Mathematics, Pusan National University

Numerical Linear Algebra

Lecture 28. QR Algorithm without Shifts

Taehyeong Kim
th_kim@pusan.ac.kr

November 16, 2020

Content

The QR Algorithm

Unnormalized Simultaneous Iteration

Simultaneous Iteration

Simultaneous Iteration $\Leftrightarrow$ QR Algorithm

Convergence of the QR Algorithm

The most basic version of the QR factorization seems impossibly simple.
A is still real and symmetric, with real eigenvalues λ_j and orthonormal eigenvectors q_j.

Algorithm 28.1. "Pure" QR Algorithm

$A^{(0)} = A$
for $k = 1, 2, \ldots$
  $Q^{(k)} R^{(k)} = A^{(k-1)}$ — QR factorization of $A^{(k-1)}$
  $A^{(k)} = R^{(k)} Q^{(k)}$ — Recombine factors in reverse order

Figure 1: "Pure" QR Algorithm

The first step of "Pure" QR Algorithm is

$$\left[\begin{array}{c|c|c|c} a_1 & a_2 & \cdots & a_n \end{array}\right] = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_n \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

$$\begin{aligned} a_1 &= r_{11} q_1, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ a_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n. \end{aligned}$$

The QR Algorithm

Like the Rayleigh quotient iteration, the QR algorithm for real symmetric matrices converges cubically. But not in the “Pure” QR Algorithm.
We apply three modifications of “Pure” QR Algorithm.

1. Before starting the iteration, A is reduced to tridiagonal form, as dicussed in Lecture 26.
2. Instead of $A^{(k)}$, a shifted matrix $A^{(k)} - \mu^{(k)} I$ is factored at each step, where $\mu^{(k)}$ is some eigenvalue estimate.
3. Whenever possible, and inparticular whenever an eigenvalue is found, the problem is “deflated” by breaking $A^{(k)}$ into submatrices.

The QR Algorithm

A QR algorithm incorporating these modifications has the following outline.

Algorithm 28.2. "Practical" QR Algorithm

$(Q^{(0)})^T A^{(0)} Q^{(0)} = A$	$A^{(0)}$ is a tridiagonalization of A
for $k = 1, 2, \ldots$	
Pick a shift $\mu^{(k)}$	e.g., choose $\mu^{(k)} = A_{mm}^{(k-1)}$
$Q^{(k)} R^{(k)} = A^{(k-1)} - \mu^{(k)} I$	QR factorization of $A^{(k-1)} - \mu^{(k)} I$
$A^{(k)} = R^{(k)} Q^{(k)} + \mu^{(k)} I$	Recombine factors in reverse order

If any off-diagonal element $A_{j,j+1}^{(k)}$ is sufficiently close to zero,

set $A_{j,j+1} = A_{j+1,j} = 0$ to obtain

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} = A^{(k)}$$

and now apply the QR algorithm to A_1 and A_2.

The QR Algorithm

This algorithm, the QR algorithm with well-chosen shifts, has been the standard method for computing all the eigenvalues of a matrix since the early 1960s. Only in the 1990s has a competitor emerged, the divide-and-conquer algorithm described in Lecture 30. Tridiagonalization was discussed in Lecture 26, shifts are discussed in the next lecture, and deflation is not discussed further in this book. For now, let us confine our attention to the "pure" QR algorithm and explain how it finds eigenvalues.

Unnormalized Simultaneous Iteration

Our approach will be to relate the QR algorithm to another method called simultaneous iteration, whose behavior is more obvious.
Idea : applying the power iteration to several vectors at once.
Suppose we starts with a set of n linearly independent vectors $v_1^{(0)}, \ldots, v_n^{(0)}$. It seems plausible that just $A^k v_1^{(0)}$ converges as $k \to \infty$ to the eigenvector corresponding to the largest eigenvalue of A in absolute value, the space $\langle A^k v_1^{(0)}, \ldots, A^k v_n^{(0)} \rangle$ should converge to the space $\langle q_1, \ldots, q_n \rangle$ spanned by the eigenvectors $q_1, \ldots, q_n$ of A corresponding to the n largest eigenvalues in absolute value.

Unnormalized Simultaneous Iteration

Define $V^{(0)}$ to be the $m \times n$ initial matrix

$$V^{(0)} = \left[\begin{array}{c|c|c} v_1^{(0)} & \cdots & v_n^{(0)} \end{array} \right], \tag{1}$$

and define $V^{(k)}$ to be the result after k application of A:

$$V^{(k)} = A^k V^{(0)} = \left[\begin{array}{c|c|c} v_1^{(k)} & \cdots & v_n^{(k)} \end{array} \right]. \tag{2}$$

For column spave of $V^{(k)}$, compute a reduced QR factorization of $V^{(k)}$:

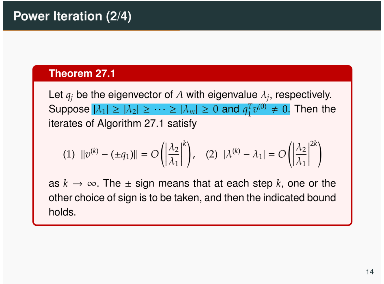
$$\hat{Q}^{(k)} \hat{R}^{(k)} = V^{(k)} \quad \text{where} \quad \hat{Q}^{(k)} : m \times n, \quad \hat{R}^{(k)} : n \times n. \tag{3}$$

It seems plausible that columns of $\hat{Q}^{(k)}$ converge to the eigenvectors $\pm q_1, \ldots, \pm q_n$ as $k \longrightarrow \infty$.

Unnormalized Simultaneous Iteration

If we expand $v_j^{(0)}$ and $v_j^{(k)}$ in the eigenvectors of A, we have

$$v_j^{(0)} = a_{1j}q_1 + \cdots + amjq_m$$
$$v_j^{(k)} = \lambda_1^k a_{1j}q_1 + \cdots + \lambda_m^k amjq_m$$



Power Iteration (2/4)

Theorem 27.1

Let q_j be the eigenvector of A with eigenvalue λ_j, respectively. Suppose $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_m| \geq 0$ and $q_1^T v^{(0)} \neq 0$. Then the iterates of Algorithm 27.1 satisfy

$$(1)\ \ \|v^{(k)} - (\pm q_1)\| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right), \quad (2)\ \ |\lambda^{(k)} - \lambda_1| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right)$$

as $k \to \infty$. The $\pm$ sign means that at each step k, one or the other choice of sign is to be taken, and then the indicated bound holds.

14

By these assumptions, we can assume the following:

All the leading principal minors of $\hat{Q}^T V^{(0)}$ are nonsingular.

Unnormalized Simultaneous Iteration

All the leading principal minors of $\hat{Q}^T V^{(0)}$ are nonsingular.

Definition 2.1

The **leading principal submatrix** of order k of an $n \times n$ matrix is obtained by deleting the last $n - k$ rows and column of the matrix

Definition 2.2

The determinant of a leading principal submatrix is called the **leading principal minor** of A.

Principal minors can be used in definiteness tests.

Unnormalized Simultaneous Iteration

Theorem 2.3

A matrix is

- *positive definite if and only if all its leading principal minors are positive.*
- *negative definite if and only if its odd principal minors are negative and its even principal minors are positive.*
- *indefinite if one of its kth order leading principal minors is negative for an even k or if there are two odd leading principal minors that have different signs.*

And it is equivalent to "leading determinant" in [1].

Test 3	$S = A^{\mathrm{T}}A$ **for a matrix** A **with independent columns**
Test 4	**All the leading determinants** $D_1, D_2, \ldots, D_n$ **of** S **are positive**
Test 5	**All the pivots of** S **are positive** (in elimination)

Unnormalized Simultaneous Iteration

Theorem 2.4

Suppose that the iteration (1),(2) *and* (3) *is carried out and that assumptions are sarisfied. Then as* $k \to \infty$, *the columns of the matrices* $\hat{Q}^{(k)}$ *converges linearly to the eigenvectors of* A*:*

$$\|q_k^{(k)} - \pm q_j\| = O(C^k) \tag{4}$$

for each j *with* $1 \le j \le n$, *where* $C < 1$ *is the constant* $\max\limits_{1 \le k \le n} \frac{|\lambda_{k+1}|}{\lambda_k}$. *As in the theorems of the last lecture, the* $\pm$ *sign means that at each step* k, *one or the other choice of sign is to be taken, and then the indicated bound holds.*

Unnormalized Simultaneous Iteration

Proof.

Extend $\hat{Q}$ to $m \times m$ orthogonal matrix Q of eigenvector of A, then $A = Q\Lambda Q^T$.

$$Q = \left[\begin{array}{c|c|c|c|c|c} q_1 & \cdots & q_n & q_{n+1} & \cdots & q_m \end{array}\right], \Lambda = \left[\begin{array}{ccc|ccc} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_n & & & \\ \hline & & & \lambda_{n+1} & & \\ & & & & \ddots & \\ & & & & & \lambda_m \end{array}\right]$$

Then we have

$$V^{(k)} = A^k V^{(0)} = Q\Lambda^k Q^T V^{(0)} = \hat{Q}\hat{\Lambda}^k \hat{Q}^T V^{(0)} + O(|\lambda_{n+1}|^k)$$

as $k \to \infty$.

Proof.

By assumptions, $\hat{Q}^T V^{(0)}$ is nonsingular, so

$$\begin{aligned} V^{(k)} &= \hat{Q}\hat{\Lambda}^k \hat{Q}^T V^{(0)} + O(|\lambda_{n+1}|^k) \\ &= (\hat{Q}\hat{\Lambda}^k \hat{Q}^T V^{(0)} + O(|\lambda_{n+1}|^k))(\hat{Q}^T V^{(0)})^{-1} \hat{Q}^T V^{(0)} \\ &= (\hat{Q}\hat{\Lambda}^k + O(|\lambda_{n+1}|^k))\hat{Q}^T V^{(0)} \end{aligned}$$

Since $\hat{Q}^T V^{(0)}$ is nonsingular, the column space of this matrix is the same as the column space of

$$\hat{Q}\hat{\Lambda}^k + O(|\lambda_{n+1}|^k).$$

It is clear that this column space converges linearly to that of $\hat{Q}$. We omit the details.

Unnormalized Simultaneous Iteration

Proof.

Since we assume that

$$\text{All the leading principal minors of } \hat{Q}^T V^{(0)} \text{ are nonsingular.}$$

It follows that the argument above also applies to leading subsets of the columns of $V^{(k)}$ and $\hat{Q}$. In each case we conclude that the space spanned by the indicated columns of $V^{(k)}$ converges linearly to the space spanned by the corresponding columns of $\hat{Q}$. From this convergence of all the successive column spaces, together with the definition of the QR factorization (3) and (4). □

Simultaneous Iteration

Since $v_1^{(k)}, \ldots, v_n^{(k)}$ is highly ill-conditioned basis of $\langle v_1^{(k)}, \ldots, v_n^{(k)} \rangle$, orthonormalize at each step.

Algorithm 28.3. Simultaneous Iteration

Pick $\hat{Q}^{(0)} \in \mathbb{R}^{m \times n}$ with orthonormal columns.
for $k = 1, 2, \ldots$
$\quad Z = A\hat{Q}^{(k-1)}$
$\quad \hat{Q}^{(k)}\hat{R}^{(k)} = Z$ — reduced QR factorization of Z

From the form of this algorithm, it is clear that the column spaces of $\hat{Q}^{(k)}$ and $Z^{(k)}$ ate the same, both being equal to the column space of $A^k\hat{Q}^{(0)}$. Thus new algorithm convergences the same conditions as the old one.

Simultaneous Iteration

Theorem 3.1

Algorithm 28.3 generates the same matrices $\hat{Q}^{(k)}$ as the iteration (1),-(3) consider in Theorem 2.3, and under the same assumptions, it converges as described in that theorem.

Simultaneous Iteration ⇔ QR Algorithm

Now we can explain the QR algorithm. It is equivalent to simultaneous iteration applied to a full set of $n = m$ initial vectors, namely, the identity, $\hat{Q}^{(0)} = I$.
Here are the three formulas that define simultaneous iteration with $\underline{Q}^{(0)} = I$, followed by a fourth formula that we shall take as a definition of an $m \times m$ matrix $\underline{A}^{(k)}$. And here are the three formulas that define the pure QR algorithm.

Simultaneous Iteration

$$\underline{Q}^{(0)} = I, \tag{5}$$

$$Z = A\underline{Q}^{(k-1)}, \tag{6}$$

$$Z = \underline{Q}^{(k)} R^{(k)}, \tag{7}$$

$$A^{(k)} = (\underline{Q}^{(k)})^T A \underline{Q}^{(k)}. \tag{8}$$

Unshifted QR Algorithm

$$A^{(0)} = A, \tag{9}$$

$$A^{(k-1)} = Q^{(k)} R^{(k)}, \tag{10}$$

$$A^{(k)} = R^{(k)} Q^{(k)}, \tag{11}$$

$$\underline{Q}^{(k)} = Q^{(1)} Q^{(2)} \cdots Q^{(k)}. \tag{12}$$

Simultaneous Iteration ⇔ QR Algorithm

We can now exhibit the equivalence of these two algorithms.

Theorem 4.1

The process (5) *and* (9) *generate identical sequences of matrices* $\underline{R}^{(k)}, \underline{Q}^{(k)}$ *and* $A^{(k)}$, *namely, those defined by the QR factorization of the* k*th power of* A,

$$A^k = \underline{Q}^{(k)} \underline{R}^{(k)}, \tag{13}$$

together with the projection

$$A^{(k)} = (\underline{Q}^{(k)})^T A \underline{Q}^{(k)}. \tag{14}$$

Simultaneous Iteration ⇔ QR Algorithm

Proof.

We use mathematical induction.
(k=0) Its trivial.
Since $A^0 = \underline{Q}^{(0)} = \underline{R}^{(0)} = I$ and $A^{(0)} = A$, (13) and (14) are immediate.
Consider now the case $k \geq 1$ for simultaneous iteration.

$$\begin{aligned} A^k &= A\underline{Q}^{(k-1)}\underline{R}^{(k-1)} \quad (\because A^k = \underline{Q}^{(k)}\underline{R}^{(k)}) \\ &= \underline{Q}^{(k)}R^{(k)}\underline{R}^{(k-1)} \quad (\because A\underline{Q}^{(k-1)} = \underline{Q}^{(k)}R^{(k)}) \\ &= \underline{Q}^{(k)}\underline{R}^{(k)} \quad (\because \underline{R}^{(k)} = R^{(k)}R^{(k-1)}\cdots R^{(1)}). \end{aligned}$$

Proof.

On the other hand, consider the case $k \geq 1$ for the QR algorithm.

$$
\begin{aligned}
A^k &= A\underline{Q}^{(k-1)}\underline{R}^{(k-1)} \quad (\because A^k = \underline{Q}^{(k)}\underline{R}^{(k)}) \\
&= \underline{Q}^{(k-1)}A^{(k-1)}\underline{R}^{(k-1)} \quad (\because A^{(k)} = (\underline{Q}^{(k)})^T A\underline{Q}^{(k)} \Rightarrow \underline{Q}^{(k)}A^{(k)} = A\underline{Q}^{(k)}) \\
&= \underline{Q}^{(k)}\underline{R}^{(k)} \quad (\because A^{(k-1)} = Q^{(k)}R^{(k)})
\end{aligned}
$$

Finally, we can verify (14) by the sequence

$$
\begin{aligned}
A^{(k)} &= (Q^{(k)})^T A^{(k-1)} Q^{(k)} \quad (\because A^{(k-1)} = Q^{(k)}R^{(k)} \text{ and } A^{(k)} = R^{(k)}Q^{(k)}) \\
&= (\underline{Q}^{(k)})^T A\underline{Q}^{(k)} \quad (\because \text{ inductive hypothesis})
\end{aligned}
$$

□

Convergence of the QR Algorithm

We can now say a great deal about the convergence of the unshifted QR algorithm.

$$A^k = \underline{Q}^{(k)} \underline{R}^{(k)} \quad (13) \qquad\qquad A^{(k)} = (\underline{Q}^{(k)})^T A \underline{Q}^{(k)} \quad (14)$$

- (13) explains why the QR algorithm can be expected to find eigenvectors.
  - It constructs orthonormal basis for successive powers A^k
- (14) explains why algorithm finds eigenvalues.
  - It follows that the diagonal elements of $A^{(k)}$ are Rayleigh quotients of A corresponding to the columns of $\underline{Q}^{(k)}$.
  - As those columns converge to eigenvectors, the Rayleigh quotients converge to the corresponding eigenvalues.
  - Meanwhile, it implies that the off-diagonal elements of $A^{(k)}$ correspond to generalized Rayleigh quotients involving approximations of distinct eigenvectors of A on left and the right.

Convergence of the QR Algorithm

Theorem 5.1

Let the pure QR algorithm be applied to a real symmetrix matrix A whose eigenvalues satisfy $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$ and whose corresponding eigenvector matrix Q has all nonsingular leading principal minors. Then as $k \to \infty$, $A^{(k)}$ converges linearly with constant $\max_k \frac{|\lambda_{k+1}|}{|\lambda_k|}$ to $\mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and $\underline{Q}^{(k)}$ converges at the same rate to Q.

Reference

[1] Gilbert Strang. *Linear algebra and learning from data*. Wellesley-Cambridge Press, 2019.

Thank you!