Comm. Korean Math. Soc. **12** (1997), No. 1, pp. 145–155

INTER-CONVERSION BETWEEN THE POWER AND ARNOLDI'S METHODS

PIL SEONG PARK

ABSTRACT. We present a couple of tools that can be used in the solution of nonsymmetric eigenvalue problems. The first one allows us to convert power iterates into Arnoldi's results so that a few eigenpairs are easily obtainable. The other converts Arnoldi's results into power iterates to simulate the power method and improve the result. Suggestions for application are also given.

1. Introduction

Many problems in science and engineering require computation of one or a few eigenvalues with the largest modulus of a large sparse nonsymmetric real matrix. From the numerical point of view, nonsymmetric problems are substancially more difficult to solve than the symmetric ones.

For accelerating linear iterative methods for solving large nonsymmetric linear systems, extensive work has been devoted to polynomial-based acceleration techniques, the basis of which is the power method. Most of them use Chebyshev polynomials to accelerate convergence(e.g., see [2, 4, 5, 6, 10]).

Saad [8] adapted Manteuffel's linear system algorithm to nonsymmetric eigenvalue problems $A\mathbf{x} = \lambda\mathbf{x}$ using a Chebyshev polynomial over an ellipse containing all of the eigenvalues except the desired one. Combining with Arnoldi's method, the algorithm can also be used to obtain several eigenpairs simultaneously, if the ellipse is chosen not to include the corresponding eigenvalues. However, because of its inherent problem, like its symmetric counterpart Lanczos method, Arnoldi's method

Received January 22, 1996. Revised November 2, 1996.
1991 AMS Subject Classification: 65F15, 65F25.
Key words and phrases: Arnoldi's method, the power method, inter-conversion, nonsymmetric eigenvalue problems, Krylov subspace.

itself is subject to severe cancellation errors. In addition, to check convergence and to determine when to stop, iterative algorithms normally require some power iterates anyhow.

In this paper, we introduce a couple of tools that may be used in various ways to get better performance of some algorithms for nonsymmetric eigenproblems. That is, we first devise a method that can compute exact Arnoldi's results from power iterates so that it is far less sensitive to cancellation errors because it uses orthogonal triangularization. Then we devise a method for the reverse way so that the Arnoldi's result can be used to simulate the power method and improve the obtained solution. Both of them do not require costly matrix-vector multiplication with the original large matrix but some small amount of work with matrices far smaller than the original.

We first review Arnoldi's method briefly. Then we derive the two tools that can be used to enhance convergence of the algorithms for nonsymmetric eigenvalue problems. Suggestions for application are given at the end.

NOTATIONS. We denote the identity matrix of order n by I_n. Sometimes we use superscripts in parentheses to denote the related dimension. For example, $\lambda_i^{(m)}$ is the ith eigenvalue of some $m \times m$ matrix, and $\mathbf{e}_j^{(m)}$ is the jth unit vctor of dimension m.

2. Arnoldi's method

Arnoldi's method, first introduced in [1], is a nonsymmetric variant of Lanczos algorithm. The algorithm generates a sequence of upper Hessenberg matrices whose eigenpairs are good approximations to a few of those of the original matrix A.

Starting with some initial vector $\mathbf{v}_1$ of Euclidean norm 1, the method generates a finite sequence of vectors by the recursion

$$h_{j+1,j}\mathbf{v}_{j+1} = A\mathbf{v}_j - \sum_{i=1}^{j} h_{i,j}\mathbf{v}_i, \quad j = 1, \ldots, m$$

where h_{ij}'s are chosen so that $\mathbf{v}_{j+1} \perp \mathbf{v}_i$, $i = 1, \ldots, j$ and $||\mathbf{v}_{j+1}||_2 = 1$. The algorithm stops for $j \leq m$ if $h_{j+1,j} = 0$.

The following theorem is well-known(for example, see [7]).

THEOREM 1. *Let* $V_m = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$.

(1) *The set* $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ *computed by the algorithm is an orthonormal basis of the Krylov subspace* $\mathcal{K}_m = span\{\mathbf{v}_1, A\mathbf{v}_1, \dots, A^{m-1}\mathbf{v}_1\}$.

(2) *The matrix* $H_m = V_m^T A V_m$ *is an upper Hessenberg matrix with elements* h_{ij}.

(3) *The Ritz values of* A *in* $\mathcal{K}_m$ *are the eigenvalues* $\lambda_i^{(m)}$ *of* H_m, *and the Ritz vectors are* $V_m \mathbf{y}_i^{(m)}$, *where* $\mathbf{y}_i^{(m)}$ *is an eigenvector of* H_m *associated with* $\lambda_i^{(m)}$

(1) is true in exact arithmetic, and $A, V, H \in \mathbf{R}^{n\times n}$ satisfies an exact Arnoldi's relation

$$(1) \qquad H = V^T A V \quad \text{or} \quad AV = VH.$$

However in finite precision calculation, Arnoldi's method is subject to severe cancellation errors and the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ can be far from orthonormal. Hence the algorithm should be stopped far before n.

It is quite simple to check step by step whether the desired accuracy is attained and to stop by checking the residual norm by using the formula

$$||(A - \lambda^{(m)} I)\phi^{(m)}|| = h_{m+1,m} |\mathbf{e}_m^{(m)^T} \mathbf{y}^{(m)}|,$$

which is a direct consequence of the following equality derived from the algorithm:

$$AV_m = V_m H_m + h_{m+1,m} \mathbf{v}_{m+1} \mathbf{e}_m^{(m)^T}.$$

A major limitation of Arnoldi's method is that its cost and storage requirements increase drastically as the number of steps m increases. One simple way to overcome the storage problem is to use an *iterative Arnoldi's method* [7].

The amount of work for each Arnoldi iteration is roughly m matrix-vector multiplications and the work needed to solve a Hessenberg eigenvalue problem of size $m \times m$. Hence m should be much smaller than the size of the original problem so that the amount of work for the step is negligibly small.

3. Inter-conversion between the power and Arnoldi's methods

In this section, we discuss how to convert power iterates into Arnoldi's results and vice versa without additional costly matrix-vector multiplications. For this purpose, we need introduce

LEMMA 1. *Let* $\mathbf{x} \in \mathbf{R}^n$, $\sigma = \pm||\mathbf{x}||_2$, *and suppose that* $\mathbf{x} \neq -\sigma\mathbf{e}_1$. *Let*

$$\mathbf{u} = \mathbf{x} + \sigma\mathbf{e}_1, \quad \pi = \frac{1}{2}||\mathbf{u}||_2^2.$$

If $U = I - \pi^{-1}\mathbf{u}\mathbf{u}^T$, *then* $U\mathbf{x} = -\sigma\mathbf{e}_1$.

PROOF. [9]
We take σ to have the same sign as the first component of $\mathbf{x}$. The matrix U annihilates all the components of the vector $\mathbf{x}$ except the first one. It is known as a Householder transformation, and has the property that it is symmetric, orthogonal, and hence involutory(i.e. $U^2 = I$). As a result, we have

LEMMA 2. *Let* $U \in \mathbf{R}^{(n-j+1)\times(n-j+1)}$ *be a Householder transformation as defined above. Premultiplying a vector* $\mathbf{z} \in \mathbf{R}^n$ *whose last* $n-j+1$ *entries are the vector* $\mathbf{x}$ *by the matrix*

$$U_j \equiv \begin{bmatrix} I_{j-1} & 0 \\ 0 & U \end{bmatrix}$$

annihilates all the components of $\mathbf{z}$ *below the* j*th one, others above it being unchanged.*

The matrix U_j is symmetric, orthogonal, and involutory too. By premultiplying by appropriate U_j, $j = 1, \ldots, m$ consecutively, any matrix with m columns can be orthogonally upper-triangularized.

Getting Arnoldi's Hessenberg form from power iterates

Suppose we have applied m power iterations with the matrix $A \in \mathbf{R}^{n\times n}$ to the initial vetor $\mathbf{v} \in \mathbf{R}^n$, and let the results be $\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \ldots, A^m\mathbf{v}$. Assume $m \ll n$. We construct an upper Hessenberg matrix $H^{(m)} \in \mathbf{R}^{m\times m}$ satisfying the Arnoldi's relation (1).

Let $M \in \mathbf{R}^{n\times m}$ be the matrix whose columns consist of the power iterates, i.e.

$$M = [\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^{m-1}\mathbf{v}]$$

and let

$$K \equiv [A\mathbf{v}, A^2\mathbf{v}, \dots, A^{m-1}\mathbf{v}, \mathbf{v}].$$

Then $K = MP$, or equivalently,

$$M = KP^T \tag{2}$$

where $P = [\mathbf{e}_2, \dots, \mathbf{e}_m, \mathbf{e}_1] \in \mathbf{R}^{m\times m}$ is a permutation matrix. Consider another matrix

$$K' \equiv [A\mathbf{v}, A^2\mathbf{v}, \dots, A^m\mathbf{v}] = AM. \tag{3}$$

Since all the columns, except the last one, of K and K' are the same, orthogonal triangularization on K and K' yields

$$U_m U_{m-1} \cdots U_2 U_1 [A\mathbf{v}, A^2\mathbf{v}, \dots, A^{m-1}\mathbf{v}, \mathbf{v}] = R,$$
$$U'_m U_{m-1} \cdots U_2 U_1 [A\mathbf{v}, A^2\mathbf{v}, \dots, A^{m-1}\mathbf{v}, A^m\mathbf{v}] = R',$$

where U_j, $j = 1, \dots, m$ and U'_m are as defined in Lemma 2, and R and R' are upper triangular. Letting $Q^T = U_m U_{m-1} \cdots U_2 U_1$ and $Q'^T = U'_m U_{m-1} \cdots U_2 U_1$, we obtain

$$Q^T K = R, \qquad Q'^T K' = R'$$

or

$$K = QR, \qquad K' = Q'R', \tag{4}$$

where $Q, Q' \in \mathbf{R}^{n\times n}$ and $R, R' \in \mathbf{R}^{n\times m}$. These are the QR factorizations(e.g., see [3]) of K and K'.

From (2), (3), and (4),

$$K' = AM = AKP^T = AQRP^T. \tag{5}$$

In addition, since

$$Q' = U_1 U_2 \dots U_{m-1} U'_m = Q U_m U'_m,$$

K' can also be written by (4) as

$$K' = Q'R' = QU_m U'_m R'. \tag{6}$$

Since only the first m rows of R and R' are nonzero, the QR factorizations of K and K' can be written as

$$K = \overline{QR}, \quad K' = \overline{Q'R'},$$

where $\overline{Q}, \overline{Q'} \in \mathbf{R}^{n\times m}$ consist of the first m columns of Q and Q', and $\overline{R}, \overline{R'} \in \mathbf{R}^{m\times m}$ consist of the first m rows of R and R', respectively. Hence, from (5) and (6),

$$A\overline{QR}P^T = \overline{Q}W\overline{R'},$$

where W is the $m \times m$ principal submatrix of $U_m U'_m$. Hence

$$A\overline{Q} = \overline{Q}W\overline{R'}P\overline{R}^{-1}.$$

Lettting

$$H^{(m)} = W\overline{R'}P\overline{R}^{-1}, \tag{7}$$

THEOREM 2. *The matrix $H^{(m)}$ defined above is the upper Hessenberg matrix that can be obtained by Arnoldi's method.*

PROOF. Since $\overline{Q}$ has orthonormal columns, $H^{(m)}$ satisfies the Arnoldi's relation. Now it remains to show that $H^{(m)}$ is upper Hessenberg. Since $\overline{R}$ and $\overline{R'}$ are upper triangular, so is $\overline{R}^{-1}$. Since $P = [\mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_m, \mathbf{e}_1]$, P is upper Hessenberg and so is $\overline{R'}P\overline{R}^{-1}$. Note that W is a diagonal matrix because it is the $m \times m$ principal submatrix of $U_m U'_m$ by Lemma 2. Hence their product $H^{(m)}$ must be upper Hessenberg.

To compute $H^{(m)}$, we should not form the matrix as is. Instead, postmultiplying (7) by $\overline{R}$ and taking transpose, we get

$$\overline{R}^T H^{(m)^T} = P^T \overline{R'}^T W$$

since W is diagonal. After computing the righthand side, we only have to solve the lower triangular system m times to obtain $H^{(m)^T}$.

R and R' are the results of QR factorization. But since only the last colums of K and K' are different, we consider the following $n \times (m+1)$ augmented matrix

$$\overline{K} = [K, A^m\mathbf{v}] = [A\mathbf{v}, A^2\mathbf{v}, \dots, A^{m-1}\mathbf{v}, \mathbf{v}, A^m\mathbf{v}].$$

After upper-triangularization of the first $m-1$ columns by premultiplying by appropriate $U_1, \dots, U_{m-1} \in \mathbf{R}^{n\times n}$, let $U, U' \in \mathbf{R}^{(n-m+1)\times(n-m+1)}$ be the Householder transformations that introduce zeros below the mth position in the last two columns of the resulting matrix respectively. Let $\mathbf{x} = (x_i), \mathbf{x}' = (x_i{}') \in \mathbf{R}^{(n-m+1)}$ be the subvectors that consist of the last $n-m+1$ entries of the last two columns respectively. Let

$$\sigma = sign(x_1)||\mathbf{x}||_2, \qquad \sigma' = sign(x_1{}')||\mathbf{x}'||_2.$$

By premultiplying the resulting last two columns by U and U', we get R and R'.

Now by Lemma 1, we can write

$$U = I - c\mathbf{u}\mathbf{u}^T \quad \text{where} \quad \mathbf{u} = \mathbf{x} + \sigma\mathbf{e}_1, \quad c = 2/||\mathbf{u}||_2^2,$$
$$U' = I - c'\mathbf{u}'\mathbf{u}'^T \quad \text{where} \quad \mathbf{u}' = \mathbf{x}' + \sigma'\mathbf{e}_1, \quad c' = 2/||\mathbf{u}'||_2^2.$$

Hence

$$\begin{aligned} UU' &= (I - c\mathbf{u}\mathbf{u}^T)(I - c'\mathbf{u}'\mathbf{u}'^T) \\ &= I - c\mathbf{u}\mathbf{u}^T - c'\mathbf{u}'\mathbf{u}'^T + cc'(\mathbf{u}^T\mathbf{u}')\mathbf{u}\mathbf{u}'^T \end{aligned}$$

and its $(1,1)$-component is

$$w \equiv (UU')_{11} = 1 - cu_1u_1 - c'u'_1u'_1 + cc'(\mathbf{u}^T\mathbf{u}')u_1u'_1$$