Reading Notes [2]

Zhengbo Zhou*

January 14, 2023

Abstract

The authors try to implement a Jacobi algorithm which utilizes the benefit of fastness of low precision. It first computes the approximate eigendecomposition in low precision, then orthogonalize it using MGS approach and apply to original matrix as a preconditioner, and finally compute the eigensystem of preconditioned matrix.

The paper also provides the bound on distance of low precision eigenmatrix and its orthogonal QR factor, a bound on off(Preconditioned Matrix), and a sufficient condition for the Jacobi algorithm to have quadratic convergence that related to $\mathrm{Egap}(A)$ and $\mathrm{Egap}(Q^TAQ)$. The key element is the preconditioning method. Notice that the quadratic convergence is reaily presented by [1], and by preconditioning, the preconditioned matrix is automatically satisfies the condition presented in [1]. We can study the way of constructing $\|Z - Q\|$ and way of constructing error bounds.

1 Summary and Outline of Paper

Lemma 1.1. *Lemma 2.1 Let $A \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(A) = n$. Suppose the MGS method computes the approximate QR factorization $A \approx \widehat{Q}\widehat{R}$ in precision v, where $\widehat{R} \in \mathbb{R}^{n\times n}$ is upper triangular and $\widehat{Q} \in \mathbb{R}^{m\times n}$. Then there exist constants $\eta_i \equiv \eta_i(m,n)$ for $i = 1,2,3$ such that*

$$\|A - \widehat{Q}\widehat{R}\| \le \eta_1\|A\|v, \quad \|\widehat{Q}^T\widehat{Q} - I_n\| \le \eta_2\kappa(A)v,$$

and $\widehat{Q} + \delta\widehat{Q}$ is orthogonal with $\|\delta\widehat{Q}\| \le \eta_3\kappa(A)v$, where $\kappa(A) = \sigma_1(A)/\sigma_{\min}(A)$ is the condition number of A.

Lemma 1.2. *Lemma 2.2 Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. The computed symmetric eigenvalue decomposition (SED) $A \approx \widehat{P}\widehat{\Lambda}\widehat{P}^T$ with $\widehat{P} \in \mathbb{R}^{n\times n}$ and $\widehat{\Lambda} = \mathrm{diag}\left(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_n\right) \in \mathbb{R}^{n\times n}$ via any eigensolver in LAPACK or EISPACK in precision v is nearly the exact symmetric Schur decomposition of $A + E$, i.e.,*

$$A + E = (\widehat{P} + \delta\widehat{P})\widehat{\Lambda}(\widehat{P} + \delta\widehat{P})^T,$$

where $\|E\| \le p(n)\|A\|v$ and $\widehat{P} + \delta\widehat{P}$ is orthogonal with $\|\delta\widehat{P}\| \le p(n)v$. Here, $p(n)$ is a modestly growing function of n.

*Department of Mathematics, University of Manchester, Manchester, M13 9PL, England (zhengbo.zhou@postgrad.manchester.ac.uk).

Lemma 1.3. *Lemma 2.3 Let $A \in \mathbb{R}^{m\times n}$ be a real matrix ($m \geq n$). The computed SVD $A \approx \widehat{U}\widehat{\Sigma}\widehat{V}^T$ with $\widehat{U} \in \mathbb{R}^{m\times m}$, $\widehat{V} \in \mathbb{R}^{n\times n}$, and $\widehat{\Sigma} = \text{diag}\,(\widehat{\sigma}_1, \ldots, \widehat{\sigma}_n) \in \mathbb{R}^{m\times n}$ via any SVD solver in LAPACK, LINPACK or EISPACK in precision v is nearly the exact SVD of $A + E$, i.e.,*

$$A + E = (\widehat{U} + \delta\widehat{U})\widehat{\Sigma}(\widehat{V} + \delta\widehat{V})^T$$

where $\|E\| \leq p(m,n)\|A\|v$ and $\widehat{U} + \delta\widehat{U}$ and $\widehat{V} + \delta\widehat{V}$ are both orthogonal with $\|\delta\widehat{U}\| \leq p(m,n)v$ and $\|\delta\widehat{V}\| \leq p(m,n)v$. Here, $p(m,n)$ is a modestly growing function of m and n.

Lemma 1.4. *Lemma 2.4 If A and $A + E$ are $n \times n$ real symmetric matrices, then*

$$|\lambda_k(A + E) - \lambda_k(A)| \leq \|E\|,$$

for $k = 1, \ldots, n$.

Lemma 1.5. *Lemma 3.3 Let $A^{(k)}$ be the matrix A after k Jacobi updates in Algorithm 3.2 with $A^{(0)} = A$. Suppose*

$$\min_{\lambda_i(A^{(0)})\neq\lambda_j(A^{(0)})} \left|\lambda_i\left(A^{(0)}\right) - \lambda_j\left(A^{(0)}\right)\right| \geq d > 0.$$

Moreover, if we reach a point that $\|\,\text{off}\left(A^{(k)}\right)\|_F < d/(4\sqrt{2})$, then we have

$$\|\,\text{off}\left(A^{(k+N)}\right)\|_F \leq \frac{\sqrt{34/9}}{d}\|\,\text{off}\left(A^{(k)}\right)\|_F^2$$

where $N = n(n-1)/2$

2 Outline of Proofs

Lemma 2.1. *Suppose that $Z \in \mathbb{R}^{n\times n}$ in Step 1 of Algorithm 4.1 is computed by any eigensolver in LAPACK or EISPACK in precision v and $Q \in \mathbb{R}^{n\times n}$ in Step 2 of Algorithm 4.1 is computed by using the MGS method to Z in precision $\omega(\omega \ll v)$. Then there exist constants $h_i \equiv h_i(n)$ for $i = 1, 2$ such that*

$$\|Z - Q\|_F \leq h_1 v + h_2\omega$$

Outline of proof.

- Rewrite $\|Z - Q\|_F = \|QR + F_1 - Q\| \leq \|Q\|\|R - I\|_F + \|F_1\|$.
- Construct bounds of $\|Q\|$ and $\|F_1\|$ by given theorems.
- Rewrite $\|R - I\|_F = \sqrt{\sum_{i=1}^n (r_{ii} - 1)^2 + \sum_{i<j} r_{ij}^2}$.
- Construct bounds for r_{ii} and r_{jj}.
- Notice $1/\|R^{-1}\| \leq |r_{ii}| \leq \|R\|$, hence we need to construct bounds on $\|R\|$ and $\|R^{-1}\|$.
- Moreover, notice $|r_{ij}| \leq \|R^{-1} - R^T\| \leq \|I - R^T R\|\|R^{-1}\|$. Hence we need to construct bounds on $\|I - R^T R\|$ by using the orthogonality of $Z + \delta Z$. Why? The first inequality

- Finally assembly these together, we have $\|Z - Q\| \leq h_1 v + h_2 w$, if Z is orthogonal at precision v and Q is its orthogonal QR factor computed by MGS at precision w.

□

Theorem 2.2. *Theorem 4.2 Suppose that $A = ZDZ^T$ is computed by any eigensolver in LAPACK or EISPACK in precision v and $Q \in \mathbb{R}^{n\times n}$ is computed by using the MGS method to Z in precision $\omega(\omega \ll v)$. Then there exists a constant $\gamma_1 \equiv \gamma_1(n)$ such that*

$$\| \operatorname{off}\left(Q^T A Q\right) \|_F \leq \gamma_1 \|A\| v$$

Proof. Outline of proof

- Notice the important inequality: $\|\operatorname{off}(Q^T A Q)\|_F \leq \|Q^T A Q - D\|_F$[1].
- In order to relate this theorem with lemma 2.1, we write

$$\|Q^T A Q - D\|_F \leq \|(Q-Z)^T A (Q-Z)\|_F + 2\|(Q-Z)^T A Z\|_F + \|Z^T A Z - D\|_F$$

- We have already construct a bound on $\|Q - Z\|_F$ by lemma 2.1. Remain to construct bound on $\|Z^T A Z - D\|_F$.
- Notice the expression $D = (Z + \delta Z)^T (A + E)(Z + \delta Z)$ from lemma 1.2. Since we have the bound for $\|Z\|$, $\|\delta Z\|$, $\|E\|$ and $\|Q - Z\|$, we can construct the finalized bound.

□

The following two theorems gives the sufficient condition(s) on the quadratic convergence of apply Jacobi algorithm to the preconditioned matrix $T^{(0)} = Q^T A Q$.

Theorem 2.3. *Theorem 4.3 Let $T^{(k)}$ be the matrix T after k Jacobi updates in Algorithm* 4.1 *with $T^{(0)} = Q^T A Q$. Suppose $\omega \ll v$ and*

$$\min_{\lambda_i(T^{(0)}) \neq \lambda_j(T^{(0)})} \left| \lambda_i\left(T^{(0)}\right) - \lambda_j\left(T^{(0)}\right) \right| \geq d > 0$$

If *$\gamma_1 \|A\| v < d/(4\sqrt{2})$ for some constant $\gamma_1 = \gamma_1(n)$, then we have*

$$\| \operatorname{off}\left(T^{(N)}\right) \|_F \leq \frac{\sqrt{34/9}}{d} \| \operatorname{off}\left(T^{(0)}\right) \|_F^2,$$

where $N = n(n-1)/2$.

Outline of proof. We know that from previous theorem that, there exist a γ_1, such that

$$\|\operatorname{off}(Q^T A Q)\|_F = \|\operatorname{off}(T^{(0)})\|_F \leq \gamma_1 \|A\| v.$$

If we have $\gamma_1 \|A\| v \leq d/(4\sqrt{2})$, then we have

$$\|\operatorname{off}(T^{(0)})\|_F \leq d/(4\sqrt{2})$$

apply the Lemma 1.5 with $k = 0$, we have

[1] This is due to the off operator remove all the diagonals, whereas $Q^T A Q - D$ will not result all the diagonal entries to be zero. In fact, this is true for all diagonal matrices D, and the equality is attained for $D = \operatorname{diag}(\operatorname{diag}(Q^T A Q))$.

- $\min_{\lambda_i(T^{(0)})\neq\lambda_j(T^{(0)})}\left|\lambda_i(T^{(0)}) - \lambda_j(T^{(0)})\right| \geq d$ (by assumption)
- $\|\text{off}(T^{(0)})\|_F \leq d/4\sqrt{2}$ by assumption and previous theorem.

Hence we have the quadratic convergence. □

Theorem 2.4. *Theorem 4.4 Let $T^{(k)}$ be the matrix T after k Jacobi updates in Algorithm 4.1 with $T^{(0)} = Q^T AQ$. Suppose $\omega \ll v$ and*

$$\min_{\lambda_i(A)\neq\lambda_j(A)}\left|\lambda_i(A) - \lambda_j(A)\right| \geq d > 0.$$

If $2\gamma_2\|A\|\omega \leq \rho_1 d$ for some constant $\gamma_2 = \gamma_2(n)$ and $0 < \rho_1 < 1$, and $\gamma_1\|A\|v < (1-\rho_1)\, d/(4\sqrt{2})$ for some constant $\gamma_1 = \gamma_1(n)$, then we have

$$\left\|\text{off}\left(T^{(N)}\right)\right\|_F \leq \frac{\sqrt{34/9}}{(1-\rho_1)\, d}\left\|\text{off}\left(T^{(0)}\right)\right\|_F^2$$

Proof. Let $d_A \leq \min_{\lambda_i(A)\neq\lambda_j(A)}\left|\lambda_i(A) - \lambda_j(A)\right|$, and suppose we have

$$2\gamma_1\|A\|w \leq \rho_1 d_A$$

for some $0 \leq \rho_1 \leq 1$, then we have

$$(1-\rho_1)d_A \leq d_{T^{(0)}}. \tag{2.1}$$

The sufficient condition for the algorithm to converge that related to $d_{T^{(0)}}$ is the following:

$$\gamma_1\|A\|v \leq \frac{d_{T^{(0)}}}{4\sqrt{2}}. \tag{AIM}$$

From equation (2.1), we have

$$\frac{(1-\rho_1)d_A}{4\sqrt{2}} \leq \frac{d_{T^{(0)}}}{4\sqrt{2}}. \tag{2.2}$$

In order to satisfies (AIM), we need

$$\gamma_1\|A\|v \leq \frac{(1-\rho_1)d_A}{4\sqrt{2}}.$$

Hence if we precondition the matrix A by $Q^T AQ$, then by Thm 2.3,

$$\|\text{off}(Q^T AQ)\|_F = \|\text{off}(T^{(0)})\|_F \leq \gamma_1\|A\|v \leq \frac{(1-\rho_1)d_A}{4\sqrt{2}} \leq \frac{d_{T^{(0)}}}{4\sqrt{2}},$$

which ensures the quadratic convergence by relating to the eigengap of A rather than $Q^T AQ$. □

3 Supplements

In the proof of this lemma, there is a statement:

Claim 3.1

Suppose Z and Q are the matrices from Lemma 4.1, then

$$\lambda_{\max}(Z^TZ) \leq 1 + (2\|\delta Z\|\|Z\| + \|\delta Z\|^2), \tag{3.1}$$
$$\lambda_{\min}(Z^TZ) \geq 1 - (2\|\delta Z\|\|Z\| + \|\delta Z\|^2). \tag{3.2}$$

We will present a proof of this:

Proof. We need the following result. From Lemma 1.1, we have $Z + \delta Z$ is orthogonal. Moreover, we have

$$\|Z\| \leq \|Z + \delta Z\| + \delta Z = 1 + \|\delta Z\|. \tag{3.3}$$

For Eq. (3.1). From (3.3), we have $\|Z\| - \|\delta Z\| \leq 1$. Then taking the square of both sides, we have

$$\|Z\|^2 - 2\|Z\|\|\delta Z\| + \|\delta Z\|^2 \leq 1$$

Moving around the equation, we have

$$\lambda_{\max}(Z^TZ) = \|Z\|^2 \leq 1 + 2\|Z\|\|\delta Z\| - \|\delta Z\|^2 \leq 1 + 2\|Z\|\|\delta Z\| + \|\delta Z\|^2.$$

For Eq. (3.2). Noticing the following lemma

Lemma 3.1. *If A and $A + E$ are $n \times n$ real symmetric matrices, then*

$$|\lambda_k(A + E) - \lambda_k(A)| \leq \|E\|$$

for $k = 1, 2, \ldots, n$.

Let us consider $(Z + \delta Z)^T(Z + \delta Z) = Z^TZ + Z^T\delta Z + \delta Z^TZ + \delta Z^T\delta Z$. If we let $A + E = (Z + \delta Z)^T(Z + \delta Z)$ and $A = Z^TZ$, then apparently, this A and $A + E$ satisfies the condition of above lemma, then

$$\begin{aligned} 1 = \lambda_{\min}((Z + \delta Z)^T(Z + \delta Z)) &\leq \lambda_{\min}(Z^TZ) + \|Z^T\delta Z + \delta Z^TZ + \delta Z^T\delta Z\| \\ &\leq \lambda_{\min}(Z^TZ) + (2\|Z\|\|\delta Z\| + \|\delta Z\|^2) \\ \implies \lambda_{\min}(Z^TZ) &\geq 1 - (2\|Z\|\|\delta Z\| + \|\delta Z\|^2) \end{aligned}$$

which proves the claim. □

References

[1] H. P. M. van Kempen. On the quadratic convergence of the special cyclic Jacobi method. *Numerische Mathematik*, 9(1):19–22, 1966. (Cited on p. 1.)

[2] Zhiyuan Zhang and Zheng-Jian Bai. A mixed precision Jacobi method for the symmetric eigenvalue problem, 2022. (Cited on pp. 1, 2, 3, 4, and 5.)