

[WORK IN PROGRESS]

Report on Paper "A Chebshev-Davidson Algorithm For Large
Symmetric Eigenproblems"

by Yunkai Zhou and Yousef Saad

Sourabh Antani

Chapter 1

Introduction

This is a paper report on the paper 'A Chebyshev-Davidson Algorithm For Large Symmetric Eigenproblems' by Yunkai Zhou and Yousef Saad [1]. All the sections in this report are taken or paraphrased from [1] and not original work. The structure of this report also follows the structure of the paper so that it is easy to follow.

Considering standard eigenvalue problem $Au = \lambda u$, with $A \in \mathbb{R}^{n \times n}$ is symmetric, n is large and smallest eigenvalues are wanted.

The Davidson approach sacrifices the attractive Krylov subspace structure, at cost of having to compute eigenpairs and associated residual vectors of a projection matrix at each outer iteration. The advantage, however, is that Davidson approach can augment the subspace by a new vector potentially much better than one based on strict Krylov subspace structure. The augmentation vector added to subspace at each step comes from solution of the correction-equation $\text{diag}(A) - \mu I)t = -r$, where $r = Ax - \mu x$ is the residual vector corresponding to a Ritz pair (μ, x) and t is the augmentation vector to be computed.

This paper, however, presents an alternate approach by using Chebyshev polynomial based intervalwise filtering instead of solving correction equations.

1.1 Rayleigh-Ritz Method

The Rayleigh-Ritz Method [2] is a method commonly applied to approximate the eigen-pairs of a matrix by projecting it onto a smaller dimension matrix.

To approximate the eigenproblem $Ax = \lambda x$, where $A \in \mathbb{C}^{N \times N}$, solve the eigenproblem $V^*AVy_i = \mu_i y_i$ where $V \in \mathbb{C}^{N \times m}$, $m < N$ is an orthonormal matrix. μ_i and Vy_i are then the approximations to the eigenpair of A provided the subspace with basis given by columns of v contains the corresponding eigenvector.

1.2 Advantages of polynomial filtering

This section discusses three observations.

The first observation is polynomial filtering argument: For a symmetric matrix A with the eigendecomposition $A = Q\Lambda Q^T$, any polynomial $\psi(s) : \mathbb{R} \rightarrow \mathbb{R}$ satisfies $\psi(A)v = Q\psi(\Lambda)Q^T v$, $\forall v \in \mathbb{R}^n$.

The second observation is on the fast local convergence of Jacobi-Davidson is because it retains the approximate direction of RQI in the basis of projection subspace. Assuming (μ, x) being the Ritz pair that best approximation of the wanted eigenvalue, and x of unit length, t can be obtained by simplified problem $(I - xx^T)(A - \mu I)t = r$ where the original method required $x \perp t$, $(I - xx^T)(A - \mu I)(I - xx^T)t = r$. Thus dropping the orthogonality constraint and removing the right projection matrix would still maintain the approximation. The left projection matrix is crucial in maintaining the RQI direction in the direction of JD.

The third observation is that one can improve the global convergence by choosing the polynomial $\psi(s)$ which magnifies directions corresponding to an interval containing the wanted eigenvalues and simultaneously, dampens the unwanted eigenvalues. This minimizes the chance that the wanted eigenvalue will be missed. Davidson type methods may miss the wanted eigenvalues because the correction equation resembles shift inverse formulations and chosen shift greatly affects the eigenvalue computed. Thus Chebyshev filtering is a more robust alternative.

1.3 Chebyshev polynomial filtering

The observations in the previous section suggest that polynomials which significantly magnify the lowest end of a wanted interval and dampens unwanted intervals at the same time can be used as a filter to improve global convergence. This method integrates Chebyshev polynomial filtering into the varying dimension Davidson type algorithm, reducing the required subspace much smaller than Lanczos-type method.

The Chebyshev polynomials of the first kind are given by

$$C_k(t) = \begin{cases} \cos(k \cos^{-1}(t)), & -1 \leq t \leq 1 \\ \cosh(k \cosh^{-1}(t)), & |t| > 1 \end{cases}$$

$$C_0(t) = 1, c_1(t) = t, C_{k+1}(t) = 2tC_k(t) - C_{k-1}(t), t \in \mathbb{R}$$

The filtering is based on the property of the Chebyshev polynomial that its magnitude grows rapidly outside the interval $[-1, 1]$. Suppose that the spectrum of A is contained in $[a_0, b]$ and we want to dampen the interval $[a, b]$ for $a > a_0$, then we need to simply map $[a, b]$ into $[-1, 1]$ by an affine mapping (e.g. scaling, rotation etc.), thus the eigenvalues closer to a_0 will be mapped away from $[-1, 1]$ and hence, magnified while the eigenvalues in $[a, b]$ will be dampened. This is the desired filtering. The Chebyshev iteration described above is presented in the algorithm 1 below

1 Need to add algorithm

Algorithm 1: Chebyshev filter

An iteration of the algorithm is equivalent to computing

$$y = p_m(A)x, \quad \text{where } P_m(t) = C_m\left(\frac{t - c}{e}\right)$$

The σ 's are used for scaling. The three-term recurrence using $p_m(A)$ yields

$$x_{j+1} = \frac{2}{e}(A - cI)x_j - x_{j-1}, j = 1, 2, \dots, m-1, \quad \text{given an } x_0 \text{ and } x_1 = (A - cI)x_0$$

This is equivalent to a power iteration of the form

$$\begin{pmatrix} x_{j+1} \\ x_j \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{2}{e}(A - cI) & -I \\ I & 0 \end{pmatrix}}_{\mathcal{B}} \begin{pmatrix} x_j \\ x_{j-1} \end{pmatrix}$$

The eigenvalues of the non-symmetric matrix \mathcal{B} are complex and of modulus 1, except those corresponding to eigenvalues of A that are less than a which are real and magnitude larger than 1. Thus a scaling step is recommended with $\bar{x}_j = \frac{C_j[\frac{2}{e}(A - cI)]}{C_j[\frac{2}{e}(a_0 - cI)]} x_0$, where $C_j[\frac{2}{e}(a_0 - cI)]$ is the scaling factor. Since this is only used for scaling, a rough estimate of a_0 is sufficient. We can start by choosing a value smaller than a . Subsequent steps can choose the smallest Ritz value.

1.4 Chebyshev polynomial accelerated Davidson method

The pseudocode for the Chebyshev-Davidson method is presented in the algorithm 2 below.

Some points worth noting the natural indexing scheme employed by the algorithm offers the following advantages. The deflation of the converged eigenvectors is handled by indexing the columns of the projection basis V . No extra storage for the converged eigenvectors is necessary and the restarting is simplified by indexing.

Some other comments that the authors make are: It is critical that the upperbound estimate used in the algorithm truly bounds the eigenvalues of A from above, otherwise the larger eigenvalues will be magnified, leading to wrong convergence. The choice of lower bound is also very important but there is some flexibility.

1 Need to add algorithm

Algorithm 2: Chebyshev-Davidson method

1.5 Analysis

The analysis section applies several simplifications in order to understand the convergence of the algorithm. Assume that $\lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$ are eigen values of A with corresponding eigenvectors being q_1, q_2, \dots, q_n . The algorithm adaptively changes the interval to be dampened in each iteration and the matrix applied in step j is

$$p_m^{(j)}(A) = C_m^{(j)}((A - c_j I)/e_j)$$

The first simplification is to keep the the interval to dampen fixed, i.e. the matrix to be used in each step is the same, so the index j in the above equation can be done away with. Next, assume that no restart is used.

Further simplification is achieved by reducing the dimensionality of the algorithm to one. Thus, algorithm now becomes a standard power method with matrix $p_m(A)$. Now, one can easily see that the convergence is now dependent upon the ratio of the two largest eigenvalues. Now setting the interval to be dampened as $[a, b]$ where $\lambda_1 < a$, we know that the dominant eigen value of the matrix is $C_m((\lambda_1 - c)/e)$ and algorithm would converge to the first eigenvector q_1 with convergence factor

$$\rho = \frac{\max_{j>1} |C_m((\lambda_j - c)/e)|}{C_m((\lambda_1 - c)/e)} < 1$$

Now, if we increase the dimension to two, the simplified method has a Krylov interpretation. Assume that the first basis vector for the subspace is $p_m(A)x$. The second vector then would be $p_m(A)x_1$ where x_1 is the approximate (Ritz) eigenvector obtained by projecting A onto the one-dimensional space spanned by the first vector and hence is a multiple of $p_m(A)x$. Thus $K_2 = \text{span}\{p_m(A)x, p_m(A)x_1\} = \text{span}\{p_m(A)x, p_m^2(A)x\}$.

Taking it a step further, the third step of the algorithm will inject a vector of the form $p_m(A)x_2$ where x_2 is the Ritz eigenvector computed by projecting A onto the subspace K_2 . Hence x_2 must be linear combination of the vectors of K_2 . Hence we can write $x_2 = \alpha_1 p_m(A)x + \alpha_2 p_m^2(A)x$. Thus the 3rd vector of the subspace would be of the form $\alpha_1 p_m^2(A)x + \alpha_2 p_m^3(A)x$ and the basis of the subspace K_3 can be written as $K_3 \equiv \text{span}\{p_m(A)x, p_m^2(A)x, p_m^3(A)x\}$. This can be extended to an arbitrary step j .

1.5.1 Chebyshev-Davidson method is not a Krylov subspace method

Proposition 5.1 in the paper [1] states that assuming the simplifications stated above, i.e fixed filtering interval and no restart, the Chebyshev-Davidson method is mathematically equivalent to a Rayleigh-Ritz process applied to A using the Krylov subspace $K_j(p_m(A)x)$. Thus if one generated the orthogonal basis V_j of the Krylov subspace K_j and computed the eigenvalues of $V_j^T A v_j$, these eigenvalues will be identical with those of the simplified algorithm above. However, This is NOT a Krylov subspace method because the projection uses A and not $p_m(A)$. Noteworthy, here, is that the eigenvectors of A are the eigenvectors of $p_m(A)$.

Next, if we bring to light, the fact that the Chebyshev-Davidson algorithm adapts the filters by adjusting the bounds of the interval to be dampened at each iteration. Thus $K_2 = \text{span}\{p_m^{(1)}(A)x, p_m^{(2)}(A)x\}$

and

$$K_3 = \text{span}\{p_m^{(1)}(A)x, p_m^{(2)}p_m^{(1)}(A)x, \alpha_1 p_m^{(3)}p_m^{(1)}(A)x + \alpha_2 p_m^{(3)}(A)x p_m^{(2)}(A)x p_m^{(1)}(A)x\}$$

Infact the last term of the Krylov subspace J_j contains $\Phi_j(t) = \Pi_i = 1^k p_{mj}^{(k)}(t)$ with a non-zero coefficient.

Bibliography

- [1] Y. Zhou and Y. Saad. A Chebyshev-Davidson algorithm for large symmetric eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 29(3):954–971, 2007.
- [2] L.N. Trefethen and D. Bau. *Numerical Linear Algebra*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1997.