MANCHESTER 1824

The University of Manchester

NLA Group Manchester

Computing Accurate Eigenvalues using the Preconditioned Jacobi Algorithm

Zhengbo Zhou
Department of Mathematics
The University of Manchester

zhengbo.zhou@student.manchester.ac.uk

The 30th Biennial Numerical Analysis Conference, Glasgow

Joint work with Nick Higham, Françoise Tisseur and Marcus Webb.

Accuracy of the Eigenvalues

$A \in \mathbb{R}^{n \times n}$ symmetric positive definite.

Accuracy of the Eigenvalues

$A \in \mathbb{R}^{n \times n}$ symmetric positive definite.
Interested in the **relative forward error**

Accuracy of the Eigenvalues

$A \in \mathbb{R}^{n \times n}$ symmetric positive definite.
Interested in the **relative forward error**

Tridiagonalization based methods

$$\frac{|\lambda_i(A) - \widehat{\lambda}_i(A)|}{\lambda_i(A)} \leq p(n)\, u\, \kappa(A).$$

- $p(n)$ = low deg. poly., u = working precision.
- $\kappa(A)$ = 2 norm condition number of A.
- $\lambda_i(A), \widehat{\lambda}_i(A)$ = ith largest exact, computed eigenvalue.

Accuracy of the Eigenvalues

$A \in \mathbb{R}^{n\times n}$ symmetric positive definite.
Interested in the **relative forward error**

Tridiagonalization based methods

$$\frac{|\lambda_i(A) - \widehat{\lambda}_i(A)|}{\lambda_i(A)} \leq p(n)\, u\, \kappa(A).$$

Jacobi algorithm is "more accurate"

Demmel & Veselić (1992).
By implementing a specific stopping criterion,

$$\frac{|\lambda_i(A) - \widehat{\lambda}_i(A)|}{\lambda_i(A)} \leq p(n)\, u\, \kappa_S(A)$$

$$\kappa_S(A) = \kappa(DAD),\ D = \mathrm{diag}(a_{ii}^{-1/2})$$

Preconditioned Jacobi Algorithm

Drawback: $O(n^3)$ flops with a large constant.

Preconditioned Jacobi Algorithm

Drawback: $O(n^3)$ flops with a large constant.

Motivation: **(Hari, 1991)** A smaller $\mathrm{off}(A) = \|A - \mathrm{diag}(a_{ii})\|_F$ leads to a faster convergence.

Properties of the preconditioner $\widetilde{Q}$:

- $\mathrm{off}(\widetilde{Q}^T A \widetilde{Q})$ is smaller,
- orthogonal at u, and
- cheap to compute.

Preconditioned Jacobi Algorithm

Drawback: $O(n^3)$ flops with a large constant.

Motivation: **(Hari, 1991)** A smaller $\mathrm{off}(A) = \|A - \mathrm{diag}(a_{ii})\|_F$ leads to a faster convergence.

Properties of the preconditioner $\widetilde{Q}$:

- $\mathrm{off}(\widetilde{Q}^T A \widetilde{Q})$ is smaller,
- orthogonal at u, and
- cheap to compute.

Overview of the talk

- How to construct such a preconditioner?
- Do the computed eigenvalues have high relative accuracy?

Construction of Preconditioner I

Key idea: Exploiting a low precision u_ℓ ($u < u_\ell$).

Construction of Preconditioner I

Key idea: Exploiting a low precision u_ℓ ($u < u_\ell$).

Approach 1: Orthogonalization method

Z (2022); Zhang & Bai (2022).

1. Compute an eigenvector matrix Q_ℓ. (u_ℓ)
2. Orthogonalize Q_ℓ to $\widetilde{Q}$. (u)

Construction of Preconditioner II

Approach 2: Modified tridiagonalization method

Higham, Tisseur, Webb & Z (2025)

1. Perform Tridiag at u_ℓ. Store Householder vectors and construct transformation matrix at u. T_ℓ, Q_T
2. Apply any eigensolver to T_ℓ at u. Q_S
3. Obtain $\widetilde{Q} = Q_T Q_S$ at u.

Construction of Preconditioner II

Approach 2: Modified tridiagonalization method

Higham, Tisseur, Webb & Z (2025)

1. Perform Tridiag at u_ℓ. Store Householder vectors and construct transformation matrix at u. T_ℓ, Q_T
2. Apply any eigensolver to T_ℓ at u. Q_S
3. Obtain $\widetilde{Q} = Q_T Q_S$ at u.

Remaining question: *What is the size of* $\mathrm{off}(\widetilde{Q}^T A \widetilde{Q})$*?*

Reduction of Off–diagonals

Zhang & Bai (2022): Using approach 1 + MGS

$$\text{off}(\widetilde{Q}^T A \widetilde{Q})/\|A\|_F \leq p(n)\, u_\ell.$$

We generalized this result to

- approach 1 + HHQR
- approach 1 + NS, and
- approach 2.

Reduction of Off–diagonals

Zhang & Bai (2022): Using approach 1 + MGS

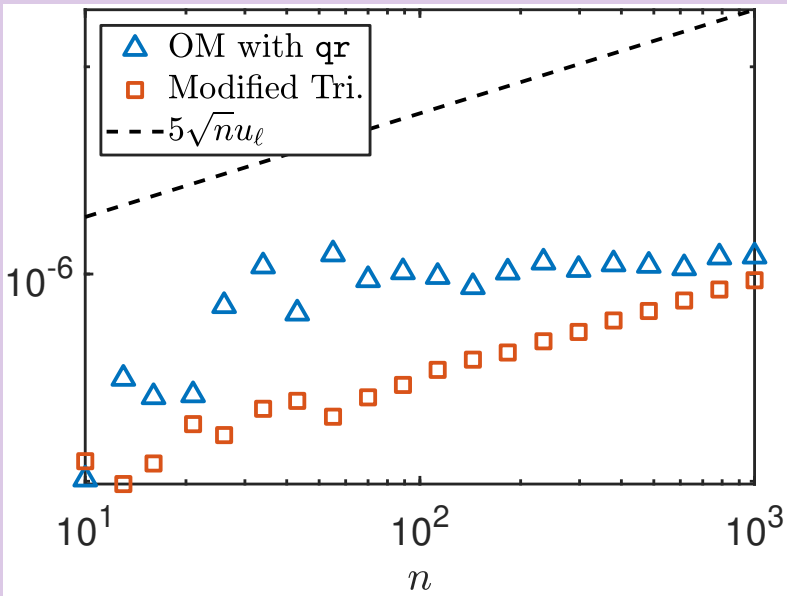
$$\mathrm{off}(\widetilde{Q}^T A \widetilde{Q})/\|A\|_F \leq p(n)\, u_\ell.$$

We generalized this result to

- approach 1 + HHQR
- approach 1 + NS, and
- approach 2.

- $\kappa(A) = 10^8$
- Eigenvalues are geometrically distributed.
- (u_ℓ, u) = (single, double).

$\mathrm{off}(\widetilde{Q}^T A \widetilde{Q})/\|A\|_F$



High Precision Preconditioning

Even more: *Compute* $\widetilde{Q}^T A \widetilde{Q}$ *at* u_h *instead of* u.

High Precision Preconditioning

Even more: *Compute* $\widetilde{Q}^T A \widetilde{Q}$ *at* u_h *instead of* u.

Proposed algorithm: MP3Jacobi

Given a SPD $A \in \mathbb{R}^{n \times n}$.

- Construct preconditioner $\widetilde{Q}$. (u_ℓ, u)
- Obtain preconditioned matrix $\widetilde{A} = \widetilde{Q}^T A \widetilde{Q} \rightsquigarrow \widetilde{A}_{\text{comp}}$. (u_h)
- Apply Jacobi to $\widetilde{A}_{\text{comp}}$. (u)

High Precision Preconditioning

Even more: *Compute* $\widetilde{Q}^T A \widetilde{Q}$ *at* u_h *instead of* u.

Proposed algorithm: MP3Jacobi

Given a SPD $A \in \mathbb{R}^{n \times n}$.

- Construct preconditioner $\widetilde{Q}$. (u_ℓ, u)
- Obtain preconditioned matrix $\widetilde{A} = \widetilde{Q}^T A \widetilde{Q} \rightsquigarrow \widetilde{A}_{\text{comp}}$. (u_h)
- Apply Jacobi to $\widetilde{A}_{\text{comp}}$. (u)

Theorem (Higham, Tisseur, Webb & Z, 2025)

$$\frac{|\widehat{\lambda}_i(\widetilde{A}_{\text{comp}}) - \lambda_i(A)|}{\lambda_i(A)} \leq p(n)\, u\, \kappa_S(\widetilde{A}).$$

Remarks I

Why u_h?

u_h is used to control the effect of rounding errors appearing when applying the preconditioner.

Remarks I

Why u_h?

u_h is used to control the effect of rounding errors appearing when applying the preconditioner.

Decompose the relative forward error

$$\frac{|\widehat{\lambda}_i(\widetilde{A}_{\text{comp}}) - \lambda_i(\widetilde{A}_{\text{comp}})|}{\lambda_i(A)} + \frac{|\lambda_i(\widetilde{A}_{\text{comp}}) - \lambda_i(\widetilde{A})|}{\lambda_i(A)} + \frac{|\lambda_i(\widetilde{A}) - \lambda_i(A)|}{\lambda_i(A)}$$

Remarks I

Why u_h?

u_h is used to control the effect of rounding errors appearing when applying the preconditioner.

Decompose the relative forward error

$$\frac{|\widehat{\lambda}_i(\widetilde{A}_{\text{comp}}) - \lambda_i(\widetilde{A}_{\text{comp}})|}{\lambda_i(A)} + \frac{|\lambda_i(\widetilde{A}_{\text{comp}}) - \lambda_i(\widetilde{A})|}{\lambda_i(A)} + \frac{|\lambda_i(\widetilde{A}) - \lambda_i(A)|}{\lambda_i(A)}$$

- Write $\widetilde{A}_{\text{comp}} = \widetilde{A} + \Delta\widetilde{A}$. The second term can be bounded by $p(n)u\kappa_S(\widetilde{A})$ only if we have $|\Delta\widetilde{a}_{ij}|/\sqrt{\widetilde{a}_{ii}\widetilde{a}_{jj}} \approx O(u)$.

Remarks I

Why u_h?

u_h is used to control the effect of rounding errors appearing when applying the preconditioner.

Decompose the relative forward error

$$\frac{|\widehat{\lambda}_i(\widetilde{A}_{\text{comp}}) - \lambda_i(\widetilde{A}_{\text{comp}})|}{\lambda_i(A)} + \frac{|\lambda_i(\widetilde{A}_{\text{comp}}) - \lambda_i(\widetilde{A})|}{\lambda_i(A)} + \frac{|\lambda_i(\widetilde{A}) - \lambda_i(A)|}{\lambda_i(A)}$$

- Write $\widetilde{A}_{\text{comp}} = \widetilde{A} + \Delta\widetilde{A}$. The second term can be bounded by $p(n)u\kappa_S(\widetilde{A})$ only if we have $|\Delta\widetilde{a}_{ij}|/\sqrt{\widetilde{a}_{ii}\widetilde{a}_{jj}} \approx O(u)$.
- *Key point*: matrix-matrix multiplication is not forward accurate! We need to use u_h to control this ratio.

Remarks II

Why $\kappa_S(\widetilde{A})$?

Key point: $\kappa_S(\widetilde{A})$ can be significantly smaller than $\kappa_S(A)$ and $\kappa(A)$.

Remarks II

Why $\kappa_S(\widetilde{A})$?

Key point: $\kappa_S(\widetilde{A})$ can be significantly smaller than $\kappa_S(A)$ and $\kappa(A)$.

How small can $\kappa_S(\widetilde{A})$ be?

If $\mathrm{off}(\widetilde{A})$ is small such that $\mathrm{off}(\widetilde{A}) < 0.5 \times \min_i \widetilde{a}_{ii}$, then $\kappa_S(\widetilde{A}) < 3$.

Remarks II

Why $\kappa_S(\widetilde{A})$?

Key point: $\kappa_S(\widetilde{A})$ can be significantly smaller than $\kappa_S(A)$ and $\kappa(A)$.

How small can $\kappa_S(\widetilde{A})$ be?

If $\mathrm{off}(\widetilde{A})$ is small such that $\mathrm{off}(\widetilde{A}) < 0.5 \times \min_i \widetilde{a}_{ii}$, then $\kappa_S(\widetilde{A}) < 3$.

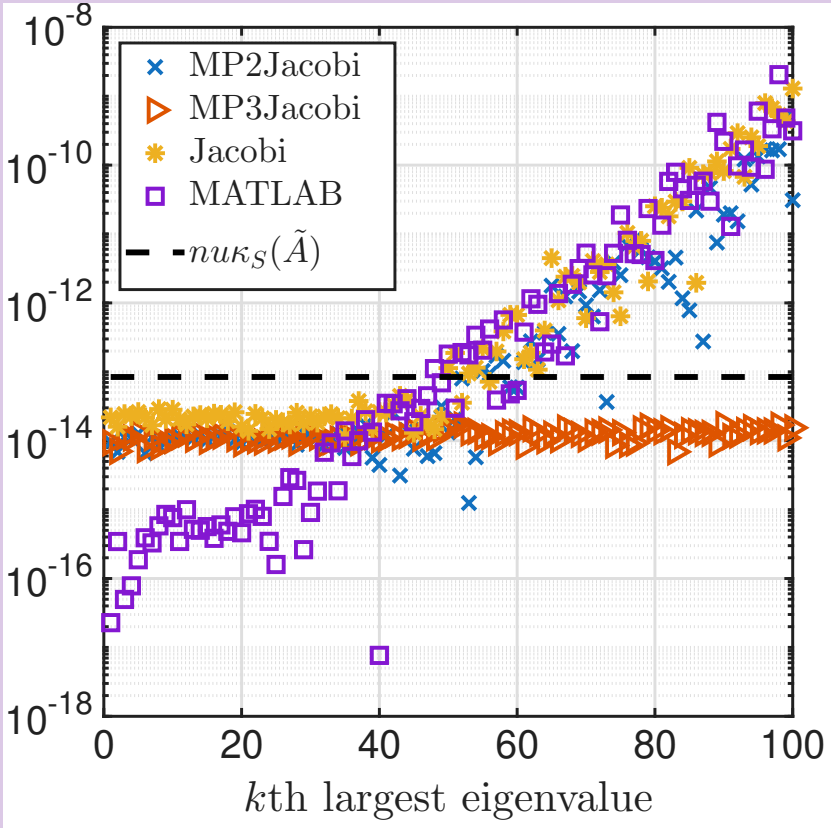
Matrix type	$\kappa(A)$	$\kappa_S(A)$	$\kappa_S(\widetilde{A})$
`hilb(7)`	5e8	2e8	3
`pascal(15)`	3e15	6e12	1e4

Experiment I

Setup:

- Random matrix $A \in \mathbb{R}^{100\times 100}$ SPD.
- $\kappa(A) = 10^8$.
- Geometrically distributed eigenvalues.
- (u_ℓ, u, u_h) = (single, double, quadruple).
- **MP2Jacobi**: $u_h = u$.

Relative forward error

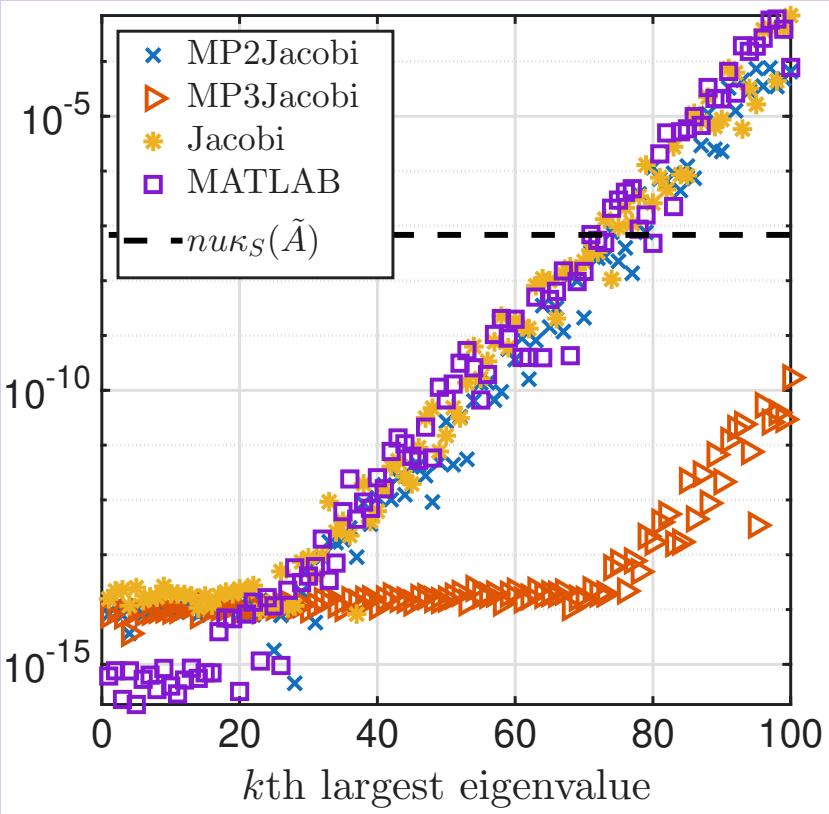


Experiment II

Setup:

- Random matrix $A \in \mathbb{R}^{100\times 100}$ SPD.
- $\kappa(A) = 10^{14}$.
- Geometrically distributed eigenvalues.
- (u_ℓ, u, u_h) = (single, double, quadruple).
- **MP2Jacobi**: $u_h = u$.

Relative forward error

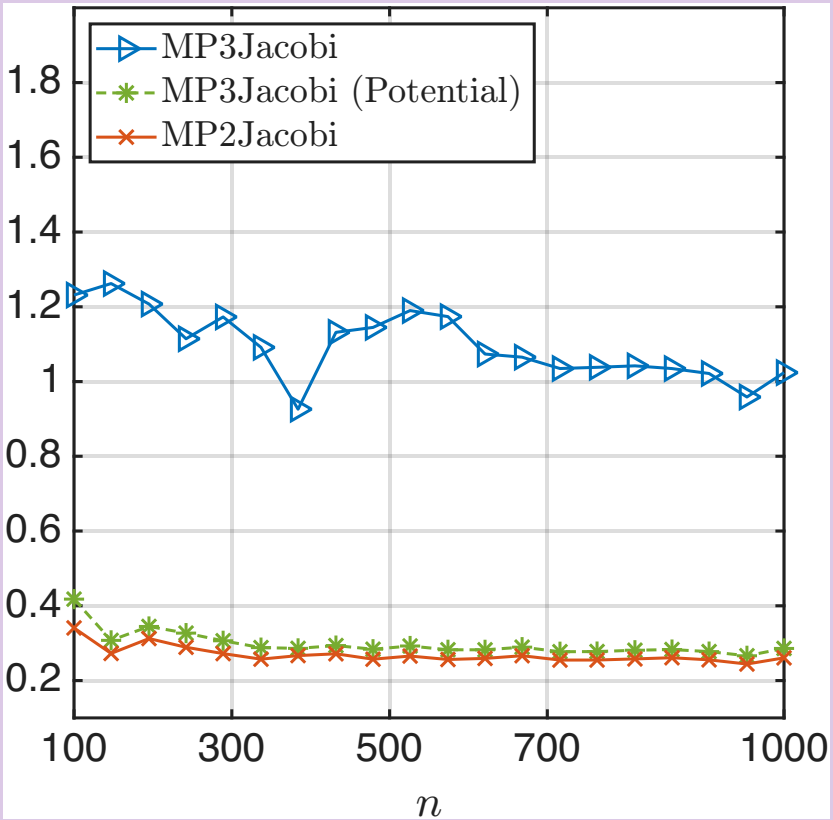


Experiment III

Setup:

- Random matrix $A \in \mathbb{R}^{n \times n}$ SPD, n is varying.
- $\kappa(A) = 10^8$.
- Geometrically distributed eigenvalues.
- (u_ℓ, u, u_h) = (single, double, quadruple).
- **MP3Jacobi (Potential)**: Applying the preconditioner in MP3Jacobi takes only 100 units more time than in MP2Jacobi.

Rel. timing against Jacobi



Summary

We proposed and analyzed:

- An alternative way to construct a preconditioner for the Jacobi algorithm.
- A mixed-precision preconditioned Jacobi algorithm with much more accurate computed eigenvalues. The cost will be two matrix multiplications at u_h.
- Preprint is available on https://arxiv.org/abs/2501.03742.

COMPUTING ACCURATE EIGENVALUES USING A MIXED-PRECISION JACOBI ALGORITHM

NICHOLAS J. HIGHAM*, FRANÇOISE TISSEUR†, MARCUS WEBB†, AND ZHENGBO ZHOU†‡

Summary

We proposed and analyzed:

- An alternative way to construct a preconditioner for the Jacobi algorithm.
- A mixed-precision preconditioned Jacobi algorithm with much more accurate computed eigenvalues. The cost will be two matrix multiplications at u_h.
- Preprint is available on https://arxiv.org/abs/2501.03742.

COMPUTING ACCURATE EIGENVALUES USING A MIXED-PRECISION JACOBI ALGORITHM

NICHOLAS J. HIGHAM*, FRANÇOISE TISSEUR†, MARCUS WEBB†, AND ZHENGBO ZHOU†‡

Thanks for your listening!

References I

- James Demmel and Krešimir Veselić. Jacobi's method is more accurate than QR. *SIAM Journal on Matrix Analysis and Applications*, 13(4):1204–1245, 1992.
- Vjeran Hari.On sharp quadratic convergence bounds for the serial Jacobi methods. *Numerische Mathematik*, 60(1):375–406, 1991.
- Zhengbo Zhou. A mixed precision eigensolver based on the Jacobi algorithm. M.Sc. thesis, The University of Manchester, Manchester, UK, September 2022.
- Zhiyuan Zhang and Zheng-Jian Bai. A mixed precision Jacobi method for the symmetric eigenvalue problem. arXiv:2303.03547, November 2022.

References II

- Nicholas J. Higham, Françoise Tisseur, Marcus Webb, and Zhengbo Zhou. Computing accurate eigenvalues using a mixed-precision Jacobi algorithm. arXiv:2501.03742, June 2025.