on w and it is an orthogonal projector, thus the name Variable Projection.

In the original paper we developed the theory and described a detailed algorithm that included a tight implementation of Levenberg-Marquardt for this problem. That required formulas for the differentiation of the pseudo-inverse and projectors, that were included. We also implemented the method on a code (VARPRO) that was widely distributed and that was shown to solve problems that were hitherto very difficult for solvers that considered all the variables as independent, the so called joint problem. A natural approach, alternating the minimization between the two sets of variables, did not provide all the benefits of VARPRO:

- Elimination of the linear parameters reduces the number of initial guesses required.

- VARPRO was proven to lead to better conditioned problems and to transform problems with many local minima into ones with well defined ones.
- VARPRO was proven to always converge faster that using the same optimization for the joint problem. This includes convergence of VARPRO versus divergence for the joint minimization.

Basically, (2.2) is a nonlinear equality constraint. VARPRO honors it by construction. Clearly, methods that ignore this constraint and consider all the parameters as independent do it at its own peril.

Many different applications have separable models. Noticeably exponential fitting, both real and complex, is a pervading and classical one in which VARPRO excels [12]. I only quote a few applications that are specially interesting or rewarding:

- One of the most significant applications is to Nuclear Magnetic Resonance (NMR) Imaging and Spectroscopy.
- In 2005, G. Fleming published a paper with the title: "What can Lattice Quantum Chromo-Dynamics theorists learn from NMR spectroscopists?", and the answer was VARPRO.
- In recent years exploration geophysicists have discovered the power of VARPRO, which is now appearing as a hot tool in the holy grail of full waveform inversion [13].

3 Fast Vandermonde Solvers and Multi-Dimensional Extensions

The original paper [1] described an $O(n^2)$ algorithm for solving linear systems with Vandermonde matrices that took 1/3 operations less that the best known algorithm to that date. The motivation for work in this area was related to the next topic in our list: deferred corrections, where high order approximations to linear differential operators were required. The following paper with A. Bjorck [2] refined this algorithm and connected it to Newton's interpolation algorithm. An Algol code was included and tested, showing the remarkable stability of the new algorithm.

It is worth recalling that conditioning is problem related. It is quite well known that monomial basis and consequently Vandermonde matrices are fairly ill conditioned and this certainly affects Gaussian elimination. However, NJ Higham [4] proved that our algorithm is not so affected by this ill-conditioning if the defining parameters are appropriately chosen. As we

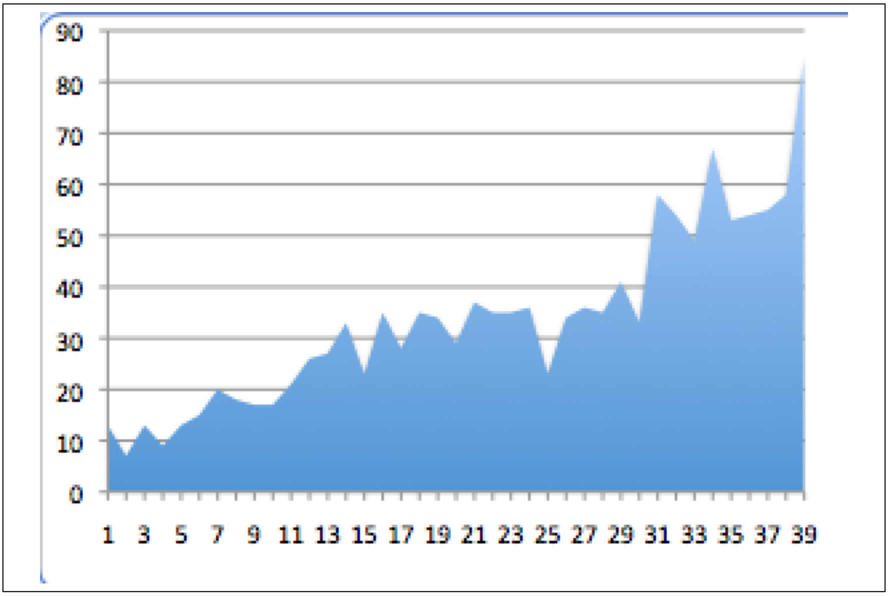

Figure 2.1: Citations/Year - Year 1 corresponds to 1979, as reported in Scholar.

showed experimentally in some approximation problems, the algorithm performs accurately for sizes that are large enough for most application, while Gaussian elimination blows up at low orders.

With this one-dimensional basis we went to consider higher dimensions in order to apply deferred corrections to partial differential equations. There are two standard ways to do this and we considered both:

1. Tensor products [8].

2. Approximations in simplices [5].

At the time, nothing had been done relating to the manipulation of tensor products on a computer. It existed though a solid theoretical basis of multilinear algebra and ample experience of physicists on the manipulation of tensors. We were interested on Kronecker products of matrices and that is what the paper addresses. It relates to our need of high order differentiation polynomial based formulas in many variables leading to Vandermonde matrices and their generalizations.

What we came up with was an economical algorithm to multiply a Kronecker product of matrices (with any number of factors) by a vector using only the factors and three loops. Observe that Kronecker products of matrices result in matrices where the orders are the product of the orders of the factors and they are highly redundant, so our algorithm takes into account this structure and works only with the factors: the large Kronecker product matrix is never formed.

The algorithm was rigorously described in terms of multilinear algebra and it consisted essentially of a backward recursion that reduced a problem with k factors to several with $k-1$ factors, eventually reaching dimension one, where we could apply the fast Vandermonde solver previously developed. The proof is heavy going and although we explicitly showed an implementation of the most important three dimensional case, it was still difficult to read. Carl de Boor recommended it for tensor products of splines, but after having to explain our paper several times he decided to write a simplified version [3].

The approximation in simplices is even more difficult. It was stated as some sub-matrices of Kronecker products and again a backward recursion provided the solution. But here there was an additional problem. While one dimensional Vandermonde's are nonsingular iff the defining parameters are different, this is not enough in more than one dimension. So, not only we had to consider the correct number of interpolant points, but we also had to find positions that lead to a nonsingular system for the determination of the weights.

It turns out that kth order polynomial approximation on simplices (triangles in 2D) corresponds to polynomials whose exponents sum to k (while for tensor products we use all the terms with exponents $\leq k$), which are $k*(k+1)/2$. On a triangle, the corresponding number of points can be obtained by taking k levels with $1, 2, ..., k$ points in each level. If these points are distinct within a level, our constructive algorithm proves that the set is well-posed and the linear system for determining the weights is solvable.

Incidentally, we feel that this could be a good tool for creating discontinuous polynomial elements of any order in any dimension for Galerkin type finite element solvers.

4 Deferred Corrections for Differential Equations

In our first paper on the subject [6] we offered a proof of the convergence of a one step difference correction in the sense of L Fox. From there on followed a series of papers, including my PhD Thesis (University of Wisconsin, 1967) that generalized and applied the idea to different problems, mostly boundary value problems for ordinary differential equations, but also including initial value problems and nonlinear elliptic problems. In [7] we considered a general functional analysis presentation that used asymptotic expansions to prove the convergence of an algorithm in which calculations of order k lead to higher order approximations in a recursive manner, thus providing the basis for an adaptive order algorithm fairly early in the game.

Later on we developed the idea of equidistributed meshes for BVPODE problems and that provided all the tools necessary to construct an adaptive boundary value problem solver for first order systems of nonlinear differential equations, similar to what existed for initial value problems. A family of public domain codes (PASVA) was developed combining these ideas at a time where satisfactory software for these type of problems was inexistent.

Deferred corrections can be simply described in the abstract functional analysis setting with the following elements:

- A general nonlinear differential equation problem: $F(x) = 0$.
- A low order discretization: $F_h^2(x_h^2) = 0$ (the exponent indicates the order and the subindex h indicates that we are working on a mesh).
- After obtaining x_h^2, a high order residual is calculated: $R_h^4 = D_h[F(x_h^2) - F_h^4(x_h^2)]$.
- The order of approximation is enhanced by a correction, using the original low order discretization and the higher order residual: $F_h^2(x_h^4) = -R_h^4$.

This process can be reiterated. Having an asymptotic expansion of the residual helps in deriving the higher order approximations, using the tools developed in the previous section, but it is not essential.

5 Seismic Ray Tracing, Geological Modeling and Inverse Problems

The last code of the PASVA series was PASVA4 [9], which included a number of additional features, such as interior discontinuities and additional algebraic parameters. This generality was required to attack efficiently the problem of two-point seismic ray tracing in seismology and oil prospecting that we started considering in 1976.

Given a model of the Earth, a source and a set of receivers, the seismic ray tracing problem consists of following the path of the energy from the source to each receiver as it travels through an inhomogeneous earth. A good model of the Earth can be described as an assemblage of different types of rocks, separated by contact surfaces. For our system we assumed that both the slowly varying rock properties within a block and the discontinuity surfaces were modeled by different types of smooth approximants, such as B-splines. This was acceptable given that the wavelengths involved only provided resolution to about 100 ms.

Mathematically, each ray is determined by solving a system of 7 ordinary differential equations, with initial, end and, because of the discontinuities, interior boundary conditions. Although the location of the discontinuities is not known, by an appropriate setting they can be included in the solution of the equations and that is why we need all the properties of PASVA4 for this problem.

However, this is only part of the history. In a practical application we will have millions of shots and receiver pairs, so that an elaborated algorithm needs to be devised to automate the process and make it efficient, including parallelization in large clusters. In fact, we were one of the first to introduce this paradigm in the oil industry [10], with actual implementations.

This becomes even more essential when the ray tracer is used in an inverse process. In fact, the seismic survey is used to infer the material properties and geometry of the rocks. Ray tracing is the forward simulation step on a guessed earth model. An iterative optimization is used to find the best model that fits the data obtained in the survey. The most straightforward data would be times of travel of different arrivals as obtained in the different receivers. This turns out to be far too complicated and human time consuming, so other more practical variants are used.

The actual phenomena that we are modeling is wave propagation in an elastic heterogeneous media. Ray tracing was used in the old days because there was not sufficient computational power to use full wave solvers. Nowadays, although still very expensive, full waveform inversion is the preferred approach. Unfortunately, the resulting optimization problem is very tricky and multimodal. It turns out, as we indicated earlier, that VARPRO helps to mollify the optimization and it is finding more applications in this area every day.

6 Least Squares Data Fitting With Applications

This is a book published in 2013 and that is receiving some attention. In a compact way travels over many topics in linear and nonlinear least squares, describing problems and methods of solution in a terse style, while connecting to relevant literature. In that respect is neither a conventional texbook nor an encyclopedia of the subject, but rather a consultation source. One of the main features are the chapters on applications, that include some large scale practical ones.

References

[1] On the construction of discrete approximations to linear differential expressions, V Pereyra, C. Ballester, Math. Comp. **21**: 297-302 (1967).

[2] Solution of Vandermonde systems of equations", Ake Bjorck, V. Pereyra, Math. Comp. **24**:893-904 (1970)

[3] Efficient computer manipulation of tensor products. C de Boor, TOMS **5**:173-182 (1979).

[4] Stability analysis of algorithms for solving confluent Vandermonde-like systems. Nicholas J Higham, SIAM Journal on Matrix Analysis and Applications **11**:23-41 (1990).

[5] Numerical differentiation and the solution of multidimensional Vandermonde systems, V Pereyra, G. Galimberti, Math. Comp. **24**:357-362 (1970).

[6] The difference correction method for nonlinear two point boundary value problems of class M, V Pereyra, Rev. U.M.A. **22**:184-201 (1965).

[7] Iterated deferred corrections for nonlinear operator equations, V Pereyra, Numer. Math. **10**:316-323 (1967).

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

[9] PASVA4: An ordinary boundary solver for problems with discontinuous interfaces and algebraic parameters, V Pereyra, M. Lentini, Mat. Apl. e Comp. **2**:103-118 (1983).

[10] A parallel network computer approach for 3-D geophysical modeling, V Pereyra, J.C. Meza and M. Koshy. SANDIA Report SAND92-8467J.UC-405. SEG Extended Abstracts (1992).

[11] Mesh selection for discrete solution of boundary problems in ordinary differential equations, V Pereyra, E. G. Sewell, Numer. Math. **23**:261-268 (1975).

[12] Exponential Data Fitting and its Applications, edited by V Pereyra, Godela Scherer (2010).

[13] Variable projections for separable nonlinear least squares problems since 2002. V Pereyra, G Scherer, in preparation (2018).