SIAM REVIEW
Vol. 31, No. 1, pp. 50–102, March 1989

SOME HISTORY OF THE CONJUGATE GRADIENT AND LANCZOS ALGORITHMS: 1948-1976*

GENE H. GOLUB† AND DIANNE P. O'LEARY‡

Abstract. This paper gives some of the history of the conjugate gradient and Lanczos algorithms and an annotated bibliography for the period 1948-1976.

Key words. conjugate gradient algorithm, Lanczos algorithm, variable metric algorithms

AMS(MOS) subject classifications. 65F10, 65F15, 65H10, 65F03

1. Introduction. The conjugate gradient and Lanczos algorithms for solving linear systems of equations and eigenproblems represent very important computational innovations of the early 1950s. These methods came into wide use only in the mid-1970s. Shortly thereafter, vector computers and massive computer memories made it possible to use these methods to solve problems which could not be solved in any other way. Since that time, the algorithms have been further refined and have become a basic tool for solving a wide variety of problems on a wide variety of computer architectures. The conjugate gradient algorithm has also been extended to solve nonlinear systems of equations and optimization problems, and this has had tremendous impact on the computation of unconstrained and constrained optimization problems.

The purpose of this paper is to trace some of the history of the conjugate gradient and Lanczos algorithms during the period from their original development to their widespread application in the mid-1970s.

It is not the purpose of this paper to give the definitive derivation of the algorithms and their properties; for such information, the reader is referred to the references in the bibliography as well as more recent treatments such as *Matrix Computations* by G. H. Golub and C. F. Van Loan (The Johns Hopkins University Press, Baltimore, Maryland, 1983, Chapters 9 and 10). It is necessary, however, to establish notation to make the differences among the variants of the algorithm more clear. This is our first task.

The conjugate gradient algorithm can be thought of as a method for minimizing a function $\frac{1}{2}(x, Ax) - (x, b)$ where A is an $n \times n$ matrix (or an operator on a Hilbert space) and x and b are vectors in the domain and range spaces, respectively. The minimizer of this function satisfies the equation $Ax = b$ if A is self-adjoint and positive definite, so the algorithm provides a means for solving linear equations. It is characterized by the property of stepping at each iteration in the direction of the component of the gradient A-conjugate to the preceding step direction, and by the property of finite termination under exact arithmetic. The residual at step k can be expressed as the product of an optimal polynomial in A with the initial residual, and thus the distribution of eigenvalues is

* Received by the editors June 8, 1987; accepted for publication (in revised form) April 19, 1988.

† Computer Science Department, Stanford University, Stanford, CA 94305. This author's research was partly supported by National Science Foundation Grant DCR84-12314.

‡ Computer Science Department and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This author's research was partly supported by Air Force Office of Scientific Research Grant AFOSR-82-0078.

exploited; indeed, the method will converge in k or fewer iterations if the operator has k distinct eigenvalues. For simplicity, we will take the initial guess $x_0 = 0$, giving an initial residual, or negative gradient, of $r_0 = b$, and we take this as our first step direction p_0 as well. The algorithm is

$$x_{k+1} = x_k + \alpha_k p_k,$$

$$r_{k+1} = r_k - \alpha_k A p_k,$$

$$p_{k+1} = r_{k+1} + \beta_k p_k,$$

$$\alpha_k = \frac{(r_k, r_k)}{(p_k, Ap_k)}, \quad \beta_k = \frac{(r_{k+1}, r_{k+1})}{(r_k, r_k)}.$$

Variants of the algorithm arise from using different choices of the inner product, from computing the residual directly from its definition as

$$r_{k+1} = b - Ax_{k+1},$$

and from using different, but mathematically equivalent, formulas for the parameters, such as

$$\alpha_k = \frac{(r_k, p_k)}{(p_k, Ap_k)}, \quad \beta_k = -\frac{(r_{k+1}, Ap_k)}{(p_k, Ap_k)}.$$

An important variant can be derived by adding a preconditioning operator M to the formulation, applying the algorithm to the equivalent problem of minimizing $\frac{1}{2}(M^{1/2}y, AM^{1/2}y) - (M^{1/2}y, b)$, and then changing back to the variable $x = M^{1/2}y$. From one point of view, conjugate gradients is a convergence acceleration method for an underlying iteration $My_{k+1} = Ny_k + b$ where $A = M - N$, and this has been a very fruitful insight.

Another equivalent version of the algorithm is formed by eliminating the vectors p in the equations above, giving the three-term recurrence relation

$$x_{k+1} = x_{k-1} + \omega_{k+1}(r_k / \rho_k + x_k - x_{k-1}),$$

$$r_{k+1} = r_{k-1} + \omega_{k+1}(-Ar_k / \rho_k + r_k - r_{k-1}),$$

$$\omega_{k+1} = \left[1 - \frac{(r_k, r_k)}{(r_{k-1}, r_{k-1})} \frac{\rho_{k-1}}{\rho_k} \frac{1}{\omega_k}\right]^{-1}, \qquad \rho_k = \frac{(r_k, Ar_k)}{(r_k, r_k)},$$

with the definitions $x_{-1} = x_0$ and $\omega_0 = 0$.

The idea behind the Lanczos algorithm for determining eigenvalues of a matrix can be discussed easily using the three-term recurrence relation above. By forming a matrix R_k whose columns are the first k residual vectors normalized to length 1, we can derive the relation

$$AR_k = R_k T_k + r_k e_k^T,$$

where T_k is a tridiagonal matrix of dimension k whose elements can be computed from the ρ_j and ω_j values, and e_k is the kth unit vector. The residual vectors are mutually orthogonal, so a full n steps of the algorithm yield a similarity transformation of the matrix A into tridiagonal form if the residual does not become zero prematurely. The intermediate matrices T_k have interlacing eigenvalues, however, and even for small k, some eigenvalues of A may be well approximated. Lanczos used the recurrence relations as a convenient way of constructing characteristic polynomials of the matrices T_k, from which approximations to eigenvalues could be computed.

The algorithm can be extended to minimization of nonquadratic functions $f(x)$. In this case we define the residual vector r_k to be the negative gradient of f evaluated at x_k, and the matrix A, the Hessian matrix for f, changes at each iteration. The alternate formulas no longer give equivalent algorithms, and much study has been given to appropriate choices.

Our discussion does not even hint at the richness of the algorithms: acceleration procedures, convergence properties, and the interplay among the complementary views of the quadratic algorithm as a minimization procedure, as a linear system solver, and as a similarity transformation. The remainder of the paper, devoted to the history of this family of algorithms, focuses on discoveries of these ideas and others. In §2, we trace the early developments at the National Bureau of Standards. Some of the key developments involved in making the algorithms practical are summarized in §3. Section 4 gives information on the organization of the annotated bibliography, §5 is devoted to acknowledgments, and the bibliography and author index follow.

2. Early developments. The original development of this family of algorithms was carried out by a core of researchers including Cornelius Lanczos and Magnus Hestenes at the Institute for Numerical Analysis in the National Applied Mathematics Laboratories at the United States National Bureau of Standards (NBS) in Los Angeles, and Eduard Stiefel of the Eidg. Technische Hochschule Zürich. The first communication among the NBS researchers and Stiefel concerning this algorithm seems to have been at a Symposium on Simultaneous Linear Equations and the Determination of Eigenvalues, held at INA in August 1951, as discussed in Stiefel (1952). Further perspective on the interplay between the researchers at the National Bureau of Standards can be found in Forsythe, Hestenes, and Rosser (1951), Hestenes and Stiefel (1952), Rosser (1953), and Todd (1975). The National Bureau of Standards developments can be traced through internal quarterly project reports of the National Applied Mathematics Laboratories. The following is a condensation of some of the information in those reports. In some cases, work can be attributed to a single person, in others, only to a project directed by a manager or group of managers.

2.1 April - June, 1949.

Project: Determination of Characteristic Values of Matrices. Manager: C. Lanczos.

Lanczos was investigating eigenvalue algorithms in this and other projects and was preparing to write up the work.

Project: Approximate Solution of Sets of Arbitrary Simultaneous Algebraic Equations. Manager: C. Lanczos.

A "method for solving equations was investigated." The method is steepest descent applied to minimizing the residual. "At present the investigation is directed to the possibility of increasing the convergence of the successive reductions, by replacing A by $\bar{A} = \gamma I + A$ with suitably chosen γ."

This was the quarter in which Hestenes was hired (although in his position at the University of California in Los Angeles since 1947 he had already had contact with the NBS group), but there seems to be no explicit mention of his activities.

2.2 July, 1949 - June, 1951.

Lanczos seems to have been working on the eigenvalue algorithm and other things, and the Lanczos (1950) manuscript was submitted and accepted for publication in the *Journal of Research of the National Bureau of Standards* in the last quarter of 1949. Hestenes seems to have been working on the Hestenes-Karush project and variational problems, among other things. Both were participating in common seminars and project meetings. There is no mention of the conjugate gradient algorithm.

2.3 July - September, 1951.

Project: Solution of Sets of Simultaneous Algebraic Equations and Techniques for the Inversion and Iteration of Matrices. Managers: Forsythe, Hestenes, Lanczos, Motzkin, Rosser, Stein.

"Experimental work with the finite step methods described by M. R. Hestenes in a paper entitled 'Iterative methods for solving linear equations' was initiated by G. E. Forsythe and M. L. Stein." "Dr. E. Stiefel and Dr. M. R. Hestenes are writing a joint paper on extensions and implications of the methods described in the papers presented by J. Barkley Rosser, E. Stiefel and M. R. Hestenes at the Symposium on Simultaneous Linear Equations and the Determination of Eigenvalues held August 23-25, 1951, at the INA. ... For the extensive work of C. Lanczos on the solution of linear algebraic equations, see the description of the algorithm which he devised."

Project: Studies in Applied Mathematics. Managers: Lanczos, Rosser, van der Corput.

This describes the nucleus of his "Solution of systems of linear equations by minimized iterations" paper.

Project: Calculation of Eigenvalues, Eigenvectors, and Eigenfunctions of Linear Operators.

Experiments were conducted on applying Newton's method to the characteristic polynomial obtained from using conjugate gradients ($r-p$ version) on a symmetric positive definite matrix.

2.4 October - December, 1951.

Project: Solution of Sets of Simultaneous Algebraic Equations and Techniques for the Inversion and Iteration of Matrices. Managers: Forsythe, Hestenes, Lanczos, Motzkin, Rosser, Stein.

"The joint exposition by E. Stiefel and M. R. Hestenes of their 'finite iteration' procedure is almost finished."

Lanczos was working on a method to solve $Ax = b$ by finding the large eigenvalues and corresponding eigenvectors of $A' = \lambda_{max} I - A$, where λ_{max} is the largest eigenvalue of A. He applied a Chebyshev iteration to eliminate components of the residual corresponding to large eigenvalues of A, and resolved components corresponding to small eigenvalues using the eigeninformation for A'. The method was recommended for problems with multiple right-hand-sides.

Project: Variational Methods

Hestenes and Stein completed a study of algorithms for minimizing $(Ax-b)^* H(Ax-b)$. Hayes developed convergence theorems applicable to Rayleigh-Ritz and conjugate gradients for solving linear boundary value problems.

2.5 January - March, 1952.

The Lanczos (1952) manuscript was accepted for publication in the *Journal of Research of the National Bureau of Standards.*

2.6 April - June, 1952.

Project: Solution of Sets of Simultaneous Algebraic Equations and Techniques for the Inversion and Iteration of Matrices Managers: Forsythe, Hestenes, Lanczos, Lehmer, Motzkin.

The Hestenes and Stiefel (1952) manuscript was completed and accepted for publication. "It gives a full description of a wide class of methods of 'conjugate directions,' which includes as special cases both Gaussian elimination and the Hestenes-Lanczos-Stiefel method of 'conjugate gradients.' The latter is a finite iterative scheme which appears practically and theoretically to be a most attractive machine method."

"To test the stability of the conjugate gradient method in regard to rounding-off errors, symmetric systems of linear equations in 12 unknowns were solved on the IBM equipment. In order to know the eigenvalues, an orthogonal matrix was constructed, so that for any given set of eigenvalues a symmetric matrix could be found. The experiments were carried out on three machines [sic] with the following ratios of the largest to the smallest eigenvalue: 4.9, 100, 5000. The computing machine which was used for these experiments had a fixed decimal point and was allowed to work with 10 digits. By shifting, at least seven digits were carried through the computations. For the smallest ratio an answer with seven correct digits was reached in 11 steps. For the ratio 100 six correct digits in the 15th step were obtained. In the third case a good answer has not yet been found since the shifting caused considerable difficulties. The experiments showed the necessity of using floating operations for this method."

Experiments with the nonsymmetric formulas of matrices of dimension 8 gave convergence in less than or equal to 8 steps, even on singular problems, using the SWAC with 8 ½ digits in the arithmetic, obtaining 7-8 correct digits at termination.

Hayes was finishing work on the application of the "method given by E. Stiefel and M. R. Hestenes" to linear boundary value problems. Lanczos was working on solution of large-scale linear systems by Chebyshev polynomials.

2.7 July - September, 1952.

The conjugate gradient algorithm is often called the "Hestenes and Stiefel" algorithm in the reports of numerical experiments and other activities.

3. Key developments related to the algorithms.

3.1 The early papers. The first presentation of conjugate direction algorithms seems to be by Fox, Huskey, and Wilkinson (1948) who considered them as direct methods, and by Forsythe, Hestenes, and Rosser (1951), Hestenes and Stiefel (1952), and Rosser (1953) who discuss the NBS research. The conjugate gradient algorithm was described in Hestenes (1951), Lanczos (1952), Hestenes and Stiefel (1952), Stiefel (1952), Stiefel (1953), Curtiss (1954), Hestenes (1955), Hestenes (1956), and Stiefel (1957). Hestenes, Lanczos, and Stiefel clearly considered the algorithm to be a full n step direct method for certain problems, but also suggested its use as an iterative algorithm requiring fewer than n steps for well-conditioned problems and possibly more than n steps for ill-conditioned ones. Computational results for the conjugate gradient algorithm were presented in Stiefel (1953), Fischbach (1956), Stiefel (1958), and Engeli, Ginsburg, Rutishauser and Stiefel (1959), as well as in Hestenes and Stiefel (1952). Preconditionings, filtering, or change of inner product were considered in Fischbach (1956), Stiefel (1958), and in Engeli et al. (1959). Hayes (1954) and Altman (1959) discuss the conjugate gradient algorithm in Hilbert space.

Computations using the Lanczos algorithm were given in Lanczos (1950) and Rosser, Lanczos, Hestenes, and Karush (1951). Complete reorthogonalization was recommended by Brooker and Sumner (1956), Gregory (1958), and Wilkinson (1958).

3.2 Work in the 1960s. In this period, the conjugate gradient algorithm began to acquire a mixed reputation. It was still regarded as a standard algorithm, as evidenced by its inclusion in the Handbook Series (Ginsberg (1963)), in anthologies such as Beckman (1960), and in review articles in control (Paiewonsky (1965) and Westcott (1969)), chemistry (Wilde (1965)), and pattern recognition (Nagy (1968)).

Frank (1960) tested the algorithm on a matrix with eigenvalues related to the Chebyshev polynomials, the hardest case for conjugate gradients, and reported slow convergence. Applications in structural analysis by Livesley (1960) were unsuccessful, although Bothner-By and Naar-Colin (1962) were satisfied with their results in analysis of chemical spectra, and Feder (1962) recommended the algorithm in lens design. Campbell (1965) studied ocean circulation and Wilson (1966) solved optimal control problems with the aid of conjugate gradients. Pitha and Jones (1967) were other users of the algorithm.

Work was also being done on understanding the s-dimensional steepest descent algorithm, which produces the same sequence of iterates as the conjugate gradient algorithm restarted every s steps. References include Khabaza (1963), Forsythe (1968), and Marchuk and Kuznecov (1968).

Ideas that would eventually lead to successful preconditioned conjugate gradient algorithms were being developed. Dufour (1964) applied conjugate gradients to problems in geodesy and discussed several important ideas, including extensions to least squares problems with equality constraints, preconditioning, and elimination of half of the unknowns using a Schur complement. Varga (1960) suggested a sparse partial factorization of a matrix as a splitting operator for Chebyshev acceleration, and Dupont, Kendall, and Rachford (1968), Dupont (1968), and Stone (1968) also considered sparse factorizations. Other ideas related to preconditioning were discussed by Frank (1960) (polynomial filtering), Wachspress and Habetler (1960) (diagonal scaling), Habetler and Wachspress (1961) (Chebyshev acceleration of SSOR), Ehrlich (1964) (Chebyshev acceleration of block SSOR), Gunn (1964) (Chebyshev acceleration of ADI and ADI on a simpler operator), and Evans (1968) (Chebyshev acceleration of matrix splittings).

Wachspress (1963) used ADI as a preconditioner to conjugate gradients to obtain a very efficient algorithm.

Antosiewicz and Rheinboldt (1962), Nashed (1965), Daniel (1967), Horwitz and Sarachik (1968), Hestenes (1969), and Kawamura and Volz (1969) discussed the conjugate gradient algorithm in Hilbert space, and Kratochvil (1968) studied the algorithm for a class of operators on Banach spaces.

A very important advance in the solution of nonlinear equations and optimization algorithms was made in the development of methods which can solve many such problems effectively without evaluation of the derivative matrix. The first algorithms in this class, which reduce to conjugate gradients on quadratic functions, were presented by Feder (1962), Powell (1962), Fletcher and Powell (1963) building on work of Davidon (1959), Fletcher and Reeves (1964), Shah, Buehler, and Kempthorne (1964), and Broyden (1965). Polak and Ribiere (1969), Polyak (1969), Zoutendijk (1960), Sinnott and Luenberger (1967), Pagurek and Woodside (1968), Luenberger (1969), and Miele, Huang, and Heideman (1969) solved constrained problems using conjugate gradients.

The theory of Kaniel (1966) greatly increased the understanding of the convergence properties of the conjugate gradient and Lanczos methods.

Causey and Gregory (1961), Wilkinson (1962), Wilkinson (1965), and Yamamoto (1968) gave practitioners further insight into causes of failure in the Lanczos algorithm for nonsymmetric problems. The algorithm was used in applications problems in infrared spectra (Eu (1968)), scattering theory (Garibotti and Villani (1969)), network analysis (Marshall (1969)), and nuclear shell analysis (Sebe and Nachamkin (1969)).

3.3 The early 1970s. Although it is clear from the discussion above that the conjugate gradient and Lanczos algorithms were widely used in the 1960s, the numerical analysis community was not satisfied with the understanding of the algorithms or with their speed. Preconditioning techniques were not widely known (although much development had been done on splitting techniques), and it was in the early 1970s that key developments were made in making preconditioned algorithms practical.

The paper of Reid (1971) drew the attention of many researchers to the potential of the algorithm as a iterative method for sparse linear systems. It was a catalyst for much of the subsequent work in conjugate gradients.

The dissertation of Paige (1971), with publications as Paige (1972) and (1976), served the analogous purpose for the Lanczos algorithm by providing, among other things, the first step to an understanding of the loss of orthogonality of the Lanczos vectors, thus giving the key to the development of stable algorithms that did not require complete reorthogonalization. This made the Lanczos algorithm practical for large sparse problems by reducing storage and computation time. Developments along this line were made by Takahasi and Natori (1971-72) and Kahan and Parlett (1976).

Preconditioning techniques, although discussed in the 1960s, now became widely used. Axelsson (1972) suggested preconditioning conjugate gradients by a scaled SSOR operator. Other preconditionings were discussed by Evans (1973), Bartels and Daniel (1974), Chandra, Eisenstat, and Schultz (1975), Axelsson (1976), Concus, Golub, and O'Leary (1976), Douglas and Dupont (1976) and by Meijerink and van der Vorst (partial factorizations) in work that reached journal publication in 1977.

Paige and Saunders (1975) provided the first stable extension of the conjugate gradient algorithm to indefinite matrices. Concus and Golub (1976) considered a class of nonsymmetric matrices.

The block Lanczos algorithm was developed in Cullum and Donath (1974) and Underwood (1975).

Applications of the conjugate gradient algorithm, such as those by De and Davies (1970), Kamoshida, Kani, Sato, and Okada (1970), Kobayashi (1970), Powers (1973), Wang and Treitel (1973), Dodson, Isaacs, and Rollett (1976), and Konnert (1976) and of the Lanczos algorithm, such as those by Chang and Wing (1970), Emilia and Bodvarsson (1970), Weaver and Yoshida (1971), Whitehead (1972), Harms (1974), Hausman, Bloom, and Bender (1975), Ibarra, Vallieres, and Feng (1975), Platzman (1975), Cline, Golub, and Platzman (1976), and Kaplan and Gray (1976), also continued during this period. The Lanczos algorithm was rediscovered by Haydock, Heine, and Kelley (1972) and (1975) and applied to determining energy states of electrons.

3.4 Preconditioning. The *word* "preconditioning" is used by Turing (1948) and by then seems to be standard terminology for problem transformation in order to make solution easier. The first application of the word to the idea of improving the convergence of an iterative method may be by Evans (1968), and Evans (1973) and Axelsson (1974) apply it to the conjugate gradient algorithm. The *idea* of preconditioning the conjugate gradient algorithm is much older than this, as noted above, being perhaps implicit in the original conjugate gradient papers, somewhat more explicit in Hestenes (1956), and actually used in Engeli et al. (1959). Wachspress (1963) seems to be the first to use an iterative algorithm for discretized partial differential equations (ADI) as a preconditioner for the conjugate gradient algorithm.

4. The form of the annotated bibliography. The references are arranged alphabetically by author within year of publication. Each paper is given one or more "Classification Codes":

A applications
C conjugate direction/gradient algorithms for linear systems

E eigenproblems
L Lanczos algorithm for eigenproblems
N nonlinear conjugate gradient algorithms
P preconditioning
S matrix splittings

An author index follows the bibliography.

A reader should keep in mind several warnings. Because of publication delays, alphabetization, and mixing of journal publications with technical reports and dissertations, the bibliography is not completely chronological. The bibliography is not exhaustive; in particular, the references to nonlinear versions of the algorithm represent only a sample of the work done in this period, and references to literature in languages other than English are quite incomplete. The annotation for each paper only gives information relevant to the conjugate gradient and Lanczos algorithms and to preconditioning, and thus may not provide a complete summary of the work.

Quotations in the annotations are excerpts from the work itself. In works concerning applications to partial differential equations, the parameter h denotes the stepsize in the discretization.

5. Acknowledgments. Many people provided encouragement, references, corrections, and additions for this project, and we are very grateful to them. Included in this list are Åke Björck, Paul Concus, James Daniel, Howard Elman, L. Ehrlich, D. J. Evans, Roland Freund, S. Hammarling, B. P. Il'in, B. Kamgar-Parsi, David Kincaid, David Luenberger, Stephen Nash, Chris Paige, Tom Patterson, M. J. D. Powell, Axel Ruhe, Michael Saunders, Paul Saylor, Horst Simon, G. W. Stewart, John Todd, Henk van der Vorst, Eugene Wachspress, Olof Widlund, Ralph Willoughby, H. Woźniakowski, and David Young. The bibliography of Kammerer and Nashed (1972) was also very helpful. The project reports for the National Applied Mathematics Laboratory were obtained through the library of the Center for Applied Mathematics of the National Bureau of Standards (recently renamed the National Institute for Standards and Technology), in Gaithersburg, Maryland.

Annotated Bibliography

1948

1. /C/ Fox, L., H. D. Huskey, and J. H. Wilkinson (1948) "Notes on the Solution of Algebraic Linear Simultaneous Equations," *Quart. J. of Mech. and Appl. Math.* 1, pp. 149-173.

   Presents a "method of orthogonal vectors" as a direct method involving forming an A-conjugate set by Gram-Schmidt orthogonalization and expanding the solution vector in this basis.

2. /P/ Turing, A. M. (1948) "Rounding-off Errors in Matrix Processes," *Quart. J. of Mech. and Appl. Math.* 1, pp. 287-308.

   Introduces a quantitative measure of conditioning. "There is a very large class of problems which naturally give rise to highly ill-conditioned equations [an example being a polynomial fit in two dimensions with data in a small region]. In such a case the equations might be improved by a differencing procedure, but this will not necessarily be the case with all problems. Preconditioning of equations in this way will always require considerable liaison between the experimenter and the computer, and this will limit its applicability" (p. 299).

1950

3. /CEL/ Lanczos, C. (1950) "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators," *J. Res. Nat. Bur. Standards* 45, pp. 255-282.

Gives a polynomial expansion which can be used to solve the eigenvalue problem and develops recurrence relations for the polynomials. Notes that the recurrence is sensitive to round-off, and develops an alternate one, based on the principle of choosing the combination of previous vectors which makes the norm of the resulting vector as small as possible, achieving a three-term recurrence for the polynomials. Derives a bi-orthogonalization algorithm for finding eigenvalues of nonsymmetric matrices, and derives an algorithm with a single set of vectors for symmetric matrices. Uses the vectors to generate a set of polynomials which accumulate the characteristic polynomial of the original matrix. Recognizes that fewer than n steps may be needed to obtain a subset of the eigenvalues. Presents applications to eigenvalue problems in differential equations and integral operators. "The present investigation contains the results of years of research in the fields of network analysis, flutter problems, vibration of antennas, solution of systems of linear equations, encountered by the author in his consulting and research work for the Boeing Airplane Co., Seattle, Wash. The final conclusions were reached since the author's stay with the Institute for Numerical Analysis, of the National Bureau of Standards." "The literature available to the author showed no evidence that the methods and results of the present investigation have been found before. However, A.M. Ostrowski of the University of Basle and the Institute for Numerical Analysis informed the author that his method parallels the earlier work of some Russian scientists; the references given by Ostrowski are: A. Krylov, *Izv. Akad. Nauk SSSR 7, 491 to 539 (1931);* N. Luzin, *Izv. Akad. Nauk SSSR 7, 903 to 958 (1931).* On the basis of the reviews of these papers in the *Zentralblatt*, the author believes that the two methods coincide only in the point of departure. The author has not, however, read these Russian papers."

4. /EL/ Milne, W. E. (1950) "Numerical Determination of Characteristic Numbers," *J. Res. Nat. Bur. Standards* 45, pp. 245-254.

Approximates eigensystem of an ordinary differential equation by using a related partial differential equation and discretizing using finite differences. Derives a related trigonometric expansion whose roots determine the eigenvalues of the finite difference system. Relates the method to Lanczos (1950).

1951

5. /EL/ Arnoldi, W. E. (1951) "The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem," *Quarterly of Appl. Math.* 9, pp. 17-29.

Derives the nonsymmetric Lanczos algorithm as a Galerkin method with the left and right vectors bi-orthogonal, reducing the matrix to tridiagonal form and proposes its use as an iterative method for n steps or fewer. Derives a new algorithm with the left and right vectors equal and orthogonal, reducing the matrix to upper Hessenberg form. Suggests using several steps of the power method to get a starting vector for either algorithm.

6. /C/ Forsythe, G. E., M. R. Hestenes, and J. B. Rosser (1951) "Iterative Methods for Solving Linear Equations," *Bull. Amer. Math. Soc.* 57, p. 480.

(Abstract for Summer Meeting in Minneapolis, Sept. 4-7, 1951, quoted in its entirety) "Several iterative methods are given for solving the equations $Ax=b$, where A is a given matrix and b is a vector. These methods appear to be particularly adapted to high speed computing machines. They have the property that if there were no round-off error the solution would be obtained in at most n steps where n is the rank of A. In the singular case the least square solution is obtained. At each iteration the problem is started anew. Accordingly there is no accumulation of errors. In the hermitian case the method is based on the following result. If $A,B>0$ are hermitian matrices which commute then the system $b, Ab, \cdots, A^n b$ may be replaced by a set of B-orthogonal vectors by the algorithm $z_0=b$, $z_1=z_0-a_0Az_0$, $z_{i+1}=b_iz_i-a_iAz_i+c_{i-1}z_{i-1}$. (Received July 23, 1951)"

7. /C/ Hestenes, Magnus R. (1951) *Iterative Methods for Solving Linear Equations*, NAML Report 52-9, July 2, 1951, National Bureau of Standards, Los Angeles, California.

(Superceded by Hestenes and Stiefel (1952). Reprinted in *J. of Optimization Theory and Applications* 11 (1973), pp. 323-334.) "The methods presented are an outgrowth of discussions with Forsythe, Lanczos, Paige, Rosser, Stein, and others. For the positive Hermitian case it is convenient to separate the methods into two types. The first [the three-term recurrence form of conjugate gradients] is a method which is my interpretation of the suggestions made by Forsythe and Rosser. The second [the x-r-p version] is one which grew out of my discussion of the problem with Paige. The two methods are equivalent and yield the same estimates at each stage." If the first algorithm is used, recommends three-term recurrence for x with direct calculation of the residual vector. Gives alternate formulas for α and β and relates the parameters in the two methods. Shows finite termination, orthogonality of the residuals, and a bound on α. Discusses the case A positive semi-definite and recommends normal equations for nonsymmetric problems. Relates the algorithm to conjugate direction methods and constructs the formula for the inverse of A. Derives the property that the kth iterate of the algorithm minimizes a quadratic function over a k dimensional subspace. Gives a 4×4 example.

8. /EL/ Hestenes, Magnus R. and William Karush (1951a) "A Method of Gradients for the Calculation of the Characteristic Roots and Vectors of a Real Symmetric Matrix," *J. Res. Nat. Bur. Standards* 47, pp. 45-61.

Uses the iteration $x_{k+1}=x_k-\alpha p(x_k)$, with $p(x)=Ax-\mu(x)x$ and $\mu(x)$ the Rayleigh quotient. Analyzes the method using the Lanczos polynomials and the symmetric tridiagonal matrix similar to A.

9. /EL/ Hestenes, Magnus R. and William Karush (1951b) "Solutions of $Ax=\lambda Bx$," *J. Res. Nat. Bur. Standards* 49, pp. 471-478.

Extends the Hestenes-Karush (1950) algorithm to the generalized eigenvalue problem.

10. /C/ Hestenes, Magnus R. and Marvin L. Stein (1951) *The Solution of Linear Equations by Minimization*, NAML Report 52-45, December 12, 1951, National Bureau of Standards, Los Angeles, California.

(Reprinted in *J. of Optimization Theory and Applications* 11 (1973), pp. 335-359.) Proposes solving $Ax=b$ by minimizing $(b-Ax)^* H(b-Ax)$, where H is Hermitian positive definite. Studies iterations of the form $x_{k+1}=x_k+\alpha_i p_i$, and derives conditions on α_i and p_i to guarantee convergence. Notes that steepest descent, steepest descent with non-optimal α_i, Gauss-Seidel, SOR, block SOR, n-step methods such as those "investigated by Lanczos, Hestenes, and Stiefel", and other algorithms are special cases.

11. /EL/ Karush, W. (1951) "An Iterative Method for Finding Characteristic Vectors of a Symmetric Matrix," *Pacific J. Math.* 1, pp. 233-248.

Suggests an algorithm equivalent to taking s steps of the Lanczos algorithm, finding the minimizing eigenvector approximation, and iterating, making reorthogonalization less critical than in the Lanczos algorithm. References Lanczos (1950), but does not really draw the relationship.

12. /ELP/ Rosser, J. B., C. Lanczos, M. R. Hestenes, and W. Karush (1951) "Separation of Close Eigenvalues of a Real Symmetric Matrix," *J. Res. Nat. Bur. Standards* 47, pp. 291-297.

Solves a difficult 8×8 eigenvalue problem by the Lanczos (1950) algorithm by a "hand computer" in 100 hours (This method "seems best adapted for use by a hand computer using a desk computing machine.") and by the Hestenes and Karush (1951) method (fixed α) on an IBM Card-Programmed Electronic Calculator. ("Considerable time was spent by Karush in becoming familiar with the machine, so that it is difficult to say just how long the computation would require of an experienced operator. Probably 3 or 4 days would be ample.") Suggests polynomial preconditioning to increase the separation of the eigenvalues.

1952

13. /CP/ Hestenes, Magnus R. and Eduard Stiefel (1952) "Methods of Conjugate Gradients for Solving Linear Systems," *J. Res. Nat. Bur. Standards* 49, pp. 409-436.

"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."

For A symmetric and positive definite: develops conjugate gradients as an iterative method noting that x_{n+1} is often considerably better than x_n although earlier convergence may occur. Gives the $x-r-p$ version of the algorithm and notes that $||x-x^*||$ and $||x-x^*||_{A^{1/2}}$ are monotonically decreasing although the residual norm may oscillate. Gives formulas for obtaining characteristic roots from the recurrences. Proves algebraic and geometric properties of conjugate direction and conjugate gradient algorithms and references Fox, Huskey, and Wilkinson (1948). Gives an algorithm in which the residual norm is monotonic which modifies the x iterates from conjugate gradients. Gives some round-off analysis and recommends smoothing the initial residual. Gives an end correction procedure in case orthogonality is lost. Investigates other normalizations for the direction vectors.

For A symmetric semidefinite: notes that conjugate gradients can obtain a least squares solution.

For general A: uses A^*A-type algorithm.

Also presents the conjugate direction and conjugate gradient algorithms applied to $MAx=Mb$ and gives the examples $M=I$ and $M=A^*$. Shows that conjugate directions with unit vectors applied to a symmetric matrix is equivalent to Gauss elimination. Gives a conjugate direction example in which $\|x-x^*\|$ is monotonically increasing at intermediate steps. Describes a duality between orthogonal polynomials and n-dimensional geometry. Gives the 3-term recurrence relations for the residual polynomials. Notes the relation to the Lanczos (1950) algorithm for computing characteristic polynomials and that the conjugate gradient parameters can be computed by continued fraction expansion of a ratio of polynomials in A. Recommends computational formulas $\alpha=r^Tp/p^TAp$ and $\beta=-r^TAp/p^TAp$. Gives numerical examples and notes that the largest system yet solved involved 90 iterations on 106 difference equations.

14. /CL/ Lanczos, Cornelius (1952) "Solution of Systems of Linear Equations by Minimized Iterations," *J. Res. Nat. Bur. Standards* 49, pp. 33-53.

"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." "The present investigation is based on years of research concerning the behavior of linear systems, starting with the author's consulting work for the Physical Research Unit of the Boeing Airplane Company, and continued under the sponsorship of the National Bureau of Standards." Applies the Lanczos (1950) algorithm to solving nonsymmetric systems of linear equations by generating a double set of vectors (equivalently, polynomials) $p_k=p_k(A)b$ with leading coefficient 1 so that $\|p_k\|$ is minimal, and $q_k=q_k(A)b$ with constant coefficient 1 so that $\|q_k\|$ is minimal. Shows that the p and p^* sequences are biorthogonal and that the q sequences (saving the vectors) can be used to construct minimal residual solutions for the original system and others involving the same matrix. Advocates complete reorthogonalization or periodic restart to reduce the accumulation of error. Recommends scaling symmetric matrices to make diagonal elements 1 and nonsymmetric matrices to make column norms 1. If A has real nonnegative eigenvalues, recommends a two-phase algorithm, "purifying" the right hand side by Chebyshev polynomial iteration designed to damp out the components corresponding to large eigenvalues and then running the minimized iteration algorithm on the remainder. Notes that this can also be used to give smooth approximate solutions to nearly singular problems by terminating the second phase when a correction vector becomes too large.

15. /CLP/ Stein, Marvin L. (1952) "Gradient Methods in the Solution of Systems of Linear Equations," *J. Res. Nat. Bur. Standards* 48, pp. 407-413.

Reports on numerical experiments on a preconditioned form of steepest descent (but preconditioning is not used for acceleration). Also converts linear system to an eigenvalue problem and applies algorithm of Hestenes and Karush (1951).

16. /C/ Stiefel, Eduard (1952) "Über einige Methoden der Relaxationsrechnung," *Zeitschrift für angewandte Mathematik und Physik* 3, pp. 1-33.

(A version of this paper was presented at the NBS conference in August, 1951.) Surveys Jacobi- and Gauss-Seidel-type methods and steepest descent. Defines "simultaneous relaxation" as adding a linear combination of several vectors to the current guess. Notes that the parameters are easy to calculate if the directions are conjugate. Defines a "n-step iteration" (conjugate gradients) and notes that it can also be used to solve other linear systems with the directions already generated, to invert matrices, and to solve eigenvalue problems as Lanczos (1950) does. Uses the 5-point operator as a model problem, and provides numerical experiments. "Note added in proof: 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. An internal preliminary report for the National Bureau of Standards was given by M. R. Hestenes in August, 1951 (N.A.M.L. Report 52-9)."

17. /C/ Stiefel, Eduard (1952-53) "Ausgleichung ohne Aufstellung der Gaussschen Normalgleichungen," *Wissenschaftliche Zeitschrift der Technischen Hochschule Dresden* 2, pp. 441-442.

Proposes a conjugate direction algorithm for solving least squares problems which uses $A^T r$ as the initial direction, and keeps the directions AA^T-conjugate. (This algorithm later became known as the LSCG algorithm.)

1953

18. /P/ Forsythe, George E. (1953) "Solving Linear Algebraic Equations Can Be Interesting," *Bull. Amer. Math. Soc.* 59, pp. 299-329.

"With the concept of 'ill-conditioned' systems $Ax=b$ goes the idea of 'preconditioning' them. Gauss and Jacobi made early contributions to this subject [referring to the trick of adding an extra equation to a least squares system]. . . . A convenient means of preconditioning is to premultiply the system with a matrix B, so that one has to solve $BAx=Bb$" (p. 318). Gives two examples: $B=A^T$, giving the normal equations, and B being the operator generated by Gaussian elimination, so that BA is upper triangular.

19. /ELP/ Hestenes, Magnus R. (1953) "Determination of Eigenvalues and Eigenvectors of Matrices," in *Simultaneous Linear Equations and the Determination of Eigenvalues*, ed. L. J. Paige and Olga Taussky, Applied Mathematics Series 29, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., pp. 89-94.

Surveys methods used at NBS for symmetric eigenvalue computation: power algorithm, steepest descent on the Rayleigh quotient, two forms of a "generalization" of the Lanczos (1950) algorithm which uses a preconditioning matrix that commutes with A to obtain the recursion for the characteristic polynomial, and a block method. Also discusses the generalized eigenvalue problem.

20. /C/ Rosser, J. Barkley (1953) "Rapidly Converging Iterative Methods for Solving Linear Equations," in *Simultaneous Linear Equations and the Determination of Eigenvalues*, ed. L. J. Paige and Olga Taussky, Applied Mathematics Series 29, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., pp. 59-64.

Describes the conjugate direction algorithm in its general form, but states that the identity preconditioner is convenient. "Through the use of colloquia and discussion groups, nearly all scientific members of the Institute have made some sort of contribution to the problem. Accordingly, it is impossible to assign complete credit for the results disclosed herein to a single person or a few persons. However, certain members of the staff have given concentrated attention to the problem over an extended period and are primarily responsible for the results noted herein. In alphabetical order, these are G. E. Forsythe, M. R. Hestenes, C. Lanczos, T. Motzkin, L. J. Paige, and J. B. Rosser."

21. /EL/ Rutishauser, H. (1953) "Beiträge zur Kenntnis des Biorthogonalisierungs-Algorithmus von Lanczos," *Zeitschrift für angewandte Mathematik und Physik* 4, pp. 35-56.

Proves that there is a starting vector for the Lanczos algorithm which generates m vectors when the degree of the characteristic polynomial is m. Advocates a method of making the co-diagonal elements small when the eigenvalues are real, thus improving the convergence of algorithms to find eigenvalues of the bidiagonal matrix. Gives bounds for the eigenvalues. Relates the algorithm to a system of differential equations.

22. /P/ Shortley, George (1953) "Use of Tschebyscheff-Polynomial Operators in the Numerical Solution of Boundary-Value Problems," *J. of Appl. Phys.* 24, pp. 392-396.

Uses Chebyshev acceleration of the Jacobi algorithm for solving difference approximations to elliptic partial differential equations.

23. /CEL/ Stiefel, Eduard (1953) "Some Special Methods of Relaxation Technique," in *Simultaneous Linear Equations and the Determination of Eigenvalues*, ed. L. J. Paige and Olga Taussky, Applied Mathematics Series 29, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., pp. 43-48.

Presents the conjugate gradient algorithm as a minimization procedure and gives results of numerical experiments on Laplace's equation on a 3×3 grid and calculation of an Airy stress function for the profile of a dam (139 unknowns, 100 hours of computation on the Zurich Relay Calculator). Notes that the residuals are orthogonal and may be used to calculate eigenvalues. "The resulting procedure is similar to that suggested by Lanczos[1950]."

1954

24. /CP/ Curtiss, J. H. (1954) "A Generalization of the Method of Conjugate Gradients for Solving Systems of Linear Algebraic Equations," *Math. Tables and Aids to Comp.* 8, pp. 189-193.

Develops conjugate gradient algorithm for solving nonsymmetric systems by applying it to $BATA^TB^T$. Explains that $B=I, T=A^{-1}$ gives the Hestenes and Stiefel (1952) algorithm, $B=A^T, T=(A^TA)^{-1}$ gives the Hestenes and Stiefel least squares-type algorithm, and $B=I, T=I$ gives the Craig (1955) algorithm.

25. /C/ Forsythe, A. I. and G. E. Forsythe (1954) "Punched-Card Experiments with Accelerated Gradient Methods for Linear Equations," in *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, ed. Olga Taussky, Applied Mathematics Series 39, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., pp. 55-69.

Runs the Motzkin-Forsythe algorithm (steepest descent with an occasional accelerating step) on 6×6 examples, concluding that it is twice as fast as consistently under-relaxing steepest descent and much faster than steepest descent alone. Notes that the Hestenes and Stiefel (1952) methods seem to supercede these.

26. /C/ Hayes, R. M. (1954) "Iterative Methods of Solving Linear Problems on Hilbert Space," in *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, ed. Olga Taussky, Applied Mathematics Series 39, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., pp. 71-103.

Extends the conjugate direction algorithms to Hilbert space and proves linear convergence for the conjugate gradient algorithm for general operators and superlinear convergence for operators of the form $I+K$ where K is completely continuous.

27. /EL/ Stecin, I. M. (1954) "The Computation of Eigenvalues Using Continued Fractions," *Uspekhi matem. nauk* 9 No. 2(60), pp. 191-198.

Discusses Lyusternik's idea for finding the eigenvalues of a symmetric matrix or operator by transforming a series $c_0/z + c_1/z^2 + \cdots c_n/z^n$ (where $c_k=(A^kb,b)$) to a continued fraction whose coefficients are found by Chebyshev's method of moments. Discusses one particular algorithm for doing this, and notes that this is similar to the method of Lanczos orthogonalization.

28. /SP/ Young, David (1954) "On Richardson's Method for Solving Linear Systems with Positive Definite Matrices," *J. of Math. and Physics* 32, pp. 243-255.

Gives convergence results and optimal choice of acceleration parameters for Richardson's method. Compares with SOR and gradient methods.

1955

29. /C/ Craig, Edward J. (1955) "The *N*-Step Iteration Procedures," *J. of Math. and Physics* 34, pp. 64-73.

Discusses a set of conjugate direction methods, including the conjugate gradient algorithm, the algorithm built on A^*A-conjugate directions, and the (new) algorithm built on directions which are A^* times a set of orthogonal vectors. Notes that the last two algorithms can be used on symmetric or nonsymmetric matrices.

30. /P/ Forsythe, G. E. and E. G. Straus (1955) "On Best Conditioned Matrices," *Proceedings of the Amer. Math. Soc.* 6, pp. 340-345.

Studies the problem of minimizing the 2-norm condition number of T^*AT where A is Hermitian positive definite and T is in a class of regular linear transformations. As a special case, determines the optimal diagonal preconditioning matrix to be the one which makes the resulting diagonal elements equal to 1.

31. /CEP/ Hestenes, Magnus R. (1955) "Iterative Computational Methods," *Communications on Pure and Applied Mathematics* 8, pp. 85-96.

Gives a description of the conjugate gradient algorithm in general form and notes that it can be used to solve singular consistent problems. Discusses the eigenvalue problem, but not Lanczos' algorithm. "The terminology 'conjugate gradient' was suggested by the fact that p_i is the gradient of F, apart from a scale factor, on the linear manifold conjugate to $p_0, p_1, \cdots, p_{i-1}$, that is, orthogonal to $[Ap_0, \cdots, Ap_{i-1}]$."

32. /S/ Sheldon, John W. (1955) "On the Numerical Solution of Elliptic Difference Equations," *Math. Tables and Aids to Comp.* 9, pp. 101-112.

Presents the SSOR algorithm

33. /CL/ Stiefel, E. (1955) "Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme," *Comm. math. helv.* 29, pp. 157-179.

Surveys a family of algorithms for solving linear systems. Views steepest descent as Euler's method on the descent trajectory. Establishes one-to-one correspondence between the family of algorithms and sequences of polynomials satisfying $R_i(0)=1$. Gives the Lanczos polynomials and conjugate gradients as example of such a sequence, with polynomials orthogonal with respect to a discrete distribution function. Studies iterations based on the distribution $\lambda^{\alpha}(1-\lambda)^{\beta}$. Gives numerical results for Poisson equation with constant right-hand side.

1956

34. /EL/ Brooker, R. A. and F. H. Sumner (1956) "The Method of Lanczos for Calculating the Characteristic Roots and Vectors of a Real Symmetric Matrix," *Proc. Inst. Elect. Engrs.* B.103 Suppl., pp. 114-119.

Gives expository treatment of Lanczos algorithm. Recommends Jacobi method for small problems, Lanczos with reorthogonalization for large ones.

35. /CEL/ Crandall, Stephen H. (1956) *Engineering Analysis: A Survey of Numerical Procedures*, McGraw-Hill, New York.

Gives textbook description of conjugate gradient and Lanczos algorithms. "The usefulness of these methods for actual calculation is still being evaluated. . . . There is, however, no denying the mathematical elegance of the methods."

36. /ACP/ Fischbach, Joseph W. (1956) "Some Applications of Gradient Methods," in *Proceedings of the Sixth Symposium in Applied Mathematics (1953)*, McGraw-Hill, New York, pp. 59-72.

Discusses and experiments with conjugate gradients for computing the inverse of a matrix, for solving a two-point boundary value problem, and for solving a mildly nonlinear differential equation after a close approximation to the solution is obtained. "All those who have carried out computations by the method of conjugate gradients have observed that the $(N+1)$st step is usually better than the Nth and represents an improvement which overcomes rounding-off error. Frequently $2N$ steps are better than N. . . . One possible way of reducing the error growth is to change the metric (change definition of scalar product) so that the matrix is better conditioned."

37. /CP/ Hestenes, Magnus R. (1956a) "The Conjugate-Gradient Method for Solving Linear Systems," in *Proceedings of the Sixth Symposium in Applied Mathematics (1953)*, McGraw-Hill, New York, pp. 83-102.

Derives conjugate direction and conjugate gradient algorithms in general form, minimizing a function with an arbitrary inner product matrix, and having a preconditioning matrix. Notes that the conjugate gradient parameters can be bounded in terms of generalized eigenvalues. Discusses the standard conjugate gradient algorithm and the minimum error norm form. Shows that every n-step iterative method can be reproduced by a conjugate direction method. "From a mathematical point of view [the original Hestenes and Stiefel algorithm] represents the general case in the sense that every conjugate gradient algorithm can be reduced to this form by a change of variable or by a simple change of the original system to be solved." Notes that no essential changes are required to extend to Hilbert space.

38. /C/ Hestenes, Magnus R. (1956b) "Hilbert Space Methods in Variational Theory and Numerical Analysis," in *Proceedings of the International Congress of Mathematicians 1954* 3, North-Holland, Amsterdam, pp. 229-236.

Studies properties of quadratic forms in Hilbert space. Describes conjugate gradients as a minimization method on the error function, summarizing results of Hayes (1954).

39. /CEL/ Lanczos, Cornelius (1956) *Applied Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey.

Discusses use of the p,q Lanczos (1950) algorithm for finding eigenvalues and eigenvectors. Notes that the large eigenvalues are approximated quickly, and the small eigenvalues could be determined by "preliminary inversion of the matrix." Suggests use of Chebyshev polynomial transformation of the matrix to determine eigenvalues in an intermediate range.

1957

40. /C/ Stiefel, E. (1957) "Recent Developments in Relaxation Techniques," in *Proceedings of the International Congress of Mathematicians 1954* 1, North-Holland, Amsterdam, pp. 384-391.

Defines a "relaxation process" as one which reduces a measure of the error at each step. Notes that for symmetric positive definite matrices, Gauss-Seidel, Gauss elimination (considered as an iteration), and gradient methods are relaxation processes. Develops the optimal polynomial property for conjugate gradients.

1958

41. /AEL/ Gregory, R. T. (1958) "Results Using Lanczos' Method for Finding Eigenvalues of Arbitrary Matrices," *J. Soc. Industr. Appl. Math* 6, pp. 182-188.

Uses Lanczos (1950) algorithm for complex non-Hermitian matrices with double precision arithmetic, scaling of vectors, and full re-orthogonalization.

42. /CLP/ Lanczos, C. (1958) "Iterative Solution of Large-Scale Linear Systems," *J. Soc. Industr. Appl. Math* 6, pp. 91-109.

Discusses the effect of ill-conditioning and right-hand-side measurement errors on the accuracy of solutions to linear systems with symmetric coefficient matrices. Analyzes nonsymmetric ones through the symmetric system of size $2n$. Estimates largest eigenvalue by refinement of power method, and scales the matrix by it. Then applies iteration based on Chebyshev polynomials and matrix of dimension $n+2$.

43. /CEL/ Stiefel, Eduard L. (1958) "Kernel Polynomials in Linear Algebra and Their Numerical Applications," in *Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues*, Applied Mathematics Series 49, National Bureau of Standards, U.S. Government Printing Office, Washington, D.C., pp. 1-22.

Derives relaxation algorithms by considering various sets of polynomials with value 1 at zero. Recommends a two stage process for ill-conditioned systems: filter out error components corresponding to a large but clustered set of eigenvalues and then apply conjugate gradients to remove components corresponding to a few small eigenvalues. As an example, solves Laplace's equation with constant right-hand side on a 10×10 grid with 11 Chebyshev iterations on [2,8] and 2 conjugate gradient steps, getting 4 orders of magnitude reduction in the error. Recommends solving nonlinear systems by the change of variables $AA^*y=b$. Applies kernel polynomials to the problem of eigenvalue estimation, obtaining the Lanczos (1950) algorithm, among others. "As it stands, Lanczos' algorithm can only be successful for low-order matrices with nicely separated eigenvalues. For larger matrices the rounding-off errors destroy quickly the orthogonality of the vectors. As in solving linear equations, it is necessary to find for such matrices a suitable *combination of the methods* available." Discusses polynomial transformations to emphasize certain ranges of the spectrum, and applies the Lanczos algorithm to the transformed matrix. Discusses the generation of orthogonal polynomials by the quotient-difference algorithm, including the variant corresponding to the Lanczos algorithm.

44. /AEL/ Wilkinson, J. H. (1958) "The Calculation of Eigenvectors by the Method of Lanczos," *Computer J.* 1, pp. 148-152.

Uses reorthogonalization on symmetric Lanczos and reorthogonalization plus double precision on unsymmetric Lanczos. Notes that the latter is a very powerful algorithm.

1959

45. /C/ Altman, Mieczyslaw (1959) "On the Convergence of the Conjugate Gradient Method for Non-Bounded Linear Operators in Hilbert Space," in *Approximation Methods in Functional Analysis*, Lecture Notes, California Institute of Technology, pp. 33-36.

Proves convergence of conjugate gradients for a self-adjoint, positive definite linear operator, with domain a dense linear space, satisfying $(Au,u) \geq k(u,u)$ for some positive constant k and all u in the domain.

46. /N/ Davidon, W. C. (1959) *Variable Metric Method for Minimization*, Report ANL-5990, Argonne National Laboratory, Argonne, Illinois.

Derives a method (the DFP method, further developed in Fletcher and Powell (1963)) meant to "improve the speed and accuracy with which the minima of functions can be evaluated numerically" compared to such methods as conjugate gradients, steepest descent, and Newton-Raphson. Proposes guessing the inverse Hessian matrix (symmetric and positive definite), and generating a search direction equal to this matrix times the negative gradient. Uses either this direction plus an orthogonal correction, or a line search along this direction, to determine the next iterate. Then modifies the inverse Hessian approximation according to the Quasi-Newton condition using a rank-one or rank-two update. Discusses the initial Hessian approximation and the incorporation of linear constraints. Discusses in an appendix a simplified rank-one updating procedure.

47. /ACP/ Engeli, M., Th. Ginsburg, H. Rutishauser, and E. Stiefel (1959) *Refined Iterative Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*, Birkhauser Verlag, Basel/Stuttgart.

Stiefel: Solves self-adjoint partial differential equations by variational formulation, not by differential equation itself.

Rutishauser: Surveys gradient methods (Richardson one and two, steepest descent, Frankel (second order Richardson), Chebychev, hypergeometric relaxation, conjugate gradients, conjugate residual). Considers "combined methods," including:

1. conjugate gradients with Chebyshev (attributed to Lanczos (1952)): smooth the residual with Chebyshev polynomial over range of high eigenvalues, then apply conjugate gradients. It is noted that the high eigenvalues are "reactivated" by conjugate gradients, and the method is not recommended.

2. "Replace the system given by another system with the same solution but with a coefficient matrix A of smaller P-condition number." (P-condition number = condition number in the 2-norm.) Polynomial conditioning is explicitly considered and is attributed to Stiefel (1958) in the case of eigencomputation. The iterations are called "inner" for the polynomial and "outer" for conjugate gradients, and Chebyshev-conjugate gradients is recommended.

Notes that methods such as those of Richardson and Frankel can also be used for eigencomputations. Gives the conjugate gradient tridiagonal matrix in non-symmetric 3-term form. Notes round-off difficulties, and recommends proceeding more or less than n steps, as long as the residual remains small. Recommends comparing the approximations from two tridiagonal matrices (same initial vector, different number of steps, or different initial vector) to determine convergence. Also discusses determining the eigenvalues of the original matrix from the conjugate gradient-Chebyshev method.

Ginsburg: Gives results of numerical experiments. For a finite difference problem with 70 variables, needs approximately 10 conjugate gradient iterations with preconditioning by a 10th-order Chebyshev polynomial. Compares with steepest descent, conjugate gradient, and other methods on this and other examples.

Engeli: Surveys relaxation methods.

Conclusions: For moderate condition problems, use relaxation. For bad conditioning, use conjugate gradients or conjugate gradients-Chebyshev with recursive residuals. Recommends conjugate gradients over conjugate gradients-Chebyshev unless some low eigenvalues are needed.

48. /AC/ Läuchli, Peter (1959) "Iterative Lösung und Fehlerabschätzung in der Ausgleichsrechnung," *Zeitschrift für angewandte Mathematik und Physik* 10, pp. 245-280.

Develops conjugate gradients and other relaxation methods for overdetermined linear systems. Notes that finding the point in an n-dimensional subspace of R^m (spanned by the columns of C) which is closest to a point l is equivalent to solving $C^TCx=C^Tl$, but that the problem can also be formulated in terms of a basis B for the null space of C^T, representing x implicitly as $B^Tx+b=0$ (for some vector b) and avoiding normal equations. Uses Synge's method of the hypercircle to find upper and lower bounds on the sum of squared residuals. Notes that the inverse matrix can be constructed by an update at each conjugate gradient step. Provides numerical examples.

1960

49. /C/ Beckman, F. S. (1960) "The Solution of Linear Equations by the Conjugate Gradient Method," in *Mathematical Methods for Digital Computers*, ed. Anthony Ralston and Herbert S. Wilf, Wiley, New York.

Derives conjugate gradients as a conjugate direction method including flow chart, comments on error analysis, etc.

50. /CP/ Frank, Werner L. (1960) "Solution of Linear Systems by Richardson's Method," *J. Assoc. Comput. Mach.* 7, pp. 274-286.

Follows Stiefel (1958) in using Chebyshev acceleration on a partial interval; then applies conjugate gradients. Tests the algorithm on a 50×50 matrix tri(-1,2,-1) with (1,1) element modified to 1. Needs 20 conjugate gradient iterations (instead of the theoretical termination in 5) to get 5 digits of accuracy; requires the full 50 if conjugate gradients is used alone. Required 46 conjugate gradient iterations to solve 5 point difference equations for $n=361$.

51. /P/ Householder, A. S. and F. L. Bauer (1960) "On Certain Iterative Methods for Solving Linear Systems," *Numer. Math.* 2, pp. 55-59.

Discusses "methods of projection" which iterate $x=x+Yp$, where the columns of Y span a subspace and p is chosen so that the error decreases. Notes that steepest descent and relaxation techniques both fit into this framework, but does not mention conjugate gradients.

52. /AC/ Livesley, R. K. (1960) "The Analysis of Large Structural Systems," *Computer J.* 3, pp. 34-39.

Tries to apply conjugate gradients to an ill-conditioned system in structural analysis. Finds conjugate gradients ineffective because it requires n matrix multiplications and thus n tape scans, and "rounding errors show a tendency to build up to such an extent that the solution after N steps is often a worse approximation to the correct solution than the starting point." "The method was therefore abandoned in favour of an elimination process."

53. /EP/ Osborne, E. E. (1960) "On Pre-Conditioning of Matrices," *J. Assoc. Comput. Mach.* 7, pp. 338-345.

Constructs a sequence of diagonal similarity transformations to scale an irreducible matrix to increase the smallest eigenvalue relative to the matrix norm so that the eigensystem can be more easily determined.

54. /PS/ Varga, Richard S. (1960) "Factorization and Normalized Iterative Methods," in *Boundary Problems in Differential Equations*, ed. Rudolph E. Langer, University of Wisconsin Press, Madison, pp. 121-142.

"The main purpose of this article is to introduce a class of iterative methods which depend upon the direct solution of matrix equations involving matrices more general than tridiagonal matrices." Assumes that A is Stieltjes. Introduces the idea of regular splitting. Given a regular splitting, accelerates by overrelaxation, Chebyshev semi-iteration, Peaceman-Rachford (1955) algorithm, or Douglas-Rachford (1956) algorithm. Suggests normalizing factors to have unit diagonal for computational efficiency. Discusses effectiveness of successive line overrelaxation. Proposes approximate factorization of A, keeping the factors as sparse as A (the algorithm that has come to be known as incomplete Cholesky factorization with no extra diagonals). Shows that this yields a regular splitting for the 5-point operator.

55. /PS/ Wachspress, E. L. and G. J. Habetler (1960) ''An Alternating-Direction-Implicit Iteration Technique,'' *J. Soc. Industr. Appl. Math.* 8, pp. 403-424.

''Conditions'' a matrix by diagonal scaling before applying ADI.

56. /CN/ Zoutendijk, G. (1960) *Methods of Feasible Directions*, Elsevier, Amsterdam.

Uses conjugate directions to construct a finitely terminating quadratic programming algorithm.

1961

57. /AEL/ Causey, R. L. and R. T. Gregory (1961) ''On Lanczos' Algorithm for Tridiagonalizing Matrices,'' *SIAM Rev.* 3, pp. 322-328.

Discusses biorthogonal reduction to tridiagonal form. Distinguishes between fatal and nonfatal instances when the inner product between the left and right vectors vanish.

58. /S/ D'Yakonov, E. G. (1961) ''An Iteration Method for Solving Systems of Finite Difference Equations,'' *Soviet Math. Dokl.* 2, pp. 647-650.

(Dokl. Akad. Nauk SSSR 138, pp. 271-274.) Analyzes the iteration $Mx_{k+1}=Mx_k-\omega(Ax_k-b)$ where A is a finite difference approximation to an elliptic operator over the unit square and M represents several iterations of the ADI operator for the Laplacian. Gives a work estimate of $O(n \ln^2 n^{-1/2}) \ln \varepsilon$ to solve the problem with precision ε.

59. /PS/ Golub, Gene H. and Richard S. Varga (1961) ''Chebyshev Semi-Iterative Methods, Successive Overrelaxation Iterative Methods, and Second Order Richardson Iterative Methods, Parts I and II,'' *Numer. Math.* 3, pp. 147-156, 157-168.

Compares the rates of convergence of the three iterative methods of the title. Proposes applying Chebyshev acceleration to two-cyclic matrices resulting from block preconditionings, such as those derived from the block SOR splitting of Arms, Gates, and Zondek (*SIAM J.* 4, 1956, pp. 220-229). Gives applications to partial difference equations.

60. /PS/ Habetler, G. J. and E. L. Wachspress (1961) ''Symmetric Successive Overrelaxation in Solving Diffusion Difference Equations,'' *Math. of Comp.* 15, pp. 356-362.

Uses Chebyshev acceleration on Sheldon's SSOR algorithm (*J. Assoc. Comput. Mach.* 6, 1959, pp. 494-505). Shows SSOR not effective in diffusion calculations in nuclear reactor theory if the grids are too irregular. Gives algorithm to estimate parameters.

61. /C/ Martin, D. W. and G. J. Tee (1961) ''Iterative Methods for Linear Equations with Symmetric Positive Definite Matrix,'' *Computer J.* 4, pp. 242-254.

Surveys stationary iterative methods, steepest descent, and conjugate gradients including previous numerical results. Concludes that ''no single method is to be recommended for universal applications.''

62. /S/ Oliphant, Thomas A. (1961) "An Implicit, Numerical Method for Solving Two-Dimensional Time-Dependent Diffusion Problems," *Quarterly of Appl. Math.* 19, pp. 221-229.

Proposes an iterative method for nine-point finite difference approximations, using a partial factorization of the difference matrix as a splitting. Applies the algorithm to linear and nonlinear problems.

63. /EL/ Rollett, J. S. and J. H. Wilkinson (1961) "An Efficient Scheme for the Co-diagonalization of a Symmetric Matrix by Givens' Method in a Computer with a Two-level Store," *Computer J.* 4, pp. 177-180.

Notes that the resulting bidiagonal matrix for their algorithm is the same as that from the Lanczos (1950) algorithm.

64. /EL/ Strachey, C. and J. G. F. Francis (1961) "The Reduction of a Matrix to Codiagonal Form by Eliminations," *Computer J.* 4, pp. 168-176.

Notes that the Lanczos (1950) method is equivalent to an elimination method for reduction of a Hessenberg matrix to tridiagonal form.

1962

65. /C/ Antosiewicz, Henry A. and Werner C. Rheinboldt (1962) "Numerical Analysis and Functional Analysis," in *Survey of Numerical Analysis*, ed. John Todd, McGraw-Hill, New York, pp. 485-517 (Ch. 14).

Presents conjugate directions for linear self-adjoint positive definite operators on Hilbert space and proves a convergence rate.

66. /AC/ Bothner-By, Aksel A. and C. Naar-Colin (1962) "The Proton Magnetic Resonance Spectra of 2,3-Disubstituted n-Butanes," *J. of the ACS* 84, pp. 743-747.

Analyzes chemical spectra by solving a least squares problem with conjugate gradients.

67. /ACN/ Feder, Donald P. (1962) "Automatic Lens Design with a High-Speed Computer," *J. of the Optical Soc. of Amer.* 52, pp. 177-183.

Suggests conjugate gradients or DFP methods, among others, to minimize a merit function in lens design.

68. /S/ Oliphant, Thomas A. (1962) "An Extrapolation Procedure for Solving Linear Systems," *Quarterly of Appl. Math.* 20, pp. 257-265.

Generalizes the method of Oliphant (1961) to five-point operators, and allows partial factorizations of a modified difference matrix.

69. /C/ Petryshyn, W. V. (1962) "Direct and Iterative Methods for the Solution of Linear Operator Equations in Hilbert Space," *Trans. AMS* 105, pp. 136-175..

Derives minimum error method and, from it, other algorithms. Does not use the extra matrices for preconditioning.

70. /N/ Powell, M. J. D. (1962) "An Iterative Method for Finding Stationary Values of a Function of Several Variables," *Computer J.* 5, pp. 147-151.

Proposes a method which, given a starting point x_0, finds a minimizer in one direction, x_1, then minimizes in the $n-1$ dimensional hyperplane through x_1 orthogonal to the first direction, giving x_2. Then the minimizer is on the line between x_0 and x_2. The method is "not unlike the conjugate gradient method of Hestenes and Stiefel (1952)."

71. /EL/ Wilkinson, J. H. (1962) "Instability of the Elimination Method of Reducing a Matrix to Tri-Diagonal Form," *Computer J.* 5, pp. 61-70.

Relates the Lanczos (1950) algorithm to Hessenberg's method (1941 Ph.D. thesis) applied to a lower Hessenberg matrix, reducing it to tridiagonal form.

1963

72. /CEL/ Faddeev, D. K. and V. N. Faddeeva (1963) *Computational Methods of Linear Algebra,* W. H. Freeman and Co., San Francisco, California.

(Translated by Robert C. Williams from 1960 publication of State Publishing House for Physico-Mathematical Literature, Moscow.) Discusses in Chapter 4 the "method of orthogonalization of successive iterations" for finding eigenvalues of matrices, which, in the symmetric case, is the Lanczos (1950) algorithm. Discusses in Chapter 6 how to continue the algorithm for symmetric and nonsymmetric matrices in case it terminates in fewer than n steps. Discusses the use of the "A-minimal iteration algorithm," the "A-biorthogonal algorithm," steepest descent, s-dimensional steepest descent, and conjugate direction algorithms for solving linear systems.

73. /N/ Fletcher, R. and M. J. D. Powell (1963) "A Rapidly Convergent Descent Method for Minimization," *Computer J.* 6, pp. 163-168.

Derives the Davidon-Fletcher-Powell (DFP) algorithm for minimizing non-quadratic functions and accumulating an approximate Hessian matrix. References Hestenes and Stiefel (1952).

74. /CL/ Fridman, V. M. (1963) "The Method of Minimum Iterations with Minimum Errors for a System of Linear Algebraic Equations with a Symmetric Matrix," *USSR Comp. Math. and Math. Phys.* 2, pp. 362-363.

Derives a conjugate gradient method (from the Lanczos perspective) which minimizes the 2-norm of the error over the subspace $Ar^{(0)}, A^2r^{(0)}, \cdots$.

75. /C/ Ginsburg, Theo (1963) "The Conjugate Gradient Method," *Numer. Math.* 5, pp. 191-200.

(The Handbook Series Linear Algebra conjugate gradient algorithm.) Uses the 3-term recurrence version of the conjugate gradient algorithm.

76. /C/ Khabaza, I. M. (1963) "An Iterative Least-Square Method Suitable for Solving Large Sparse Matrices," *Computer J.* 6, pp. 202-206.

Proposes the s-dimensional steepest descent algorithm applied to minimization of the norm of the residual for solving linear systems. Does not recompute the parameters in subsequent iterations unless the residual begins to increase. Notes superiority to conjugate gradients and SOR on some test problems.

77. /ACPS/ Wachspress, Eugene L. (1963) "Extended Application of Alternating Direction Implicit Iteration Model Problem Theory," *J. Soc. Industr. Appl. Math.* 11, pp. 994-1016.

Uses ADI applied to the model problem as a preconditioner for conjugate gradients applied to more general problems. Gives some discussion of convergence rate as a function of mesh spacing. References Lanczos (1952) rather than Hestenes and Stiefel. References Engeli et al. (1959) for other examples of "compound iteration."

1964

78. /ACP/ Dufour, H. M. (1964) "Resolution des Systemes Lineaires par la Methode des Residus Conjugues," *Bulletin Géodésique* 71, pp. 65-87.

Derives the minimum residual and conjugate gradient algorithms and proposes their use for symmetric positive definite systems, for linear least squares problems, for least squares subject to equality constraints, and for systems resulting from block elimination of a 2×2 block matrix, leading to a Schur complement of the form $C-B^*A^{-1}B$ as the matrix in the problem. Discusses preconditioning when an approximate inverse is available. Applies the method to problems in geodesy.

79. /PS/ Ehrlich, Louis W. (1964) "The Block Symmetric Successive Overrelaxation Method," *J. Soc. Industr. Appl. Math.* 12, pp. 807-826.

Uses Chebyshev acceleration on block SSOR. Estimates rate of convergence and gives numerical results.

80. /N/ Fletcher, R. and C. M. Reeves (1964) "Function Minimization by Conjugate Gradients," *Computer J.* 7, pp. 149-154.

Generalizes conjugate gradients to nonquadratic functions by adding line searches and by taking the current gradient to be the current residual. Quadratic termination is obtained without evaluating or approximating the Hessian matrix.

81. /S/ Gunn, James E. (1964a) "The Numerical Solution of $\nabla . a \nabla u = f$ by a Semi-Explicit Alternating-Direction Iterative Technique," *Numer. Math.* 6, pp. 181-184.

Proposes and analyzes the iteration $Mx_{n+1}=Mx_n-\omega(Ax_n-b)$ where M is one step of the Peaceman-Rachford ADI iteration for the discretization of the desired operator $\nabla . a \nabla$ and the domain is rectangular. Obtains a work estimate of $O(h^{-2}\log h^{-1}\log \varepsilon^{-1})$ to reduce the error by ε.

82. /S/ Gunn, James E. (1964b) "The Solution of Elliptic Difference Equations by Semi-Explicit Iterative Techniques," *SIAM J. Numer. Anal.* 2(Series B), pp. 24-45.

Proposes and analyzes the iteration $Mx_{n+1}=Mx_n-\omega(Ax_n-b)$ where M is one step of the Peaceman-Rachford ADI iteration (variable ω) for the discrete Laplacian operator (i.e., not the matrix A), the elliptic operator is not necessarily symmetric, and the domain is rectangular. Uses Chebyshev acceleration and second-order Richardson and obtains an improved convergence result over Gunn (1964). Applies the algorithm to mildly nonlinear problems.

83. /CEL/ Householder, Alston S. (1964) *The Theory of Matrices in Numerical Analysis*, Blaisdell Publishing Co., New York.

"The Lanczos algorithm is well known in the theory of orthogonal polynomials, but Lanczos (1950) seems to have been the first to apply it to the reduction of matrices (p.28, Dover edition)." Develops Lanczos tridiagonalization in matrix form; discusses Lanczos polynomials.

84. /EL/ Parlett, Beresford (1964) "The Development and Use of Methods of LR type," *SIAM Rev.* 6, pp. 275-295.

Notes that Henrici observed that the first diagonal of the QD scheme can be found by the Lanczos (1950) algorithm; thus, QD links the power method to Lanczos' method.

85. /CN/ Powell, M. J. D. (1964) "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives," *Computer J.* 7, pp. 155-162.

Proposes an algorithm which uses n line searches per iteration to generate a new direction. Shows that, for a quadratic function, the directions are conjugate. Proposes a modification in case the n line search directions become linearly dependent. Gives numerical examples.

86. /C/ Pyle, L. Duane (1964) "Generalized Inverse Computations Using the Gradient Projection Method," *J. Assoc. Comput. Mach.* 11, pp. 422-428.

Notes that a Gram-Schmidt-based algorithm for computing generalized inverses is a conjugate direction method if the matrix is square.

87. /N/ Shah, B. V., R. J. Buehler, and O. Kempthorne (1964) "Some Algorithms for Minimizing a Function of Several Variables," *J. Soc. Industr. Appl. Math.* 12, pp. 74-92.

Introduces Partan, a method with quadratic termination in $2n-1$ steps or fewer, which generates conjugate directions. Includes preconditioning matrix in the formulation.

1965

88. /CN/ Broyden, C. G. (1965) "A Class of Methods for Solving Nonlinear Simultaneous Equations," *Math. of Comp.* 19, pp. 577-593.

Develops a family of algorithms based on satisfying the quasi-Newton condition and using rank-one or rank-two updates to the approximate derivative matrix at each iteration. Proposes three update formulas. Proposes either a step-size of one, or using the norm of the residual in a criterion for termination of the line search. Gives an Algol program and numerical results on ten test problems.

89. /AC/ Campbell, William J. (1965) "The Wind-Driven Circulation of Ice and Water in a Polar Ocean," *J. of Geophysical Research* 70, pp. 3279-3301.

Solves nonlinear equations with 780 variables using conjugate gradients on a linear system at each iteration.

90. /CN/ Fletcher, R. (1965) "Function Minimization without Evaluating Derivatives - A Review," *Computer J.* 8, pp. 33-41.

Reviews conjugate direction methods of a 1962 paper of Smith, Powell (1964) and a 1964 paper of Davies, Swann, and Campey.

91. /L/ Golub, G. and W. Kahan (1965) "Calculating the Singular Values and Pseudo-Inverse of a Matrix," *SIAM J. Numer. Anal.* 2(Series B), pp. 205-224.

Uses Lanczos' observation that the singular values of a matrix are the eigenvalues of a matrix of dimension $n+m$ with zeros on the block diagonal and A and A^* off the diagonal. Generates the bidiagonal form from Householder transformations or from the Lanczos (1950) algorithm.

92. /CN/ Nashed, M. Z. (1965) "On General Iterative Methods for the Solutions of a Class of Nonlinear Operator Equations," *Math. of Comp.* 19, pp. 14-24.

Gives a class of iterative methods for operators in Hilbert space and shows that conjugate gradients and others are first-order approximations to these methods.

93. /AN/ Paiewonsky, Bernard (1965) "Optimal Control: A Review of Theory and Practice," *AIAA J.* 3, pp. 1985-2006.

Surveys control problems and nonlinear optimization methods.

94. /CEL/ Vorobyev, Yu V. (1965) *Method of Moments in Applied Mathematics*, Gordan and Breach Science Pub., New York.

(Translated from the Russian by B. Seckler.) Discusses the Lanczos algorithm for generating an orthogonal basis and and a sequence of orthogonal polynomials for symmetric matrices. Notes its use in finding eigenvalues, and discusses Lyusternik's idea of converting a certain infinite series to a continued fraction in order to use moments to determine the coefficients of the same orthogonal polynomials that Lanczos used. References L. A. Lyusternik, "Solution of Linear Algebraic Problems by the Method of Continued Fractions," *Transactions of a Seminar on Functional Analysis*, No. 2, Voronezh (1956), pp. 85-90. Notes that the "point of departure" for Lanczos (1950) and (1952) and for Hestenes and Stiefel (1952) and Lyusternik was "the Chebyshev-Markov classical scalar problem of moments for the quadratic functional (Ax,x)," and that "the methods of Lanczos and Lyusternik were subsequently extended to completely continuous self-adjoint operators" by Karush (1952) and Stecin (1954).

95. /AN/ Wilde, D. J. (1965) "A Review of Optimization Theory," *Indust. and Eng. Chem.* 57 no. 8, pp. 19-31.

Mentions conjugate gradients and other methods.

96. /CEL/ Wilkinson, J. H. (1965) *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford.

Advocates use of Lanczos (1950) algorithm with double precision arithmetic and complete reorthogonalization. Restarts with different initial vectors if the size of the new vector deteriorates in the nonsymmetric case.

1966

97. /EN/ Bradbury, W. W. and R. Fletcher (1966) "New Iterative Methods for Solution of the Eigenproblem," *Numer. Math.* 9, pp. 259-267.

Uses conjugate gradient and DFP algorithms to solve the generalized symmetric eigenproblem by minimizing the Rayleigh quotient. Notes that the line searches can be performed exactly. Renormalizes at each step to keep the infinity norm of the iterate equal to one. Reports faster convergence with conjugate gradients except on very ill-conditioned problems, but both methods are slower than QR if many eigenvalues are desired.

98. /PS/ D'Yakonov, Ye. G. (1966) "The Construction of Iterative Methods Based on the Use of Spectrally Equivalent Operators," *USSR Comp. Math. and Math. Phys.* 6, No. 1, pp. 14-46.

(Zh. vÿchisl. Mat. mat. Fiz. 6, No. 1, pp. 12-34.) Uses spectrally equivalent operators in a Richardson iterative algorithm and analyzes convergence.

99. /CELP/ Kaniel, Shmuel (1966) "Estimates for Some Computational Techniques in Linear Algebra," *Math. of Comp.* 95, pp. 369-378.

Develops convergence bounds for conjugate gradients in terms of Chebyshev polynomials and the condition number of the matrix. Develops bounds for Lanczos (1950) method eigenvalues in terms of condition number and separations. Notes that results extend to Hilbert space. Results corrected in Belford and Kaufman (1974).

100. /EL/ Lehmann, N. J. (1966) "Zer Verwendung optimaler Eigenwerteingrenzungen bei der Lösung symmetrischer Matrizenaufgaben," *Numer. Math.* 8, pp. 42-55.

Develops a previous idea of using a set of Rayleigh quotients to estimate eigenvalues to the special case where the test vectors are those from the Lanczos (1950) recursion and determines inclusion intervals for the largest. Applies the algorithm to *tri* (−1,2,−1) for n =30. Gets good estimates for 4 eigenvalues after 8 iterations.

101. /N/ Mitter, S., L. S. Lasdon, and A. D. Waren (1966) "The Method of Conjugate Gradients for Optimal Control Problems," *Proc. IEEE* 54, pp. 904-905.

Notes that the Fletcher-Reeves (1964) method also applies in function space.

102. /AN/ Pitha, J. and R. Norman Jones (1966) "A Comparison of Optimization Methods for Fitting Curves to Infrared Band Envelopes," *Canadian J. of Chemistry* 44, pp. 3031-3050.

Concludes that DFP is more effective than a nonlinear conjugate gradient method.

103. /CEP/ Wachspress, Eugene L. (1966) *Iterative Solution of Elliptic Systems and Applications to the Neutron Diffusion Equations of Reactor Physics*, Prentice-Hall, Englewood Cliffs, New Jersey.

In Chapter 5, derives the Lanczos (1950) algorithm and "combined" algorithms (e.g., Lanczos-Chebyshev) in a way similar to Engeli et al. (1959). Notes that the algorithms can be applied to a product of two symmetric matrices. Derives the Chebyshev algorithm for real eigenvalues and for eigenvalues bounded by an ellipse in the complex plane. Discusses Lanczos' eigenvalue algorithm with initial filtering and with a polynomial in A as the operator. In Chapter 6, discusses premultiplication of the linear system by a matrix, and applying the Lanczos or Chebyshev algorithm to the transformed system. Uses ADI preconditioning as an example. Gives a rate of convergence estimate for the model problem ADI preconditioned algorithm. In Chapter 9, derives a multigrid algorithm, relating the idea of contracting the basis to Lanczos projection, and performs numerical experiments indicating improvement over the Golub-Varga two-cyclic version of the Chebyshev algorithm and over SOR.

104. /AN/ Wilson, Robert (1966) "Computation of Optimal Controls," *J. of Math. Anal. and Applics.* 14, pp. 77-82.

Changes a constrained optimization problem to an unconstrained dual problem, decomposes it into subproblems, and applies conjugate gradients.

105. /AN/ Young, P. C. (1966) "Parameter Estimation and the Method of Conjugate Gradients," *Proc. IEEE* 54, pp. 1965-1967.

Uses Mitter, Lasdon, Waren (1966) version of Fletcher-Reeves (1964) algorithm for real-time process parameter estimation. "Unfortunately, the excellent characteristics of the conjugate gradients approach . . . are not maintained as the level of additive noise is increased. Considerable data averaging or 'smoothing' becomes necessary even for low noise levels, and this tends to destroy the real-time nature of the algorithm."

1967

106. /CN/ Broyden, C. G. (1967) "Quasi-Newton Methods and Their Application to Function Minimization," *Math. of Comp.* 21, pp. 368-381.

Further develops the algorithms in Broyden (1965), focusing on rank-two updates, function minimization, and linear systems with symmetric matrices.