UNIVERSITY OF MINNESOTA TWIN CITIES

Origin and development of Krylov subspace methods

Yousef Saad

University of Minnesota

75th Anniversary of Mathematics and Statistics at NIST

June 28-30, 2022

About Krylov

Aleksei Nikolaevich Krylov :

- *Born:* August 15, 1863 in Visyaga, Simbirskoy (now Ulyanovskaya), Russia
- *Died:* October 26, 1945 in Leningrad, USSR (now St Petersburg, Russia)

➤ Son of an artillery officer. Joined Maritime Academy in 1888 as student and then a teacher. Kept position for 50 years.

➤ Work: shipbuilding, magnetism, artillery, mathematics, astronomy, geodesy

➤ Held important positions and became quite influential in USSR

1931: Krylov developed a new method for computing characterstic polynomials

Main idea: Let v_1 be a nonzero vector of grade n and consider

$$\boxed{v_i = Av_{i-1}, \ i = 2, \cdots, n+1}$$

➤ If: $\boxed{p_n(t) = t^n - \mu_{n-1}t^{n-1} - \cdots - \mu_1 t - \mu_0}$ == characteristic polynomial of A then $p_n(A)v_0 = 0 \rightarrow$

$$v_{n+1} - \mu_{n-1}v_n - \cdots - \mu_1 v_2 - \mu_0 v_1 = 0$$

➤ Method: express $v_{n+1} = A^n v_1$ as a lin. combination of $v_1, v_2, \cdots, v_n$
Requires solving an ill-conditioned system

➤ Alternative viewpoint: Transform A into companion form. Consider the basis $V = [v_1, \cdots, v_n]$. Then

$$A[v_1 \vdots v_2 \vdots \cdots \vdots v_n] = [v_1 \vdots v_2 \vdots \cdots \vdots v_n] \times \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & \mu_0 \\ 1 & 0 & 0 & \ldots & 0 & \mu_1 \\ & 1 & 0 & \ldots & 0 & \mu_2 \\ & & . & \ldots & . & \vdots \\ & & & 1 & 0 & \mu_{n-2} \\ & & & & 1 & \mu_{n-1} \end{bmatrix} \rightarrow \boxed{AV = VH}$$

1941: K. Hessenberg developed a method to transform A into what we now call Hessenberg form. Strong relation to Krylov.

Basic idea: at j-th step define v_{j+1} as:

$h_{j+1,j}v_{j+1} = Av_j - \sum_{i=1}^{j} h_{ij}v_i$

s.t. $v_{j+1} \perp e_1, e_2, \ldots, e_j$ & $e_{j+1}^T v_{j+1} = 1$.

Can add pivoting (Wilkinson)

$$\begin{aligned} h_{ij} &= e_i^T(Av_j), \; i = 1 : j \\ v &= Av_j - \textstyle\sum_{i=1}^{j} h_{ij}v_i \\ h_{j,j+1} &= e_{j+1}^T v \\ v_{j+1} &= v/h_{j+1,j} \end{aligned}$$

The Arnoldi method

1951: Arnoldi developed another method to transform A into Hessenberg form. Linear independence via orthogonality

For $j = 1 : m$ Do:

$$\begin{aligned} h_{ij} &= v_i^T(Av_j), \; i = 1 : j \\ v &= Av_j - \textstyle\sum_{i=1}^{j} h_{ij} v_i \\ h_{j,j+1} &= \|v\| \\ v_{j+1} &= v / h_{j+1,j} \end{aligned}$$

EndDo

- Note: Assumes $m < n$ $\rightarrow$ No longer a reduction to Hessenberg form
- Instead: Arnoldi adopts the viewpoint of projection (Galerkin) methods
- Above $\equiv$ classical Gram-Schmidt – Better: use modified Gram-Schmidt
- In Arnoldi's article: $h_{j+1,j} \equiv 1$ [no scaling]

➤ Note: most of Arnoldi's paper was about the Lanczos method [1950, J.R. NBS] and its interpretation as a Galerkin process

THE PRINCIPLE OF MINIMIZED ITERATIONS IN THE SOLUTION OF THE MATRIX EIGENVALUE PROBLEM[1]

BY

W. E. ARNOLDI

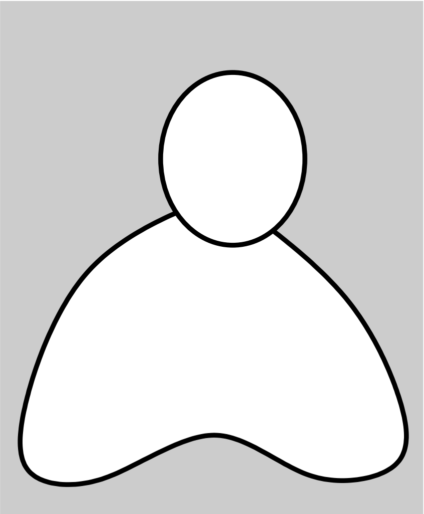
Hamilton Standard Division, United Aircraft Corporation, East Hartford, Conn.

An interpretation of Dr. Cornelius Lanczos' iteration method, which he has named "minimized iterations", is discussed in this article, expounding the method as applied to the solution of the characteristic matrix equations both in homogeneous and non-homogeneous form. This interpretation leads to a variation of the Lanczos procedure which may frequently be advantageous by virtue of reducing the volume of numerical work in practical applications. Both methods employ essentially the same algorithm, requiring the generation of a series of orthogonal functions through which a simple

➤ Only in section 6 do we see Arnoldi's method - again viewed as a Galerkin approach

➤ ... followed by a description of his method.

About Arnoldi: Walter Edwin Arnoldi:



- *Born:* December 14, 1917 in New York (NY)
- *Died:* October 5, 1995, Hartford (CT)

➤ Degree in Mech. Engin. from Stevens Institute of Tech., then Masters from Harvard (1939?) - Worked for United Technologies Corp. from 1939 until retirement in 1977.

Galerkin approach for eigenvalue problems

Given: Subspace K with orthonormal basis $V = [v_1, v_2, \cdots, v_m]$

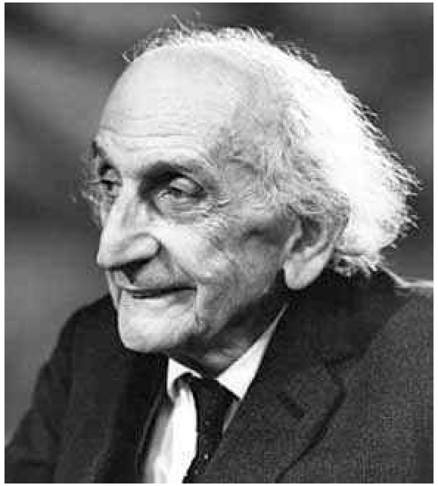
Problem: find an approximate eigenpair $\tilde{\lambda} \in \mathbb{C}$ and $\tilde{u} = Vy$ such that $V^H(A - \tilde{\lambda}I)\tilde{u} = 0 \quad \rightarrow$ $\boxed{(V^H AV - \tilde{\lambda}I)y = 0}$

- Krylov subspace methods $\mathbf{span}\{V\} = \mathbf{span}\{v_1, Av_1, \cdots, A^{m-1}v_1\}$
- Arnoldi procedure to get the v_i's
- Note: In Hermitian case $V^H AV$ is Hermitian $\rightarrow$ Hessenberg matrix becomes tridiagonal $\rightarrow$ simplification $\rightarrow$ Lanczos Algorithm (1950)
- Lanczos arrived at his algorithm via rather fascinating route

➤ Acknowledgment section of the 1950 paper by C. Lanczos:

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 author expresses his heartfelt thanks to C. K. Stedman, head of the Physical Research Unit of the Boeing Airplane Co. and to J. H. Curtiss, Acting Director of the Institute for Numerical Analysis, for the generous support of his scientific endeavors.

About Lanczos



Cornelius Lanczos (Born Kornél Löwy)

- *Born:* February 2, 1893 in Székesfehérvár, Hungary
- *Died:* June 25, 1974 in Budapest, Hungary

➤ Had a rather turbulent life. Often forced to move

> *Budapest ('15), Freiburg ('21), Frankfurt ('24), Berlin ('28), Frankfurt ('29), Purdue Univ. '31, back to Germany '31, Back to Purdue '32, Boeing '44 and '46. NBS in Los Angeles '49, Dublin '52*

➤ Work showed deep insight rooted in approximation theory (Influence of Fejér?) – Viewpoint: Polynomial approximations.

1942 He developed (along with G. C. Danielson) what is now known as the FFT - witout realizing the O(N log N) cost.

1949–1952 Institute for Numerical Analysis at NBS [Olga Taussky-Todd, John Todd, George Forsythe were colleagues]

- Work on Lanczos algorithm and CG-like methods from this period
- Had to leave the US during McCarthy era.

1952 Offer (from Schrödinger who fled Austria in 1933) to head Theoretical Physics Department at the Dublin Institute for Advance Study in Ireland

1974 Died in Budapest during a visit to the Eötvös Lóránd University

- See: `https://mathshistory.st-andrews.ac.uk/Biographies/Lanczos/`

Iterative methods for linear systems: Relaxation

- Initial Problem: $\boldsymbol{b - Ax = 0}$

➤ Modify i-th component of current x to make $\boldsymbol{e_i^T(b - Ax^{new}) = 0}$

➤ Then repeat with another i, ..., until convergence

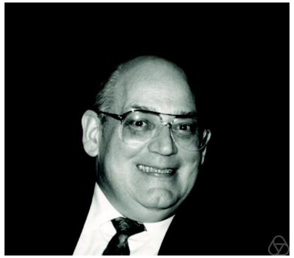
➤ Idea first developed by Gauss ca 1817, then Jacobi (1850), Seidel (1874)

➤ Can be viewed as a sequence of one-dimensional projection methods

➤ 1950's – 1960's: Frankel, Young, Varga, ...

➤ Jacobi, Gauss-Seidel, Successive-Over-Relaxation remained state of the art up to early 1970s.

➤ Read what Richard Varga writes in 'Matrix Iterative Analysis' [1962]:



As an example of the magnitude of problems that have been successfully solved on digital computers by cyclic iterative methods, the Bettis Atomic Power laboratory of the Westinghouse Electric Corporation had in daily use in 1960 a two-dimensional program which would treat as a special case, Laplacean-type matrix equations of order 20,000. Adds as a footnote: *... Even more staggering is Bettis' use of a 3-Dimensional program called "TNT-1", which treats coupled matrix equations of order 108,000.*

- State of the art in 1960: solving a P.D.E. system with $\approx 100,000$ eqns
- Could do this in fraction of a second on a laptop today

One-dimensional projection processes

- To solve system $\boldsymbol{Ax = b}$. Let $x \equiv$ current iterate and $\boxed{\boldsymbol{r = b - Ax}}$
- Select a new d (search subspace) and a new e (constraint subspace)
- New iterate: $\boxed{\tilde{x} := x + \alpha d}$ Let $\boxed{\tilde{r} = b - A\tilde{x}}$
- 'Petrov-Galerkin' condition: $\boxed{\tilde{r} \perp e \rightarrow r - Ad \perp e}$ $\rightarrow$ $\boxed{\alpha = \frac{(r,e)}{(Ad,e)}}$
- Repeat process until convergence ...

- Steepest descent: $d = e = r$ [Cauchy 1847 (nonlin.) Kantorovitch 1945]

- Minimal residual $d = r$, $e = Ar$

- Residual norm steepest descent $d = r$, $e = A^T r$

1937 Kaczmarz method: $d = A^T e_i$, $e = e_i$ for $i = 1, \cdots, n$.

- Kaczmarz method equivalent to Gauss-Seidel for

$$\boxed{AA^T u = b \text{ with } (x = A^T u)}$$

- Very popular in the 70s for Computer Tomography [ART method]

1938 Cimmino's method == Jacobi method for $A^T A x = A^T b$

Polynomial iteration

- Richardson iteration + some one-dim projection methods are of the form

$$x_{k+1} = x_k + \beta_k r_k$$

1950 Frankel considers a 'second-order' iteration:

$$x_{k+1} = x_k + \beta_k d_k \text{ where } d_k = r_k - \alpha_k d_{k-1}$$

- With constant coefficients $\longrightarrow$ Chebyshev iteration.
- Many papers adopted an 'approximation theory' viewpoint

Krylov methods take off: The CG algorithm

- Magnus Hestenes [UCLA] and Eduard Stiefel [ETH, Zürich] developed the method of *Conjugate Gradients* independently

M. Hesteness

E. Stiefel

- Article:

Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards, 1952.

From the 1952 CG article:

"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. (...) 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."

- Lanczos developed a similar method [different notation and viewpoint:]

Article: *Solution of systems of linear equations by minimized iterations*, Nat. Bur. Standards (1952).

C. Lanczos

- In effect: $\equiv$ Minimal Residual method
- Note: Same journal, same institution (INA, NBS)
- Lanczos' came out in July '52, Hestenes and Stiefel's in Dec. '52.

About Hestenes

Magnus Rudolph Hestenes:

- *Born:* Born February 13, 1906, Bricelyn, Minnesota
- *Died:* May 31, 1991, Los Angeles, CA

1932 Ph.D. at the University of Chicago

1947 Professorship at UCLA - kept position until retirement in '73

➤ Associated with INA (NBS) [listed as 'UCLA liaison' member]

➤ Work: calculus of variations, optimal control, gradient-type methods for linear systems and eigenvalue problems

Aftermath of CG article:

- Early on: CG Method viewed as an unstable, direct method.

1959 Engeli, Ginsburg, Rutishauser, Stiefel: viewed CG as iterative procedure.

1971 Chris Paige analyzes the Lanczos algorithm (PhD thesis) for eigenvalue problems. This and later work by Parlett played an important role in reviving the algorithm

1972 Detailed study by John Reid on CG as iterative procedure

Krylov methods: the 'nonsymmetric' period

- Lanczos [MR paper 1952] : shows a method that is essentially the BiCG algorithm - then says : *let us restrict our attention to symmetric case ... (Normal eqns.)* A pity!

- Forward to 1976: Fletcher introduces BiCG. Later: CGS [Sonneveld, 1989] Bi-CGSTAB [van der Vorst, 1992], QMR, TFQMR [Freund, 1991], ...

Krylov methods: the ‘nonsymmetric’ period

- Lanczos [MR paper 1952] : shows a method that is essentially the BiCG algorithm - then says : *let us restrict our attention to symmetric case ... (Normal eqns.)* A pity!

- Forward to 1976: Fletcher introduces BiCG. Later: CGS [Sonneveld, 1989] Bi-CGSTAB [van der Vorst, 1992], QMR, TFQMR [Freund, 1991], ...

- “Orthogonal projection” track: ORTHOMIN [Vinsome, 1976], GCR [Axelsson, Vinsome, Eisenstat Elman & Schultz,] [ORTHODIR, ORTHORES, Young et al.], GMRES [YS-Schultz 1986], ...

- + Theory [Faber Manteuffel theorem, 1984], Convergence [Eisenstat-Elman-Schultz, 1983], Greenbaum-Strakos, ...

Preconditioning

- Idea of preconditioning quite old – Golub & O'Leary (1989) trace term 'preconditioning' back to Turing [1948]

1937 Cesari used polynomial precond. to speed-up Richardson iter.

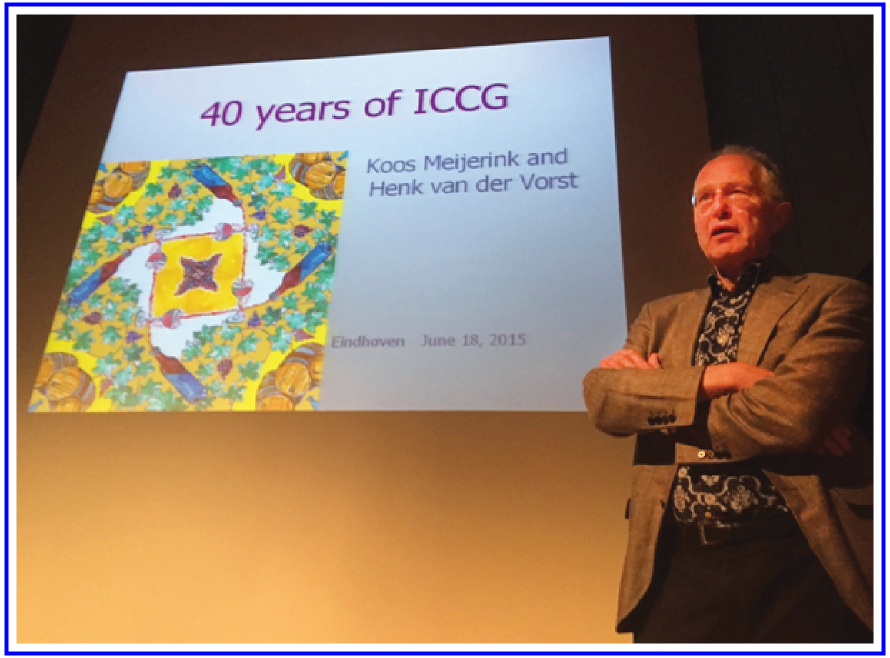
1952 Polynomial preconditioning again by Lanczos - later Stiefel [1959]

1953 Forsythe uses the term explicitly: *With the concept of "Ill conditioned" systems $Ax = b$ goes the idea of "preconditioning" them. Gauss [1823] and Jacobi [1845] made early contributions to this subject."*

1963 Wachpress uses ADI preconditioned CG

- Idea of ILU or Approx. factorizations: Buleev '60, Oliphant '62; Stone (SIP) '68; Dupont Kendall Rachford '68; Axelsson '72,..., then *ICCG*

Preconditioning: the impact of ICCG



Dinner speech given by Meijerink and van der Vorst, authors of ICCG paper, at ‘Preconditioning 2015’

Henk van der Vorst

➤ Incomplete Cholesky Conjugate Gradient (ICCG) paper appeared in 1977 (Math. Comp). Work done much earlier. Paper had a tremendous impact.

Concluding remarks: past and future of Krylov methods

NIST and Institute of Numerical Analysis played a major role in development of modern Krylov methods

➤ Detailed account by Magnus Hestenes and John Todd *"Mathematicians learning to use computers. The Institute for Numerical Analysis, UCLA, 1947-1954."* (Special publication by NIST) Exerpt on the CG:

The conjugate gradient algorithm has, as a subroutine, an algorithm which is equivalent to an orthogonalization routine developed by Lanczos. The conjugate gradient routine therefore can be derived from the results given by Lanczos. Using his orthogonalization routine, Lanczos devised a "Method of Minimized Iterations" for solving a system of linear equations. This method is a variant of the original conjugate gradient routine. Credit should also be given to Rosser, Forsythe, Karush, and Paige for the development of the conjugate gradient routine, because the routine was also an outgrowth of their efforts. Rosser and Stiefel presented the conjugate gradient method at the August 23–25, 1951 Symposium on Simultaneous Linear Equations and the Determination of Eigenvalues, which was a part of the Semicentennial Celebration of NBS.

➤ C.G. said to be one of *Top 10* algorithms of the 20th century

- In era of data-based sciences, Krylov methods are finding numerous uses
- In this context: Lanczos algorithm is as important as CG

- In many ways, Krylov subspaces are optimal for dimension reduction
- Extremes: *Randomization* on one end and *standard Krylov* on the other
- In between: Block Krylov .. or multiple Krylov (combine results from a few Krylov subspaces)
- A safe prediction: *Many more uses and extensions to come!*