UMIACS-TR-91-47 March 1991
CS-TR 2641

Lanczos and Linear Systems

G. W. Stewart*

ABSTRACT

Lanczos's major contributions to the numerical solution of linear equations are contained in two papers: "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators" and "Solutions of Linear Equations by Minimized Iterations," the second of which contains the method of conjugate gradients. In this note we retrace Lanczos's journey from Krylov sequences to conjugate gradients.

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-87-0188.

Lanczos and Linear Systems

G. W. Stewart

ABSTRACT

Lanczos's major contributions to the numerical solution of linear equations are contained in two papers: "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators" and "Solutions of Linear Equations by Minimized Iterations," the second of which contains the method of conjugate gradients. In this note we retrace Lanczos's journey from Krylov sequences to conjugate gradients.

Introduction

Lanczos's major contributions to the numerical solution of linear equations are contained in two papers: "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators" and "Solutions of Linear Equations by Minimized Iterations" [6, 7]. The first paper is the usual reference the Lanczos algorithm, and as such it has been surveyed elsewhere in this collection. However, according to Lanczos the second paper "is a natural sequel to the previous publication and depends on the previous findings." For this reason, the papers must be treated together.

If we restrict ourselves to linear equations, Lanczos's contributions can be summarized as follows.

1. The use of a dependency among the vectors of a Krylov sequence to solve the associated linear system.

2. A recursive algorithm to construct the dependency from a sequence of Hankel systems.

3. The use of the intermediate solutions to solve a linear system. The denouement of this development is the method of conjugate gradients.

4. The use of Chebyshev polynomials in the solution of linear systems.

Unfortunately, this outline suggests an orderly progression that is not found in the papers themselves. Reading through them, one gets the impression that Lanczos was struggling, not very successfully, to organize a number of insights that had crystalized over a fairly short period of time. The result is a rambling style, in which the basic ideas keep getting in each other's way.

The laxity of Lanczos's exposition makes it impossible to give a concise summary that is fully representative of his ideas. To strike a balance between brevity and accuracy, we will pretend that Lanczos was solely concerned with the solution of the symmetric, positive-definite system

$$\mathbf{A}\mathbf{y} = \mathbf{b}, \tag{1}$$

though in fact he worked with nonsymmetric systems and systems of the form

$$\mathbf{y} - \lambda\mathbf{A}\mathbf{y} = \mathbf{b}. \tag{2}$$

We will also streamline his notation.

The Solution in Terms of Krylov Sequences

The starting point of Lanczos's development is the sequence of vectors in the matrix

$$\mathbf{K}_k = (\mathbf{b}\ \mathbf{A}\mathbf{b}\ \mathbf{A}^2\mathbf{b} \ldots \mathbf{A}^k\mathbf{b}) \equiv (\mathbf{b}_0\ \mathbf{b}_1\ \mathbf{b}_2 \ldots\ \mathbf{b}_k).$$

Today we call this sequence a Krylov sequence.[1] Lanczos notes that there must be a smallest integer m for which $\mathbf{K}_m$ is not of full column rank. Hence there must be a vector $\mathbf{g} = (g_0, g_1, \ldots, g_{m-1}, 1)^{\mathrm{T}}$ such that

$$\mathbf{K}_m\mathbf{g} = 0.$$

To say the same thing, if we set

$$g(x) = g_0 + g_1x + \cdots + g_{m-1}x^{m-1} + x^m, \tag{3}$$

then the vector $\hat{\mathbf{g}} = g(\mathbf{A})\mathbf{b}$ is zero.[2]

[1] Krylov [5] used this sequence, as Lanczos will do, to compute the minimal polynomial of $\mathbf{A}$. In a footnote, Lanczos cites Krylov's paper and adds, "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." Householder [4] gives a detailed treatment of Krylov's method and its relation to other algorithms for the algebraic eigenvalue problem.

[2] The conventions introduced here will be used throughout this survey. If $\mathbf{p}$ is a vector, $p(x)$ will denote the polynomial formed in analogy with (3), and $\hat{\mathbf{p}}$ will denote $p(\mathbf{A})\mathbf{b}$. If the polynomial p is monic, we shall say that the vector $\mathbf{p}$ is also monic.

Now, from the minimality of m and the nonsingularity of $\mathbf{A}$ it follows that $g_0 \neq 0$. Let

$$h(x) = \frac{g(x) - g_0}{g_0 x}.$$

Then it is easily verified that $\hat{\mathbf{h}} \equiv h(\mathbf{A})\mathbf{b} = \mathbf{A}^{-1}\mathbf{b}$. In other words, the equation

$$\mathbf{y} = \mathbf{K}_{m-1}\mathbf{h} \tag{4}$$

expresses the solution of (1) as a linear combination of the vectors of the Krylov sequence.

Lanczos goes on to show that the zeros of g are eigenvalues of $\mathbf{A}$ and to express the corresponding eigenvectors in terms of the Krylov sequence. We will not follow this development but will move on to his method for finding the integer m.

Hankel Systems and Orthogonal Polynomials

Lanczos observes that the matrix $\mathbf{K}_m$ is degenerate if and only if the matrix $\mathbf{M}_m = \mathbf{K}_m^{\mathrm{T}}\mathbf{K}_m$ is singular. To determine the first m for which $\mathbf{M}_m$ is singular, he considers the sequence of equations

$$\mathbf{M}_k\mathbf{p}_k = r_k\mathbf{e}_{k+1}, \qquad k = 1, 2, \ldots \tag{5}$$

where $\mathbf{p}_k$ is monic and $\mathbf{e}_{k+1}$ denotes the $(k+1)$th coordinate vector. The scalar r_k is regarded as an unknown, providing the degree of freedom missing in the monic vector $\mathbf{p}_k$. If $\mathbf{M}_{k-1}$ is nonsingular, this equation has a unique solution. The desired m, then, is the first value of k for which r_k turns out to be zero, and in this case $\mathbf{g} = \mathbf{p}_m$.

In order to solve the systems (5), Lanczos takes advantage of the fact that the matrices $\mathbf{M}_k$ are Hankel matrices; i.e. m_{ij} depends only $i + j$ (Lanczos calls such a system a *recurrent* set of equations). Setting

$$\mathbf{N}_k = \mathbf{K}_k^{\mathrm{T}}\mathbf{A}\mathbf{K}_k,$$

he introduces the systems

$$\mathbf{N}_k\mathbf{q}_k = s_k\mathbf{e}_{k+1}, \tag{6}$$

and derives a double recursion to generate the vectors $\mathbf{p}_k$ and $\mathbf{q}_k$. He calls the scheme the *progressive algorithm*.

The solutions of (5) and (6) will play a central role in Lanczos's development, and it may help the reader to know in advance what they are. The vectors $\hat{\mathbf{p}}_k =$

$p(\mathbf{A})\mathbf{b}$ are the vectors that would result from orthogonalizing the Krylov sequence $\mathbf{b}, A\mathbf{b}, \ldots$. The vectors $\hat{\mathbf{q}}_k$ are the vectors that would result from orthogonalizing the same sequence with respect to $\mathbf{A}$; i.e., $\hat{\mathbf{q}}_k^{\mathrm{T}}\mathbf{A}\hat{\mathbf{q}}_l = 0$ for $k \neq l$. This kind of orthogonalization is commonly called *conjugation* with respect to $\mathbf{A}$. The orthogonality relations are easily established from (5) and (6) respectively (the trick is to pad the vectors $\mathbf{p}$ and $\mathbf{q}$ with zeros so that they are all of the same dimension).[3]

The Method of Minimized Iterations

Lanczos now points out that expansions such as (4) have the defect that with increasing k the Krylov vectors tend to lie in the space of the dominant eigenvectors, and information about the smaller eigenvalues is lost. In order to circumvent this difficulty, he proposes to replace the Krylov sequence with a sequence $\hat{\mathbf{p}}_k$, defined by the recurrence $\hat{\mathbf{p}}_0 = \mathbf{b}$,

$$\hat{\mathbf{p}}_{k+1} = \mathbf{A}\hat{\mathbf{p}}_k - \alpha_k\hat{\mathbf{p}}_k - \beta_k\hat{\mathbf{p}}_{k-1} - \gamma_k\hat{\mathbf{p}}_{k-2} - \cdots, \tag{7}$$

where the coefficients $\alpha_k, \beta_k, \gamma_k, \ldots$ are chosen to minimize the length of $\hat{\mathbf{p}}_{k+1}$. He then points out that the vectors $\hat{\mathbf{p}}_k$ are orthogonal and satisfy a three term recurrence relation; i.e., the γ's and all higher coefficients in (7) are zero. Finally, he writes down the corresponding three-term recurrence for the polynomials $p_k(x)$ and asserts without proof that they are the same as the p-polynomials generated by (5).

The rest of the first paper is devoted to the use of the p-polynomials to solve the eigenvalue problem. The application to linear systems is found in the second paper, to which we now turn.

The Conjugate Gradient Method

Lanczos begins by observing that problem of solving the system (1) is easier than the problem of finding all the eigenvectors of $\mathbf{A}$ and therefore will have a simpler solution. After repeating the characterization of the polynomial $p_k(x)$ as minimizing the length of $\hat{\mathbf{p}}_k$, he goes on to assert that among all monic polynomials

[3] Lanczos's progressive algorithm also provides the wherewithal for a fast Hankel solver in the spirit of Levinson's fast Toeplitz solver [8]. It is unlikely that Lanczos knew of Levinson's work. Although Lanczos is not one for searching the literature, he is punctilious about acknowledging priorities.

$q_k(x)$ minimizes the length of $\hat{\mathbf{q}}_k/q_k(0)$. A consequence is that if we set

$$s_k(x) = \frac{q_k(x) - q_k(0)}{q_k(0)x}$$

then $\hat{\mathbf{s}}_k$ is a minimum residual solution; i.e., the the linear combination of the first k Krylov vectors that minimizes the residual $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$.

But Lanczos has other fish to fry. First he recasts the double recurrence between the solutions of (5) and (6) in the form

$$\begin{aligned} \hat{\mathbf{p}}_{k+1} &= \rho_k \hat{\mathbf{p}}_k + \mathbf{A}\hat{\mathbf{q}}_k, \\ \hat{\mathbf{q}}_{k+1} &= \sigma_k \hat{\mathbf{q}}_k + \hat{\mathbf{p}}_{k+1}, \end{aligned}$$

where

$$\rho_k = -\frac{\hat{\mathbf{p}}_k^{\mathrm{T}} \hat{\mathbf{q}}_k}{\hat{\mathbf{p}}_k^{\mathrm{T}} \hat{\mathbf{p}}_k}$$

and

$$\sigma_k = \frac{\hat{\mathbf{p}}_k^{\mathrm{T}} \mathbf{A} \hat{\mathbf{p}}_{k+1}}{\hat{\mathbf{p}}_k^{\mathrm{T}} \mathbf{A} \hat{\mathbf{q}}_k} = \frac{\hat{\mathbf{p}}_{k+1}^{\mathrm{T}} \hat{\mathbf{p}}_{k+1}}{\hat{\mathbf{p}}_k^{\mathrm{T}} \mathbf{A} \hat{\mathbf{q}}_k}.$$

He then considers the expansion

$$\mathbf{y} = \sum_{i=0}^{n-1} \eta_i \hat{\mathbf{q}}_i,$$

and shows that $\eta_i = -1/p_{i+1}(0)$. Thus a sequence of approximate solutions can be generated in the form

$$\mathbf{y}_k = \sum_{i=0}^{k-1} \eta_i \hat{\mathbf{q}}_i$$

which can be updated as the $\hat{\mathbf{p}}_i$ and $\hat{\mathbf{q}}_i$ are generated. The resulting algorithm is easily seen to be the usual method of conjugate gradients with a slightly altered scaling and with the initial approximation $\mathbf{y}_0 = 0$.[4] Lanczos goes on to point out that the process can be started with any residual and that the sequence of residual vectors is given by $\mathbf{r}_{k+1} = -\eta_k \hat{\mathbf{p}}_{k+1}$.

[4]The name "conjugate gradients" is due to Hestenes and Stiefel [3]. The gradients are the residuals, which are actual gradients of an associated minimization problem. The $\hat{\mathbf{q}}_i$ are the conjugate gradients, obtained by conjugating the residuals.

Chebyshev Purification

Lanczos notes that owing to rounding errors it may be necessary to reorthogonalize the sequence $\{\hat{\mathbf{p}}_k\}$, an uneconomical process since all the vectors must be retained. He therefore suggests reducing the number of iterations by purifying the starting vector of components along the larger eigenvalues of $\mathbf{A}$.

To accomplish this goal, Lanczos uses Chebyshev polynomials, but not in the usual way. Instead he constructs an iteration polynomial with the property that the residual is given by $f_{k+1}(\mathbf{A})\mathbf{b}$, where $f(x)$ is a polynomial that exhibits a downward trend as x varies from 0 to an estimate of the largest eigenvalue of A. This iteration becomes part of a two stage algorithm, in which the Chebyshev-based iteration is used to reduce the contributions of the largest eigenvalues and conjugate gradients is then used to eliminate the contributions of the small eigenvalues. Lanczos recommends this procedure for ill-posed problems, in which the contributions of the smaller eigenvalues can actually magnify the error in the solution.

Lanczos's two step procedure is not used today, though the practice of *proconditioning* a system before using the method of conjugate gradients is a common. In general the effects of rounding error on the method are complicated and still imperfectly understood.[5]

Conclusions

Because Lanczos starts from first principles and only specializes when necessary, these papers still repay the effort required to read them. For example, Lanczos does not confine himself to symmetric matrices, but works initially with biorthogonal and biconjugate systems of polynomials. Moreover, his approach makes clear the relation of the method of conjugate gradients to the Lanczos algorithm — something of a mystery to many people — and of both to orthogonal polynomials. Lanczos speaks with rare good sense about the systems of equations arising from ill-posed problems, and it would be rash to suggest that we have advanced very far beyond him today.

There remains the question of who should be credited with the discovery of the method of conjugate gradients. Lanczos, Hestenes [2], and Stiefel [9] each published the method at about the same time in single-authored papers. It is interesting that they approach the method from different angles. We have already followed Lanczos's journey from Krylov sequences to conjugate gradients. Stiefel

[5] The most complete error analysis of the method to date is that of Greenbaum [1].

derives the method from minimizing properties of polynomial iterations. Hestenes gives no indication of how he discovered the method; he simply writes down the formulas and proves that they work. Although Stiefel's paper is clearly independent of the other two, there must have been some interaction between Lanczos and Hestenes, since at the time they both were together at the National Bureau of Standards and Hestenes acknowledges discussions with a number of people, including Lanczos. However, Lanczos specifically notes the independence of both Hestenes and Stiefel's derivations. I am inclined to let Lanczos have the last word and call the algorithm the conjugate gradient method of Hestenes, Lanczos, and Stiefel — in alphabetical order.

Acknowledgement

I am indebted to Dianne O'Leary for her useful comments.

References

[1] A. Greenbaum. Behavior of slightly perturbed lanczos conjugate gradient recurrences. *Linear Algebra and Its Applications*, 113:7–63, 1989.

[2] M. R. Hestenes. Iterative methods for solving linear equations. Report 52-9, NAML, 1951. Reprinted in Journal of Opeimization Theory and Applications, Volume 11, pp. 323–334, 1973.

[3] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[4] A. S. Householder. *The Theory of Matrices in Numerical Analysis.* Dover Publishing, New York, 1964. Originally published by Ginn Blaisdell.

[5] A. N. Krylov. O Čislennom rešenii uravnenija, kotorym v techničeskih voprasah opredeljajutsja častoy malyh kolebaniĭ material'nyh. *Izv. Adad. Nauk SSSR otd. Mat. Estest.*, pages 491–539, 1931. Cited in [4].

[6] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45:255–282, 1950.

[7] C. Lanczos. Solutions of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49:33–53, 1952.

[8] N. Levinson. The Wiener RMS (Root Mean Square) error criterion in filter design and prediction. *Journal of Mathematics and Physics*, 25:261–278, 1946.

[9] E. Stiefel. Über einige Methoden der Relaxationsrechnung. *Zeitschrift für angewandte Mathematik und Physik*, 3:1–33, 1952.