Some History of Conjugate Gradients and Other Krylov Subspace Methods

October 2009

SIAM Applied Linear Algebra Meeting 2009

Dianne P. O'Leary

Some History of Conjugate Gradients

Dianne P. O'Leary

Computer Science Dept. and
Institute for Advanced Computer Studies
University of Maryland
oleary@cs.umd.edu

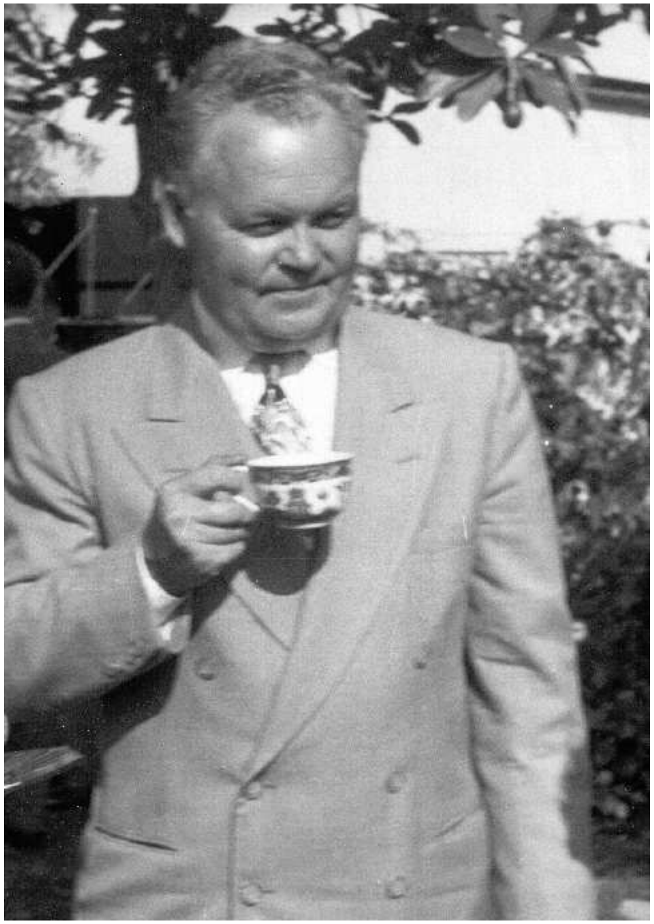
http://www.cs.umd.edu/users/oleary

A Tale of 2 Cities

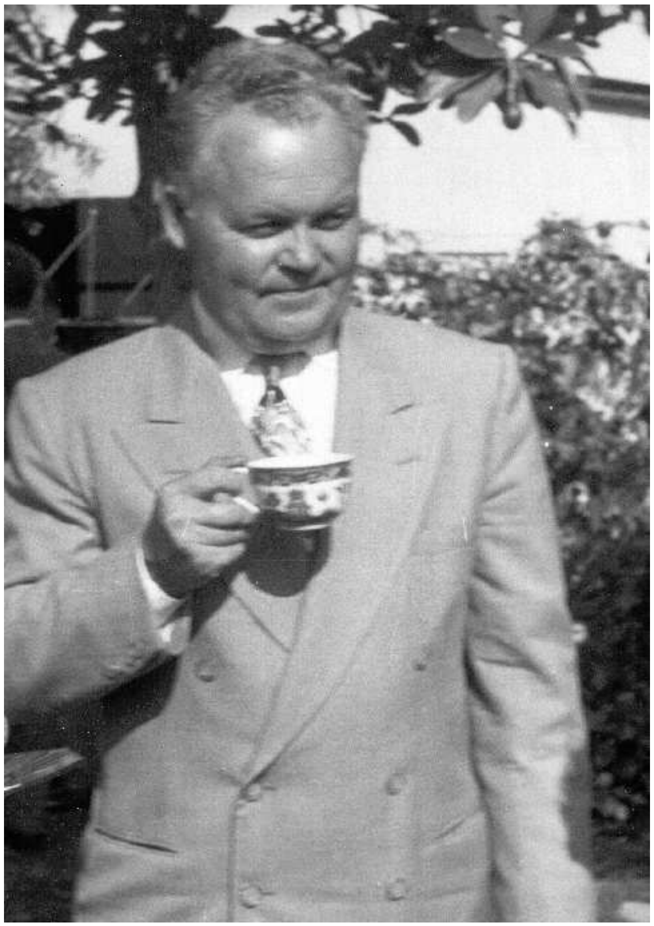
Los Angeles, California

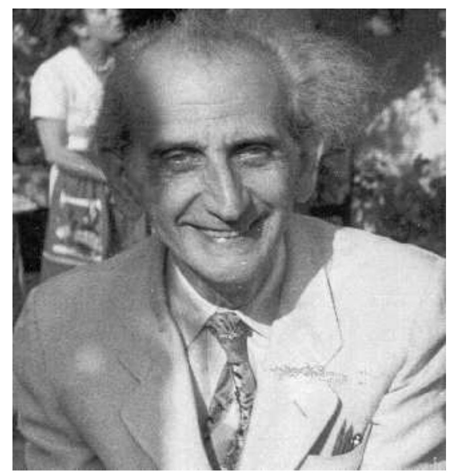
Zürich, Switzerland

... and a Tale of 2 Men

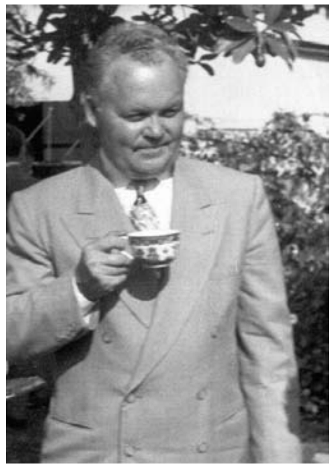


… or maybe 3 Men





… or maybe 5 Men



The Conjugate Gradient Algorithm

Notation

- We solve the linear system
$$Ax^* = b$$
where $A \in \mathcal{R}^{n \times n}$ is symmetric positive definite and $b \in \mathcal{R}^n$.
- We assume, without loss of generality, that our initial guess for the solution is
$$x^{(0)} = 0\,.$$
- We denote the Krylov subspace of dimension m by
$$\mathcal{K}_m(A, b) = \mathrm{span}\{b, Ab, \ldots, A^{m-1}b\}\,.$$
- Then the conjugate gradient algorithm chooses $x^{(k)} \in \mathcal{K}_k(A, b)$ to minimize $(x - x^*)^T A(x - x^*)$.
  - work per iteration: 1 matrix-vector multiply and a few dot products and combinations of vectors.
  - storage: 3 vectors, plus original data.

Context for the Conjugate Gradient Algorithm

Hestenes and Stiefel (1952) presented the conjugate gradient algorithm in the *Journal of Research of the NBS*.

Context: Two existing classes of algorithms

- Direct methods, 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.
- 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.

Ideal algorithm:

- finite termination.
- if stopped early, would give a useful approximation.

Hestenes & Stiefel account of how the paper came to be written

"The method of conjugate gradients was developed independently by E. Stiefel of the Institute of Applied Mathematics at Zürich 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. Hoschstrasser, and M. Stein."

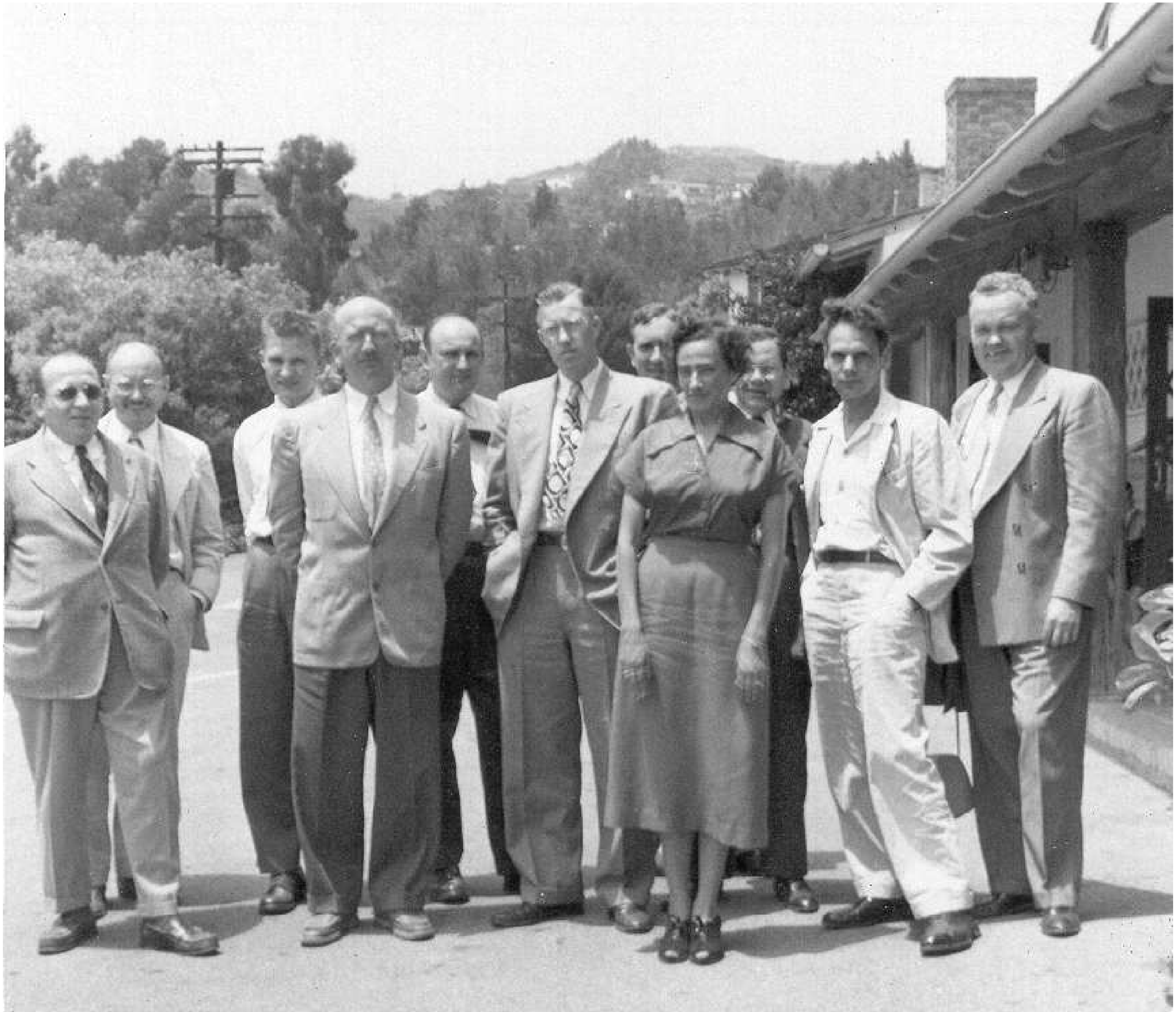


Figure 1: Group of NBS Institute for Numerical Analysis researchers in 1950, including Mark Kac, Edward J. McShane,J. Barkley Rosser, Aryeh Dvoretzky, George G. Forsythe, John Todd, Olga Taussky-Todd, Everett C. Yowell (?), Wolfgang R. Wasow, and Magnus R. Hestenes (photo: NIST).

Two distinct voices in the paper:

- Hestenes:
  - variational theory and optimal control
  - 1936: developed an algorithm for constructing conjugate bases,
  - discouraging numerical experience by George Forsythe in using steepest descent for solving linear systems.
- Stiefel:
  - relaxation algorithms
  - continued fractions
  - qd algorithm

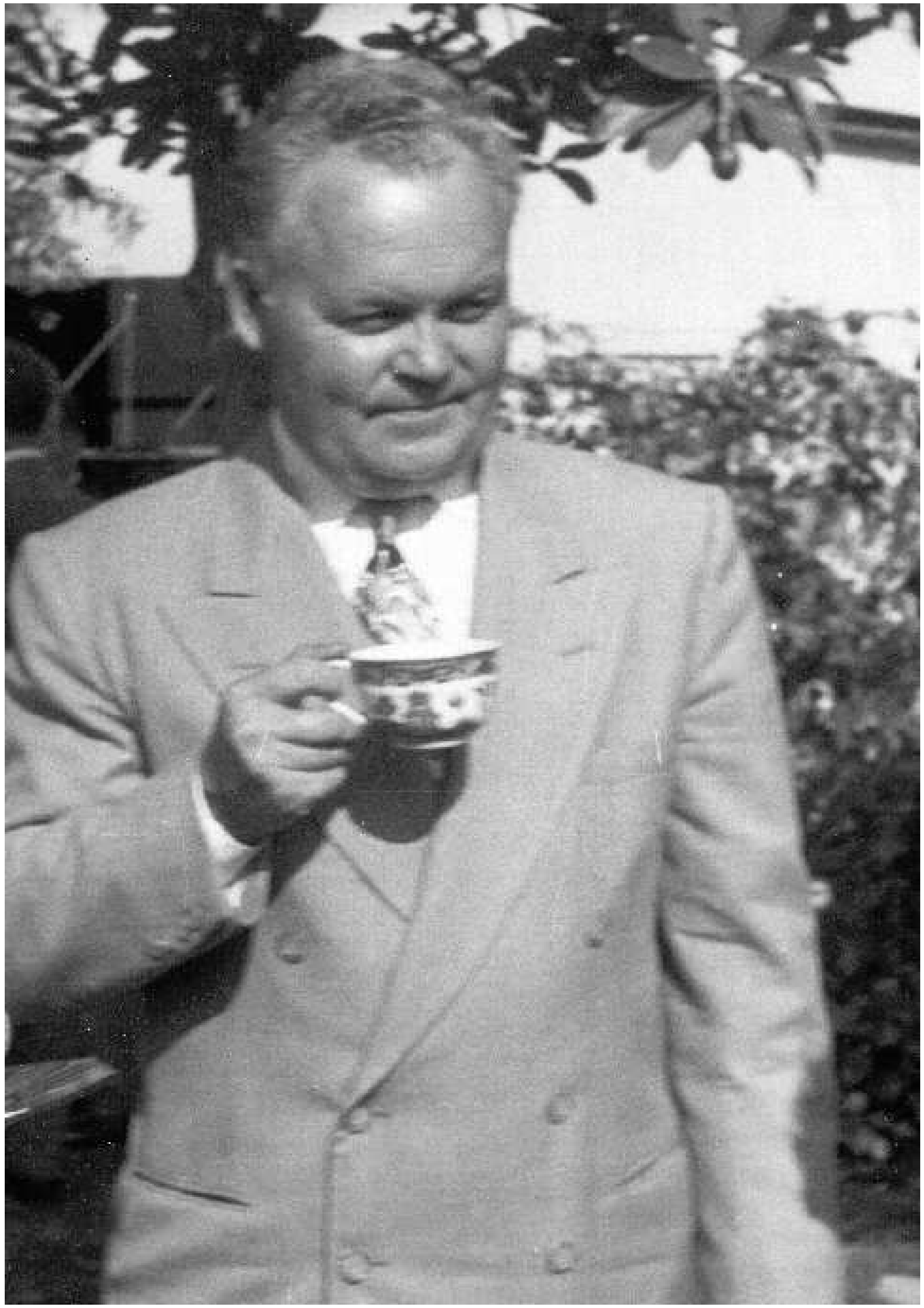


Figure 2: Magnus Hestenes (photo: NIST)

Magnus R. Hestenes

- born in Bricelyn, Minnesota, in 1906.
- undergraduate: St. Olaf College; graduate: University of Wisconsin and the University of Chicago.
- faculty appointment at Univ. of Chicago, but left in 1947 for UCLA, where he taught until his retirement.
- had 34 Ph.D. students and was a well-loved adviser and teacher, known for his nurturing kindness toward his very junior colleagues.
- chaired the Mathematics Department, directed the university computing center, served as vice president of the American Mathematical Society.
- held appointments with the Rand Corporation, the Institute for Defense Analyses, and the IBM Watson Research Center.
- associated with NBS from 1949 to 1954, when the INA was transferred from NBS to UCLA.

- best known for publications on the problem of Bolza, a famous paper on quadratic forms in Hilbert space.
- remained scientifically active until his death in 1991, concentrating in his later years on the method of multipliers.

E. Stiefel

Eduard Stiefel

- born in 1909 in Zürich, Switzerland.
- spent virtually his entire career at the Eidgenssischen Technischen Hochschule (ETH) in Zürich, first as a student of mathematics and physics, and then, following his habilitation degree in 1943, as a professor.
- early work was in topology, eventually studying the geometry and topology of Lie groups.
- founded in 1948 the Institut für Angewandte Mathematik (Institute for Applied Mathematics) at ETH, in collaboration with Heinz Rutishauser and Ambros P. Speiser.
- was a visionary who realized the enormous significance of the new computing technology and the impact it would have on mathematics and science and made ETH a center for computation.

- best known works include substantial contributions in computational linear algebra, quadrature, and approximation theory before turning his attention to mechanics and celestial mechanics late in his life.
- died in 1978, a few months short of his 70th birthday.

What was Lanczos' Role?

- 1950 paper: on an "iteration method" for the eigenproblem.
- Section 7: "method of minimized iterations". Develops an iteration based on choosing coefficients α_{kj} so that the norm of

$$v_{k+1} = Av_k - \sum_{j<k+1} \alpha_{kj} v_j$$

  is minimized.
- Shows that this yields an orthogonal basis (for what we now call the Krylov subspace) and a 3-term recurrence.
- develops a biorthogonalization algorithm when A is nonsymmetric.
- uses the recurrences to solve eigenproblems.

Figure 4: Cornelius Lanczos and Mrs Arnold D. Hestenes? (photo: NIST)

Cornelius Lanczos (born Kornél Löwy)

- began and ended his life peacefully in Hungary, but his life pivoted on three exiles.
- was born in 1893, the eldest son of a Jewish lawyer.
- attended Jewish elementary school, Catholic secondary school, and the University of Budapest.
- Ph.D. work concerning special relativity received some attention by Einstein, but political turmoil and Jewish quotas in Hungary caused Lanczos to move to Germany.
- continued his work in physics, and published an integral formulation of the Schrödinger equation just before Schrödinger published his differential equation formulation. Lanczos' paper was misinterpreted for many years.
- spent a year as Einstein's assistant and married a German, Maria Rupp.

- fled to the US In 1931, with a visiting position at Purdue University. Maria, with tuberculosis, stayed behind.
- continued his work in physics, but increasingly focused on mathematical techniques, and he developed the Tau method for approximating functions by telescoping series.
- Maria died in 1939, and he brought his son to the US.
- became known as an excellent teacher, eventually writing many popular and ground-breaking textbooks, celebrated for their clarity and their use of vivid examples. His approach to computation was unique for the time: he worked with calculator, not slide rule, and this led him to novel algorithmic approaches.
- derived the FFT in the early 1940s, although the idea did not become widely known until the popularization of the (equivalent) Cooley-Tukey FFT algorithm.

- worked on the NBS Mathematical Tables Project during a leave from Purdue, worked at Boeing Aircraft Company in Seattle, and then joined the Institute for Numerical Analysis of the NBS.
- investigated algorithms for solving linear systems of equations and eigenvalue problems.
- turmoil at NBS following politicized investigations, and Lanczos came under investigation for allegedly being a communist sympathizer.
- a third exile, in 1952 and then permanently in 1954, to the Dublin Institute for Advanced Studies.
- maintained his ties to Dublin, although he traveled and lectured extensively. His expository skills were much renowned, and he won the Chauvenet Prize for mathematical exposition in 1960.
- married Ilse Hildebrand, and refocused on physics, including the geometry of space-time, although publications in mathematics continued.
- died of a heart attack on his second visit back to Hungary, in 1974.

Why weren't these algorithms discovered earlier?

Why weren't these algorithms discovered earlier?

- When Hestenes worked on conjugate bases in 1936, he was advised by a Harvard professor that it was too obvious for publication

Why weren't these algorithms discovered earlier?

- When Hestenes worked on conjugate bases in 1936, he was advised by a Harvard professor that it was too obvious for publication
- CG doesn't work on slide rules.

Why weren't these algorithms discovered earlier?

- When Hestenes worked on conjugate bases in 1936, he was advised by a Harvard professor that it was too obvious for publication
- CG doesn't work on slide rules.
- CG has little advantage over Gauss elimination for computation with calculators.

Why weren't these algorithms discovered earlier?

- When Hestenes worked on conjugate bases in 1936, he was advised by a Harvard professor that it was too obvious for publication
- CG doesn't work on slide rules.
- CG has little advantage over Gauss elimination for computation with calculators.
- CG is not well suited for a room of human computers – too much data exchange.

Why weren't these algorithms discovered earlier?

- When Hestenes worked on conjugate bases in 1936, he was advised by a Harvard professor that it was too obvious for publication
- CG doesn't work on slide rules.
- CG has little advantage over Gauss elimination for computation with calculators.
- CG is not well suited for a room of human computers – too much data exchange.
- The motivation would have to be an appropriate computational engine.

Why were these algorithms discovered then?

Hestenes, Lanczos, and Stiefel all had shiny, brand new toys:

- the SWAC for Hestenes and Lanczos,
- the Z4 for Stiefel.

The men behind these machines were the muses for the discovery of the CG algorithm.

Konrad Zuse (photo: Comp. Museum) http://www.computerhistory.org/collections/accession/102630811

Konrad Zuse, 1980 (photo: Comp. Museum) `http://www.computerhistory.org/collections/accession/102630803`

Konrad Zuse

- born in Berlin 1910.
- built the world's first program-controlled (Turing-complete) computer in 1941.
- degree in civil engineering from Technische Hochschule Berlin-Charlottenburg in 1935.
- worked as a design engineer for an aircraft company, but resigned to work on his machine.
- built the Z1 in his parents' living room. The machine and the blueprints were blown up during an air raid. Designed the first high-level programming language, 1948, but it was never implemented.
- Founded a computer manufacturing company in 1946 to build the Z4.
- IBM took an option on his 1937 patents in 1946.

- Stiefel discovered in 1949 that this amazing Z4 machine was sitting in the small alpine village of Neukitchen, Germany. He traveled there and arranged for the machine to be rented and moved to ETH.
- Developed a theory that the universe is running on a grid of computers.
- Died in 1995.

"I do not have art studies but I also do not have computer science studies"

Zuse painting (photo: epimag.com)
`http://www.epemag.com/zuse/part8b.htm`

Harry Huskey (photo: Comp. Museum)
http://www.computermuseum.li/Testpage/SWAC-1950.htm

Harry Huskey

- born in North Carolina in 1916, and grew up in Idaho.
- Ms and PhD from Ohio State University
- taught math at Univ. of Pennsylvania.
- worked (part time) on ENIAC in 1945.
- visited NPL for a year, working on the Pilot ACE with Turing.
- also involved in EDVAC and SEAC (Standards Eastern).
- designed and built the SWAC 1949-1953.
- designed the G15 for Bendix Aviation; 1st personal computer?
- faculty member at UC Berkeley and UC Santa Cruz.
- supervised PhD thesis of Niklaus Wirth (1963).
- retired in 1986 and lives in South Carolina.
- 4 children; married 2nd wife (1st died) in 1994.
- Fellow of the ACM.

The Z4 and the SWAC

Z4 at ETH (photo: Zuse homepage) `http://user.cs.tu-berlin.de/~zuse/Konrad_Zuse/en/rechner_z4.html`

Z4, Deutsche Museum (photo: Clemens Pfeiffer)
http://commons.wikimedia.org/wiki/File:
Zuse-Z4-Totale_deutsches-museum.jpg

Z4 memory (photo: Zuse homepage)

`http://www.epemag.com/zuse/part7a.htm`

Z4 programming unit (photo: Zuse homepage)
`http://www.epemag.com/zuse/part7b.htm`

The SWAC and Harry Huskey 1950 (photo: Comp. Museum)
`http://www.computermuseum.li/Testpage/SWAC-1950.htm`

	SWAC vs.	Z4
Designer	Harry Huskey	Konrad Zuse
Operational	1950	1945
	Los Angeles	Germany
Purpose	Interim (until Univac arrival)	first of a commercial series
Size	20x30 sq feet	1000 kilograms
Cost	$200,000	30,000 SFR rent for 1950-55
Completion time	19 months	4 years (bombings)
Technology	2300 vacuum tubes	2500 relays, 21 stepwise relays
	(18,000 in 1948 ENIAC)	mechanical memory
Memory	256 37-bit words	64 32 bit words
Operations	+,-,*,compare,input,output,	+,-,*,/,sqrt,...
	data extraction (mask&shift)	
Speed of add:	64 microsec.	400,000 microsec.
speed of multiply:	384 microsec.	
	fastest in world in 1950	descendant of 1st in world
Input	Punch tape / cards	punch tape

The Legacy of the 1950s Work on the CG Algorithm

The Scope of the 1952 Paper

Assume that A is symmetric positive definite.

- Section 3: a terse presentation of the formulas.
  - Later motivation: the recurrences lead to a sequence of approximations that converge monotonically in the sense of reducing an error function.
  - a sequence of polynomials can be constructed in order to find the eigensystem of the matrix.
- direct method, special case of conjugate directions: finite termination.
- alternate computational formulas
- geometric arguments about optimization properties.
- use as iterative method: solves 106 "difference equations" in 90 iterations. (By 1958: 10x10 grid Laplace equation in 11 Chebyshev iterations + 2 cg.)

- monotonicity properties.
- round-off error analysis. 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.
- smoothing initial residual.
- remedy for loss of orthogonality.
- solution if A is rank deficient.
- solution if A is nonsymmetric (normal equations).
- algebraic formulation of preconditioning.
- relation to Gauss elimination and orthogonal polynomials.
- relation to Lanczos algorithm and continued fractions.

The authors remind us of 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), “Infolge der beschränkten Speicherkapazität der Maschine dauerte ein Zyklus etas 2 h 20 m.” (Stiefel 1952 *ZAMP*)
- to solve systems as large as dimension 12 on an IBM card-programmed calculator,
- to solve small systems on the SWAC (Standards Western Automatic Computer), which had only 256 words of memory and could do 16,000 operations per sec.

Early recognition

- The algorithm garnered considerable early attention but went into eclipse in the 1960s, as naïve implementations were unsuccessful on some of the ever-larger problems that were being attempted.
- Work by John Reid in the early 1970s drew renewed attention to the algorithm, and since then it has been an intense topic of research.

Current Status

- Today is the standard algorithm for solving linear systems involving large, sparse, symmetric (or Hermitian) matrices.

Recent Recognition of the Algorithm

- Science Citation Index lists over 1500 citations to the Hestenes & Stiefel 1952 paper, over 500 since 2000.
- In 2009 it has over 60 citations, in journals of chemical physics, magnetic resonance, signal processing, geosciences, artificial intelligence, polimery, bioinformatics, physiology, magnetics, drug metabolism, heat transfer, simulation, medical imaging, etc.
- NIST celebrated its centennial by picking its 100 most significant achievements. Among them:
  - ASCII
  - a highly-successful consumer information series
  - creation of Bose-Einstein condensation
  - the Conjugate Gradient Algorithm
  - Lanczos' eigenvalue algorithm

- *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 20th Century, citing in particular the pioneering work of Hestenes, Stiefel, and Lanczos.

Major open research questions

- Effective preconditioning.
- Mitigating the effects of restarting for the nonsymmetric versions (GMRES).

References

1. Computer Museum History Center, "Historical Zuse Computer Z23," Mountain View, CA.
`http://www.computerhistory.org/projects/zuse_z23`

2. William R. Davis et al., eds., *Cornelius Lanczos Collected Published Papers with Commentaries,* North Carolina State University Publishers, Raleigh, 1998, volume 5.

3. Gene H. Golub and Dianne P. O'Leary, "Some history of the conjugate gradient and Lanczos algorithms: 1948-1976," SIAM Review 31 (1989) 50-102.

4. Ruth B. Horgan, SWAC Coding Guide, NBS Report 2047, 1952.
`http://www.bitsavers.org/pdf/nbs/swac/SWAC_CodingGuide_Nov52.pdf`

5. Harry D. Huskey, “SWAC Standards Western Automatic Computer: The Pioneer Day Session at NCC July 1978” transcript, IEEE Annals of the History of Computing 19(2), 1997, 51-61. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=586073

6. David R. Lide, ed., *A Century of Excellence in Measurements, Standards, and Technology – A Chronicle of Selected NBS/NIST Publications 1901-2000,* NIST Special Publication 958, 2001, pp. 81-85.

7. Wikipedia, Harry Huskey, Konrad Zuse,
http://en.wikipedia.org/wiki/Harry_Huskey
http://en.wikipedia.org/wiki/Konrad_Zuse

8. Horst Zuse, “The Life and Work of Konrad Zuse,” EPE Online
http://www.epemag.com/zuse/

9. Horst Zuse, “Konrad Zuse Homepage,”
http://user.cs.tu-berlin.de/~zuse/Konrad_Zuse/