Analysis of Accelerating Algorithms for the Restarted Arnoldi Iteration

Akira Nishida, Yoshio Oyanagi
Department of Information Science, Faculty of Science,
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113 Japan
e-mail: nishida@is.s.u-tokyo.ac.jp

Keywords: nonsymmetric eigenvalue problems, restarted Arnoldi iteration, polynomial acceleration

ABSTRACT

We present an approach for the acceleration of the restarted Arnoldi iteration for the computation of a number of eigenvalues of the standard eigenproblem $Ax = \lambda x$. This study applies the Chebyshev polynomial to the restarted Arnoldi iteration and proves that it computes necessary eigenvalues with far less complexity than the QR method. We also discuss the dependence of the convergence rate of the restarted Arnoldi iteration on the distribution of spectrum.

BACKGROUND

In the past five years, there have been great progress in the further developments of the methods for the standard eigenproblem. Arnoldi's method, which had the defect of increasing computational complexity per iteration step, was much improved by Saad [8] with the explicitly restarting technique, by which the dimensions of the searchspaces can be kept modest. Although the restarted Arnoldi iteration is quite effective, the dimension of the subspace is inevitably large, in particular when the wanted eigenvalues are clustered. Moreover it favors the convergence on the envelope of the spectrum. In this paper, we use the convex hull proposed for the solution of the nonsymmetric linear system to accelerate the convergence of the restarted Arnoldi iteration.
The algorithm of the explicitly restarted Arnoldi iteration is summarized in Table 1. The choice of subspace dimension m is usually a tradeoff between the length of the reduction that may be tolerated and the rate of convergence. The accuracy of the Ritz values typically increases as m does. For most problems, the size of m is determined experimentally.
Suppose A is diagonalizable with eigenpairs (u_j, λ_j) for $j = 1, ..., n$. If $\psi(\cdot)$ is some polynomial and we expand the current starting vector x_1 in terms of the basis of eigenvectors, then

$$\psi(A)x_1 = u_1\psi(\lambda_1)\zeta_1 + \cdots + u_n\psi(\lambda_n)\zeta_n$$

Assuming that the eigenpairs (u_i, λ_i) are ordered so that the wanted k ones are at the beginning of the expansion, we seek a polynomial such that

$$\max_{i=k+1,\ldots,n} |\psi(\lambda_i)| < \min_{i=1,\ldots,k} |\psi(\lambda_i)|.$$

Components in the direction of unwanted eigenvectors are dumped.
The acceleration techniques and hybrid methods presented by Saad [8] attempt to improve the explicitly restarted Arnoldi iteration by approximately solving this min-max problem. Motivated by Manteuffel's scheme [6], Saad proposed the use of Chebyshev polynomials. A Chebyshev polynomial $\psi(A)$ on an ellipse containing the unwanted Ritz values is applied to the restart vector in an attempt

to accelerate convergence of the original explicitly restarted Arnoldi iteration. The polynomial is applied with the use of the familiar three-term recurrence.

The choice of ellipses as enclosing regions in Chebyshev acceleration may be overly restrictive and ineffective if the shape of the convex hull of the unwanted eigenvalues bears little resemblance with an ellipse. This has spurred much research in which the acceleration polynomial is chosen so as to minimize an L_2 norm of the polynomial ψ on the boundary of the convex hull of the unwanted eigenvalues with respect to some suitable weight function ω. The only restriction with this technique is that the degree of the polynomial is limited because of cost and storage requirements. This, however, is overcome by compounding low degree polynomials. The stability of the computation is enhanced by employing a Chebyshev basis. It has been shown that the least-squares based method for solving linear systems is competitive with the ellipse based methods and are more reliable [7, 8].

For convenience we can always normalize the polynomial so that $\psi(\lambda_1) = 1$. The desired polynomial satisfying the above constraint can be sought in the form $\psi_n(\lambda) \equiv 1 - \lambda s_n(\lambda)$. By the maximum principle, the maximum modulus of $|1 - \lambda s_n(\lambda)|$ is found on the boundary of some region H of the complex plane that includes the spectrum of A and it is sufficient to regard the problem as being defined on the boundary. We use the least squares residual polynomial minimizing the L_2 norm $\| 1 - \lambda s_n(\lambda) \|_w$ with respect to some weight $w(\lambda)$ on the boundary of H [8]. Suppose that the $\mu + 1$ points $h_0, h_1, \cdots, h_\mu$ constitute the vertices of H. On each edge E_ν, $\nu = 1, \cdots, \mu$, of the convex hull, we choose a weight function $w_\nu(\lambda)$. Denoting by c_ν the center of the νth edge and by d_ν the half width, i.e., $c_\nu = (h_\nu + h_{\nu-1})/2$, $d_\nu = (h_\nu - h_{\nu-1})/2$, the weight function on each edge is defined by $w_\nu(\lambda) = 2|d_\nu^2 - (\lambda - c_\nu)^2|^{-\frac{1}{2}}/\pi$. The inner product on the space of complex polynomials is defined by $\langle p, q \rangle = \sum_{\nu=1}^{\mu} \int_{E_\nu} p(\lambda)\overline{q(\lambda)} w_\nu(\lambda)|d\lambda|$. An algorithm using explicitly the modified moments $\langle t_i(\lambda), t_j(\lambda) \rangle$, where $\{t_j\}$ is some suitable basis of polynomials, is developed for the problem of computing the least squares polynomials in the complex plane.

We express the polynomial $t_j(\lambda)$ in terms of the Chebyshev polynomials $t_j(\lambda) = \sum_{i=0}^{j} \gamma_{i,j}^{(\nu)} T_i(\xi)$ where $\xi = (\lambda - c_\nu)/d_\nu$ is real. The expansion coefficients $\gamma_{i,j}^{(\nu)}$ can be computed easily from the three term recurrence of the polynomials $\beta_{k+1} t_{k+1}(\lambda) = (\lambda - \alpha_k) t_k(\lambda) - \delta_k t_{k-1}(\lambda)$. The problem $\min_{s \in \psi_{n-1}} \| 1 - \lambda s_n(\lambda) \|_w$ is to find $\eta = (\eta_0, \eta_1, \cdots, \eta_{n-1})^T$ of $s_n(\lambda) = \sum_{i=0}^{n-1} \eta_i t_i(\lambda)$ so that $J(\eta) = \| 1 - \lambda s_n(\lambda) \|_w$ is minimum.

Table 1. A block version of explicitly restarted Arnoldi reduction with polynomial acceleration

1. Choose $V_1 \in \mathrm{R}^{n \times r}$.
2. For $j = 1, ..., m - 1$ do
   $W_j = AV_j$
   For $i = 1, ..., j$ do
   $H_{i,j} = V_i^T W_j$; $W_j = W_j - V_i H_{i,j}$
   end for
   $Q_j R_j = W_j$; $V_{j+1} = Q_j$; $H_{j+1,j} = R_j$
   end for
3. Compute the eigenvalues of $H_m = (H_{i,j}) \in \mathrm{R}^{mr \times mr}$ and select $\{\tilde{\lambda}_1, ..., \tilde{\lambda}_r\}$ of largest real parts.
4. Stop if their Ritz vectors $\tilde{X}_0 = \{\tilde{x}_1, ..., \tilde{x}_r\}$ satisfy the convergence criteria.
5. Define the iteration polynomial $p_k(\lambda)$ of degree k by $\mathrm{Sp}(H_m) - \{\tilde{\lambda}_1, ..., \tilde{\lambda}_r\}$.
6. $\tilde{X}_k = p_k(A)\tilde{X}_0$; $Q_k R_k = \tilde{X}_k$; $V_1 = Q_k$
7. Goto 2.

APPROACH

In the previous section we described the outline of the least-squares based method on any arbitrary area. It has a difficulty on the application to other purposes due to the constraint $\psi_n(0) = 1$.
We use the fact that the eigenvalue problem does not require any such condition to the polynomial and propose a new simple algorithm to get the mini-max polynomial to accelerate the convergence of the projection method. The minimum property of the Chebyshev functions described below is important to prove the optimality of this polynomial.
Let a non-negative weight function $w(\lambda)$ be given in the interval $a \geq \lambda \geq b$. The orthogonal polynomials $p_0(\lambda), p_1(\lambda), \cdots$, when multiplied by suitable factors C, possess a minimum property: the integral $\int(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0)^2 w(\lambda)d\lambda$ takes on its least value when the polynomial in the integrand is $Cp_n(\lambda)$. The polynomial in the integrand may be written as a linear combination of the $p_i(\lambda)$, in the form $(Cp_n(\lambda) + c_{n-1}p_{n-1}(\lambda) + \cdots c_0)$. Since the functions $p_n(\lambda)\sqrt{w(\lambda)}$ are orthogonal, and in fact, orthogonal if the $p_i(\lambda)$ are appropriately defined, the integral is equal to $C^2 + \sum_{\nu=0}^{n-1} c_\nu^2$, which assumes its minimum at $c_0 = c_1 = \cdots = c_{n-1} = 0$.
Using the above property, we describe the new method to generate the coefficients of the ortho-normal polynomials in terms of the Chebyshev weight below. We use the three term recurrence $\beta_{n+1}p_{n+1}(\lambda) = (\lambda - \alpha_n)p_n(\lambda) - \beta_n p_{n-1}(\lambda)$, where $p_i(\lambda)$ satisfies the ortho-normality. Because of the condition of the use of the Chebyshev polynomial $\psi_n(\lambda) = \sum_{i=0}^{n} \gamma_{i,n}^{(\nu)} T_i[(\lambda - c_\nu)/d_\nu]$, the constraints $\langle\psi_0, \psi_0\rangle = 2\sum_{\nu=1}^{\mu} |\gamma_{0,0}^{(\nu)}|^2 = 1$, $\langle\psi_1, \psi_1\rangle = \sum_{\nu=1}^{\mu}[2|\gamma_{0,1}^{(\nu)}|^2 + |\gamma_{1,1}^{(\nu)}|^2] = 1$, and $\langle\psi_0, \psi_1\rangle = 2\sum_{\nu=1}^{\mu} \gamma_{0,0}^{(\nu)}\overline{\gamma}_{1,1}^{(\nu)} = 0$ must hold. Moreover each expansion of $\psi_i(\lambda)$ at each edge must be consistent.
Using the three term recurrence of the Chebyshev polynomials, a similar recurrence $\beta_{k+1}\psi_{k+1}(\lambda) = (\lambda - \alpha_k)\psi_k(\lambda) - \delta_k\psi_{k-1}(\lambda)$ on $\psi_i(\lambda)$ holds. Denoting ξ_ν by $\xi_\nu = (\lambda - c_\nu)/d_\nu$, the equation can be rewritten as

$$\beta_{k+1}\psi_{k+1}(\lambda) = (d_\nu\xi + c_\nu - \alpha_k)\sum_{i=0}^{k} \gamma_{i,k}^{(\nu)} T_i(\xi) - \delta_k \sum_{i=0}^{k-1} \gamma_{i,k-1}^{(\nu)} T_i(\xi).$$

From the relations $\xi T_i(\xi) = [T_{i+1}(\xi) + T_{i-1}(\xi)]/2,\ i > 0$ and $\xi T_0(\xi) = T_1(\xi)$, it is expressed by

$$\sum \gamma_i \xi T_i(\xi) = \frac{1}{2}\gamma_1 T_0(\xi) + (\gamma_0 + \frac{1}{2}\gamma_2)T_1(\xi) + \cdots + \frac{1}{2}(\gamma_{i-1} + \gamma_{i+1})T_i(\xi) + \cdots + \frac{1}{2}(\gamma_{n-1} + \gamma_{n+1})T_n(\xi),$$

where $\gamma_{n+1} = 0$, and arranged into

$$\beta_{n+1}\psi_{n+1}(\lambda) = d_\nu[\frac{\gamma_{1,n}^{(\nu)}}{2}T_0(\xi) + (\gamma_{0,n}^{(\nu)} + \frac{\gamma_{2,n}^{(\nu)}}{2})T_1(\xi) + \cdots + \sum_{i=2}^{n} \frac{1}{2}(\gamma_{i-1,n}^{(\nu)} + \gamma_{i+1,n}^{(\nu)})T_i(\xi)]$$

$$+(c_\nu - \alpha_n)\sum_{i=0}^{n} \gamma_{i,n}^{(\nu)} T_i(\xi) - \delta_n \sum_{i=0}^{n-1} \gamma_{i,n-1}^{(\nu)} T_i(\xi) \qquad (T_{-1} = T_1).$$

Comparing the equation with $\psi_{n+1}(\lambda) = \sum_{i=0}^{n+1} \gamma_{i,n+1}^{(\nu)} T_i(\xi)$, we find the following relations

$$\beta_{n+1}\gamma_{0,n+1}^{(\nu)} = \frac{1}{2}d_\nu\gamma_{1,n}^{(\nu)} + (c_\nu - \alpha_n)\gamma_{0,n}^{(\nu)} - \delta_n\gamma_{0,n-1}^{(\nu)},$$

$$\beta_{n+1}\gamma_{1,n+1}^{(\nu)} = d_\nu(\gamma_{0,n}^{(\nu)} + \frac{1}{2}\gamma_{2,n}^{(\nu)}) + (c_\nu - \alpha_n)\gamma_{1,n}^{(\nu)} - \delta_n\gamma_{1,n-1}^{(\nu)},$$

and

$$\beta_{n+1}\gamma_{i,n+1}^{(\nu)} = \frac{d_\nu}{2}[\gamma_{i+1,n}^{(\nu)} + \gamma_{i-1,n}^{(\nu)}] + (c_\nu - \alpha_n)\gamma_{i,n}^{(\nu)} - \delta_n\gamma_{i,n-1}^{(\nu)}$$

$$i = 2, ..., n+1 \qquad (\gamma_{-1,n}^{(\nu)} = \gamma_{1,n}^{(\nu)}, \quad \gamma_{i,n}^{(\nu)} = 0 \quad i > n).$$

Using the relation $\beta_{k+1}\psi_{k+1}(\lambda) = (\lambda - \alpha_k)\psi_k(\lambda) - \delta_k\psi_{k-1}(\lambda)$ and the orthogonality of the Chebyshev polynomials, we derive

$$\beta_{k+1} = \langle \psi_{k+1}, \psi_{k+1} \rangle^{1/2} = \sum_{\nu=1}^{\mu} \int_{E_\nu} \psi_{k+1}\overline{\psi_{k+1}} w_\nu(\lambda)|d\lambda| = \sum_{\nu=1}^{\mu} {\sum_{i=0}}'^{k+1} \gamma_{i,k+1}^{(\nu)} \overline{\gamma_{i,k+1}^{(\nu)}}$$

where we denote by $\sum_{i=0}'^{n} a_i = 2a_0 + \sum_{i=1}^{n} a_i$. α and δ are computed similarly:

$$\alpha_k = \langle \lambda\psi_k, \psi_k \rangle = \sum_{\nu=1}^{\mu} (c_\nu {\sum_{i=0}}'^{k} \gamma_{i,k}^{(\nu)} \overline{\gamma_{i,k}^{(\nu)}} + d_\nu {\sum_{i=0}}'^{k} \gamma_{i,k}^{(\nu)} \overline{\gamma_{i+1,k}^{(\nu)}}), \quad \delta_k = \langle \lambda\psi_k, \psi_{k-1} \rangle = \sum_{\nu=1}^{\mu} d_\nu \upsilon_\nu$$

where $\upsilon_\nu = \gamma_{1,k}^{(\nu)} \overline{\gamma_{0,k-1}^{(\nu)}} + (\gamma_{0,k}^{(\nu)} + \frac{1}{2}\gamma_{2,k}^{(\nu)}) \overline{\gamma_{1,k-1}^{(\nu)}} + \sum_{i=2}^{k-1} \frac{1}{2}(\gamma_{i-1,k}^{(\nu)} + \gamma_{i+1,k}^{(\nu)}) \overline{\gamma_{i,k-1}^{(\nu)}}$.

EVALUATION

Denoting by n, nz, m, r, k respectively the order of the matrix, its number of nonzero entries, the number of block Arnoldi steps, the number of required eigenvalues, and the degree of the Chebyshev polynomial, the block Arnoldi method costs $\sum_{j=1}^{m}\{2r\,nz + 4nr^2j + 2r(r+1)n\} = 2rm\,nz + 2mr(mr + 2r + 1)n$ flops. $10r^3m^3$ flops are required for the computation of the eigenvalues of H_m of order mr by the QR method, $r^3\mathcal{O}(m^2)$ for the corresponding eigenvectors by the inverse iteration, and $2kr\,nz + \mathcal{O}(n)$ for the Chebyshev iteration [3, 9]. The computation of the coefficients costs approximately $\mathcal{O}(\mu k^2)$ flops, where μ is the number of the vertices of the convex hull. Table 2 shows that the complexity of the orthogonality-based method is roughly $O(n^2)$, while that of the QR method is $O(n^3)$.

We solved some test problems from the Harwell-Boeing sparse matrix collection [2], the spectral portraits of which are shown in Figure 1, using the block Arnoldi iteration. Ho's algorithm was used for reference. Table 3 and Table 4 indicate that our algorithm shows better performance than Ho's method in the cases where the moduli of the necessary eigenvalues are considerably larger than those of the unnecessary eigenvalues.

Table 5 shows the comparative results on the `ARNCHEB` package [1], the `ARPACK` software package [5], and the Harwell Subroutine Library code `EB13` [4]. `ARNCHEB` provides the subroutine `ARNOL`, which implements the explicitly restarted Arnoldi iteration and the Chebyshev polynomial acceleration. `EB13` implements the similar algorithm and also uses Ho's Chebyshev polynomial acceleration. `ARPACK` provides subroutine `DNAUPD` that implements the implicitly restarted Arnoldi iteration.

From the results of Table 5, we can derive the strong dependency of the polynomial acceleration on the distribution of spectrum. Figure 1 indicates that the non-clustered distribution of spectra causes the slow convergence, since the approximate spectra may completely differ from the accurate ones. Although `ARNCHEB` gives reasonable results for computing a single eigenpair, it can struggle on problems for which several eigenvalues are requested. `ARPACK` displays monotonic consistency and is generally faster and more dependable for small convergence tolerances and large departures from normality. However, its restarting strategy can be more expensive.

CONCLUSION

We simplified the computation of the least-squares polynomial which minimizes its norm on the boundary of the convex hull enclosing unwanted eigenvalues, using the minimum property of the orthogonal polynomials. The number of floating point operations rapidly increases with the size of the subspace dimension m and it indicates that we need to take m as small as possible if we want to avoid QR to become a bottleneck. Although some problems are to be solved, the validity of our method was confirmed by the experiments using the Harwell-Boeing Sparse Matrix Collection. A more detailed analysis of the precision and the complexity of the methods is required.

Table 2. Random matrices of order 50, for the cases of $\lambda_{\max} = 2, 1.5$, and 1.1, while the distribution of the other eigenvalues is $\Re\mathrm{e}\,\lambda \in [0,1]$, and $\Im\mathrm{m}\,\lambda \in [-1,1]$. m, n_c, and n_{iter} denote the order of the Arnoldi method, the maximum order of the Chebyshev polynomials, and the number of the iterations, respectively. CPU times (in seconds) by HP9000/720.

maximum	orthogonality-based					Arnoldi				QR	
eigenvalues	n_{iter}	m	n_c	error	time	n_{iter}	m	error	time	error	time
2	2	5	15	3.6E-15	0.38	2	15	8.9E-16	0.57	5.1E-15	1.87
1.5	3	5	20	3.0E-15	0.70	3	15	3.7E-15	0.82	3.6E-15	1.85
1.1	5	10	20	2.9E-14	1.6	1	50	7.5E-13	3.93	5.2E-15	18.8

*Table 3. Test problems from CHEMWEST, a library in the Harwell-Boeing Sparse Matrix Collection, which was extracted from modeling of chemical engineering plants. The results by Ho's algorithm (right) versus those by the orthogonality-based method (left), with size of the basis 20, degree of the polynomial 20, and block size 1, respectively, are listed. * denotes the algorithm fails to converge. CPU time by Alpha Station 600 5/333.*

problem	WEST0497		WEST0655		WEST0989		WEST2021	
order of matrix	497		655		989		2021	
number of entries	1727		2854		3537		7353	
number of multiplications	924	440	275	120	13751	*	767	320
number of restarts	14	10	3	2	162	*	12	7
CPU time (sec.)	0.37	0.22	0.17	0.12	8.71	*	1.28	0.67

Table 4. Test problems from TOLOSA extracted from fluid-structure coupling (flutter problem). Size of the basis, degree of the polynomial, and block size are 20, 20, 1, respectively.

order of matrix	2000		4000		6000		8000		10000	
number of entries	5184		8784		12384		15984		19584	
number of multiplications	589	240	393	180	236	140	393	380	236	80
number of restarts	7	4	5	3	3	2	5	7	3	1
CPU time (sec.)	0.83	0.43	1.24	0.70	1.23	0.85	2.57	2.81	2.14	0.97

*Table 5. CPU times by IBM RS/6000 3BT and matrix-vector products for computing the right-most eigenvalues of WEST2021 (left) and PORES2 of order 1224 (right). * denotes convergence not reached within $2000m$ matrix-vector products. We denote by r the block size and by m the subspace dimension.*

Algorithm	r=1,m=8	r=5,m=20
EB12	*	98/20930
ARNCHEB	8.6/3233	71/15921
EB13	17/4860	18/4149
ARPACK	3.7/401	2.1/167

Algorithm	r=1,m=12	r=4,m=20
EB12	0.6/423	9.1/2890
ARNCHEB	3.4/1401	4.7/1712
EB13	0.4/119	1.3/305
ARPACK	0.5/90	1.3/151

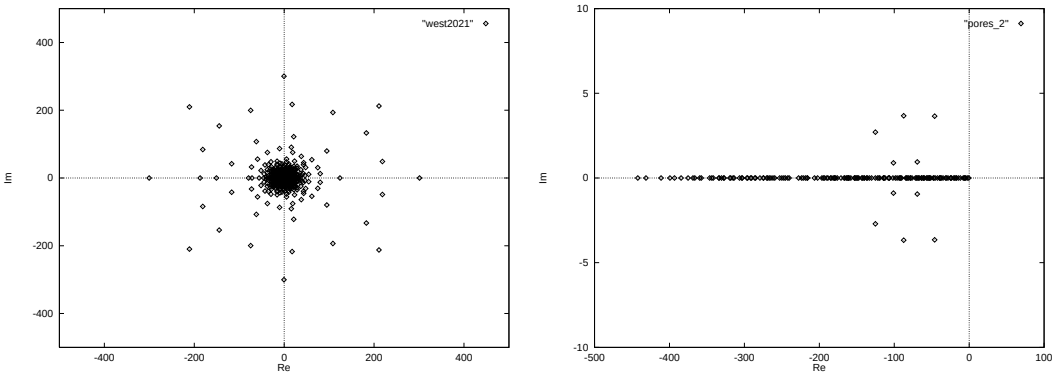

Figure 1. Spectral Portraits of WEST2021 and PORES2

REFERENCES

[1] T. Braconnier. *The Arnoldi-Tchebycheff algorithm for solving nonsymmetric eigenproblems*. Technical Report TR/PA/93/25, CERFACS, Toulouse, 1993.

[2] I. S. Duff, R. G. Grimes, and J. G. Lewis. *Sparse matrix test problems*. ACM Trans. Math. Softw., 15:1–14, 1989.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1996.

[4] R. B. Lehoucq and J. A. Scott. *An evaluation of software for computing eigenvalues of sparse nonsymmetric matrices*. Technical Report MCS-P547-1195, Argonne National Laboratory.

[5] R. B. Lehoucq and D. C. Sorensen. *Deflation techniques for an implicitly restarted Arnoldi iteration*. SIAM Journal on Matrix Analysis and Applications, 17(4), 1996.

[6] T. A. Manteuffel. *The Tchebychev iteration for nonsymmetric linear systems*. Numer. Math., 28:307–327, 1977.

[7] A. Nishida and Y. Oyanagi. *A Polynomial Acceleration of the Projection Method for Large Nonsymmetric Eigenvalue Problems*. 1996 SIAM Annual Meeting, Kansas City, USA, Jul. 1996.

[8] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, Manchester, 1992.

[9] M. Sadkane. *A block Arnoldi-Chebyshev method for computing the leading eigenpairs of large sparse unsymmetric matrices*. Numer. Math., 64:181–193, 1993.