Numerical methods for eigenvalues and eigenvectors part II, MAT 3110 UiO

Håkon Hoel

Fall 2023

Table of Contents

QR iteration

Recall that for $A \in \mathbb{R}^{n\times n}_{sym}$, the power iteration

$$x^{(m)} = Ax^{(m-1)} \tag{1}$$

$$z^{(m)} = \frac{x^{(m)}}{\|x^{(m)}\|_2} \tag{2}$$

$$\lambda^{(m)} = (z^{(m)})^T A z^{(m)} \tag{3}$$

computes the dominant eigenvalue of A, and an eigenvector in the corresponding eigenspace.

QR iteration is an extension of power iteration that computes all eigenpairs of A simultaneously.

Algorithm 1: QR iteration

Data: $A \in \mathbb{R}^{n \times n}$

Set $A_0 = A$.

for $m = 0, 1, 2, \ldots$ **do**

Compute a QR factorization

$$Q_m R_m = A_m$$

for orthogonal Q_m and upper triangular R_m.

Assign value

$$A_{m+1} \leftarrow R_m Q_m.$$

end

Algorithm 2: QR iteration

Data: $A \in \mathbb{R}^{n \times n}$
Set $A_0 = A$.
for $m = 0, 1, 2, \ldots$ **do**

$$\text{Compute} \quad Q_m R_m = A_m$$

$$A_{m+1} \leftarrow R_m Q_m.$$

end

Note:

- $R_m = Q_m^T A_m \implies A_{m+1} = Q_m^T A_m Q_m$ are all similarity transformations.
- Which by iteration implies that

$$A_{m+1} = (Q_0 Q_1 \ldots Q_m)^T A \underbrace{(Q_0 Q_1 \ldots Q_m)}_{=: Q^{(m)}}.$$

- So if for some reason $Q^{(m)} = [v_1 \; v_2 \; \ldots v_n]$ (eigenvectors of A), then

$$A_{m+1} = (Q^{(m)})^T A Q^{(m)} = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

and the method has converged: $Q^{(r)} = Q^{(m)}$ and $A_r = A_{m+1}$ for all $r \geq m$.

Theorem (Convergence of QR iteration)

Let $A \in \mathbb{R}^{n\times n}_{sym}$ *with eigenvalue decomposition*

$$A = Q\Lambda Q^T, \qquad \textit{with} \quad \Lambda = diag(\lambda_1, \ldots, \lambda_n),$$

where Q *is orthogonal and eigenvalues are non-repeating and satisfying*

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0.$$

Then the QR iteration sequence $A_m = R_m Q_m = (Q^{(m)})^T A Q^{(m)}$ *satisfies*

$$\lim_{m\to\infty} A_m = \Lambda \quad \textit{and} \quad Q^{(m)} \approx [\pm v_1, \ \pm v_2, \ldots, \pm v_n] \quad \textit{for} \quad m \gg 1.$$

Example (Application of QR iteration)

Consider

$$A = \begin{bmatrix} 0.7491 & 1.5494 & 0.7901 \\ 1.5494 & 0.3120 & 1.0222 \\ 0.7901 & 1.0222 & 1.2022 \end{bmatrix}$$

```
%%Matlab implementation of QR iteration
Qm = eye(length(A));
for m=0:10
    [Q,R] = qr(A);
    A = R*Q;
    Qm = Qm*Q; % for eigenvectors Q^{(m)}
end
```

Results

$$A_5 = \begin{bmatrix} 2.9973 & 0.0087 & 0.0000 \\ 0.0087 & -1.0705 & 0.0002 \\ 0.0000 & 0.0002 & 0.3366 \end{bmatrix}, \quad A_{10} = \begin{bmatrix} 2.9973 & -0.0001 & -0.0000 \\ -0.0001 & -1.0705 & 0.0000 \\ -0.0000 & 0.0000 & 0.3366 \end{bmatrix}$$

and

$$Q^{(5)} = \begin{bmatrix} 0.5919 & -0.6042 & -0.5335 \\ 0.5598 & 0.7844 & -0.2672 \\ 0.5799 & -0.1405 & 0.8025 \end{bmatrix}, \quad Q^{(10)} = \begin{bmatrix} -0.5906 & 0.6053 & -0.5336 \\ -0.5614 & -0.7832 & -0.2671 \\ -0.5796 & 0.1418 & 0.8024 \end{bmatrix}$$

Compare to reference eigenvalues of A:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 2.9973 \\ -1.0705 \\ 0.3366 \end{bmatrix}$$

and eigenvectors

$$[v_1 \ v_2 \ v_3] = \begin{bmatrix} 0.5906 & 0.6053 & 0.5336 \\ 0.5614 & -0.7832 & 0.2671 \\ 0.5796 & 0.1418 & -0.8024 \end{bmatrix}$$

The theorem extends to any diagonalizable $A \in \mathbb{R}^{n\times n}$ with real-valued non-repeating eigenvalues. Then A_m converges to an upper triangular matrix with diagonal $(\lambda_1, \ldots, \lambda_n)$.

Example (Application of the QR iteration method)

Consider

$$A = \begin{bmatrix} 1.271 & -6.409 & 9.208 \\ -2.875 & -25.668 & 38.705 \\ -2.120 & -20.259 & 30.397 \end{bmatrix}$$

```
%%Using same Matlab implementation
for m=0:10
    [Q,R] = qr(A);  % Q_{m}*R_{m} = A_m
    A = R*Q;        % A_{m+1} = R_m*Q_m
end
```

Example (Application of the QR iteration method)

The matrix is diagonalizable:

$$A = TDT^{-1}, \text{ with } \quad T = \begin{bmatrix} 0.298 & 0.938 & 0.055 \\ 0.756 & 0.245 & 0.820 \\ 0.582 & 0.244 & 0.569 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and QR iterations yield:

$$A_0 = A = \begin{bmatrix} 1.271 & -6.409 & 9.208 \\ -2.875 & -25.668 & 38.705 \\ -2.120 & -20.259 & 30.397 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 3.340 & 0.175 & -58.656 \\ -0.444 & 1.915 & 13.242 \\ 0.010 & -0.002 & 0.745 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 3.175 & 0.432 & 59.897 \\ -0.237 & 1.898 & -5.395 \\ -0.003 & 0.001 & 0.927 \end{bmatrix}$$

$$\vdots$$

$$A_{10} = \begin{bmatrix} 3.005 & 0.756 & 59.787 \\ -0.007 & 1.995 & 6.519 \\ -0.000 & 0.000 & 1.000 \end{bmatrix}$$

Complex-valued eigenvalues

- When there is a complex-valued pair of eigenvalues

$$\lambda_j = \overline{\lambda}_{j+1},$$

  then $|\lambda_j| = |\lambda_{j+1}|$. Not covered by Theorem 1.1.
- A_m then will converge towards a **block upper triangular** matrix. The diagonal of A_m will contain a 2×2 block matrix that has eigenvalues λ_j and λ_{j+1}.

As example, consider

$$A = \begin{bmatrix} 30 & -18 & 5 \\ 15 & 9 & -5 \\ 9 & -27 & 24 \end{bmatrix} \qquad \text{with} \quad \sigma(A) = \{27 + 9i, 27 - 9i, 9\}.$$

QR iteration with start matrix $A_0 = A$ yields

$$A_1 = \begin{bmatrix} 24.112 & -20.311 & 25.109 \\ 6.315 & 34.734 & -6.260 \\ -1.941 & 5.921 & 4.154 \end{bmatrix}, \qquad A_5 = \begin{bmatrix} 35.799 & -14.881 & -19.641 \\ 10.604 & 18.236 & -23.492 \\ -0.016 & 0.053 & 8.965 \end{bmatrix}$$

and

$$A_{10} = \begin{bmatrix} 22.957 & -20.934 & 30.563 \\ 4.650 & 31.043 & -0.604 \\ 0.000 & -0.000 & 9.000 \end{bmatrix}$$

where

$$\sigma\left(\begin{bmatrix} 22.957 & -20.934 \\ 4.650 & 31.043 \end{bmatrix}\right) = \{26.999 + 9.000i, 26.999 - 9.000i\}$$

Tridiagonalization of A

- Every QR iteration step of a full matrix A_m costs $\mathcal{O}(n^3)$ operations.
- Can be substantially improved by first tridiagonalizing A through preconditioning A_0:

$$A_0 = QAQ^T$$

  where orthogonal matrix Q is a product of Householder transformations.

Tridiagonalization of A part II

For symmetric matrices, one can find such Q s.t.

$$A = A^T = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \overset{QAQ^T}{\longmapsto} \begin{bmatrix} * & * & & & \\ * & * & * & & \\ & * & * & * & \\ & & * & * & * \\ & & & * & * \end{bmatrix}$$

Efficiency gain: Cost of QR iterations starting from preconditioned A_0 costs $\mathcal{O}(n)$.

Householder transformation

A matrix $H \in \mathbb{R}^{n \times n}$ on the form

$$H = I - \frac{2}{v^T v} v v^T, \qquad \text{with} \quad c \in \mathbb{R} \text{ and } v \in \mathbb{R}^n$$

is called a Householder matrix/transformation.

Properties:

- Orthogonality: $H^T H = I$
- And for any vector $x \in \mathbb{R}^n_*$, one can find a vector $v \in \mathbb{R}^n$ s.t.

$$Hx = \begin{bmatrix} \neq 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{(first element is non-zero, others set to zero)}$$

The choice $\quad v = x + c e_1 \quad$ with $\quad c = \begin{cases} \|x\|_2 & \text{if} \quad x_1 \geq 0 \\ -\|x\|_2 & \text{if} \quad x_1 < 0, \end{cases} \implies Hx = -c e_1.$

Tridigonalizaiton of symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 2 & 1 & 3 \\ 1 & 5 & 0 & 2 & 2 \\ 2 & 0 & 3 & 1 & 1 \\ 1 & 2 & 1 & 6 & 0 \\ 3 & 2 & 1 & 0 & 7 \end{bmatrix}$$

Step 1: Find $H \in \mathbb{R}^{4\times 4}$ such that last three elements of first row is removed, mean for

$$x = \begin{bmatrix} a_{21} \\ a31 \\ a_{41} \\ a_{51} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \end{bmatrix} \quad \text{we seek } H \text{ s.t.} \quad Hx = \begin{bmatrix} \neq 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} =$$

Pick $v = x + \|x\|_2 e_1 = (1 + \sqrt{15}, 2, 1, 3)^T$ and set

$$H = I_4 - \frac{2}{v^T v} vv^T = \begin{bmatrix} -0.2582 & -0.5164 & -0.2582 & -0.7746 \\ -0.5164 & 0.7881 & -0.1060 & -0.3179 \\ -0.2582 & -0.1060 & 0.9470 & -0.1590 \\ -0.7746 & -0.3179 & -0.1590 & 0.5231 \end{bmatrix}$$

Tridigonalizaiton of symmetric matrix II

Then (by two slides back),

$$Hx = -\|x\|_2 e_1 = -\sqrt{15} e_1$$

And for

$$H_{(5,4)} := \begin{bmatrix} 1 & 0^T \\ 0 & H \end{bmatrix} \in \mathbb{R}^{5\times 5}, \qquad 0 \in \mathbb{R}^{4\times 1}$$

We obtain

$$H_{(5,4)} A H_{(5,4)}^T = \begin{bmatrix} 4.0000 & -3.8730 & 0 & -0.0000 & 0 \\ -3.8730 & 7.8667 & 2.0243 & -0.2788 & 0.4545 \\ 0 & 2.0243 & 4.1788 & 0.2387 & 0.8876 \\ -0.0000 & \color{red}{-0.2788} & 0.2387 & 4.9440 & -2.0823 \\ 0 & \color{red}{0.4545} & 0.8876 & -2.0823 & 4.0106 \end{bmatrix}$$

Motivation for above structure:

- Multiplying with $H_{(5,4)}$ from left preserves/leaves elements in first row unchanged.
- Thereafter by $H_{(5,4)}^T$ from right preserves elements in first column (and zeros tail elements in first row by same reason $H_{(5,4)}$ zeroed tail elements in first column).

Step 2: Find $H \in \mathbb{R}^{3\times3}$ that zeros last two terms in column 2 of matrix $H_{(5,4)}AH_{(5,4)}^T$, i.e. of vector

$$x = \begin{bmatrix} 2.0243 \\ -0.2788 \\ 0.4545 \end{bmatrix}$$

Choose

$$v = x + \|x\|_2 e_1 \quad \text{and} \quad H = I - \frac{2}{v^T v} vv^T.$$

Set

$$H_{(4,3)} = \begin{bmatrix} I_2 & 0^T \\ 0 & H \end{bmatrix} \in \mathbb{R}^{5\times5}, \qquad 0 \in \mathbb{R}^{3\times2}$$

and obtain

$$H_{(4,3)}H_{(5,4)}AH_{(5,4)}^T H_{(4,3)}^T = \begin{bmatrix} 4.0000 & -3.8730 & -0.0000 & -0.0000 & -0.0000 \\ -3.8730 & 7.8667 & -2.0934 & 0.0000 & -0.0000 \\ -0.0000 & -2.0934 & 4.6161 & 0.2827 & -1.0331 \\ -0.0000 & 0.0000 & 0.2827 & 4.9360 & -1.9441 \\ -0.0000 & 0.0000 & -1.0331 & -1.9441 & 3.5812 \end{bmatrix}$$

After another such step, finding $H_{(3,2)}$ that zeros last element in third column of latest matrix and leaves first three rows unchanged, we have a tridiagonal symmetric matrix

$$A_0 = \underbrace{H_{(3,2)} H_{(4,3)} H_{(5,4)}}_{Q} A \underbrace{H_{(5,4)}^T H_{(4,3)}^T H_{(3,2)}^T}_{Q^T}$$

$$= \begin{bmatrix} 4.0000 & -3.8730 & -0.0000 & 0.0000 & -0.0000 \\ -3.8730 & 7.8667 & -2.0934 & -0.0000 & -0.0000 \\ -0.0000 & -2.0934 & 4.6161 & -1.0711 & 0.0000 \\ 0.0000 & 0.0000 & -1.0711 & 4.6654 & -2.0181 \\ -0.0000 & 0.0000 & 0 & -2.0181 & 3.8519 \end{bmatrix}$$

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if there exists an invertible matrix $T \in \mathbb{R}^{n\times n}$, s.t.

$$D = T^{-1}AT,$$

where D is a diagonal matrix.

A sufficient condition for A being diagonalizable is that the matrix has n distinct eigenvalues
(then eigenvectors are linearly independent and $T = [v_1 \ v_2 \ \dots v_n]$ will diagonalize A).

Perturbed eigenvalue problem

Problem: we seek eigenvalues of matrix A, but we have/suspect some measurement errors ΔA.

Question: how does eigenvalues of A relate to eigenvalues of $A + \Delta A$?

Theorem (Bauer–Fike)

Consider a diagonalizable matrix $A = T \Lambda T^{-1} \in \mathbb{R}^{n \times n}$ *with* $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, *and a perturbation* $\Delta A \in \mathbb{R}^{n \times n}$. *Then for any eigenvalue of the perturbed matrix* $\mu \in \sigma(A + \Delta A)$, *it holds that*

$$\min_{\lambda \in \sigma(A)} |\mu - \lambda| \leq \underbrace{\|T\|_2 \|T^{-1}\|_2}_{=:\kappa_2(T)} \|\Delta A\|_2, \tag{4}$$

Definition (Absolute condition number)

We define

$$\frac{\min_{\lambda\in\sigma(A)} |\mu - \lambda|}{\|\Delta A\|_2} \qquad \left(= \frac{\textit{error output}}{\textit{error input}}\right)$$

as the absolute condition number of $\mu = \lambda(A + \Delta A)$

The result in theorem becomes:

$$\frac{\textit{error output}}{\textit{error input}} = \min_{\lambda\in\sigma(A)} \frac{|\mu - \lambda|}{\|\Delta A\|_2} \leq \kappa_2(T).$$

When $\kappa_2(T) \gg 1$, *we refer to the perturbation eigenvalue problem as ill-conditioned.*

Remarks on inequality

$$\min_{\lambda \in \sigma(A)} |\mu - \lambda| \leq \underbrace{\|T\|_2 \|T^{-1}\|_2}_{=:\kappa_2(T)} \|\Delta A\|_2, \tag{5}$$

Remark

a) *The upper bound for absolute condition number of the eigenvalue problem, $\kappa_2(T)$, only depends on T, and, perhaps surprisingly, not explicitly on eigenvalues of A.*

b) *For symmetric A, we can find an orthogonal T. This implies that $\kappa_2(T) = 1$. (Eigenvalue perturbation is therefore a well-conditioned problem for such matrices.)*

Ill-posed perturbation problems

If eigenvalues are repeating,or clustered very closely together, then perturbation problem can become ill-posed.
Consider the non-diagonalizable matrix (not covered by Bauer–Fike theorem):

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{with repeated eigenvalue} \quad \lambda = 1.$$

$$\text{For} \quad A + \Delta A = \begin{bmatrix} 1 & \varepsilon \\ 1 & 1 \end{bmatrix} \qquad \text{with eigenvalues} \quad 1 \pm \sqrt{\varepsilon}.$$

$$\text{Input/perturbation error } \|\Delta A\|_2 = \varepsilon \ll \sqrt{\varepsilon} = \text{output error.}$$

$$\text{So that} \quad \frac{\text{output error}}{\text{input error}} = \varepsilon^{-1/2}.$$

Summary

- QR iteration, method for computing full spectrum of A
- Tridiagonalization of A can speed up QR iteration,
- Householder transformations, useful for tridigonalization
- Bauer–Fike theorem, an upper bound on how much $\sigma(A + \Delta A)$ varies from $\sigma(A)$.