Conjugate Gradients and Related KMP Algorithms: The Beginnings*

Dianne P. O'Leary†

November 7, 1995

Abstract

In the late 1940's and early 1950's, newly available computing machines generated intense interest in solving "large" systems of linear equations. Among the algorithms developed were several related methods, all of which generated bases for Krylov subspaces and used the bases to minimize or orthogonally project a measure of error. These methods include the conjugate gradient algorithm and the Lanczos algorithm. We refer to these algorithms as *the KMP family* and discuss its origins, emphasizing research themes that continue to have central importance.

1 Introduction

The conjugate gradient algorithm is now close to 45 years old. As this algorithm and related ones have grown in importance and widespread use, some mythology has developed regarding the early history of the family. The purpose of this overview is to discuss the origins of this family of algorithms, emphasizing research themes that continue to have central importance in current research.

One difficulty in talking about these algorithms is in terminology. Our discussion concerns certain methods that generate a basis for a *Krylov subspace*, the subspace spanned by the vectors b, Ab, ..., $A^{k-1}b$, where A is a given matrix and b is a given vector. "Conjugate gradient family" is too restrictive, since it implies a notion of conjugacy that is lost in some of the nonsymmetric algorithms. "Krylov subspace family" is too broad, since it rightly includes stationary iterative methods such as SOR, and nonstationary methods such as the Chebyshev semi-iterative method. The "Lanczos family" appears to exclude Arnoldi-based algorithms. One characterization of these methods is that they use the Krylov subspace either to minimize an error function (e.g., conjugate gradients), to project the original operator (e.g., the Lanczos algorithm for computing eigenvalues), or to project the error in the solution into a progressively smaller-dimensional subspace (e.g., the Arnoldi method). For lack of a better term, we will use the phrase *KMP* family to denote those *K*rylov-subspace generating algorithms that either *M*inimize an error or *P*roject an operator or error. This includes the methods of Lanczos, Arnoldi, conjugate gradients, GMRES, CGS, QMR, and an alphabet soup of other algorithms.

The following sections discuss some of the key early steps in the development of the KMP family. A reader interested in recent developments could begin with [24], [3], and the papers in this volume.

*This work was supported by the National Science Foundation under Grant CCR-95-03126

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (oleary@cs.umd.edu).

2 The Beginnings

Researchers in the late 1940's and early 1950's were concerned with developing effective algorithms for solving basic linear algebra problems (systems of linear equations and eigenvalue computations) on the computing machines available after the end of World War II. These machines represented a revolutionary step forward, but had severe limits in power. Perhaps the most significant limitation was the amount of memory. A branch of the United States National Bureau of Standards (NBS), the Institute for Numerical Analysis (INA) within the National Applied Mathematics Laboratories, had been established in Los Angeles to study means by which important computational problems could be solved on such machines. The list of people associated with the INA includes Olga Taussky, John Todd, and others mentioned later in this section.

2.1 KMP Methods: Direct vs. Iterative

One myth concerning the origins of the conjugate gradient algorithm is that it was developed as a direct method rather than an iterative one. History does not provide much support for this view, however.

The study of iterative methods for eigenvalue problems and for linear systems was a central research theme at the INA, and a landmark paper was published by Cornelius Lanczos in 1950 in the *Journal of Research of the National Bureau of Standards* [36]. In this work, Lanczos developed a three-term recurrence relation for matrix polynomials (equivalently, for a basis for a Krylov subspace) and a biorthogonalization method for finding eigenvalues of nonsymmetric matrices.

Lanczos' idea was further developed by W. E. Arnoldi, who reinterpreted the algorithm and then derived a new one, reducing a general matrix to upper Hessenberg form [1]. Arnoldi presented the algorithm as iterative in nature, and discussed filtering the right-hand side before generating the Krylov sequence.

The earliest report of NBS activity on the conjugate gradient algorithm for solving systems of equations involving symmetric positive definite matrices was in an abstract for the Summer Meeting of the American Mathematical Society held in Minneapolis in September, 1951 [21]. This discussion of the three-term recurrence form of conjugate gradients is coauthored by George Forsythe, Magnus Hestenes, and J. Barkley Rosser. Hestenes wrote a technical report on this work in July, 1951 [31]. In it, he acknowledged discussions with "Forsythe, Lanczos, Paige, Rosser, Stein, and others." He specifically credited Forsythe and Rosser for the 3-term recurrence and L. J. Paige for the usual 2-term recurrence. The conjugate direction algorithm is somewhat older, discussed in a 1948 paper by Leslie Fox, H. D. Huskey, and Jim Wilkinson [22].

Meanwhile, Eduard Stiefel of E.T.H., Zurich, visited the INA, and presented a paper on the conjugate gradient algorithm at a workshop in August, 1951. His description of the "n-step iteration" was published in 1952 and noted the connection with the Lanczos (1950) work [49].

Hestenes and Stiefel decided to combine their efforts, and they published *the* conjugate gradient paper in 1952. They described conjugate gradients as "an iterative method that terminates" [32, p.410]. They devoted several pages to showing the relation between conjugate gradients and Gaussian elimination, but the bulk of the paper focused on the iterative properties, such as monotonicity of various error measures, methods for smoothing the initial residual, remedies for loss of orthogonality, the algebraic framework for preconditioning, and the relation to the Lanczos (1950) algorithm and continued fractions.

They also discussed determining the solution when the matrix A is rank deficient. They reported the use of the algorithm to solve 106 "difference equations" in 90 iterations. By 1958, researchers had solved Laplace's equation over a 10×10 grid in 11 Chebyshev iterations plus 2 conjugate gradient iterations [17].

Lanczos also published a paper in 1952 concerning iterative methods for linear systems [37]. He developed the use of his 1950 algorithm, and developed 2-term recurrences of biorthogonal polynomials.

The conjugate gradient iteration was extended to Hilbert space in the Ph.D. dissertation of R. M. Hayes, a U.C.L.A. student working at the INA [29]. Hayes established a linear convergence rate for general operators, with superlinear convergence if the operator is of the form identity plus a completely continuous operator. The roots of the idea of clustering eigenvalues to improve the convergence of conjugate gradients seem to date from here.

Clearly early researchers of the KMP family recognized the algorithms' usefulness as iterative methods and devoted considerable thought to *preconditioning* – either by rescaling the matrix or by filtering the initial error by the use of other iterative methods as Lanczos suggested in 1952 [37, p.45].

2.2 Assigning Credit

Considerable controversy has arisen regarding proper credit for the conjugate gradient algorithm. Lanczos's papers clearly focus on KMP algorithms for nonsymmetric matrices, while Hestenes and Stiefel limit their attention to Hermitian positive definite matrices. This extra restriction enabled the use of many additional important properties.

The claims of Lanczos, Hestenes, and Stiefel do not really overlap. Lanczos says in 1952 [37, p.53], "The latest publication of Hestenes [1951] and of Stiefel [1952] is closely related to the p, q algorithm of the present paper, although developed independently and from different considerations." Stiefel's claim in 1952 [49, p.23] was that, "After writing up the present work, I discovered on a visit to the Institute for Numerical Analysis (University of California) that these results were also developed somewhat later by a group there." Hestenes and Stiefel give a more complete citation in the 1952 paper [32, pp.409-410]:

> The method of conjugate gradients was developed independently by E. Stiefel of the Institute of Applied Mathematics at Zurich and by M. R. Hestenes with the cooperation of J. B. Rosser, G. Forsythe, and L. Paige of the Institute for Numerical Analysis, National Bureau of Standards. The present account was prepared jointly by M. R. Hestenes and E. Stiefel during the latter's stay at the National Bureau of Standards. The first papers on this method were given by E. Stiefel [1952] and M. R. Hestenes [1951]. Reports on this method were given by E. Stiefel and J. B. Rosser at a Symposium on August 23-25, 1951. Recently, C. Lanczos [1952] developed a closely related routine based on his earlier paper on eigenvalue problem. Examples and numerical tests of the method have been by R. Hayes, U. Hochstrasser, and M. Stein.

All of this supports Hestenes' memories in 1987 [30]: "I believe it was done in the following order. 1. Stiefel because he had carried out some large experiments which surely took place more than a month before he came to UCLA. I invented it within a month of his arrival. 2. Hestenes. 3. Lanczos. He is third because he would have been talking about it prior to my invention of the routine. I am sure that when he saw my paper he said to himself, 'I knew it all along'. The remarkable thing is that it took two years of study of iterative methods at INA before the cg-algorithm was devised."

It is clear that each of these people came upon the conjugate gradient algorithm from different considerations: Hestenes, from his long-time interest in conjugacy, dating back to a joint paper with G. D. Birkhoff in the 1930's; Stiefel from "simultaneous relaxation," adding a linear combination of several vectors to the current iterate; and Lanczos from polynomial recurrences. This multitude of roots for the KMP family is key to its richness of theoretical properties and its computational usefulness.

3 The 1960's

Another myth about the KMP family is that the algorithms were forgotten in the 1960's, discarded because they could not compete with Gauss elimination as a direct method. Although the algorithms were not much in favor among numerical mathematicians, they did achieve considerable success in applications in spectral analysis [6], lens design [19], geodesy [16], polar circulation [7], infrared spectral analysis [45], optimal control [48], collision theory [18], structural analysis [23], pattern recognition [39], power system load flow [53], optimal flight paths [5], nonrelativistic scattering [25], network analysis [35], and nuclear shell computation [47].

Some important work was also occurring in extending the usefulness of the KMP family. Wachspress showed the effectiveness of the algorithm preconditioned by alternating direction implicit methods (ADI) in solving discretizations of partial differential equations [52]. Golub and Kahan discussed the use of the Lanczos algorithm in computing the singular value decomposition [26]. Fletcher and Reeves extended the conjugate gradient algorithm to minimization of non-quadratic functions [20], opening an important area of research. Somewhat earlier, Davidon [14] had developed the first algorithm in the quasi-Newton family, but it had not yet been recognized as a relative of the conjugate gradient algorithm that generates the same sequence of iterates when applied to quadratic minimization.

Major steps in understanding the convergence behavior of the conjugate gradient algorithm were made by Kaniel [34] and Daniel [12]. Although both papers contained some errors [4, 13, 8], the standard bounds on the convergence rate for conjugate gradients, derived using Chebyshev polynomials, can be found here.

Thus the 1960's brought considerable progress in the use and understanding of the KMP family.

4 The 1970's

The key paper reviving the interest of numerical analysts in the investigation of KMP methods was a 1970 presentation, at an Oxford conference sponsored by the Institute of Mathematics and Applications, by John Reid of the U.K. Atomic Energy Research Establishment, Harwell [46]. Reid compared the numerical performance of several variants of the conjugate gradient algorithm and emphasized the algorithm's usefulness on well-conditioned problems.

This stimulated interest among a number of researchers, and activity in the area blossomed. It was clear that progress was needed on several fronts: preconditioners that would turn ill-conditioned problems into well-conditioned ones, stabilized forms of the algorithms, and extensions to broader problem classes.

The revival of interest in preconditioning produced the important paper of J. Meijerink and Henk van der Vorst [38], available in preprint form in the early 1970's. They developed an algorithm for computing an incomplete LU factorization of an M-matrix (as did Richard Varga in a 1960 paper [51]). This work inspired the hope of having a library of preconditioners that would apply to broad problem classes. Preconditioning was also discussed by Owe Axelsson [2], by Paul Concus, Gene Golub, and Dianne O'Leary [10], and for nonlinear problems by Jim Douglas and Todd Dupont [15].

Two important extensions of the conjugate gradient algorithm were given during this period. Paul Concus, Gene Golub, and Olof Widlund solved problems in which the Hermitian part of the matrix was positive definite and could be used as a preconditioner [9, 54]. Chris Paige and Michael Saunders showed how to compute iterates in case a matrix was indefinite [44], resulting in SYMMLQ and related algorithms.

Research on the eigenvalue problem also contributed to making the KMP algorithms more useful for solving linear systems. Chris Paige [43, 40, 41, 42], Beresford Parlett and W. Kahan [33], and Jane Cullum and Ralph Willoughby [11] all made important contributions to understanding the behavior of the Lanczos recursion under inexact arithmetic. Gene Golub, Richard Underwood, and Jim Wilkinson developed a block form of the Lanczos algorithm [28, 50] that later inspired the development of block KMP algorithms.

5 Closing Comments

Further information on the early history of the KMP family, as well as a much more complete bibliography, can be found in [27].

Perhaps the most important lesson to be learned in scanning the early literature on KMP methods is that the themes that motivated the earliest work continue to this day to be basic questions that have not been fully answered.

- The influence of inexact arithmetic led the researchers of the 1950's to develop and test various forms of the iterations, and researchers of the 1990's still discuss how to handle near-breakdown of the nonsymmetric iterations and the influence of loss of othogonality of the basis vectors.

- Important aspects of the convergence behavior of the conjugate gradient iteration were understood relatively early; important questions still remain open concerning its behavior on discretizations of ill-posed problems, even though Lanczos himself had a partial understanding [37]. Convergence behavior of the nonsymmetric iterations is not nearly as well understood.

- Preconditioning remains the key to effective use of the KMP family of algorithms, and the search for preconditioners effective on broad problem classes continues.

Future work in all of these areas will continue to be inspired by the fundamental questions raised by the earliest researchers.

6 Acknowledgements

I am grateful to Loyce Adams and Larry Nazareth for their excellent work in organizing the conjugate gradient workshop and for inviting this survey.

References

[1] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quarterly of Appl. Math., 9 (1951), pp. 17–29.

[2] O. Axelsson, *A generalized SSOR method*, BIT, 12 (1972), pp. 443–467.

[3] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the Solution of Linear Systems*, SIAM, Philadelphia, 1993.

[4] G. G. Belford and E. H. Kaufman,Jr., *An application of approximation theory to an error estimate in linear algebra*, Math. of Comp., 28 (1974), pp. 711–712.

[5] B. L. Bierson, *A discrete-variable approximation to optimal flight paths*, Astronautica Acta, 14 (1969), pp. 157–169.

[6] A. A. Bothner-By and C. Naar-Colin, *The proton magnetic resonance spectra of 2,3-disubstituted n-butanes*, J. of the ACS, 84 (1962), pp. 743–747.

[7] W. J. Campbell, *The wind-driven circulation of ice and water in a polar ocean*, J. of Geophysical Research, 70 (1965), pp. 3279–3301.

[8] A. I. Cohen, *Rate of convergence of several conjugate gradient algorithms*, SIAM J. Numer. Anal., 9 (1972), pp. 248–259.

[9] P. Concus and G. H. Golub, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, in Computing Methods in Applied Sciences and Engineering (2nd Internat. Symposium, Versailles 1975), Part I, R. Glowinski and J. L. Lions, eds., Springer Lecture Notes in Econ. and Math. Systems 134, New York, 1976, pp. 56–65.

[10] P. Concus, G. H. Golub, and D. P. O'Leary, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, in Sparse Matrix Computations, J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1976, pp. 309–332.

[11] J. Cullum and R. A. Willoughby, *Lanczos and the computation in specified intervals of the spectrum of large, sparse real symmetric matrices*, in Sparse Matrix Proceedings 1978, I. S. Duff and G. W. Stewart, eds., SIAM, Philadelphia, 1979, pp. 220–255.

[12] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967a), pp. 10–26.

[13] J. W. Daniel, *A correction concerning the convergence rate for the conjugate gradient method*, SIAM J. Numer. Anal., 7 (1970b), pp. 277–280.

[14] W. C. Davidon, *Variable metric method for minimization*, tech. rep., ANL-5990, Argonne National Laboratory, Argonne, Illinois, 1959.

[15] J. Douglas, Jr. and T. Dupont, *Preconditioned conjugate gradient iteration applied to Galerkin methods for a mildly-nonlinear Dirichlet problem*, in Sparse Matrix Computations, J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1976, pp. 333–348.

[16] H. M. Dufour, *Resolution des systemes lineaires par la methode des residus conjugues*, Bulletin Géodésique, 71 (1964), pp. 65–87.

[17] M. Engeli, T. Ginsburg, H. Rutishauser, and E. Stiefel, *Refined Iterative Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*, Birkhauser Verlag, Basel/Stuttgart, 1959.

[18] B. E. Eu, *Method of moments in collision theory*, J. Chem. Phys., 48 (1968), pp. 5611–5622.

[19] D. P. Feder, *Automatic lens design with a high-speed computer*, J. of the Optical Soc. of Amer., 52 (1962), pp. 177–183.

[20] R. Fletcher and C. M. Reeves, *Function minimization by conjugate gradients*, Computer J., 7 (1964), pp. 149–154.

[21] G. E. Forsythe, M. R. Hestenes, and J. B. Rosser, *Iterative methods for solving linear equations*, Bull. Amer. Math. Soc., 57 (1951), p. 480.

[22] L. Fox, H. D. Huskey, and J. H. Wilkinson, *Notes on the solution of algebraic linear simultaneous equations*, Quart. J. of Mech. and Appl. Math., 1 (1948), pp. 149–173.

[23] R. L. Fox and E. L. Stanton, *Developments in structural analysis by direct energy minimization*, AIAA J., 6 (1968), pp. 1036–1042.

[24] R. W. Freund, G. H. Golub, and N. M. Nachtigal, *Iterative solution of linear systems*, Acta Numerica, (1991), pp. 57–100.

[25] C. R. Garibotti and M. Villani, *Continuation in the coupling constant for the total K and T matrices*, Il Nuovo Cimento, 59 (1969), pp. 107–123.

[26] G. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal., 2(Series B) (1965), pp. 205–224.

[27] G. H. Golub and D. P. O'Leary, *Some history of the conjugate gradient and Lanczos algorithms: 1948-1976*, SIAM Review, 31 (1989), pp. 50–102.

[28] G. H. Golub, R. Underwood, and J. H. Wilkinson, *The Lanczos algorithm for the symmetric*

$Ax = \lambda Bx$ problem, tech. rep., Stanford University Computer Science Department, Stanford, California, 1972.

[29] R. M. Hayes, *Iterative methods of solving linear problems on Hilbert space*, in Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues, O. Taussky, ed., vol. 39 Applied Mathematics Series, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., 1954, pp. 71–103.

[30] M. Hestenes, *Private communcation*, (July 1, 1987).

[31] M. R. Hestenes, *Iterative methods for solving linear equations*, tech. rep., NAML Report 52-9, July 2, 1951, National Bureau of Standards, Los Angeles, California, 1951. Reprinted in *J. of Optimization Theory and Applications* 11 (1973), pp. 323-334.

[32] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409–436.

[33] W. Kahan and B. N. Parlett, *How far should you go with the Lanczos process?*, in Sparse Matrix Computations, J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1976, pp. 131–144.

[34] S. Kaniel, *Estimates for some computational techniques in linear algebra*, Math. of Comp., 95 (1966), pp. 369–378.

[35] K. Kawamura and R. A. Volz, *On the convergence of the conjugate gradient method in Hilbert space*, IEEE Trans. on Auto. Control, AC-14 (1969), pp. 296–297.

[36] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45 (1950), pp. 255–282.

[37] C. Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 33–53.

[38] J. Meijerink and H. A. V. der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comp., 31 (1977), pp. 148–162.

[39] G. Nagy, *Classification algorithms in pattern recognition*, IEEE Trans. on Audio and Electroacoustics, AU-16 (1968a), pp. 203–212.

[40] C. C. Paige, *Practical use of the symmetric Lanczos process with re-orthogonalization*, BIT, 10 (1970), pp. 183–195.

[41] ——, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, tech. rep., Ph. D. dissertation, University of London, 1971.

[42] ——, *Computational variants of the Lanczos method for the eigenproblem*, J. Inst. Maths. Applics., 10 (1972), pp. 373–381.

[43] ——, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Maths. Applics., 18 (1976), pp. 341–349.

[44] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[45] J. Pitha and R. N. Jones, *A comparison of optimization methods for fitting curves to infrared band envelopes*, Canadian J. of Chemistry, 44 (1966), pp. 3031–3050.

[46] J. K. Reid, *On the method of conjugate gradients for the solution of large sparse systems of linear equations*, in Large Sparse Sets of Linear Equations, Academic Press, New York, 1971, pp. 231–254.

[47] T. Sebe and J. Nachamkin, *Variational buildup of nuclear shell model bases*, Annals of Physics, 51 (1969), pp. 100–123.

[48] J. F. Sinnott, Jr. and D. G. Luenberger, *Solution of optimal control problems by the method of conjugate gradients*, in 1967 Joint Automatic Control Conference, Preprints of Papers, Lewis Winner, New York, 1967, pp. 566–574.

[49] E. Stiefel, *Über einige methoden der relaxationsrechnung*, Zeitschrift für angewandte Mathematik und Physik, 3 (1952), pp. 1–33.

[50] R. Underwood, *An iterative block Lanczos method for the solution of large sparse symmetric eigenproblems*, tech. rep., Ph.D. dissertation, Stanford University Computer Science Dept. Report STAN-CS-75-496, Stanford, California, 1975.

[51] R. S. Varga, *Factorization and normalized iterative methods*, in Boundary Problems in Differential Equations, R. E. Langer, ed., University of Wisconsin Press, Madison, 1960,

pp. 121–142.

[52] E. L. Wachspress, *Extended application of alternating direction implicit iteration model problem theory*, J. Soc. Industr. Appl. Math., 11 (1963), pp. 994–1016.

[53] Y. Wallach, *Gradient methods for load-flow problems*, IEEE Trans. on Power Apparatus and Systems, PAS-87 (1968), pp. 1314–1318.

[54] O. Widlund, *A Lanczos method for a class of nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 15 (1978), pp. 801–812.