

Methods of Conjugate Gradients for Solving Linear Systems

The advent of electronic computers in the middle of the 20th century stimulated a flurry of activity in developing numerical algorithms that could be applied to computational problems much more difficult than those solved in the past. The work described in this paper [1] was done at the Institute for Numerical Analysis, a part of NBS on the campus of UCLA [2]. This institute was an incredibly fertile environment for the development of algorithms that might exploit the potential of these new automatic computing engines, especially algorithms for the solution of linear systems and matrix eigenvalue problems. Some of these algorithms are classified today under the term *Krylov Subspace Iteration*, and this paper describes the first of these methods to solve linear systems.

Magnus Hestenes was a faculty member at UCLA who became associated with this Institute, and Eduard Stiefel was a visitor from the Eidgenössischen Technischen Hochschule (ETH) in Zürich, Switzerland. At this time, there were two commonly used types of algorithms for solving linear systems. The first, like Gauss elimination, modified a tableau of matrix entries in a

systematic way in order to compute the solution. These methods were finite, but required a rather large amount of computational effort with work growing as the cube of the number of unknowns. The second type of algorithm used “relaxation techniques” to develop a sequence of iterates converging to the solution. Although convergence was often slow, these algorithms could be terminated, often with a reasonably accurate solution estimate, whenever the human “computers” ran out of time.

The ideal algorithm would be one that had finite termination but, if stopped early, would give a useful approximate solution. Hestenes and Stiefel succeeded in developing an algorithm with exactly these characteristics, the method of conjugate gradients.

The algorithm itself is beautiful, with deep connections to optimization theory, the Pad table, and quadratic forms. It is also a computational gem, the standard algorithm used today to solve large sparse systems of equations, involving symmetric (or Hermitian) positive definite matrices, for which matrix modification methods are impractical.



Fig. 1. Wolfgang Wasow (left) and Magnus Hestenes (right).



Fig. 2. Eduard Stiefel [From *Zeitschrift für angewandte Mathematik und Physik* **30**, 139 (1979); used with permission].

The INA group was quite collegial, and many of the ideas in this paper have roots in joint discussions among many participants. Hestenes and Stiefel summarize the development this way [1, 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 by 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 [1950]. Examples and numerical tests of the method have been by R. Hayes, U. Hoshstrasser, and M. Stein.”

The papers referred to in this summary are those of Stiefel [3], Hestenes [4], and Lanczos [5,6], the last of which is discussed elsewhere in this volume.

It is interesting to hear two distinct voices in this paper of Hestenes and Stiefel. Hestenes came to the

algorithm from the point of view of variational theory and optimal control. In 1936 he had developed an algorithm for constructing a set of mutually conjugate basis vectors, but was advised by a Harvard professor that it was too obvious for publication [7]. Yet this background, plus discouraging numerical experience by George Forsythe in using the steepest descent algorithm for solving linear systems, was grist for the mill in the development of the conjugate gradient algorithm. Stiefel, on the other hand, had a strong orientation toward relaxation algorithms, continued fractions, and the qd-algorithm, and he developed conjugate gradients from this viewpoint.

The technical contributions of the paper begin in Section 3, with a rather terse algebraic presentation of the formulas of the conjugate gradient algorithm. Motivation comes later, with the observation that the recurrences lead to a sequence of approximations that converge monotonically in the sense of reducing an error function. At the same time, a sequence of polynomials can be constructed in order to find the eigen-system of the matrix. In the next section, the authors explain that the algorithm is a special case of a conjugate direction algorithm, and it is this property that yields finite termination. Algebraic properties are further developed in Section 5, which presents some alternate computational formulas. Sections 6 and 7 are more geometric, discussing optimization properties of the solution estimates. Section 8 turns to practical considerations: will round-off errors on electronic computers render this theoretically beautiful algorithm useless? The rather conservative analysis led to the conclusion that although round-off certainly hurts the algorithm, reasonable precautions and end corrections could overcome this in most cases. Section 9 explored different normalization options. In Section 10, “extensions” were proposed; in particular, the authors discussed how to solve nonsymmetric problems by considering the normal equations, and how to choose a preconditioning matrix in order to make the method applicable to broader classes of matrices. Section 11 discusses the construction of conjugate bases, and in Sections 12 and 13, the authors note that there is a close relation between conjugate gradients and Gauss elimination. Other relations are explored in later sections: continued fractions and the theory of orthogonal polynomials (Sections 14, 15, 17, and 18), and eigenvalue computations (Section 16).

The paper concludes with some numerical examples demonstrating the usefulness of the algorithm on a matrix of dimension 4 using exact arithmetic, a well-conditioned matrix of dimension 6 with some rounding error, and an ill-conditioned matrix of dimension 3, in which the conjugate gradient solution was about as



Fig. 3. Researchers associated with the NBS Institute for Numerical Analysis (1950). From left to right: Mark Kac, Edward J. McShane, J. Barkley Rosser, Aryeh Dvoretzky, George E. Forsythe, Olga Taussky-Todd, Wolfgang R. Wasow, and Magnus R. Hestenes.

accurate as that from Gauss elimination. The authors remind us how difficult numerical computations were in the middle of the 20th century: the algorithm had been used to solve 106 difference equations on the Zuse computer at ETH (with a sufficiently accurate answer obtained in 90 iterations); to solve systems as large as dimension 12 on an IBM “card programmed calculator”; and to solve small systems on the SWAC (Standards Western Automatic Computer), which had only 256 words of memory [2, Appendix C].

Hestenes and Stiefel’s paper remains the classic reference on this algorithm, explaining the use of the algorithm as both a finitely terminating method and as a way of obtaining an approximate solution if halted before termination. It is such a rich paper that it contains the roots of virtually all of the advances in this area up to the present time: for example, preconditioning to accelerate convergence of the algorithm, variations that can be used to solve systems of equations involving nonsymmetric matrices, and evaluation of computational variants in order to choose the most numerically stable one.

The algorithm garnered considerable early attention but went into eclipse in the 1960s as naive implementations were unsuccessful on some of the ever-larger problems that were being attempted. Work by John Reid [8] in the early 1970s drew renewed attention to the algorithm, and since then it has been an intense topic of research. Today it is the standard algorithm for solving linear systems involving large, sparse, symmetric (or Hermitian) matrices. Strategies for constructing preconditioning matrices, which reduce the number of iterations, remain a major research theme, as well as understanding the convergence properties of some non-symmetric variants. Some of the early history of the algorithm is discussed in [9].

Despite the fact that the algorithm has become a standard topic in numerical textbooks, the original paper is still widely read and cited; *Science Citation Index* lists over 800 citations between 1983 and 1999, and it is clear that this algorithm, and the non-symmetric variants developed to complement it, will remain the standard algorithms for solving sparse linear systems for the foreseeable future.

The conjugate gradient algorithm was but one claim to greatness for these remarkable and multifaceted researchers.

Magnus R. Hestenes [10] was born in Brice lyn, Minnesota, in 1906. His undergraduate work was at St. Olaf College, and his graduate work at the University of Wisconsin and the University of Chicago. His first faculty appointment was at Chicago, but he left in 1947 to accept a professorship at UCLA, where he taught until his retirement. He had 34 Ph.D. students and was a well-loved advisor and teacher, known for his nurturing kindness toward his very junior colleagues. Administrative duties over the years included chairing the Mathematics Department, directing the university's computing center, and serving as vice president of the American Mathematical Society. He also held appointments with the Rand Corporation, the Institute for Defense Analyses, and the IBM Watson Research Center. His association with NBS's INA lasted from 1949 to 1954, when the INA was transferred from NBS to UCLA. His best known works include many publications on the problem of Bolza, a famous paper on quadratic forms in Hilbert space [11], and the conjugate gradient paper with Stiefel. Hestenes remained scientifically active until his death in 1991, concentrating in his later years on the method of multipliers and continuing to write and publish.

Eduard Stiefel [12] was born in 1909 in Zürich, Switzerland. He spent virtually his entire career at the Eidgenössischen Technischen Hochschule (ETH) in Zürich, first as a student of mathematics and physics, and then, following his habilitation degree in 1943, as a professor. His early work was in topology, eventually studying the geometry and topology of Lie groups. 1948 was a turning point for him, however. He founded the Institut für Angewandte Mathematik (Institute for Applied Mathematics) at ETH, in collaboration with Heinz Rutishauser and Ambros P. Speiser. Stiefel was a visionary who realized the enormous significance of the new computing technology and the impact it would have on mathematics and science. When he discovered in 1949 that a major computing engine, the Z4 of the German designer Konrad Zuse, was sitting in the small alpine village of Neukirchen, Germany, he traveled there and arranged for the machine to be rented and moved to ETH. Zuse, isolated by wartime secrecy, had independently developed computing technology that in many ways was superior to that available in the U.S. at the time [13]. Stiefel's initiative made ETH the first European university with an electronic computer, putting it in the forefront of numerical computation and computer

science. This led to several breakthrough developments by him and his colleagues, including the qd algorithm, the programming language ALGOL, and the conjugate gradient algorithm. His own interests evolved toward numerical algorithms, and he made substantial contributions in computational linear algebra, quadrature, and approximation theory before turning his attention to mechanics and celestial mechanics late in his life [14]. His technical works include over 60 journal publications and a wonderful 1960s textbook on numerical mathematics. He died in 1978, a few months short of his 70th birthday.

Computing in Science and Engineering, a publication of the IEEE Computer Society and the American Institute of Physics, named Krylov Subspace Iteration as one of the Top 10 Algorithms of the Century [15], citing in particular the pioneering work of Hestenes, Stiefel, and Lanczos. The citation reads:

*Krylov Subspace Iteration
Hestenes, Stiefel, Lanczos
Conjugate gradient methods are iterative matrix algorithms for solving very large linear systems of equations, especially efficient for sparse square matrices. Such systems arise in various application areas, such as modeling of fluid flows, reservoir engineering, mechanical engineering, semiconductor device analysis, nuclear reaction models, and electric circuit simulation. These matrices can be huge, up to millions of degrees of freedom. Modern improvements include GMRES and BI-CGSTAB.*

Prepared by Dianne P. O'Leary.

Bibliography

- [1] Magnus R. Hestenes and Eduard Stiefel, Methods of Conjugate Gradients for Solving Linear Systems, *J. Res. Natl. Bur. Stand.* **49**, 409-436 (1952).
- [2] Magnus R. Hestenes and John Todd, *NBS-INA—The Institute for Numerical Analysis—UCLA 1947-1954*, NIST Special Publication 730, National Institute of Standards and Technology, Gaithersburg, MD (1991).
- [3] Eduard Stiefel, Über einige Methoden der Relaxationsrechnung, *Z. Angew. Math. Phys.* **3**, 1-33 (1952).
- [4] M. R. Hestenes, Iterative Methods for Solving Linear Equations, NAML Report 52-9, July 2, 1951, NBS, Los Angeles, CA. Reprinted in *J. Optim. Theory Appl.* **11**, 323-334 (1973).
- [5] Cornelius Lanczos, Solution of Systems of Linear Equations by Minimized Iterations, *J. Res. Natl. Bur. Stand.* **49**, 33-53 (1952).
- [6] Cornelius Lanczos, An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators, *J. Res. Natl. Bur. Stand.* **45**, 255-282 (1950).

- [7] Magnus Hestenes, *Conjugacy and Gradients*, Princeton History Conference, Summer 1987.
- [8] 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*, J. K. Reid (ed.), Academic Press, New York (1971) pp. 231-254.
- [9] Gene H. Golub and Dianne P. O'Leary, Some History of the Conjugate Gradient and Lanczos Algorithms: 1948-1976, *SIAM Rev.* **31**, 50-102 (1989).
- [10] A. Miele, An Appreciation of Professor M. R. Hestenes, *J. Optim. Theory Appl.* **14**, 445-451 (1974).
- [11] M. R. Hestenes, Applications of the Theory of Quadratic Forms in Hilbert Space to the Calculus of Variations, *Pacific J. Math.* **1**, 525-581 (1951).
- [12] J. Waldvogel, U. Kirchgraber, H. R. Schwarz, and P. Henrici, Eduard Stiefel (1909-1978), *Z. Angew. Math. Phys.* **30**, 133-136 (1979).
- [13] Konrad Zuse, Die Rolle der ETH bei der Computerentwicklung, *Z. Angew. Math. Phys.* **30**, 399-403 (1979).
- [14] Eduard Stiefel, Comments on the Bibliography, *Z. Angew. Math. Phys.* **30**, 141-142 (1979).
- [15] Top 10 Algorithms of the Century, *Comput. Sci. Eng.* **2** (1), (2000).