J. KSIAM Vol.14, No.1, 29–34, 2010

A NEW UNDERSTANDING OF THE QR METHOD

CHOHONG MIN

DEPARTMENT OF MATHEMATICS, KYUNGHEE UNIVERSITY, SOUTH KOREA
E-mail address: chohongmin@gmail.com

ABSTRACT. The QR method is one of the most common methods for calculating the eigenvalues of a square matrix, however its understanding would require complicated and sophisticated mathematical logics. In this article, we present a simple way to understand QR method only with a minimal mathematical knowledge. A deflation technique is introduced, and its combination with the power iteration leads to extracting all the eigenvectors. The orthogonal iteration is then shown to be compatible with the combination of deflation and power iteration. The connection of QR method to orthogonal iteration is then briefly reviewed. Our presentation is unique and easy to understand among many accounts for the QR method by introducing the orthogonal iteration in terms of deflation and power iteration.

1. INTRODUCTION

Since its inception by Francis [4, 5] and Kublanovskaya [8], the QR method has been the most widely used and the most popular method for calculating the eigenvalues of a full matrix. It has been generalized to wider range of eigenvalue problems; QZ method, one of its variants, solves generalized eigenvalue problem $Ax = \lambda Bx$ [7], and a generalization of the QR method called GR method has been researched by Watkins et al [11]. For each specific application, the method has been tuned and upgraded; the restarted QR method is for comrade and fellow matrices [3], and the QR method with a balance technique is for finding the roots of polynomials [1].

Though its importance cannot be overstated, the QR method is normally deferred to graduate course, or would be just presented without enough explanation why it works. The enigma stems from the fact that the convergence proof is not trivial at all [6]. Even presenting a simple way to understand the QR method has been a research topic [10, 2]. The most accounts take the approach of explaining the orthogonal iteration and its connection to QR method [9, 6, 10], and so does this article. But our presentation is different in explaining the orthogonal iteration as successive application of the power iteration to deflated matrices. This article introduces a simple way to understand the QR method only with a minimal knoledge of mathematics.

Received by the editors November 4 2009; Accepted March 8 2010.
2000 *Mathematics Subject Classification.* 65F15.
Key words and phrases. eigenvalues, eigenvector, QR method.

2. POWER ITERATION AND DEFLATED MATRIX

Throughout this paper, we assume a matrix $A \in \mathbb{C}^{n\times n}$ to have distinct eigenvalues λ_1, λ_2, $\cdots$, λ_n with their associated eigenvectors x_1, x_2, $\cdots$, x_n. The eigenvalues are numbered in decreasing order, $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$.

Theorem 2.1. *(Power Iteration) Assume that v^0 in algorithm 1 is chosen randomly enough to have nonzero component of eigenvector x_1, then the sequence v^k satisfies*

$$\lim_{k\to\infty} dist\left(v^k, span(x_1)\right) = 0.$$

Proof. By the assumption, $v^0 = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ with $a_1 \neq 0$. v^k in algorithm 1 is parallel to Av^{k-1}, whenever $k \geq 1$. Repeating this argument leads to the fact that v^k is parallel to A^kv^0, and $v^k = A^kv^0/\left\|A^kv^0\right\|$.

$$\begin{aligned}
A^kv^0 &= a_1\lambda_1^kx_1 + a_2\lambda_2^kx_2 + \cdots + a_n\lambda_n^kx_n \\
\lim_{k\to\infty} v^k &= \lim_{k\to\infty} \frac{a_1\lambda_1^kx_1 + a_2\lambda_2^kx_2 + \cdots + a_n\lambda_n^kx_n}{\left\|a_1\lambda_1^kx_1 + a_2\lambda_2^kx_2 + \cdots + a_n\lambda_n^kx_n\right\|} \\
&= \lim_{k\to\infty} \frac{\left(\frac{\lambda_1}{\|\lambda_1\|}\right)^k x_1 + \frac{a_2}{a_1}\left(\frac{\lambda_2}{\|\lambda_1\|}\right)^k x_2 + \cdots + \frac{a_n}{a_1}\left(\frac{\lambda_n}{\|\lambda_1\|}\right)^k x_n}{\left\|x_1 + \frac{a_2}{a_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k x_2 + \cdots + \frac{a_n}{a_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k x_n\right\|} \\
&= \lim_{k\to\infty} \left(\frac{\lambda_1}{\|\lambda_1\|}\right)^k \frac{x_1}{\|x_1\|}
\end{aligned}$$

As $k \to \infty$, v^k belongs to the $span(x_1)$, thus $\lim_{k\to\infty} dist\left(v^k, span(x_1)\right) = 0$. ☐

Algorithm 1 Power iteration : $v = power(A)$

Input : an $n \times n$ matrix A
v^0 : randomly chosen
for $k = 0, 1, 2, \cdots$ **do**
  $v^{k+1} = \frac{Av^k}{\|Av^k\|}$
end for
Output : $v = v^k$ for some large k

The above theorem shows that the sequence of power iteration will eventually belong to $span(x_1)$. Let us set a unit eigenvector $v^1 \in span\,(x_1)$ as a vector v^k in the power iteration applied to A for some large k. The next theorem suggests a way to calculate the next largest one.

Theorem 2.2. *(Deflated Matrix) Let v_1 be a unit eigenvector of A associated with λ_1, then the matrix $\left(I - v_1 v_1^T\right) A$, called the deflated matrix of A with v_1, has eigenvalues $0, \lambda_2, \cdots, \lambda_n$ with corresponding eigenvectors $v_1, \left(I - v_1 v_1^T\right) x_2, \cdots, \left(I - v_1 v_1^T\right) x_n$.*

Proof. Since $\left(I - v_1 v_1^T\right) A v_1 = \lambda_1 v_1 - \lambda_1 v_1 = 0$, 0 is an eigenvalue of the deflated matrix with eigenvector v_1. The deflated matrix annihilates the v_1 component, hence

$$\begin{aligned} \left(I - v_1 v_1^T\right) A \left(I - v_1 v_1^T\right) x_j &= \left(I - v_1 v_1^T\right) A x_j \\ &= \lambda_j \left(I - v_1 v_1^T\right) x_j . \end{aligned}$$

Note that whenever $j \neq 1$, $x_j - (v_1 \cdot x_j) v_1$ cannot be a zero vector, otherwise it contradicts the linear independence of the eigenvectors belonging two different eigenvalues. □

Corollary 2.3. *Let the power iteration, denoted by $\left(v_2^k\right)$, operate on the deflated matrix $\left(I - v_1 v_1^T\right) A$, then*

$$\lim_{k \to \infty} dist\left(v_2^k, span\left(x_1, x_2\right)\right) = 0.$$

Proof. The largest eigenvalue of the deflated matrix is λ_2 and its associated eigenvector is $x_2 - (v_1 \cdot x_2) v_1$. From theorem 2.1, as $k \to \infty$, v_2^k belongs to the space $span\left(x_2 - (v_1 \cdot x_2) v_1\right)$ which is the subspace of $span\left(x_1, x_2\right)$. □

Power iteration obtains a unit eigenvector of the eigenvalue with the largest modulus. The deflation annihilates the largest eigenvalue, and exposes the next largest one for the power iteration to pick up. In this way, we set a unit vector $v_2 \in span\left(x_1, x_2\right)$ as v^k in the power iteration applied to $\left(I - v_1 v_1^T\right) A$ for some large k. We repeatedly apply the power iteration on the deflated matrix to obtain unit length vectors $v_1, v_2, \cdots, v_n$ with the properties listed in Algorithm 2.

Algorithm 2 Successive power iterations on deflated matrices

Input : an $n \times n$ matrix A
$v_1 = power\left(A\right)$
$v_2 = power\left(\left(I - v_1 v_1^T\right) A\right)$
$\vdots$
$v_n = power\left(\left(I - v_{n-1} v_{n-1}^T\right) \cdots \left(I - v_1 v_1^T\right) A\right)$
Output : $v_1, v_2, \cdots, v_n$

Theorem 2.4. *The vectors $v_1, v_2, \cdots, v_n$ in Algorithm 2 satisfy*

$$\begin{array}{lll} v_1 & \in & span\left(x_1\right) \\ v_2 & \in & span\left(x_1, x_2\right) \\ & \vdots & \\ v_n & \in & span\left(x_1, x_2, \cdots, x_n\right) \end{array} .$$

Proof. Let $A_i = \left(I - v_i v_i^T\right) \cdots \left(I - v_1 v_1^T\right) A$ be the i^{th} deflated matrix in Algorithm 2. By Theorem 2.2, each deflation annihilates the largest eigenvalue of A_i and adds a scalar multiple of the associated eigenvector to all the other eigenvectors. Thus the largest eigenvalue in modulus of A_i is λ_{i+1}, and its associated eigenvector belongs to $span\left(v_1, \cdots, v_i, x_{i+1}\right)$, for each $i = 0, 1, \cdots, n-1$. For the matrix A_i, v_{i+1} is an eigenvector associated with the eigenvalue of the largest modulus, and $v_{i+1} \in span\left(v_1, \cdots, v_i, x_{i+1}\right)$ for each i. By mathematical induction on i, it is clear that $v_{i+1} \in span\left(x_1, \cdots, x_i, x_{i+1}\right)$ for each i. □

3. Orthogonal Iteration

Algorithm 2 sequentially computes all the eigenvectors. Even though the power iteration for v_i is not completed, the temporary value should serve as a good approximation, and Algorithm 3 combines all the power iterations in one iteration by using the approximations $v_1^k, \cdots, v_n^k$ for $v_1, \cdots, v_n$, respectively. The two algorithms are just two formulations of the successive power iterations on deflated matrices. Practically, Algorithm 3 is more efficient than Algorithm 2 in a sense that the former can propose a good approximation for the full eigenvector system in a meantime of the iteration, while the latter cannot.

Algorithm 3 Combined power iterations on recursively deflated matrices

for $k = 0, 1, 2, \cdots$ **do**

$$
\begin{aligned}
v_1^{k+1} &= A v_1^k && , \quad v_1^{k+1} = v_1^{k+1} / \left\| v_1^{k+1} \right\| \\
v_2^{k+1} &= \left(I - v_1^{k+1} \left(v_1^{k+1}\right)^T\right) A v_1^k && , \quad v_2^{k+1} = v_2^{k+1} / \left\| v_2^{k+1} \right\| \\
& && \quad \vdots \\
v_n^{k+1} &= \left(I - v_{n-1}^{k+1} \left(v_{n-1}^{k+1}\right)^T\right) \cdots \left(I - v_1^{k+1} \left(v_1^{k+1}\right)^T\right) A v_n^k && , \quad v_n^{k+1} = v_n^{k+1} / \left\| v_n^{k+1} \right\|
\end{aligned}
$$

end for

Note that the routine inside Algorithm 3 is nothing but the QR factorization that obtains the orthonormal basis $v_1^{k+1}, \cdots, v_n^{k+1}$ from the basis $Av_1^k, \cdots, Av_n^k$. Writing the vectors in columns, $V^k = \left[v_1^k, \cdots, v_n^k\right]$, Algorithm 4, called orthogonal iteration, is hence a simple restatement of Algorithm 3.

Algorithm 4 Orthogonal Iteration

for $k = 0, 1, 2, \cdots$ **do**
  $AV^k = V^{k+1} R^{k+1}$: QR factorization of the matrix AV^k
end for

Theorem 3.1. *Let* $\left(V^k\right)_{k\in\mathbb{N}}$ *be the sequence in the orthogonal iteration, then*

- $\lim_{k\to\infty} span\left(v_1^k,\cdots,v_j^k\right) = span\left(x_1,\cdots,x_j\right)$ for $j=1,2,\cdots,n$
- $\lim_{k\to\infty} v_i^k\cdot Av_j^k = 0$ if $i>j$.

Proof. The orthogonal iteration is an implementation of the successive application of the power iteration on the deflated matrices. By theorem 2.4, $\lim_{k\to\infty} dist\left(v_j^k, span\left(x_1,\cdots,x_j\right)\right)$ for each j, and the first argument follows.

As $k\to\infty$, v_j^k belongs to $span\left(x_1,\cdots,x_j\right)$, and Av_j^k belongs to the same space, since the eigenvectors are invariant. By the first argument, Av_j^k belongs to $span\left(v_1^k,\cdots,v_j^k\right)$ when $k\to\infty$. In the QR factorization, v_i^k is orthogonal to v_j^k whenever $i>j$, and $v_i^k\perp span\left(v_1^k,\cdots,v_j^k\right)$. Therefore $\lim_{k\to\infty} v_i^k\cdot Av_j^k=0$ whenever $i>j$. □

4. QR METHOD

The previous section reveals the relation between the column vectors of the orthogonal iteration and the eigenvectors. To retrieve the eigenvalues of A, let us set $A^k=\left(V^k\right)^T AV^k$. Then Theorem 3.1 states that A^k becomes upper triangular as $k\to\infty$. Since A^k is similar to A, the eigenvalues of A should appear in the diagonal of the upper triangular matrix. Thus when $A^k=\left(V^k\right)^T AV^k$ is inserted in the orthogonal iteration, the eigenvalues can be obtained. From $AV^k=V^{k+1}R^{k+1}$,

$$\begin{array}{lllll} A^k & = & \left(V^k\right)^T AV^k & = & \left(V^k\right)^T V^{k+1}R^{k+1} \\ A^{k+1} & = & \left(V^{k+1}\right)^T AV^{k+1} & = & R^{k+1}\left(V^k\right)^T V^{k+1} \end{array}.$$

Since V^k and V^{k+1} are orthogonal matrices, $\left(V^k\right)^T V^{k+1}$ is also orthogonal. Therefore the above equations can be simply written as

$$\begin{array}{lll} A^k & = & Q^{k+1}R^{k+1} \\ A^{k+1} & = & R^{k+1}Q^{k+1}, \end{array}$$

where $A^k=Q^{k+1}R^{k+1}$ is the QR factorization of A^k. The above recursive formula shows that the sequence A^k can be generated detached from the orthogonal iteration. Algorithm 5, called QR method, shows the complete procedure how to find the full eigenvalues of a general matrix. Since A^k is similar to A, the eigenvalues are preserved each k. By Theorem 3.1, the sequence A^k will eventually become upper triangular matrix, which reveals the eigenvalues on its diagonal.

ACKNOWLEDGMENTS

This research was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2008-331-C00045).

Algorithm 5 QR Method

Input : an $n \times n$ matrix A
$A^0 = A$
for $k = 0, 1, 2, \cdots$ **do**
  $A^k = Q^{k+1} R^{k+1}$
  $A^{k+1} = R^{k+1} Q^{k+1}$
end for
Output : an upper triangular matrix A^k for some large k

REFERENCES

[1] G. Ammar, D. Calvetti, W. Gragg, and L. Reichel. Polynomial zerofinders based on szego polynomials. *J. Comput. Appl. Math.*, 127:1–16, 2001.
[2] R. Burnside and P. Guest. A simple proof of the transposed qr algorithm. *SIAM Review*, 38:306–308, 1996.
[3] D. Calvetti, S. Kim, and L. Reichel. The restarted qr-algorithm for eigenvalue computation of structured matrices. *J. Comput. Appl. Math.*, 149:415–422, 2002.
[4] J. Francis. The qr transformation i. *Comput. J.*, 4:265–271, 1961.
[5] J. Francis. The qr transformation ii. *Comput. J.*, 4:332–345, 1961.
[6] G. Golub and C. Loan. *Matrix Computations*. The John Hopkins University Press, 1989.
[7] B. Kagstrom and A. Ruhe. Matrix pencils. *Springer-Verlag Lecture Notes in Mathematics*, 973, 1983.
[8] V. Kublanovskaya. On some algorithms for the solution of the complete eigenvalue problem. *USSR Comput. Math. Math. Phys.*, 3:637–657, 1961.
[9] L. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, 1997.
[10] D. Watkins. Understanding the qr algorithm. *SIAM Review*, 24:427–439, 1982.
[11] D. Watkins. Qr-like algorithms for eigenvalue problems. *J. Comput. Appl. Math.*, 123:67–83, 2000.