Lecture 9: QR algorithm

School of Mathematical Sciences, Xiamen University

1. Simultaneous iteration (SI)

- Sometimes also called *subspace iteration* or *orthogonal iteration* or *block power iteration*

Algorithm 1: Simultaneous iteration

Pick $\mathbf{Q}_n^{(0)} \in \mathbb{C}^{m\times n}$ with orthonormal columns
for $k = 1, 2, 3, \ldots,$
  $\mathbf{Z}^{(k)} = \mathbf{A}\mathbf{Q}_n^{(k-1)}$
  $\mathbf{Q}_n^{(k)}\mathbf{R}_n^{(k)} = \mathbf{Z}^{(k)}$ (QR factorization)
end

- Here is an informal analysis of this method. Assume $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$ is diagonalizable with $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_m\}$ and

$$|\lambda_1| \geq \cdots \geq |\lambda_n| > |\lambda_{n+1}| \geq \cdots \geq |\lambda_m|.$$

We have

$$\mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1}\mathbf{Q}_n^{(0)} = \lambda_n^k\mathbf{S}\mathrm{diag}\left\{\left(\frac{\lambda_1}{\lambda_n}\right)^k, \cdots, 1, \cdots, \left(\frac{\lambda_m}{\lambda_n}\right)^k\right\}\mathbf{S}^{-1}\mathbf{Q}_n^{(0)}.$$

Let

$$\begin{bmatrix}\mathbf{X}_n^{(k)} \\ \mathbf{X}_\mathrm{c}^{(k)}\end{bmatrix} := \mathrm{diag}\left\{\left(\frac{\lambda_1}{\lambda_n}\right)^k, \cdots, 1, \cdots, \left(\frac{\lambda_m}{\lambda_n}\right)^k\right\}\mathbf{S}^{-1}\mathbf{Q}_n^{(0)}.$$

Since $\left|\dfrac{\lambda_i}{\lambda_n}\right| \geq 1$ if $i \leq n$, and $\left|\dfrac{\lambda_i}{\lambda_n}\right| < 1$ if $i > n$, we get $\mathbf{X}_\mathrm{c}^{(k)}$ approaches zero like $\left|\dfrac{\lambda_{n+1}}{\lambda_n}\right|^k$, and $\mathbf{X}_n^{(k)}$ does not approach zero.

Indeed, if $\mathbf{X}_n^{(0)} := \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \end{bmatrix} \mathbf{S}^{-1}\mathbf{Q}_n^{(0)}$ has full rank (a generalization of the assumption $\alpha_1 \neq 0$ in power iteration), then $\mathbf{X}_n^{(k)}$ will have full rank too. We can prove (the proof is left as an exercise)

$$\operatorname{span}\{\mathbf{Q}_n^{(k)}\} = \operatorname{span}\{\mathbf{A}^k\mathbf{Q}_n^{(0)}\}.$$

Here, $\operatorname{span}\{\cdot\} = \operatorname{range}(\cdot)$. Write $\mathbf{S} = \begin{bmatrix} \mathbf{S}_n & \mathbf{S}_\mathrm{c} \end{bmatrix}$. Then

$$\mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1}\mathbf{Q}_n^{(0)} = \lambda_n^k(\mathbf{S}_n\mathbf{X}_n^{(k)} + \mathbf{S}_\mathrm{c}\mathbf{X}_\mathrm{c}^{(k)}).$$

Thus $\operatorname{span}\{\mathbf{Q}_n^{(k)}\}$ converges to

$$\begin{aligned} \operatorname{span}\{\mathbf{Q}_n^{(k)}\} &= \operatorname{span}\{\mathbf{A}^k\mathbf{Q}_n^{(0)}\} = \operatorname{span}\{\mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1}\mathbf{Q}_n^{(0)}\} \\ &= \operatorname{span}\{\mathbf{S}_n\mathbf{X}_n^{(k)} + \mathbf{S}_\mathrm{c}\mathbf{X}_\mathrm{c}^{(k)}\} \\ &\to \operatorname{span}\{\mathbf{S}_n\mathbf{X}_n^{(k)}\} = \operatorname{span}\{\mathbf{S}_n\}, \end{aligned}$$

the invariant subspace spanned by the first n eigenvectors.

- Note that if we follow only the first $j < n$ columns of $\mathbf{Q}_n^{(k)}$ through the iterations of the algorithm, they are *identical* to the columns that we would compute if we had started with only the first j columns of $\mathbf{Q}_n^{(0)}$ instead of n columns.

  In other words, simultaneous iteration is effectively running the algorithm for $j = 1, 2, \cdots, n$ all at the same time.

  So if *all* the first n eigenvalues have distinct absolute values, i.e.,

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|,$$

  and if *all* the leading principal submatrices of

$$\mathbf{X}_n^{(0)} := \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \end{bmatrix} \mathbf{S}^{-1} \mathbf{Q}_n^{(0)}$$

  have full rank, the same convergence analysis as before implies that the first $j \le n$ columns of $\mathbf{Q}_n^{(k)}$ converge to $\text{span}\{\mathbf{S}_j\}$.

Theorem 1

Consider running simultaneous iteration on matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$ *with* $n = m$ *and* $\mathbf{Q}_n^{(0)} = \mathbf{I}$*. If* $\mathbf{A} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^{-1}$ *is diagonalizable with*

$$\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_m\}, \quad |\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|,$$

and if all the leading principal submatrices of $\mathbf{S}^{-1}$ *have full rank, then* $\mathbf{A}^{(k)} := (\mathbf{Q}_n^{(k)})^* \mathbf{A}\mathbf{Q}_n^{(k)}$ *converges to the Schur form of* $\mathbf{A}$*. The eigenvalues will appear in decreasing order of absolute value.*

Proof: See Demmel's book: Theorem 4.8, Page 158, Applied numerical linear algebra.

- The entry $\mathbf{A}_{j,j}^{(k)}$ converges to λ_j like $\max\left(\left|\frac{\lambda_{j+1}}{\lambda_j}\right|^k, \left|\frac{\lambda_j}{\lambda_{j-1}}\right|^k\right)$.
- The block $\mathbf{A}^{(k)}(j+1:m, 1:j)$ converges to zero like $\left|\frac{\lambda_{j+1}}{\lambda_j}\right|^k$.

2. QR algorithm without shifts

Algorithm 2: "Pure" QR algorithm

$\mathbf{A}^{(0)} = \mathbf{A}$

for $k = 1, 2, 3, \ldots,$

$\quad \mathbf{Q}^{(k)}\mathbf{R}^{(k)} = \mathbf{A}^{(k-1)}$ (QR factorization)

$\quad \mathbf{A}^{(k)} = \mathbf{R}^{(k)}\mathbf{Q}^{(k)}$

end

Proposition 2

We have $\mathbf{A}^{(k)} = (\underline{\mathbf{Q}}^{(k)})^*\mathbf{A}\underline{\mathbf{Q}}^{(k)}$, *where* $\underline{\mathbf{Q}}^{(k)} := \mathbf{Q}^{(1)}\mathbf{Q}^{(2)}\cdots\mathbf{Q}^{(k)}$.

Proof.

Note that $\mathbf{A}^{(k)} = (\mathbf{Q}^{(k)})^*\mathbf{A}^{(k-1)}\mathbf{Q}^{(k)}$. □

Proposition 3

The QR *factorization of the* k*th power of* $\mathbf{A}$ *is given by*

$$\mathbf{A}^k = \underline{\mathbf{Q}}^{(k)}\underline{\mathbf{R}}^{(k)},$$

where $\underline{\mathbf{Q}}^{(k)} := \mathbf{Q}^{(1)}\mathbf{Q}^{(2)}\cdots\mathbf{Q}^{(k)}$*, and* $\underline{\mathbf{R}}^{(k)} := \mathbf{R}^{(k)}\mathbf{R}^{(k-1)}\cdots\mathbf{R}^{(1)}$.

Proof.

We use induction. For $k = 1$, $\mathbf{A} = \mathbf{A}^{(0)} = \mathbf{Q}^{(1)}\mathbf{R}^{(1)} = \underline{\mathbf{Q}}^{(1)}\underline{\mathbf{R}}^{(1)}$. Assume $\mathbf{A}^{k-1} = \underline{\mathbf{Q}}^{(k-1)}\underline{\mathbf{R}}^{(k-1)}$. Then by $\mathbf{A}^{(k-1)} = (\underline{\mathbf{Q}}^{(k-1)})^*\mathbf{A}\underline{\mathbf{Q}}^{(k-1)}$, we have

$$\mathbf{A}^k = \mathbf{A}\underline{\mathbf{Q}}^{(k-1)}\underline{\mathbf{R}}^{(k-1)} = \underline{\mathbf{Q}}^{(k-1)}\mathbf{A}^{(k-1)}\underline{\mathbf{R}}^{(k-1)} = \underline{\mathbf{Q}}^{(k)}\underline{\mathbf{R}}^{(k)}.$$

This completes the proof. $\square$

- Connection with power iteration: By $\mathbf{A}^k = \underline{\mathbf{Q}}^{(k)}\underline{\mathbf{R}}^{(k)}$, the first column of $\underline{\mathbf{Q}}^{(k)}$ is the result of applying k steps of power iteration on $\mathbf{A}$ to the vector $\mathbf{e}_1$.
- Connection with inverse iteration: By $\underline{\mathbf{Q}}^{(k)} = (\mathbf{A}^*)^{-k}(\underline{\mathbf{R}}^{(k)})^*$, the last column of $\underline{\mathbf{Q}}^{(k)}$ is the result of applying k steps of inverse iteration on $\mathbf{A}^*$ to the vector $\mathbf{e}_m$.

Theorem 4

If $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$ is diagonalizable with

$$\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_m\}, \quad |\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|,$$

and if all the leading principal submatrices of $\mathbf{S}^{-1}$ have full rank, then $\mathbf{A}^{(k)}$ computed by "pure" QR algorithm converges to the Schur form of $\mathbf{A}$. The eigenvalues will appear in decreasing order of absolute value.

This theorem is a direct result of the following lemma.

Lemma 5

The $\mathbf{A}^{(k)}$ *computed by "pure"* QR *algorithm is identical* (*we need an assumption about* QR *factorization here*) *to the matrix* $(\mathbf{Q}_n^{(k)})^*\mathbf{A}\mathbf{Q}_n^{(k)}$ *implicitly computed by running simultaneous iteration on matrix* $\mathbf{A} \in \mathbb{C}^{m\times m}$ *with* $n = m$ *and* $\mathbf{Q}_n^{(0)} = \mathbf{I}$.

Proof.

We use induction. By $\mathbf{Q}_n^{(1)} = \mathbf{Q}^{(1)}$, we have $\mathbf{A}^{(1)} = (\mathbf{Q}_n^{(1)})^*\mathbf{A}\mathbf{Q}_n^{(1)}$. Assume $\mathbf{A}^{(k-1)} = (\mathbf{Q}_n^{(k-1)})^*\mathbf{A}\mathbf{Q}_n^{(k-1)}$. From simultaneous iteration, we can write $\mathbf{A}\mathbf{Q}_n^{(k-1)} = \mathbf{Q}_n^{(k)}\mathbf{R}_n^{(k)}$. Then $\mathbf{R}_n^{(k)} = (\mathbf{Q}_n^{(k)})^*\mathbf{A}\mathbf{Q}_n^{(k-1)}$, and

$$\mathbf{A}^{(k-1)} = (\mathbf{Q}_n^{(k-1)})^*\mathbf{A}\mathbf{Q}_n^{(k-1)} = (\mathbf{Q}_n^{(k-1)})^*\mathbf{Q}_n^{(k)}\mathbf{R}_n^{(k)} = \mathbf{Q}^{(k)}\mathbf{R}^{(k)}.$$

Thus

$$\mathbf{A}^{(k)} = \mathbf{R}^{(k)}\mathbf{Q}^{(k)} = (\mathbf{Q}_n^{(k)})^*\mathbf{A}\mathbf{Q}_n^{(k-1)}(\mathbf{Q}_n^{(k-1)})^*\mathbf{Q}_n^{(k)} = (\mathbf{Q}_n^{(k)})^*\mathbf{A}\mathbf{Q}_n^{(k)}.$$

This completes the proof. □

- From earlier analysis, we know that the convergence rate of "pure" QR algorithm depends on the ratios of eigenvalues. To speed convergence, we can use shift and invert techniques.

3. QR algorithm with shifts

Algorithm 3: QR algorithm with shifts

$\mathbf{A}^{(0)} = \mathbf{A}$
for $k = 1, 2, 3, \dots,$
  Pick a shift $\mu^{(k)}$ near an eigenvalue of $\mathbf{A}$
  $\mathbf{Q}^{(k)}\mathbf{R}^{(k)} = \mathbf{A}^{(k-1)} - \mu^{(k)}\mathbf{I}$ (QR factorization)
  $\mathbf{A}^{(k)} = \mathbf{R}^{(k)}\mathbf{Q}^{(k)} + \mu^{(k)}\mathbf{I}$
end

Proposition 6

We have $\mathbf{A}^{(k)} = (\underline{\mathbf{Q}}^{(k)})^*\mathbf{A}\underline{\mathbf{Q}}^{(k)}$, *where* $\underline{\mathbf{Q}}^{(k)} := \mathbf{Q}^{(1)}\mathbf{Q}^{(2)}\cdots\mathbf{Q}^{(k)}$.

Proposition 7

We have the factorization (for $k \geq 1$)

$$(\mathbf{A} - \mu^{(k)}\mathbf{I})(\mathbf{A} - \mu^{(k-1)}\mathbf{I}) \cdots (\mathbf{A} - \mu^{(1)}\mathbf{I}) = \underline{\mathbf{Q}}^{(k)}\underline{\mathbf{R}}^{(k)},$$

where $\underline{\mathbf{Q}}^{(k)} := \mathbf{Q}^{(1)}\mathbf{Q}^{(2)} \cdots \mathbf{Q}^{(k)}$, *and* $\underline{\mathbf{R}}^{(k)} := \mathbf{R}^{(k)}\mathbf{R}^{(k-1)} \cdots \mathbf{R}^{(1)}$.

Proof.

We use induction. For $k = 1$, $\mathbf{A} - \mu^{(1)}\mathbf{I} = \mathbf{Q}^{(1)}\mathbf{R}^{(1)} = \underline{\mathbf{Q}}^{(1)}\underline{\mathbf{R}}^{(1)}$.
Assume $(\mathbf{A} - \mu^{(k-1)}\mathbf{I})(\mathbf{A} - \mu^{(k-2)}\mathbf{I}) \cdots (\mathbf{A} - \mu^{(1)}\mathbf{I}) = \underline{\mathbf{Q}}^{(k-1)}\underline{\mathbf{R}}^{(k-1)}$.
Then by $\mathbf{A}^{(k-1)} = (\underline{\mathbf{Q}}^{(k-1)})^*\mathbf{A}\underline{\mathbf{Q}}^{(k-1)}$, we have

$$\begin{aligned}
& (\mathbf{A} - \mu^{(k)}\mathbf{I}) \cdots (\mathbf{A} - \mu^{(1)}\mathbf{I}) = (\mathbf{A}\underline{\mathbf{Q}}^{(k-1)} - \mu^{(k)}\underline{\mathbf{Q}}^{(k-1)})\underline{\mathbf{R}}^{(k-1)} \\
= \; & (\underline{\mathbf{Q}}^{(k-1)}\mathbf{A}^{(k-1)} - \mu^{(k)}\underline{\mathbf{Q}}^{(k-1)})\underline{\mathbf{R}}^{(k-1)} \\
= \; & \underline{\mathbf{Q}}^{(k-1)}(\mathbf{A}^{(k-1)} - \mu^{(k)}\mathbf{I})\underline{\mathbf{R}}^{(k-1)} = \underline{\mathbf{Q}}^{(k)}\underline{\mathbf{R}}^{(k)}.
\end{aligned}$$

This completes the proof. $\square$

- Connection with shifted power iteration: By

$$(\mathbf{A}-\mu^{(k)}\mathbf{I})(\mathbf{A}-\mu^{(k-1)}\mathbf{I})\cdots(\mathbf{A}-\mu^{(1)}\mathbf{I}) = \underline{\mathbf{Q}}^{(k)}\underline{\mathbf{R}}^{(k)},$$

  the first column of $\underline{\mathbf{Q}}^{(k)}$ is the result of applying k steps of shifted power iteration on $\mathbf{A}-\mu^{(j)}\mathbf{I}$ to the vector $\mathbf{e}_1$ using the shifts $\mu^{(j)}$, $j = 1 : k$.
- Connection with shifted inverse iteration: By

$$\underline{\mathbf{Q}}^{(k)} = (\mathbf{A}-\mu^{(k)}\mathbf{I})^{-*}(\mathbf{A}-\mu^{(k-1)}\mathbf{I})^{-*}\cdots(\mathbf{A}-\mu^{(1)}\mathbf{I})^{-*}(\underline{\mathbf{R}}^{(k)})^{*},$$

  the last column of $\underline{\mathbf{Q}}^{(k)}$ is the result of applying k steps of shifted inverse iteration on $(\mathbf{A}-\mu^{(j)}\mathbf{I})^*$ to the vector $\mathbf{e}_m$ using the shifts $\mu^{(j)}$, $j = 1 : k$. If the shifts are good eigenvalue estimates, the last column of $\underline{\mathbf{Q}}^{(k)}$, i.e., $\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m$, converges quickly to a left eigenvector of $\mathbf{A}$.

- Connection with Rayleigh quotient iteration: Choose

$$\mu^{(1)} = r(\mathbf{e}_m), \qquad \mu^{(k+1)} = r(\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m),$$

as the shift at every step. The eigenvalue and eigenvector estimates $\mu^{(k+1)}$ and $\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m$ are identical to those that are computed by the Rayleigh quotient iteration on $\mathbf{A}^*$ starting with $\mathbf{e}_m$.

In the QR algorithm, the Rayleigh quotient $r(\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m)$ appears as the (m, m) entry of $\mathbf{A}^{(k)}$. So it comes for free! Actually, we have

$$\mathbf{A}_{mm}^{(k)} = \mathbf{e}_m^*\mathbf{A}^{(k)}\mathbf{e}_m = \mathbf{e}_m^*(\underline{\mathbf{Q}}^{(k)})^*\mathbf{A}\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m = r(\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m).$$

Then we can set $\mu^{(k+1)} = \mathbf{A}_{mm}^{(k)}$. This is known as the *Rayleigh quotient shift*. Assume the algorithm converges. Then $\underline{\mathbf{Q}}^{(k)}\mathbf{e}_m$ converges quadratically or cubically to an eigenvector.

- Other issues: *Wilkinson shift ...*

4. Practical issues on QR algorithm

Proposition 8

Hessenberg form is preserved by QR *algorithm.*

Proof.

For the upper Hessenberg matrix $\mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I}$, it is easy to show that there exists a QR factorization $\mathbf{Q}^{(k)}\mathbf{R}^{(k)} = \mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I}$ such that $\mathbf{Q}^{(k)}$ is upper Hessenberg. Then it is easy to confirm that $\mathbf{R}^{(k)}\mathbf{Q}^{(k)}$ remains upper Hessenberg and adding $\mu^{(k)}\mathbf{I}$ does not change this. $\square$

Proposition 9

Hermitian tridiagonal form is preserved by QR *algorithm* (*real shifts*)*.*

Proof.

Hermitian + tridiagonal = Hermitian + upper Hessenberg. $\square$

- First phase: Reduction to Hessenberg or tridiagonal form

$$\underset{\mathbf{A}}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\mathbf{Q}_1^* \cdot} \underset{\mathbf{Q}_1^*\mathbf{A}}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}} \xrightarrow{\cdot \mathbf{Q}_1} \underset{\mathbf{Q}_1^*\mathbf{A}\mathbf{Q}_1}{\begin{bmatrix} \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}}$$

$$\underset{\mathbf{Q}_1^*\mathbf{A}\mathbf{Q}_1}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\mathbf{Q}_2^* \cdot} \underset{\mathbf{Q}_2^*\mathbf{Q}_1^*\mathbf{A}\mathbf{Q}_1}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}} \xrightarrow{\cdot \mathbf{Q}_2} \underset{\mathbf{Q}_2^*\mathbf{Q}_1^*\mathbf{A}\mathbf{Q}_1\mathbf{Q}_2}{\begin{bmatrix} \times & \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \times & \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}}$$

$$\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix} \qquad \underbrace{\mathbf{Q}_{m-2}^* \cdots \mathbf{Q}_2^*\mathbf{Q}_1^*}_{\mathbf{Q}^*} \mathbf{A} \underbrace{\mathbf{Q}_1\mathbf{Q}_2 \cdots \mathbf{Q}_{m-2}}_{\mathbf{Q}} = \mathbf{H}.$$

- Second phase: generate a sequence of Hessenberg (or tridiagonal) matrices that converge to a triangular (or diagonal) form.

$$\underset{\mathbf{A} \neq \mathbf{A}^*}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 1}} \underset{\mathbf{H}}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 2}} \underset{\mathbf{T}}{\begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \\ & & & & \times \end{bmatrix}}$$

$$\underset{\mathbf{A} = \mathbf{A}^*}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 1}} \underset{\mathbf{H}}{\begin{bmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 2}} \underset{\mathbf{D}}{\begin{bmatrix} \times & & & & \\ & \times & & & \\ & & \times & & \\ & & & \times & \\ & & & & \times \end{bmatrix}}$$

★ For simplicity, in the following we only consider the real case, i.e., $\mathbf{A} \in \mathbb{R}^{m \times m}$.

4.1. Implicit Q theorem

Definition 10

An upper Hessenberg matrix $\mathbf{H}$ is unreduced if all $(j+1, j)$ entries of $\mathbf{H}$ are nonzero.

Theorem 11 (Consider the real case. The complex case is similar.)

Let $\mathbf{A} \in \mathbb{R}^{m \times m}$. Suppose that $\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \mathbf{H}$ is unreduced upper Hessenberg and $\mathbf{Q}$ is orthogonal. Then columns 2 to m of $\mathbf{Q}$ are determined uniquely (up to signs) by the first column of $\mathbf{Q}$.

- Implicit Q theorem implies that QR algorithm can be implemented cheaply on an upper Hessenberg matrix. The implementation will be *implicit* in the sense that we do not explicitly compute the QR factorization of an upper Hessenberg matrix each iteration but rather construct $\mathbf{Q}$ implicitly as a product of Givens rotations and other simple orthogonal/unitary matrices.

Proof. (Implicit Q theorem).

Suppose that $\mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \mathbf{H}$ and $\mathbf{V}^\top\mathbf{A}\mathbf{V} = \mathbf{G}$ are unreduced upper Hessenberg, $\mathbf{Q}$ and $\mathbf{V}$ are orthogonal, and the first columns of $\mathbf{Q}$ and $\mathbf{V}$ are equal. Let $(\mathbf{X})_i$ denote the ith column of $\mathbf{X}$. Let $\mathbf{W} \equiv \mathbf{V}^\top\mathbf{Q}$. By $\mathbf{G}\mathbf{W} = \mathbf{G}\mathbf{V}^\top\mathbf{Q} = \mathbf{V}^\top\mathbf{A}\mathbf{Q} = \mathbf{V}^\top\mathbf{Q}\mathbf{H} = \mathbf{W}\mathbf{H}$, we have

$$\mathbf{G}(\mathbf{W})_i = \mathbf{W}(\mathbf{H})_i = \sum\nolimits_{j=1}^{i+1} h_{ji}(\mathbf{W})_j.$$

Thus, $h_{i+1,i}(\mathbf{W})_{i+1} = \mathbf{G}(\mathbf{W})_i - \sum_{j=1}^{i} h_{ji}(\mathbf{W})_j$. Since $(\mathbf{W})_1 = \mathbf{e}_1$ and $\mathbf{G}$ is upper Hessenberg, we can use induction on i to show that $(\mathbf{W})_i$ is nonzero in entries 1 to i only; i.e., $\mathbf{W}$ is upper triangular. Since $\mathbf{W}$ is also orthogonal, then $\mathbf{W}$ is diagonal: $\mathbf{W} = \mathrm{diag}\{1, \pm 1, \cdots, \pm 1\}$, which implies

$$\mathbf{V}\mathrm{diag}\{1, \pm 1, \cdots, \pm 1\} = \mathbf{Q}.$$

$\square$

4.2. Implicit single shift QR algorithm

- To compute $\mathbf{H}^{(k)} = (\mathbf{Q}^{(k)})^\top \mathbf{H}^{(k-1)} \mathbf{Q}^{(k)}$ from $\mathbf{H}^{(k-1)}$ in the QR algorithm, we will need only to

  (1) compute the first column of $\mathbf{Q}^{(k)}$ (which is parallel to the first column of $\mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I}$ and so can be gotten just by normalizing this column vector).

  (2) choose other columns of $\mathbf{Q}^{(k)}$ so $\mathbf{Q}^{(k)}$ is orthogonal and $\mathbf{H}^{(k)}$ is unreduced Hessenberg.

- By the implicit Q theorem, we know that we will have computed $\mathbf{H}^{(k)}$ correctly because $\mathbf{Q}^{(k)}$ is unique up to signs, which do not matter. (Signs do not matter because changing the signs of the columns of $\mathbf{Q}^{(k)}$ is the same as changing $\mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I} = \mathbf{Q}^{(k)}\mathbf{R}^{(k)}$ to $(\mathbf{Q}^{(k)}\mathbf{S}^{(k)})(\mathbf{S}^{(k)}\mathbf{R}^{(k)})$, where $\mathbf{S}^{(k)} = \mathrm{diag}\{\pm 1, \pm 1, \cdots, \pm 1\}$. Then $\mathbf{H}^{(k)} = (\mathbf{S}^{(k)}\mathbf{R}^{(k)})(\mathbf{Q}^{(k)}\mathbf{S}^{(k)}) + \mu^{(k)}\mathbf{I} = \mathbf{S}^{(k)}(\mathbf{R}^{(k)}\mathbf{Q}^{(k)} + \mu^{(k)}\mathbf{I})\mathbf{S}^{(k)}$, which is an orthogonal similarity that just changes the signs of some columns and rows of $\mathbf{H}^{(k)}$.)

- To see how to use the implicit Q theorem to compute $\mathbf{H}^{(1)}$ from $\mathbf{H}^{(0)} = \mathbf{H}$, we use a 5×5 example.

1. $\mathbf{Q}_1^\top = \begin{bmatrix} c_1 & s_1 & & & \\ -s_1 & c_1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$, $\mathbf{H}_1 = \mathbf{Q}_1^\top \mathbf{H} \mathbf{Q}_1 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ + & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{bmatrix}$

2. $\mathbf{Q}_2^\top = \begin{bmatrix} 1 & & & & \\ & c_2 & s_2 & & \\ & -s_2 & c_2 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$, $\mathbf{Q}_2^\top \mathbf{H}_1 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{bmatrix}$

$$\mathbf{H}_2 = \mathbf{Q}_2^\top \mathbf{H}_1 \mathbf{Q}_2 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & + & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{bmatrix}$$

3. $\mathbf{Q}_3^\top = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & c_3 & s_3 & \\ & & -s_3 & c_3 & \\ & & & & 1 \end{bmatrix}$, $\mathbf{Q}_3^\top \mathbf{H}_2 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{bmatrix}$

$$\mathbf{H}_3 = \mathbf{Q}_3^\top \mathbf{H}_2 \mathbf{Q}_3 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & + & \times & \times \end{bmatrix}$$

4. $\mathbf{Q}_4^\top = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & c_4 & s_4 \\ & & & -s_4 & c_4 \end{bmatrix}$, $\mathbf{Q}_4^\top \mathbf{H}_3 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{bmatrix}$

$$\mathbf{H}_4 = \mathbf{Q}_4^\top \mathbf{H}_3 \mathbf{Q}_4 = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{bmatrix}$$

Altogether $\mathbf{Q}^\top \mathbf{H} \mathbf{Q} = \mathbf{H}_4$ is upper Hessenberg, where

$$\mathbf{Q} = \mathbf{Q}_1 \mathbf{Q}_2 \mathbf{Q}_3 \mathbf{Q}_4 = \begin{bmatrix} c_1 & \times & \times & \times & \times \\ s_1 & \times & \times & \times & \times \\ & s_2 & \times & \times & \times \\ & & s_3 & \times & \times \\ & & & s_4 & c_4 \end{bmatrix},$$

so the first column of $\mathbf{Q}$ is $\begin{bmatrix} c_1 & s_1 & 0 & \cdots & 0 \end{bmatrix}^\top$, which by the implicit Q theorem has uniquely determined the other columns of $\mathbf{Q}$ (up to signs). We now choose the first column of $\mathbf{Q}$ to be proportional to the first column of $\mathbf{H}^{(0)} - \mu^{(1)}\mathbf{I}$. This means $\mathbf{Q}$ is the same (up to signs) as in the QR factorization of $\mathbf{H}^{(0)} - \mu^{(1)}\mathbf{I}$.

4.3. Implicit double shift QR algorithm

- We describe how to maintain real arithmetic by shifting $\mu^{(k)}$ and $\overline{\mu^{(k)}}$ in succession:

$$\begin{aligned}
\mathbf{Q}^{(k-1/2)}\mathbf{R}^{(k-1/2)} &= \mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I} \\
\mathbf{H}^{(k-1/2)} &= \mathbf{R}^{(k-1/2)}\mathbf{Q}^{(k-1/2)} + \mu^{(k)}\mathbf{I} \\
&= (\mathbf{Q}^{(k-1/2)})^*\mathbf{H}^{(k-1)}\mathbf{Q}^{(k-1/2)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{Q}^{(k)}\mathbf{R}^{(k)} &= \mathbf{H}^{(k-1/2)} - \overline{\mu^{(k)}}\mathbf{I} \\
\mathbf{H}^{(k)} &= \mathbf{R}^{(k)}\mathbf{Q}^{(k)} + \overline{\mu^{(k)}}\mathbf{I} = (\mathbf{Q}^{(k)})^*\mathbf{H}^{(k-1/2)}\mathbf{Q}^{(k)} \\
&= (\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)})^*\mathbf{H}^{(k-1)}\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}
\end{aligned}$$

Lemma 12

We can choose $\mathbf{Q}^{(k-1/2)}$ *and* $\mathbf{Q}^{(k)}$ *such that*

(1) $\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}$ *is real,*

(2) $\mathbf{H}^{(k)}$ *is therefore real,*

(3) *the first column of* $\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}$ *is easy to compute.*

Proof. Since

$$\mathbf{Q}^{(k)}\mathbf{R}^{(k)} = \mathbf{H}^{(k-1/2)} - \overline{\mu^{(k)}}\mathbf{I} = \mathbf{R}^{(k-1/2)}\mathbf{Q}^{(k-1/2)} + (\mu^{(k)} - \overline{\mu^{(k)}})\mathbf{I},$$

we get

$$\begin{aligned}
& \mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}\mathbf{R}^{(k)}\mathbf{R}^{(k-1/2)} \\
= \; & \mathbf{Q}^{(k-1/2)}(\mathbf{R}^{(k-1/2)}\mathbf{Q}^{(k-1/2)} + (\mu^{(k)} - \overline{\mu^{(k)}})\mathbf{I})\mathbf{R}^{(k-1/2)} \\
= \; & \mathbf{Q}^{(k-1/2)}\mathbf{R}^{(k-1/2)}\mathbf{Q}^{(k-1/2)}\mathbf{R}^{(k-1/2)} + (\mu^{(k)} - \overline{\mu^{(k)}})\mathbf{Q}^{(k-1/2)}\mathbf{R}^{(k-1/2)} \\
= \; & (\mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I})^2 + (\mu^{(k)} - \overline{\mu^{(k)}})(\mathbf{H}^{(k-1)} - \mu^{(k)}\mathbf{I}) \\
= \; & (\mathbf{H}^{(k-1)})^2 - 2\mathrm{Re}(\mu^{(k)})\mathbf{H}^{(k-1)} + |\mu^{(k)}|^2\mathbf{I} \equiv \mathbf{M}.
\end{aligned}$$

Thus, $\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}\mathbf{R}^{(k)}\mathbf{R}^{(k-1/2)}$ is the QR factorization of the real matrix $\mathbf{M}$, and therefore, $\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}$, as well as $\mathbf{R}^{(k)}\mathbf{R}^{(k-1/2)}$, can be chosen real. This means that

$$\mathbf{H}^{(k)} = (\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)})^*\mathbf{H}^{(k-1)}\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}$$

also is real if $\mathbf{H}^{(k-1)}$ is real. The first column of $\mathbf{Q}^{(k-1/2)}\mathbf{Q}^{(k)}$ is proportional to the first column of

$$(\mathbf{H}^{(k-1)})^2 - 2\mathrm{Re}(\mu^{(k)})\mathbf{H}^{(k-1)} + |\mu^{(k)}|^2\mathbf{I},$$

whose sparsity pattern is $\begin{bmatrix}\times & \times & \times & 0 & \cdots & 0\end{bmatrix}^\top$. $\square$

- We provide a 6×6 example. Assume $\mathbf{H}$ is upper Hessenberg and the shifts are μ and $\overline{\mu}$.

  1. Choose an orthogonal matrix

$$\mathbf{Q}_1^\top = \begin{bmatrix} \widetilde{\mathbf{Q}}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad \widetilde{\mathbf{Q}}^\top \widetilde{\mathbf{Q}} = \mathbf{I}_3,$$

where the first column of $\mathbf{Q}_1$ is proportional to the first column of

$$\mathbf{H}^2 - 2\mathrm{Re}(\mu)\mathbf{H} + |\mu|^2 \mathbf{I},$$

so

$$\mathbf{Q}_1^\top \mathbf{H} = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ + & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}, \ \mathbf{Q}_1^\top \mathbf{H} \mathbf{Q}_1 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ + & \times & \times & \times & \times & \times \\ + & + & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}$$

2. Choose a Householder reflector $\mathbf{Q}_2^\top$, which affects only rows 2,3, and 4 of $\mathbf{Q}_2^\top \mathbf{H}_1$, zeroing out entries (3,1) and (4,1) of $\mathbf{H}_1 = \mathbf{Q}_1^\top \mathbf{H} \mathbf{Q}_1$ (this means that $\mathbf{Q}_2^\top$ is the identity matrix outside rows and columns 2 through 4):

$$\mathbf{Q}_2^\top \mathbf{H}_1 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & + & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix},$$

$$\mathbf{H}_2 = \mathbf{Q}_2^\top \mathbf{H}_1 \mathbf{Q}_2 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & + & \times & \times & \times & \times \\ 0 & + & + & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}$$

3. Choose a Householder reflector $\mathbf{Q}_3^\top$, which affects only rows 3,4, and 5 of $\mathbf{Q}_3^\top \mathbf{H}_2$, zeroing out entries (4,2) and (5,2) of $\mathbf{H}_2$ (this means that $\mathbf{Q}_3^\top$ is the identity matrix outside rows and columns 3 through 5):

$$\mathbf{Q}_3^\top \mathbf{H}_2 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & + & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}$$

$$\mathbf{H}_3 = \mathbf{Q}_3^\top \mathbf{H}_2 \mathbf{Q}_3 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & + & \times & \times & \times \\ 0 & 0 & + & + & \times & \times \end{bmatrix}$$

4. Choose a Householder reflector $\mathbf{Q}_4^\top$, which affects only rows 4,5, and 6 of $\mathbf{Q}_4^\top \mathbf{H}_3$, zeroing out entries (5,3) and (6,3) of $\mathbf{H}_2$ (this means that $\mathbf{Q}_4^\top$ is the identity matrix outside rows and columns 4 through 6):

$$\mathbf{Q}_4^\top \mathbf{H}_3 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & + & \times & \times \end{bmatrix}$$

$$\mathbf{H}_4 = \mathbf{Q}_4^\top \mathbf{H}_3 \mathbf{Q}_4 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & + & \times & \times \end{bmatrix}$$

5. Choose a Givens rotation $\mathbf{Q}_5^\top$

$$\mathbf{Q}_5^\top = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & c & s \\ & & & & -s & c \end{bmatrix}, \quad \mathbf{Q}_5^\top \mathbf{H}_4 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}$$

$$\mathbf{H}_5 = \mathbf{Q}_5^\top \mathbf{H}_4 \mathbf{Q}_5 = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}.$$

Altogether $\mathbf{Q}^\top \mathbf{H} \mathbf{Q}$ is upper Hessenberg, where

$$\mathbf{Q} = \mathbf{Q}_1 \mathbf{Q}_2 \mathbf{Q}_3 \mathbf{Q}_4 \mathbf{Q}_5 \quad \text{with} \quad \mathbf{Q}\mathbf{e}_1 = \mathbf{Q}_1 \mathbf{e}_1.$$